THESE

présentée

A L'UNIVERSITE DE GRENOBLE III

pour obtenir

LE TITRE DE DOCTEUR TROISIEME CYCLE

SPECIALITE: LINGUISTIQUE

par

Françoise EMERARD

*

SYNTHESE PAR DIPHONES ET TRAITEMENT DE LA PROSODIE

SOUTENUE: Le 11 mars 1977

DEVANT LA COMMISSION D'EXAMEN

M. G. TUAILLON Président
M. W. ENDRES
M. J. GENIN
M. L.J. BOE
} Examinateurs

THESE

présentée

A L'UNIVERSITE DE GRENOBLE III

pour obtenir

LE TITRE DE DOCTEUR TROISIEME CYCLE

SPECIALITE: LINGUISTIQUE

par

Françoise EMERARD

*

SYNTHESE PAR DIPHONES
ET
TRAITEMENT DE LA PROSODIE

SOUTENUE: Le 11 mars 1977

DEVANT LA COMMISSION D'EXAMEN

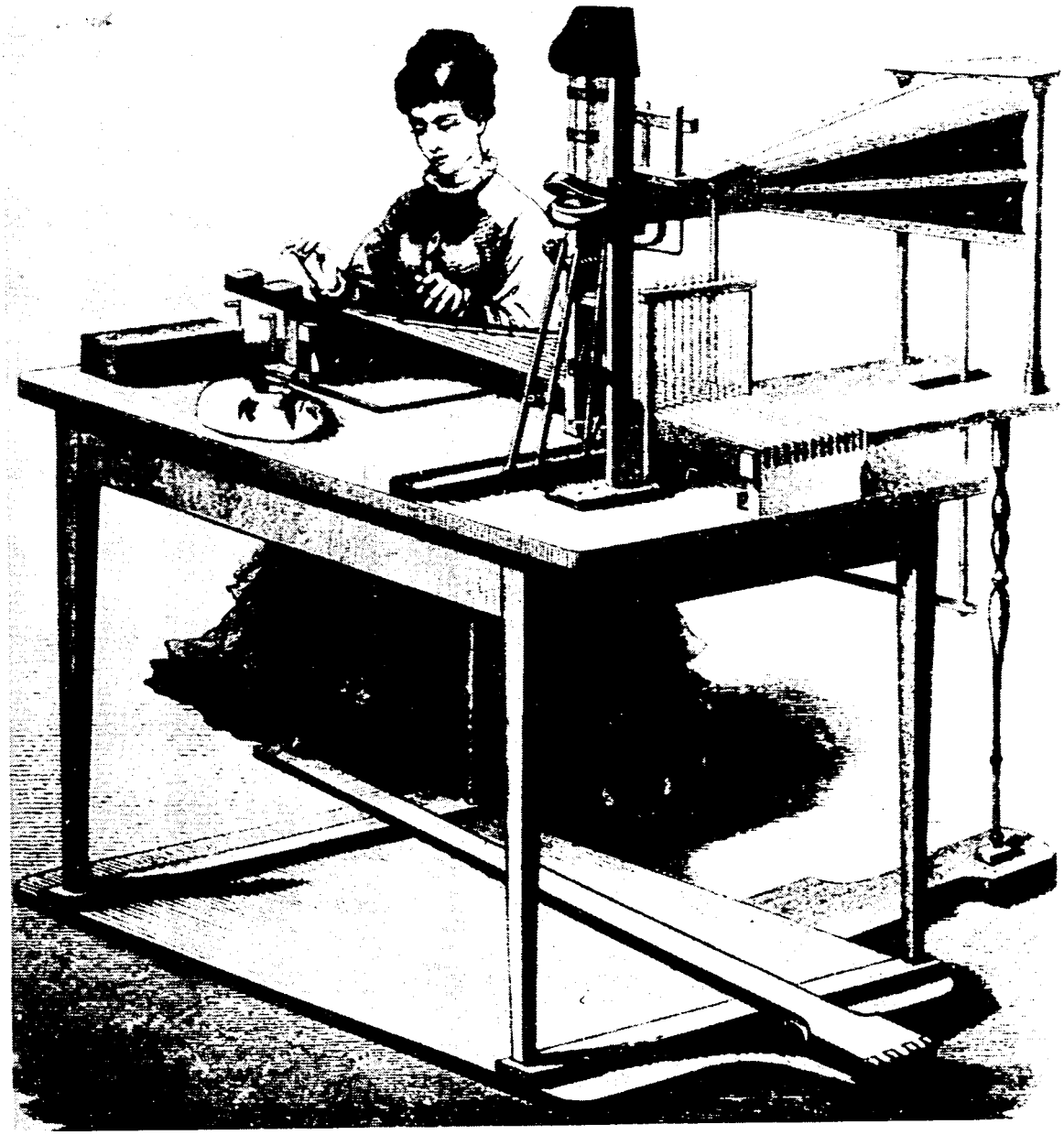
M. G. TUAILLON Président
M. W. ENDRES
M. J. GENIN } Examinateurs
M. L.J. BOE

..." et crois-moi, souvent la réalisation personnelle n'est pas l'oeuvre de l'individu seul, mais de tout son entourage".

Lidia FALCON

"Lettres à une idiote espagnole" 1975.

1835



La "machine à parler" de FABER (d'après DUDLEY et TARNOCZY 1950)

Partagée entre ma terre dauphinoise et les côtes bretonnes, j'ai beaucoup reçu des montagnes et de la mer.

Je tiens à exprimer mes profonds remerciements à Monsieur le Professeur TUAILLON que notre étude a exilé provisoirement de sa passion pour la Dialectologie et qui a bien voulu me faire l'honneur d'accepter la présidence du jury.

Je tiens à dire au Professeur W. ENDRES, Chef du Groupe de Recherches en Traitement de l'Information Appliqué à la Parole et Professeur à la Technische Hochschule de DARMSTADT combien je suis honorée de sa présence dans le jury aujourd'hui et combien j'espère que se poursuivent les échanges fructueux entre le F.T.Z. (Fernmeldetechnisches Zentralamt) de DARMSTADT, et le Département E.T.A. du CNET-LANNION.

*Que Louis-Jean BOE, chercheur à l'Institut de Phonétique de GRENOBLE trouve ici un témoignage d'amitié et de profonde reconnaissance pour son aide, ses conseils et ses encouragements qui se sont manifestés sans cesse depuis mes débuts d'étudiante à l'Institut. Je profite de cette occasion pour lui dire, ainsi qu'à Michel CONTINI et à tous ceux qui sont la vie de l'Institut, combien les étudiants dont je suis sont sensibles à la façon exemplaire dont ils vivent leur recherche et à l'intérêt qu'ils manifestent pour ceux qui partagent leur passion.**

Le travail présenté dans ce mémoire a été effectué au Centre National d'Etudes des Télécommunications de LANNION.

Que Monsieur LE MEZEC, Adjoint du Directeur Scientifique, soit remercié pour l'intérêt avec lequel il a suivi l'avancement de notre travail et pour l'interlocuteur toujours attentif qu'il a été aux difficultés des stagiaires.

Je suis vivement reconnaissante à Monsieur LORAND, alors Chef du Département Etudes et Techniques d'Acoustique , qui a bien voulu accueillir "une littéraire" dans ses laboratoires et lui faire totalement confiance. Au moment où il est appelé à de plus hautes fonctions, nous lui disons que notre souci à tous sera de continuer dans l'esprit qu'il a su insuffler au Département.

Je remercie Jacques GENIN, Chef du Groupe "Synthèse de la Parole" au Département E.T.A., de sa présence aujourd'hui dans le jury, et je lui exprime une gratitude profonde pour la façon dont il a su, en même temps que guider ce travail avec compétence, nous laisser toute liberté pour le mener à bien.

.../...

Danièle LARREUR devint mon amie et il n'est pas besoin d'en ajouter davantage. Les mots seraient trop imparfaits pour dire tout ce qu'elle m'a apporté et donné dans cette recherche dont la concrétisation est le fruit d'une collaboration de chaque instant, et dans tous les moments des jours.

Un immense merci - par l'intermédiaire de Christiane BOISSEL qui a assuré avec intelligence, efficacité et une grande gentillesse la "transcription manuscrite - dactylographique" - à tous ceux qui ont apporté avec chaleur leur compétence pour la réalisation matérielle de cet ouvrage.

Et puis, je voudrais dire à tous ceux "qui font E.T.A." - ainsi qu'à Patrick VANDAMME, Pierre et Jacquy BOYER - combien j'ai été touchée par toutes les marques d'amitié qu'ils m'ont prodiguées pendant ces deux années.

Grâce à tous et à chacun, vivre l'ambiance exceptionnelle de ce Département aura été un cadeau inestimable.

J'ai beaucoup reçu !..-

* De chaleureux remerciements vont également àDominique Vuillet et à Anne-Marie Emerard qui,en proposant leur collaboration, ont rendu possible la présentation de ce travail au jour dit.

TABLE DES MATIERES

.../...

2ÈME PARTIE : LES OPTIONS

3ÈME PARTIE : L'ANALYSE INSTRUMENTALE DES FAITS PROSODIQUES

.../...

.../...

..../...

INTRODUCTION

Depuis une trentaine d'années, la technologie des calculateurs a fait des progrès considérables. Le traitement de l'information est réalisé de plus en plus rapidement et s'étend à des champs d'application qui débordent très largement le domaine du calcul scientifique.

De ce fait, un double effort a été mené :

- pour améliorer la vitesse d'exécution des systèmes permettant l'introduction des données et la communication des résultats (organes d'entrée - sortie).

- pour simplifier le langage informatique afin qu'il puisse être utilisé par toutes les catégories d'usagers attirés par les nouvelles possibilités offertes mais peu au fait des langages de programmation.

Mais dans l'état actuel des choses, quelque soit le code symbolique utilisé pour fournir des données à traiter ou interroger un système (banque de données, gestion de stocks..) il est obligatoire de passer par une étape intermédiaire graphique. Par contre, la réponse peut déjà être communiquée sous forme vocale, grâce à la synthèse de la parole. C'est évidemment une solution séduisante : rapide et simple, elle permet (comme la télé imprimante) l'échange à distance par ligne téléphonique.

Evidemment, l'idéal serait que l'échange se réalise sous la forme d'un véritable dialogue vocal dans les deux directions et que les informations ou les questions puissent également être transmises

.../...

au calculateur sous forme vocale. Mais ceci suppose que soit résolu le problème de la reconnaissance automatique de la parole qui est beaucoup plus complexe que celui de la synthèse :

. Alors que la synthèse se contente de reproduire l'ensemble des réalisations d'un seul locuteur, la reconnaissance suppose que l'on sache identifier l'ensemble des réalisations de l'ensemble des locuteurs avec ce que cela suppose de variantes phonétiques régionales, individuelles et même momentanées.

. Si en synthèse, on peut se contenter de ne transmettre que ce qui est nécessaire et suffisant pour assurer l'intelligibilité, voire la qualité, la reconnaissance va devoir opérer à partir d'un signal qui véhicule une série d'informations sous la forme d'une structure redondante: il va falloir extraire les seules informations qui sont nécessaires et suffisantes pour assurer la communication verbale.

. S'il existe des techniques relativement simples pour passer d'une suite d'éléments discrets à un continuum sonore, l'inverse - passage du son à son équivalent phonétique - est beaucoup plus difficile. Cette opération de segmentation comporte obligatoirement une part d'artificiel (il n'existe pas de véritables unités au niveau acoustique) et pour l'opérer automatiquement et systématiquement d'une façon satisfaisante, il faut déjà avoir une petite idée de l'identité des éléments sur lesquels on opère le découpage (autrement dit pour pouvoir bien segmenter la parole en vue de sa reconnaissance, il faudrait l'avoir déjà reconnue).
D'autre part, il n'est pas évident que l'on puisse utiliser toutes les unités mises en évidence par une analyse linguistique : "la nostalgie de l'invariance et la tendance à plaquer les unités discrètes révélées par l'analyse linguistique sur la substance physique qu'est le discours réalisé semble méconnaître les principes mêmes sur lesquels s'appuie l'analyse linguistique. C'est une chose que d'admettre à titre d'hypothèse de travail, une structure linguistique hiérarchisée, qui part du trait distinctif pour aboutir à la phrase ; c'en est une

autre que de postuler un décodage qui doit d'abord et de manière absolue, récupérer le phonème" (WAJSKOP, 1970).

. Une fois résolu le problème de la segmentation, il faut reconnaître les suites de séquences de sons et leur associer une suite de mots, ce qui n'est pas trivial, compte-tenu des problèmes d'homophonie et de la non existante du mot au niveau acoustique (pour la parole continue).

. Enfin, pour avoir accès au niveau sémantique, il faut disposer d'un outil linguistique parfaitement formalisé si l'on ne veut pas se contenter d'un langage rudimentaire à la syntaxe simplifiée.

. On peut constater que cet ensemble d'opérations, décrites d'une façon séquentielle, nécessite plusieurs bouclages permettant de fonctionner par approximations successives dans le cas où une erreur à une étape donnée n'est décelée que dans l'étape suivante.

Qu'est-ce que la synthèse <u>automatique</u> de la parole ?
C'est le passage de l'écriture orthographique au son.

-La première opération va consister à savoir faire automatiquement la transcription orthographique-phonétique en adoptant un code (normatif par exemple). Cette étape que nous n'avons pas abordée, présente peu d'intérêt théorique dans la mesure où malheureusement le code écrit et le code oral (pour le français comme pour beaucoup de langues) ont suivi des évolutions divergentes. Il existe de tels systèmes et ils ne peuvent se comparer les uns aux autres que par leur économie (TEIL, 1969 ; DIVAY et GUYOMARD, 1977 à paraître).

A cette étape, et indépendamment du matériel utilisé pour réaliser la synthèse (synthétiseur à formants, vocodeur à canaux, analogue du conduit vocal) il va falloir définir une stratégie relativement empirique.
On dispose d'une façon abstraite (ensemble de symboles phonétiques) d'un système qui, théoriquement, devrait permettre, comme l'on combine

.../...

les lettres entre elles pour composer n'importe quelle phrase, de produire n'importe quelle séquence de sons.

Quand le phonéticien opère une transcription auditive (association sons-symboles phonétiques) qui d'ailleurs présuppose la connaissance de la langue, il isole des unités en nombre limité ; mais cette opération résulte de techniques qui se situent sur le plan perceptif et qui réalisent à la fois le découpage et la reconnaissance du signal acoustique.

Pour pouvoir effectuer de la synthèse, il va falloir disposer de la même façon, mais cette fois-ci non plus à un niveau abstrait (qui était le niveau phonétique) mais concrètement, d'éléments minimaux acoustiques permettant de constituer n'importe quelle phrase faute de devoir mémoriser toutes les phrases d'une même langue.
La synthèse telle que nous l'avons définie suppose une analyse préalable qui seule peut nous permettre de constituer ce stock d'unités.

. Plusieurs solutions ont été explorées pour réaliser une combinatoire économique : mots isolés dans le cas d'un vocabulaire limité (avec des règles de raccordement), syllabes, éléments correspondant à la réalisation d'un phonème et enfin diphones.
Plus le nombre d'unités sera restreint, plus les règles de composition seront complexes dans la mesure où la réalisation d'un son dépend de l'entourage dans lequel il est articulé, et ce sont justement ces interactions mutuelles qui sont à l'origine de la reconnaissance de la séquence.

Utiliser un élément correspondant au phonème suppose que l'on saura reconstituer toutes les transitions lorsqu'il sera en contact avec un autre phonème (synthèse par règles) : cette solution, si elle paraît séduisante pour un phonéticien, présente de grandes difficultés de réalisation (KELLY et GERSTMAN, 1961 ; RABINER, 1969).

Le choix du diphone (unité acoustique née de l'astuce des techniciens de la synthèse) consiste à contourner les règles de composition en intégrant dans les unités choisies, ces zones de transition et en effectuant le raccordement sur les parties stables; si le

.../...

nombre d'unités croît comme le nombre de combinaisons possibles, les règles de composition se simplifient à l'extrême : il suffit d'opérer une simple juxtaposition.

C'est à cette méthode que nous avons eu recours pour notre travail.

-La seconde étape est liée à la technique de synthèse elle-même, car dans sa pratique, la parole synthétique, et c'est là son intérêt, contient beaucoup moins d'information (au sens de la théorie de la communication) que la parole naturelle, tout en gardant son intelligibilité : elle est débarassée (plus ou moins) de sa redondance.
Il va donc falloir dans une analyse préalable, extraire des unités que nous avons choisies (les diphones) ce que l'on croît être nécessaire et suffisant pour assurer la communication (par exemple, avec un synthétiseur à formants, il faudra disposer d'une bonne dizaine de paramètres correspondants aux formants, à la nasalité, au bruit...).

Plus cette analyse sera différenciée en fonction de la nature de chaque son, plus la quantité d'information sera réduite pour reconstituer la parole à partir de ces données.

Le vocodeur comme nous le préciserons par la suite, présente une solution intermédiaire : s'il peut difficilement descendre à 1200 eb/s. comme dans le cas d'un synthétiseur à formants, il réalise automatiquement l'extraction des paramètres - grâce à l'analyseur dont il dispose - et délivre à la reconstitution, un signal parfaitement intelligible qui peut n'avoir que 2400 eb/s. ; c'est pourquoi nous avons choisi ce type de synthétiseur.

De très nombreuses études ont mis en évidence que la parole synthétique ne pouvait espérer atteindre un caractère naturel si une grande attention n'était portée au problème des faits suprasegmentaux (caractéristiques intrinsèques, accent, intonation) ; comme leur nom l'indique, ces faits se superposent aux problèmes dont nous venons de parler et cela d'une façon indépendante.

.../...

Telle qu'on l'entend, la synthèse de la parole reste pour le moment une opération relativement simple : il ne s'agit que d'associer à une suite de caractères qui sont fournis comme données d'entrée, une suite de sons ; la machine n'a à opérer aucun choix, elle réalise une simple correspondance entre écriture et son.
Pour ce qui est des faits suprasegmentaux, le problème est d'un tout autre ordre :

Mis à part la ponctuation et le découpage en mots, l'écriture ne donne aucune information sur l'accent du mot et sur l'intonation : il faudra créer de toutes pièces une organisation temporelle par modification de la durée respective des sons et par l'introduction de pauses de durée différente, ainsi qu'une structuration intonative.

On comprend que ce problème soit beaucoup plus complexe et c'est pourquoi nous nous y sommes attachée plus particulièrement.

En résumé, nous pouvons dire que la synthèse de la parole présuppose une triple analyse du signal acoustique :

1°) Une segmentation temporelle dont le résultat permet d'isoler des unités de dimension réduite.

2°) De ces unités, l'extraction de ce qui suffit à la transmission de l'information.

3°) Une modélisation permettant une reconstitution ab nihilo (par rapport aux données d'entrée du système) des faits prosodiques suffisamment souple pour pouvoir s'adapter à des constructions syntaxiques de complexités différentes.

Ces étapes franchies, il faudra mettre en mémoire toutes ces informations, (programme de transcription orthographique-phonétique, bibliothèque des éléments minimaux, règles de composition de ces sons,

.../...

règles prosodiques) et ce, en tenant compte des caractéristiques propres au synthétiseur utilisé.

Ainsi, à une suite de caractères donnés, la machine sera-t-elle en mesure d'associer une suite de sons qui seront ensuite combinés puis reconstitués sous une forme sonore avec la prosodie correspondante.

IÈRE P A R T I E

APPAREILLAGE

ET

ELEMENTS DE PAROLE

CHAPITRE I

LES METHODES ET MATERIELS DE SYNTHESE

Il s'agit ici pour nous de faire un inventaire - non exhaustif - des matériels et méthodes qui permettent de synthétiser la parole à l'aide d'un certain nombre de paramètres considérés comme pertinents et qui auront été isolés à partir d'une analyse préalable. Mais il s'agit surtout d'une part de les spécifier par rapport à l'axe méthodologique et d'autre part de concevoir, sur l'axe historique, l'explication de leur apparition.

Si l'on utilise la plus ou moins grande analogie avec nos organes phonatoires comme critère de distinction entre les divers matériels de synthèse existant, on peut opérer parmi eux un classement en trois catégories :

1°/ Les appareils qui opèrent une modification d'un signal source afin d'obtenir un spectre le plus voisin possible de celui de la parole sans référence aux mécanismes articulatoires du conduit vocal qui ont été à l'origine de la production du son.

Il s'agit essentiellement du vocodeur à canaux de DUDLEY et plus indirectement du Pattern Play Back de COOPER.

2°/ Les synthétiseurs à formants constituent une catégorie intermédiaire : ils se caractérisent par une simulation du processus de modulation du spectre de la source vocale. L'importance du rôle des formants a été mis en évidence : la source d'impulsion attaque trois ou quatre circuits résonnants dont on commande la fréquence de résonance conformément à l'image d'évolution de la fréquence du formant considéré. Mais comme dans la catégorie précédente, on ne s'intéresse pas à l'aspect articulatoire dans la production de la parole.

.../...

3°/ Les matériels dont le principe repose sur l'analyse du mode de fonctionnement des organes phonatoires. La simulation de la forme et des dimensions du conduit vocal nécessite l'introduction de paramètres de commande de nature articulatoire pour piloter le synthétiseur.

Appartiennent à cette catégorie les synthétiseurs par codage prédictif et les analogues du conduit vocal.

Quand on essaie de comprendre comment historiquement se sont développées ces différentes catégories de synthétiseurs, on s'aperçoit qu'aux deux extrêmités et tout le long de l'axe diachronique, par delà toutes les autres réalisations, se situent les recherches orientées vers la simulation du conduit vocal ; leur progression est parallèle aux progrès réalisés dans la connaissance des mécanismes phonatoires et plus précisément articulatoires (fig.7).

C'est sans doute au XVIIIe siècle que l'on peut situer les premières expériences en matière de synthèse de parole. Auparavant, la fascination exercée par une parole dont on ne sait pas "d'où elle sort" avait conduit à d'astucieux stratagèmes, mais ceux-ci ne devaient rien à la simulation et beaucoup à la dissimulation (fig. 1).

.../...

FIG. 1 - LES PREMIERS TEMPS DE LA SYNTHESE
d'après FLANAGAN J.L., 1976)

A partir de 1780, on peut mentionner toute une série de réalisations qui ne découlaient évidemment que d'une observation assez sommaire des mécanismes de l'appareil phonatoire :

* D.G. KRATZENSTEIN crée le premier modèle mécanique du conduit vocal : une source excite acoustiquement cinq résonateurs dont la forme et les dimensions permettent de produire cinq voyelles isolées (fig. 2)

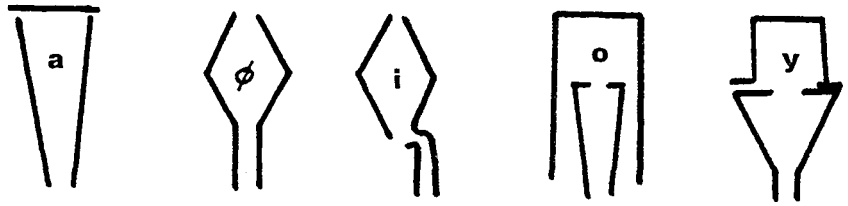


FIG. 2 - Les résonateurs de KRATZENSTEIN (d'après YOUNG, 1845)

.../...

* L'Abbé MICAL confectionne un appareil .. " les têtes parlantes"_ capable de prononcer deux phrases et dont le principe est grossièrement comparable à celui des boîtes à musique (à picots).

* Mais c'est sans doute, de cette époque, la machine parlante de VON KEMPELEN (1791) qui demeure la mieux connue. On lui doit :

  - d'avoir pris en compte le mouvement de certains organes articulatoires (la bouche en particulier) et le rôle de la glotte, constituée d'une lame vibrante excitée par un soufflet.

  - d'avoir eu l'intuition du caractère continu de la parole : on pouvait relier les sons les uns aux autres mais il fallait une habileté prodigieuse pour pouvoir manoeuvrer avec synchronisation tous les leviers et soufflets qui assuraient le fonctionnement de la machine et la succession des sons entre eux (fig.3).

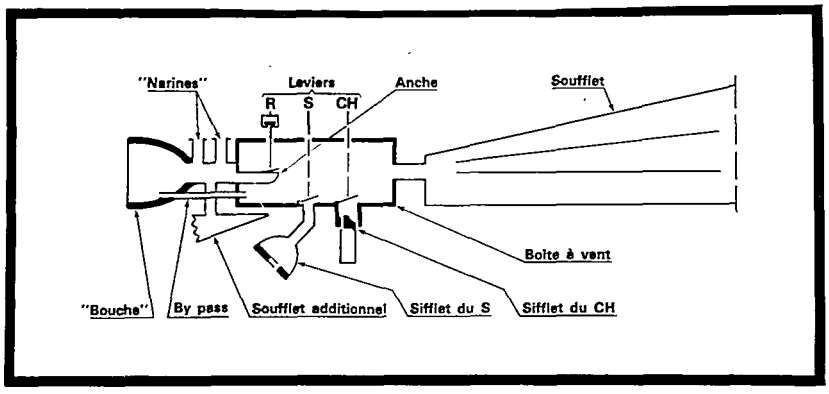


FIG.3 - Schéma de la machine de VON KEMPELEN (d'après LIENARD, 1970).

* C'est à FABER (1835) qu'on doit l'amélioration de la machine de VON KEMPELEN. Il introduit :

  - les organes articulatoires qui faisaient originellement défaut : deux mâchoires en caoutchouc et une langue articulée,
  - la modification dans la forme de la glotte.

C'est aussi à cette époque que les travaux remarquables de MÜLLER (1847) vont poser des bases sérieuses pour la connaissance du fonctionnement du larynx.

.../...

La réalisation d'analogues mécaniques du conduit vocal prend fin vers cette époque. Il faut attendre la fin du XIXe siècle et la découverte des Rayons X par W.K. RONTGEN, puis le début du XXe siècle et les progrès des développements technologiques pour voir un nouvel essor se dessiner dans la recherche sur la production et la reproduction de la parole, et dans la connaissance sur les propriétés spectrales du signal de la parole.

Les possibilités offertes par la radiographie de pouvoir mieux approcher le fonctionnement des organes phonatoires et décrire statiquement le conduit vocal pendant la production des sons "stationnaires" relancent l'intérêt d'une simulation à l'aide de paramètres articulatoires.

En 1922, STEWART réalise le premier analogue électrique de la résonance du conduit vocal.

A partir de 1931, grâce à l'apparition de l'électronique et aux progrès de l'électroacoustique, on dispose de nouveaux outils de recherche : Magnétophone - qui permet plutôt que de chercher à "créer" de la parole à la stocker pour la restituer - , sonagraphe..

Mais parallèlement, on réalise encore des simulations de la glotte à partir de larynx excisés (Tredelenburg,1937) ou de morceaux de chambre à air de bicyclette (WETHLO,1939).

En 1939, les études sur les lignes de transmission électrique conduisent DUDLEY, de la Bell Telephone, à réaliser un analogue électrique du conduit vocal : le VODER (Voice Demonstrator). Mais là encore, la continuité entre les sons n'était à peu près assurée que par l'intervention constante et habile de l'homme.

A la même époque, la multiplication du nombre des lignes téléphoniques et les problèmes de transmission que ce développement provoque, orientent les recherches vers une meilleure connaissance des processus de la parole et plus particulièrement vers l'analyse spectrale du signal de parole.

La théorie de l'Information nous apprend que la parole est très redondante sur le plan du signal(indépendamment de certaines redondances qui se situent à un niveau linguistique).

. Quand on émet une séquence parlée à raison de dix sons par seconde, l'équivalent écrit comporte une information de 50 eb/s.

.../...

. D'autre part, la transmission d'un signal de parole dans une bande de fréquences de 3 000 Hz avec un rapport signal sur bruit de 30 db requiert un débit d'information de 30 000 eb/seconde.

On s'aperçoit donc que la transmission du signal sous sa forme acoustique (ou l'équivalent codé) nécessite beaucoup plus d'information que l'équivalent graphique.Et l'on sait également que bien que le spectre des signaux de parole couvre une bande de fréquences comprise entre 60Hz et 12000Hz,la transmission 300-3400Hz suffit pour conserver à la parole une bonne intélligibilité.

* C'est dans ce but de compression de l'information et de transmission de parole à faible débit que se situe la mise au point du Vocodeur (Voice Coder) par DUDLEY (1939) : Il consiste en l'extraction dans le riche spectre de la parole des paramètres pertinents et suffisants pour rendre compte des sons.

A l'analyse, on obtient l'énergie dans 10 bandes de fréquences qui couvrent le spectre de 300 à 3400 Hz (filtres passe-bande, redressement, filtres passe bas). Un autre circuit détecte la présence ou l'absence du fondamental et donne, lorsqu'il est présent, sa fréquence (fig.4).

Schématiquement donc, le vocodeur opère à l'analyse une dissociation entre spectre et source (s).

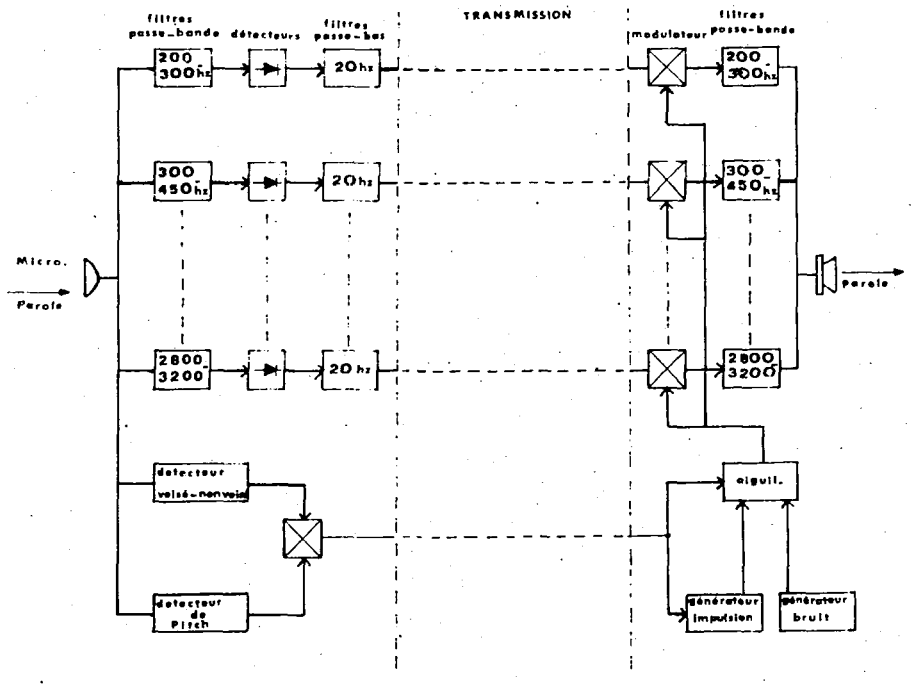


FIG. 4 - Schéma du vocodeur à canaux - D'après SCHROEDER (1966).

.../...

Toutes les informations obtenues à l'analyse sont ensuite transmises au synthétiseur dont la structure est parallèle à celle de l'analyseur : le signal d'excitation vocale (ou de bruit) est recréé et modulé en amplitude dans dix canaux correspondant à ceux de l'analyse.

Le processus d'analyse et de synthèse est effectué en temps réel, et le débit de transmission est en définitive de 2 400 éléments binaires par seconde. En conclusion, on peut dire qu'avec le vocodeur il s'agit d'un processus de décomposition et de recomposition du signal de la parole systématique dans le domaine du spectre et donc ne tenant pas compte de la nature propre de chaque son ; on peut le qualifier d'aveugle. (Nous développerons plus en détail les caractéristiques du vocodeur à canaux dans la IIe partie puisque c'est ce type de matériel que nous avons utilisé).

Dans les années 1950, deux types d'appareils ont été réalisés : le Pattern Play Back de COOPER, que l'on peut rattacher aux matériels de la première catégorie (définie en introduction du Chapitre), et un synthétiseur de type paramétrique : le synthétiseur à formants (seconde catégorie).

Ces appareils nécessitent une analyse fine et sélective du signal de la parole qui tient compte de la nature de chaque élément. La finesse de cette discrimination nécessite des opérations d'analyse très complexes qui l'ont empêché d'être automatique jusqu'à ce jour.

* Le PATTERN PLAY BACK de COOPER (1950) ou
  Relecteur de Sonagrammes.

Les sonagrammes permettent une représentation à trois dimensions du spectre de parole : le temps en abcisse, les composantes fréquentielles en ordonnée, et l'intensité globalement donnée par le degré de noirceur.

Cet outil a permis de mettre en évidence le caractère continu de la parole, c'est-à-dire le rôle des transitions de son à son (tous les travaux des Laboratoires HASKINS). Les sonagrammes permettent de suivre assez précisément l'évolution des zones de transition au cours du temps, ce qui rend possible après observation de redessiner, non pas toutes les composantes du spectre (la valeur du fondamental et par voie de conséquence des harmoniques est arbitrairement fixée dès le départ : F_0 = 120 Hz), mais l'évolution des éléments nécessaires et suffisants à la perception des sons : les indices acoustiques (fig.5).

.../...

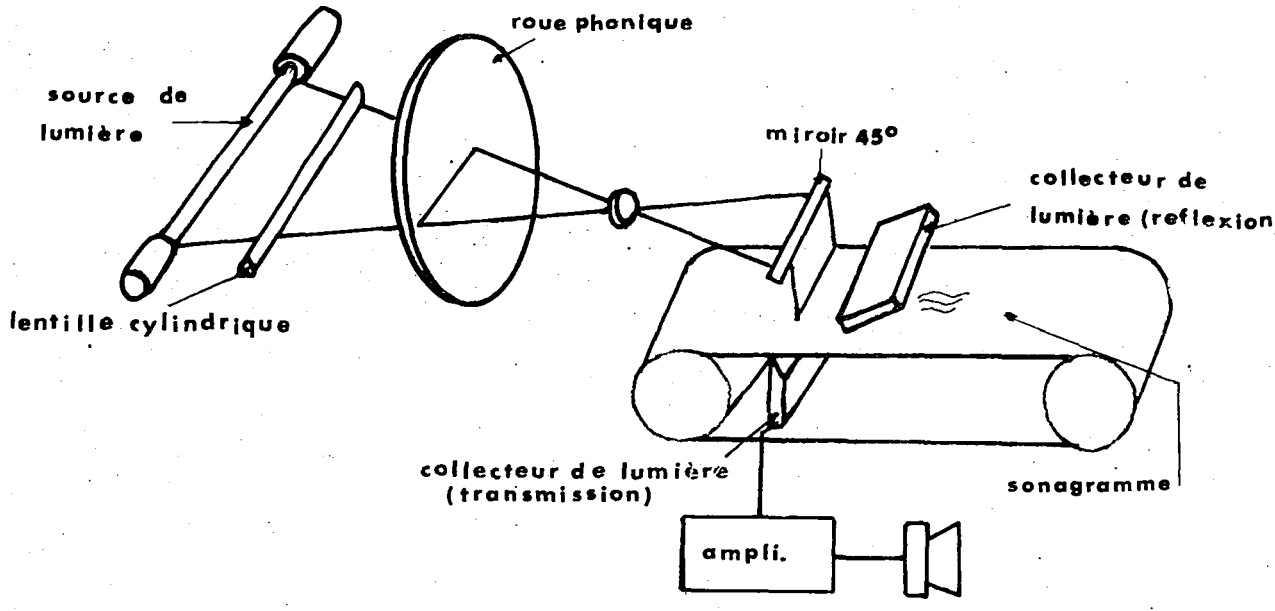


FIG. 5 - SCHEMA DU PATTERN PLAY BACK DE COOPER (d'après LIBERMAN et al, 1952).

Le dessin des formants et des zones de bruit est réalisé sur une plaque transparente. Le principe du Pattern Play Back consiste à venir lire cette configuration schématisée et à en donner l'équivalent sonore (Cooper et al, 1952). Cinquante faisceaux produits par une source lumineuse sont modulés de 120 à 6 000 Hz et sont dirigés par un miroir vers le sonagramme qui défile ; une cellule photoélectrique recueille le résultat de cette modulation et opère sa transformation en un signal électrique, lequel est amplifié puis transmis à un haut parleur.

* Les synthétiseurs à formants .

D'une part les résultats de l'analyse spectrale du signal - importance des formants et de leurs évolutions, ainsi que la théorie du locus mise en lumière par Cooper et al. (1952) -, d'autre part les possibilités de réduction du débit de l'information ont conduit à la réalisation de synthétiseurs à formants entièrement électroniques.

Il s'agit avec ce type de synthétiseur de reconstituer le spectre à partir de la structure formantique non plus à l'aide de signaux lumineux, mais à l'aide de signaux électriques. Or les formants, nous l'avons dit, résultent de la fonction de transfert du conduit vocal. C'est parce que ce matériel fait référence indirectement à la forme du conduit vocal que nous l'avons classé dans une catégorie intermédiaire.

L'analyse spectrale va permettre d'obtenir un certain nombre de paramètres dont le caractère pertinent sera testé grâce à la synthèse. La vitesse de l'évolution de ces paramètres reste faible puisqu'elle est liée à la vitesse d'évolution des organes d'articulation (de l'ordre de 20 ms).

.../...

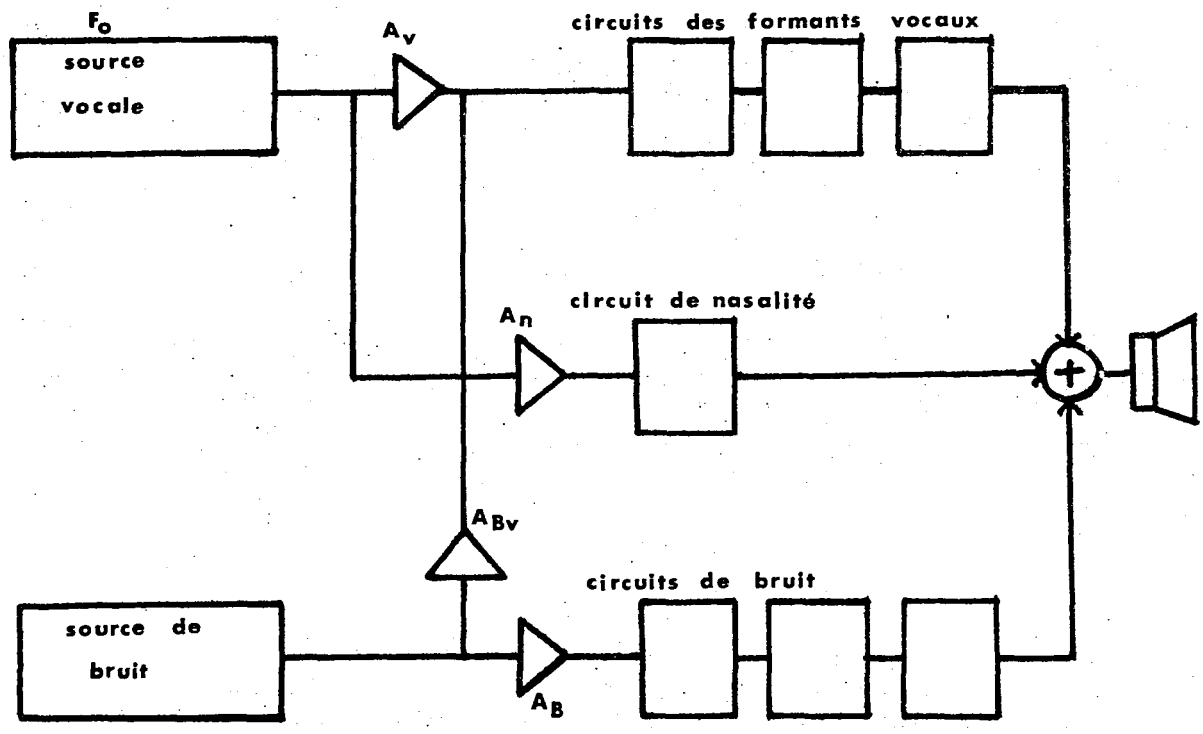


FIG.6 - SCHEMA D'UN SYNTHETISEUR A FORMANTS.

A_v Amplitude de la source vocale.
A_n Amplitude de la source vocale dans les cavités nasales.
A_{B_v} Amplitude du bruit dans les formants vocaux.
A_B Amplitude de la source de bruit.

Dans cette catégorie de synthétiseurs, le nombre de paramètres extraits est généralement compris entre 10 et 15 : (FANT et al, 1953 ; LAWRENCE, 1953 ; BEAUVIALA et al, 1968) la fréquence fondamentale, la fréquence des trois premiers formants vocaux, l'amplitude de la source vocale, l'amplitude de la source de bruit, les fréquences de résonance et d'antirésonances de bruits, l'amplitude du bruit dans les formants vocaux, l'amplitude de la source vocale dans les cavités nasales... (fig.6).

Il est possible de faire varier chacun des paramètres isolément. Le débit de transmission est faible : de l'ordre de 2 000 eb/seconde. Cependant, la synthèse n'est pas instantanée puisque le synthétiseur est

alimenté après un relevé des informations paramétriques sur des sonagrammes.

Les années 1950 marquent également, après les travaux fondamentaux de CHIBA et KAJIYAMA (1941), le début d'une période où l'on réalise à nouveau des analogues électriques du conduit vocal (DUNN 1950 ; STEVENS et al 1953) avec un système de commande qui permet de faire varier plus facilement les lieux d'articulation et les degrés d'aperture.

Puis le développement sur le marché des ordinateurs permet d'envisager la simulation numérique du conduit vocal. Dans le même temps la mise au point de la radiocinématographie (1960) donne la possibilité d'étudier les mouvements des organes phonatoires. Tous ces progrès technologiques provoquent évidemment un regain d'intérêt pour les recherches orientées vers la synthèse par simulation de l'appareil de phonation.

Signalons cependant en 1962 la réalisation du V.E.V. (Voice Excited Vocoder : le vocodeur à excitation vocale de DAVID et al), qui renoue avec le principe du vocodeur de DUDLEY : ici , à la compression optimale de la bande, on préfère une meilleure définition des composantes basses fréquences ; par conséquent on transmet intégralement (c'est-à-dire sans compression) la bande de fréquences comprise entre 250 et 940 Hz, et pour le reste, on conserve une compression du type vocodeur à canaux.

Depuis quelques années donc, les orientations des chercheurs se concentrent sur la simulation complète des processus de phonation et d'articulation par l'utilisation des paramètres articulatoires (COKER, 1967, MERMELSTEIN, 1967)comme paramètres de commande.

Deux approches semblent se dessiner :

1/ S'étant donné un modèle de production de la parole, on cherche à approximer au mieux le signal de parole en utilisant les coefficients d'un filtre numérique récursif (ceux-ci n'ont pas de correspondance directe sur le plan physiologique). On peut utiliser alors à cette fin l'analyse mathématique par codage prédictif (ATAL et al., 1971, EL MALLAWANY, 1975).

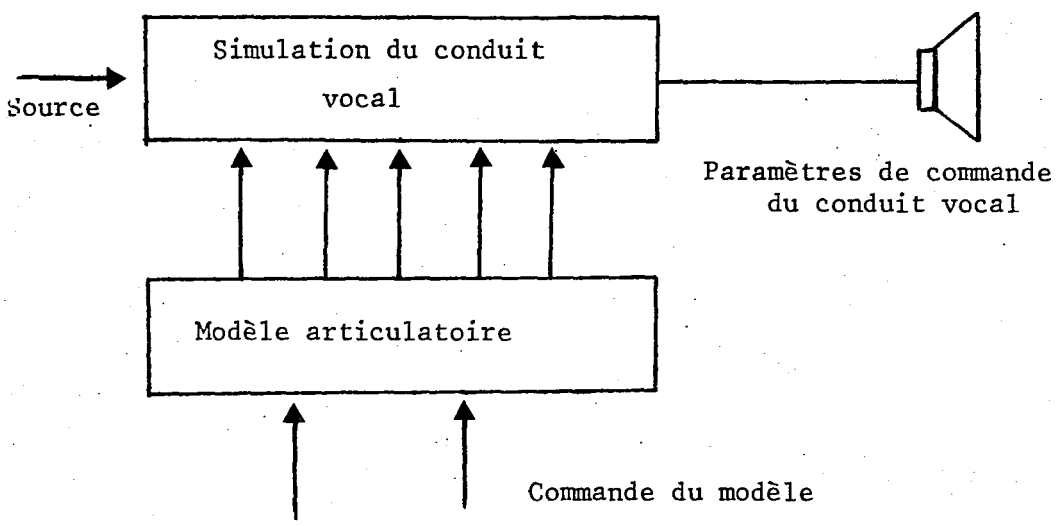
2/ Supposant connu l'ensemble des paramètres articulatoires et de leurs évolutions pendant la phonation, on cherche à réaliser par un dispositif - électronique par exemple - un analogue complet de l'appareil vocal humain (source vocale et conduit vocal) pour piloter un synthétiseur par des données représentant des paramètres significatifs sur le plan physiologique (COKER, 1967) : pression d'air dans les poumons,

.../...

tension des cordes vocales, valeur du débit injecté à l'entrée du conduit vocal (FLANAGAN, 1972; CARCAUD et al, 1976).

Cette approche soulève cependant d'énormes problèmes : en effet, étant donné la complexité des phénomènes phonatoires, on est encore loin aujourd'hui de savoir préciser tous les paramètres, de pouvoir les détecter aisément, et de savoir les simuler à tous les niveaux, c'est-à-dire non seulement acoustique et articulatoire, mais aussi musculaire, nerveux et cérébral.

On est légitimement en droit de penser que, quand cela sera, la parole obtenue sera de bonne qualité et de bonne fidélité tout en restant à faible débit : la connaissance des mouvements articulatoires (synergie) et des contraintes au niveau des organes d'articulation de la parole devraient pouvoir aboutir à l'élaboration d'un modèle permettant la commande précise d'un synthétiseur à l'aide de peu de paramètres :



Au terme de ce panorama historique sur les matériels et méthodes utilisées pour réaliser la synthèse de la parole, nous sommes consciente de ses lacunes ; mais nous pensons cependant que les autres méthodes dont nous n'avons pas parlé peuvent être grossièrement assimilées à l'une des trois catégories définies au départ. Pour une étude plus complète de ce domaine, on pourra se reporter aux publications dont nous nous sommes aidé : DUDLEY et TARNOCZY (1950), METTAS (1965), FLANAGAN (1972).

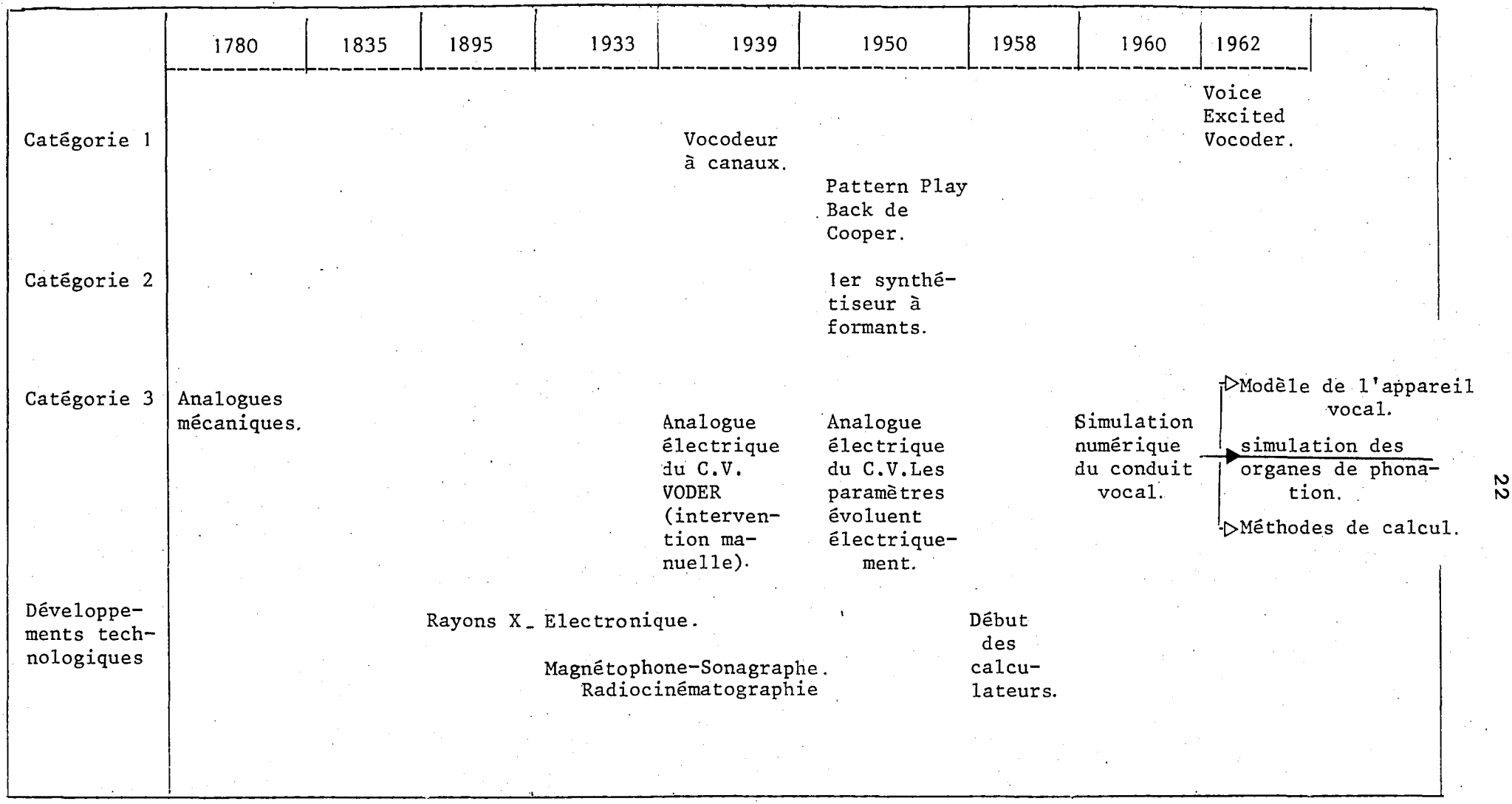


FIG.7 - EVOLUTION HISTORIQUE DES METHODES ET MATERIELS DE SYNTHESE.

CHAPITRE II

LES ELEMENTS DE PAROLE UTILISES EN SYNTHESE.

De la même façon que les matériels de synthése ont évolué vers une plus grande analogie avec notre appareil vocal à mesure que les progrès dans la connaissance des mécanismes de phonation et dans la technologie le permettaient, de la même façon le choix des éléments de parole utilisés a été fonction de ces mêmes progrès.

Nous ne reparlerons pas des premières tentatives du XVIIIe siècle parce que si l'on pouvait parler dès ce moment là de parole synthétique, les résultats n'étaient dus qu'à l'intervention indispensable d'un opérateur en l'absence duquel seuls des sons isolés pouvaient être produits.

On peut définir une double évolution dans le domaine de la synthèse de la parole : d'une part dans le mode de stockage des éléments de parole choisis, d'autre part dans le choix même des éléments de parole.

- passage d'un stockage des données de parole de type analogique à un stockage de type numérique à partir d'une compression des données.

- choix d'éléments de parole d'abord "macroscopiques" puis"microscopiques".

.../...

MODE DE STOCKAGE	ELEMENTS DE PAROLE		EXEMPLE DE REALISATIONS.
ANALOGIQUE	MACROSCOPIQUES ELEMENTS	PHRASES	"Il n'y a pas d'abonné....."
		MOTS	HORLOGE PARLANTE IBM 7770
CODAGE NUMERIQUE		MOTS	IBM 7772 URV-CNET
	MICROSCOPIQUES ELEMENTS	SYLLABES DIPHONES	ICOPHONE
		SYNTHESE PAR REGLES	

I e ETAPE : Il ne s'agit pas à proprement parler de synthèse de parole, mais d'enregistrements analogiques classiques : des phrases stéréotypées sont enregistrées soit sur un magnétophone qui permet de conserver et de reproduire la parole avec une totale fidélité, soit sur un tambour magnétique à plusieurs pistes : il suffit dans ce cas d'orienter la tête de lecture sur la piste qui contient la réponse adéquate. On a un bon exemple de cette méthode dans les phrases bien connues que dispensent les exploitants du téléphone : "Il n'y a pas d'abonné au numéro que vous avez demandé,...".

La parole reproduite est d'excellente qualité, le fonctionnement est simple et économique , mais le choix des éléments de parole limite ces procédés au nombre de phrases enregistrées au départ.

.../...

IIe ETAPE : La mise en service des ordinateurs a permis de concevoir de nouvelles procédures : on utilise maintenant des mots ou des membres de phrases que l'on stocke toujours de façon analogique sur un tambour magnétique, celui-ci étant relié à un calculateur. A chaque mot correpond une adresse ; la composition du message est réalisé par l'intermédiaire du calculateur qui établit l'adressage successif des mots du message et oriente la tête de lecture vers les mots choisis et dans l'ordre chronologique établi par l'ordonnancement du message. Plusieurs réalisations ont validé les principes de cette méthode; entre autres on peut citer l'IBM 7770, installé dans une banque et l'Audio Response System de Burroughs.(C'est aussi le principe de l'horloge parlante, mais sans calculateur). Comme précédemment, la qualité de la voix émise est excellente; cependant on doit apporter quelques restrictions concernant la méthode :

* On ne peut toujours pas parler de véritable synthèse de la parole puisque les mots du vocabulaire sont stockés sans analyse préalable ; tout au plus peut-on parler de synthèse au niveau de l'assemblage des mots en phrases .

* Pour des raisons d'encombrement et de coût, le vocabulaire de base est obligatoirement limité ; sur l'IBM 7770 par exemple, on ne peut enregistrer que 128 mots.

* Les mots du message se succèdent avec l'intonation et la durée qu'ils avaient quand ils ont été enregistrés par le locuteur ; ce caractère figé des mots donne un aspect très artificiel à la parole réalisée par concaténation.

IIIe ETAPE : Le vocodeur de DUDLEY (1939) a donné la possibilité d'effectuer le codage d'un signal de parole pour permettre sa transmission à faible débit. Si l'on opère une conversion analogique-digitale, le calculateur pourra mettre en mémoire les informations numériques du signal . On peut également effectuer le processus inverse de celui

.../...

de l'analyse : le calculateur peut restituer ces données pour alimenter un synthétiseur et recomposer la parole (conversion numérique.analogique).

Ce type de compression et de stockage a d'abord été utilisé sur des phrases - il était possible de stocker environ quatre heures de parole en mémoire sur disque magnétique - puis sur des mots : on peut citer deux des réalisations d'unités à réponse vocale (URV) qui ont utilisé le vocodeur à canaux et le stockage numérique des informations codées:

- 1'IBM 7772 (Buron,1968)
- 1'URV réalisée au CNET (PONCIN, 1970)
  (CARTIER et al,1971)

1/ L'IBM 7772 :

Un vocodeur à canaux a été utilisé ; le choix du locuteur - professionel - s'est porté sur une voix de femme ("une voix de femme a été choisie, comme réflétant le mieux, la volonté des futurs clients" ! -BURON,1968)

Les mots du vocabulaire prononcés isolément ont été enregistrés, analysés par le vocodeur, stockés, puis "retouchés" pour éliminer certaines erreurs dans la détection de la période du fondamental et pour éliminer certains échantillons dont l'information paraissait redondante, enfin pour introduire une énergie plus grande dans les composantes hautes fréquences de certains sons (f, s, ʃ) dont une part de l'information spectrale est située en dehors de la bande téléphonique.

Les messages délivrés sont le résultat de la succession des mots entre lesquels est insérée une pause. Mais il n'y a pas de programme permettant de générer la prosodie ; par conséquent, les mots qui ont été volontairement enregistrés sur un ton neutre pour pouvoir être utilisés dans n'importe quel contexte se concatènent avec les caractéristiques prosodiques réalisées au moment de l'enregistrement. Le résultat est une voix monotone dépourvue de tout naturel.

.../...

2/ L'Unité à réponse vocale réalisée au CNET.

L'application visée concernait directement les exploitants des services téléphoniques : il s'agissait dans un premier temps de permettre aux abonnés d'obtenir de façon automatique et sous forme vocale le coût de leur dernière communication téléphonique ou le contenu de leur compteur de taxation. Quant au vocabulaire de base, il est déterminé par l'application visée ; il est constitué de parties fixes, c'est-à-dire de membres de phrases dont l'occurrence est prédéterminée - ce qui a permis de les enregistrer avec la prosodie convenable : "votre compteur indique... taxes de base" et "votre dernière communication a coûté ... taxes de base". Le vocabulaire comprend également des mots nécessaires à la composition d'un nombre quelconque compris entre 1 et 999999. Ces mots, bien qu'ils soient susceptibles d'apparaître dans différents contextes sont enregistrés et stockés en un seul exemplaire ; ceci présente évidemment l'avantage de n'utiliser qu'un faible volume dans la mémoire du calculateur : 33 éléments de parole, dont le mot de liaison et, suffisent pour synthétiser n'importe lequel de ces nombres (fig. 8).

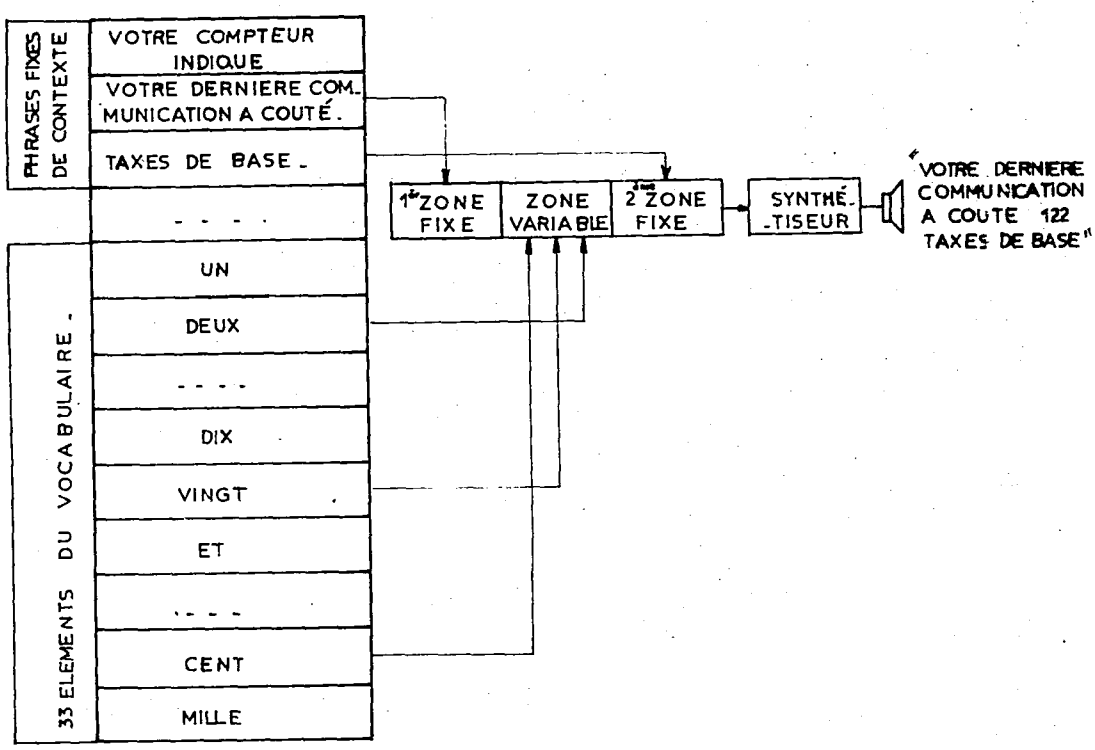


FIG. 8 - PRINCIPE DE LA SYNTHESE PAR MOTS.

.../...

Chaque mot représentant un chiffre a été enregistré isolément afin d'éliminer les influences contextuelles évidentes de chiffres extraits de nombres composés de plusieurs éléments. D'autre part, comme dans le système IBM, des retouches ont été effectuées pour améliorer le vocabulaire de base, non pas de façon automatique ou semi automatique, mais "manuellement", élément par élément, ce que permettait le vocabulaire réduit.

L'on conçoit cependant que l'économie réalisée par le procédé de codage et par le choix du vocabulaire va se payer par une programmation plus complexe pour élaborer une adaptation prosodique des nombres à leur environnement : l'élément "cinq" par exemple peut se trouver inclus dans le nombre 105 200, ou bien se situer en fin de nombre (6 205) ; dans l'un et l'autre cas, la durée et l'intonation seront bien différentes.

C'est pourquoi la modulation du débit et de l'intonation est envisagée pour améliorer l'intelligibilité, la qualité et le naturel des messages obtenus par juxtaposition de nombres simples .

Grâce au positionnement de marqueurs de rythme inscrits sur les mots dans le dictionnaire, on définit des zones d'accélération sur certains échantillons vocodeurs représentatifs du nombre : un programme permet de modifier, non pas le nombre des échantillons mais la durée de chaque échantillon : celle-ci, quand elle est accélérée dans les zones définies, passe de 25 à 12,5 ms.

La décision d'accélération est prise en fonction de la position du chiffre dans le nombre : par exemple, il est bien connu en phonétique que la syllabe finale est plus longue que les autres ; de la même façon, le dernier élément d'un nombre est plus long que l'élément imbriqué dans un nombre composé. Un ralentissement ne se justifie jamais du fait que les mots ont été prononcés de façon isolée, c'est-à-dire en dehors de tout contexte susceptible de les accélérer.

.../...

Parallèlement on tient compte de l'influence contextuelle du chiffre dans le nombre pour produire une modulation du paramètre d'intonation ; on introduit pour ce faire des "mots d'intonation" dans les éléments du dictionnaire, qui consistent en l'indication de la pente globale de Fo et de son signe sur chacun des éléments par augmentation ou diminution de la période du fondamental (on prévoit par exemple une montée de la fréquence fondamentale sur la dernière voyelle du dernier élément du nombre situé avant "taxes de base").

La réalisation de cette unité à réponse vocale a montré la possibilité d'allier à un stockage économique des données de parole, une procédure de gestion des paramètres prosodiques observés à l'analyse, et d'améliorer par là-même l'intelligibilité et la qualité de la parole de synthèse.

IVe ETAPE : Synthèse de la parole à partir "d'éléments microscopiques"

Il faut dans le domaine de la synthèse de parole à partir d'éléments de petite dimension, distinguer deux approches différentes :

La synthèse peut être réalisée :

- soit par simple juxtaposition de ces éléments : il peut s'agir de syllabes, de diphones, ou encore de segments du type voyelle-consonne-voyelle ...

- soit à partir de segments de parole que l'on assemble grâce à tout un ensemble de règles capables de reproduire les zones de transition initialement absentes - de par le principe même de la méthode - d'un segment au suivant.

Nous reparlerons des réalisations du premier type quand nous exposerons l'approche que nous avons nous-même envisagée. Quant aux réalisations qui composent la seconde catégorie, elles sont si nombreuses qu'il serait vain de vouloir en rendre compte ici. On pourra se reporter directement aux publications consacrées à l'historique des études sur la synthèse de la parole. Nous voudrions simplement dans le

.../...

cadre de cette étude, tenter d'expliquer par quel cheminement on en est venu à la conception d'une synthèse par règles.

C'est l'apparition du sonagraphe (KOENIG et al,1946) qui a rendu possible la représentation acoustique de l'analyse spectrale du signal de parole.HARRIS, C.M. (1953), utilisant la démarche expérimentale, constate l'échec d'une tentative pour réaliser de la parole continue à partir de la simple juxtaposition d'éléments minimaux.

COOPER et al. (1948) avaient mis en évidence le rôle perceptuel des zones de transition des formants d'un "phonème" à un autre, et affirmé par là_même l'inutilité d'une méthode de synthèse qui reposerait sur la succession de sons figés une fois pour toutes.

* HARRIS, C. M., est sans doute le premier à avoir tenté de produire de la parole continue à partir de sa segmentation en éléments minimaux : il enregistre des éléments de parole - "buildings blocks" - composés de la suite [consonne-voyelle-consonne] à l'intérieur de laquelle il découpe des segments de la taille du "phonème" et qu'il colle bout à bout pour réaliser des mots. Il conclut que la parole obtenue est à la fois dépourvue de naturel et inintelligible.

Il affirme donc la nécessité de réaliser plusieurs "blocks" ou images spectrales du même segment pour prendre en compte les interactions spectrales entre des éléments voisins : le [w] dans [wi] n'est pas le même que [w] dans[wu] parce que les caractéristiques formantiques de [i] et [u] sont différentes. Le nombre de "blocks" nécessaires pour représenter un seul segment ne peutpas être connu à priori, il ne peut être déterminé que par un ensemble de tests de perception.

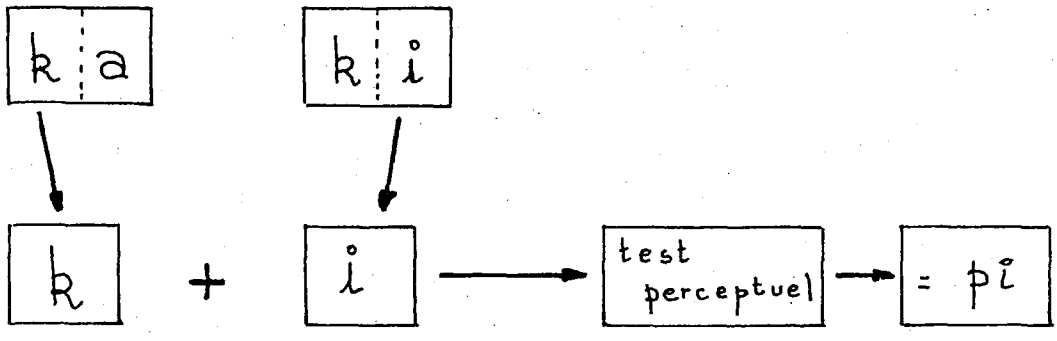
La conclusion implicite qui se dégage de cette étude est qu'une synthèse par juxtaposition d'éléments minimaux supposerait

.../...

un nombre extrêmement important de ces éléments. A la limite, on ne pourrait même plus utiliser des réalisations de "phonèmes" mais des éléments d'un même allophone, rejoignant en cela les conclusions de TWADELL (1952) : "ce sont les allophones et non les phonèmes qui sont les constantes acoustiques de la parole".

* C'est aussi à partir des années 1950 que les Laboratoires HASKINS ont entrepris des recherches sur la structure acoustique de la parole, aidés en cela par le Sonagraphe qui permet la visualisation du spectre des sons de la parole, et par le Pattern Play Back pour vérifier les hypothèses induites de l'observation.

Ces outils leur ont permis de mettre en évidence le rôle des formants pour l'identification des voyelles, et de noter la modification dans la direction des formants observée sur les voyelles quand celles-ci se situent dans un contexte consonantique (LIBERMAN, INGEMAN et al.,1959). De la même façon, ils ont pu démontrer que le spectre de l'explosion n'était pas suffisant pour différencier les consonnes occlusives :
soit l'enregistrement séparé de deux syllabes [ka] et [ki] (DELATTRE, P., 1958) ; si l'on découpe l'explosion de [k]dans [ka]et qu'on l'ajoute au début de la réalisation [i] dans [ki] , les auditeurs perçoivent non pas [ki] mais [pi] :



.../...

Ce test montre l'intervention de la voyelle subséquente pour la perception de la consonne ; d'où la conclusion suivante : les modifications des transitions (en particulier du second formant) des voyelles provoquées par l'environnement consonantique constituent un indice acoustique nécessaire à l'identification des consonnes; les zones de transition que l'on observe entre les "états stables" des unités minimales ne sont pas des phénomènes accessoires dans la production de la parole, mais participent directement et essentiellement à la reconnaissance des unités produites. COOPER et al. concluent qu'on ne peut pas trouver de réalisation phonémique à l'état libre :" on ne peut pas trouver d'invariant acoustique pour un phonème donné . "

Toutes les études menées à partir de cette époque parviennent à une conclusion identique :

Le phonème est une entité abstraite qui n'a pas d'existence acoustique, le signal de la parole se présente sous la forme d'un continuum sonore qu'il est impossible de segmenter en unités discrètes : "les indices acoustiques des phonèmes successifs sont tellement interdépendants dans la chaîne sonore que les segments de sons que l'on peut identifier ne correspondent pas à des segments au niveau phonématique"(*) (LIBERMAN et al., 1967).

Pour ces raisons, une synthèse de la parole qui reposerait sur la simple juxtaposition d'unités minimales est vouée à l'échec ; sa réussite suppose la prise en compte de tout un ensemble de règles susceptibles de reconstituer la continuité spectrale entre les éléments choisis.

Ces résultats constituent un apport fondamental dans la connaissance de la structure acoustique de la parole et de son fonctionnement linguistique . Il existe bien deux niveaux différents : le niveau acoustique et le niveau phonétique. Si cela était clair pour l'Ecole Stucturaliste de PRAGUE depuis ses premiers travaux, l'Ecole Distributionnaliste américaine (Z.S. HARRIS, 1951) était persuadée de

.../...

la justesse de l'économie de cette différence. Ils ont donné le départ à toute une série de réalisations de synthèse entreprises à partir d'éléments de parole de petite dimension et comprenant la formulation de règles de raccordement pour produire de la parole continue.(KELLY et GERSTMAN (1961) sont les premiers à avoir implanté un système de synthèse par règles sur calculateur).

L'avantage essentiel de cette méthode tient à l'économie réalisée dans le stockage des éléments de parole (la langue française dispose d'une trentaine de réalisations phonémiques), mais en contrepartie, la méthode impose une assez grande complexité au niveau de l'assemblage des "phonèmes", de par la nécessité de calculer les transitions entre les sons adjacents.

(*) .."The acoustic cues for successive phonemes are intermixed in the sound stream to such an extent that definable segments of sounds do not correspond to segments at phoneme level".

2ÈME P A R T I E

LES

OPTIONS

CHAPITRE I

LE VOCODEUR A CANAUX

Nous savons qu'à l'origine, le vocodeur a été mis au point (Dudley, 1939) pour permettre une compression du signal de parole afin de pouvoir transmettre simultanément plusieurs communications sur une même ligne téléphonique.

1/ Le spectre des signaux de parole s'étend pratiquement de 60 Hz à 12 000 Hz soit environ 8 octaves, mais l'expérience montre que la transmission de la seule bande de fréquences 300-3400 Hz est suffisante pour conserver à la voix une intelligibilité convenable et même la plupart des caractéristiques individuelles.

2/ D'autre part, les analyses spectrographiques ont permis de mettre en évidence la multiplicité des indices acoustiques de la parole ainsi que la faible vitesse de variation du spectre à cause des caractéristiques mécaniques des organes articulatoires. Ces propriétés statistiques du signal de la parole ajoutées à la capacité temporelle de perception discriminante de l'auditeur ont fait envisager la possibilité de n'extraire du signal de parole que des images spectrales moyennées sur un temps d'intégration inférieur à 25 ms.

La connaissance des problèmes à résoudre, les progrès technologiques qui ont permis d'utiliser des composants électroniques de plus en plus performants et les possibilités de codage numérique ont permis d'apporter des améliorations progressives à la qualité de la parole produite.

C'est le vocodeur à canaux réalisé au Département ETA du CNET que nous avons utilisé et que nous allons décrire maintenant (FERRIEU et PERSON, 1968; ZURCHER, 1973; ZURCHER et al, 1975).

.../...

Le vocodeur constitue donc un système de transmission de la parole à faible débit ; il comprend deux parties : une partie analyse du signal et une partie synthèse. Les deux parties sont ici reliées entre elles par un calculateur type T 1600 qui permet d'obtenir le résultat de l'analyse et d'opérer des modifications sur les données ; la partie analyse peut aussi être bouclée directement sur la partie synthèse, on obtient alors la reconstitution immédiate du signal de parole sous sa forme simplifiée.

Le principe du vocodeur consiste au départ à opérer une classification entre les sons :

* Les sons produits par une excitation périodique ; cela correspond, au niveau de la production, à la vibration des cordes vocales : le débit d'écoulement de l'air envoyé par les poumons est modulé par la vibration des cordes vocales.

- acoustiquement, ces sons présentent un spectre de raies à des fréquences harmoniques qui sont des multiples entiers de la fréquence fondamentale. L'enveloppe spectrale présente des maxima d'intensité qui proviennent du renforcement de certaines fréquences par les cavités de l'appareil vocal : on définit des zones de formants.

- sur le plan de la perception, ils correspondent à des sons voisés.

≠ Les sons produits par une source de bruit (bruits de friction ou d'occlusion).

- acoustiquement, ils présentent un spectre continu dont l'énergie provient d'un écoulement d'air turbulent le long du conduit vocal.

- phonétiquement, ils correspondent aux sons non voisés.

L'inconvénient de ce principe de détection tient au fait qu'il effectue un choix binaire : source d'excitation périodique ou source de bruit, ce qui conduit à une classification simplifiée dans laquelle ne rentrent pas les sons pour lesquels une source de bruit se superpose à la source d'excitation vocale : pour les constrictives voisées en particulier (v, z, ʒ), seul le voisement est détecté.

STRUCTURE DU VOCODEUR (fig. 9)

1/ - L'analyseur du vocodeur :

.../....

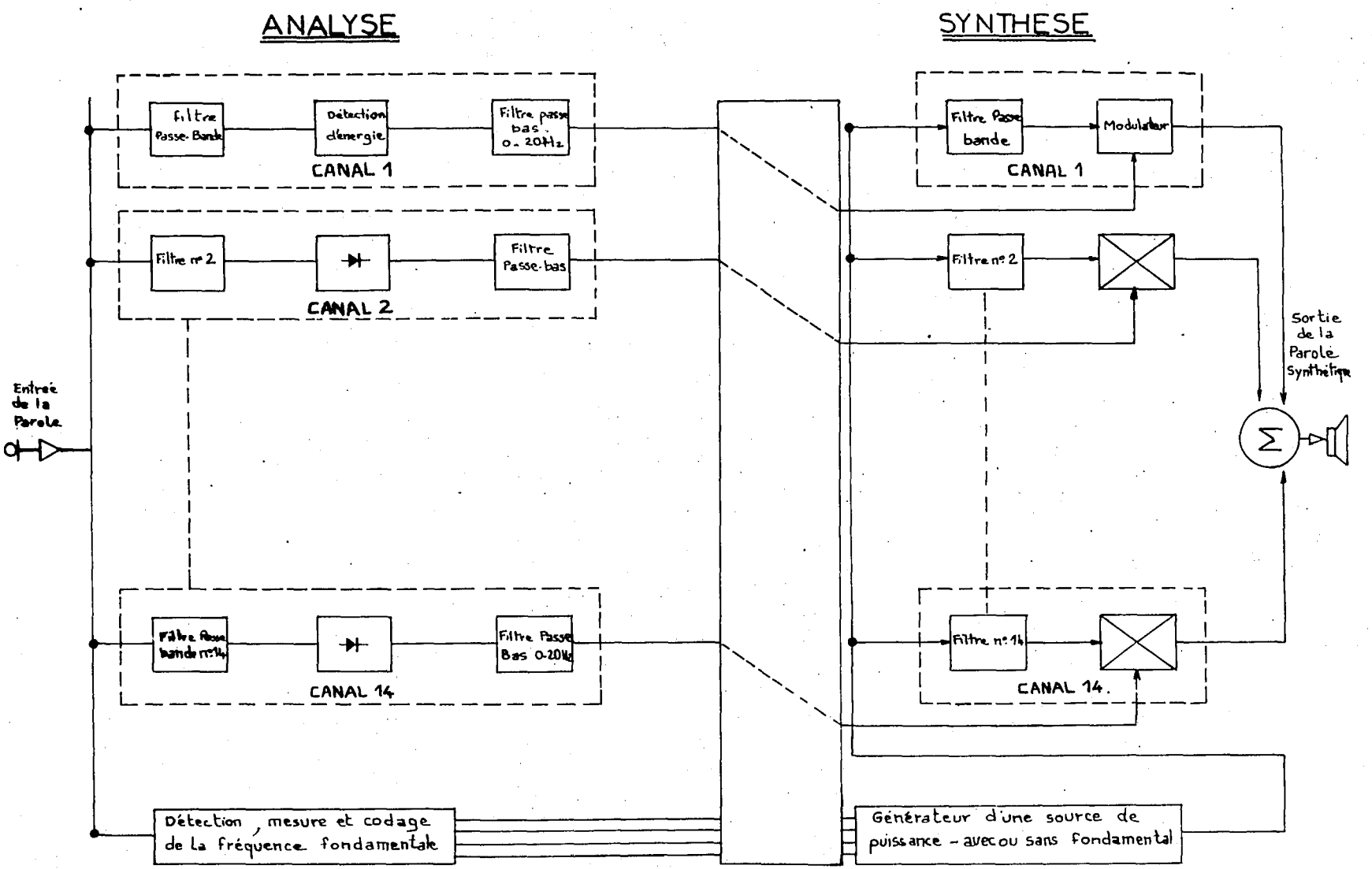


FIG.9 – SCHEMA DE PRINCIPE D'UN VOCODEUR –

Il permet d'extraire d'un signal complexe de parole deux types d'information dont on pense qu'elles suffisent à caractériser et à différencier les sons :

- information sur le spectre instantané du signal, c'est-à-dire sur la répartition de l'énergie dans l'échelle fréquentielle à un instant donné.

- information sur l'origine de la source d'énergie, c'est-à dire y-a-t-il ou non vibration des cordes vocales et si oui, quelle est la valeur de la fréquence fondamentale .

a/ Information sur le spectre instantané :

Le vocodeur que nous avons utilisé comporte quatorze filtres qui s'étendent de 250 Hz à 4300 Hz. Les bandes passantes de ces filtres sont contigües, et la largeur de chaque bande a été définie de façon à être la plus sélective possible :

1e	filtre	250 - 450 Hz
2e	filtre	450 - 650 Hz
3e	"	650 - 850 Hz
4e	"	850 - 1050 Hz
5e	"	1050 - 1300 Hz
6e	"	1300 - 1600 Hz
7e	"	1600 - 1900 Hz
8e	"	1900 - 2200 Hz
9e	"	2200 - 2500 Hz
10e	"	2500 - 2800 Hz
11e	"	2800 - 3100 Hz
12e	"	3100 - 3500 Hz
13e	"	3500 - 3900 Hz
14e	"	3900 - 4300 Hz

Cependant des études sont menées actuellement au Département ETA pour trouver une répartition plus rigoureuse des filtres, en particulier pour réduire davantage la largeur des filtres basses fréquences qui pour l'instant couvrent 200 Hz.

Les filtres se recoupent à - 6 dB ; après détection de l'énergie à la sortie de chaque filtre, un filtre passe-bas (0-18 Hz) est chargé d'éliminer les variations très rapides du niveau d'énergie.

Chacun des filtres recueille à chaque instant d'échantillonnage l'énergie contenue dans sa bande de fréquences. La fréquence d'échantillonnage a été fixée à 75 Hz, ce qui permet de connaître l'évolution temporelle de l'enveloppe spectrale toutes les 13,3 ms. Nous justifierons plus loin le choix de cette vitesse d'échantillonnage qui peut sembler à priori élevée.

.../...

Ces informations sont codées numériquement et logarithmiquement, le codage dure environ 60 µs ; l'énergie globale dans chaque canal est codée à 16 niveaux ; le pas de quantification est de 4 décibels, ce qui donne une dynamique de 60 db entre les deux niveaux extrêmes.

b/ Information sur la source :

Il ne s'agit plus de s'intéresser à la répartition de l'énergie mais d'analyser la structure fine du spectre. Cette analyse permet de différencier les sons voisés (spectre de raies) des sons non voisés (spectre de bandes continu) et de donner la valeur de la fréquence fondamentale des sons voisés.

* Dans une première réalisation (FERRIEU & PERSON,1973), un circuit de clamping détectait les crêtes d'amplitude maximum (fig.10).

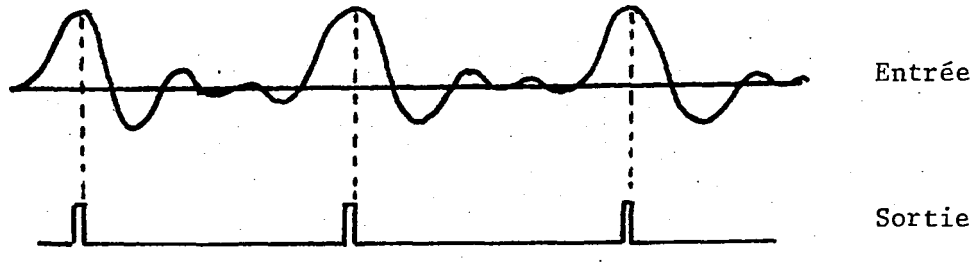


FIG : 10-Signaux d'entrée/sortie du circuit de "clamping".

Un algorithme câblé cherchait une récurrence parmi les crêtes détectées. Si une récurrence était trouvée, l'algorithme prenait la décision "son voisé" et donnait la période de la fréquence fondamentale ainsi mise en évidence. Ce procédé a bien fonctionné mais il présentait néanmoins les inconvénients suivants :

. le circuit de clamping utilisé ne permettait pas toujours de mettre en évidence les crêtes significatives de la périodicité (surtout dans les zones non stationnaires).

. une erreur sur la localisation de ces crêtes entraînait aussi une erreur sur la décision de voisement ce qui est sans doute plus grave.

* Dans une seconde réalisation (ZURCHER et al,1975), les fonctions de décision de voisement et de mesure du fondamental ont été séparées.

.../...

La décision de voisement est prise en fonction de la répartition spectrale de l'énergie et fonctionne de manière très sûre.

La mesure du fondamental est toujours faite par mise en évidence des crêtes significatives de la périodicité. Mais l'utilisation de circuits de prétraitement appropriés facilite cette opération (fig.11).

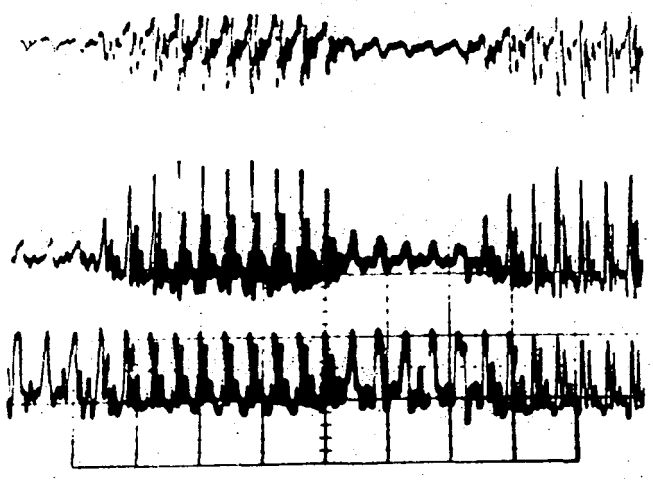


FIG : 11 - Prétraitement pour la détection des crêtes significatives (d'après ZURCHER et al, 1975)

Le signal, après un filtrage passe-bas (F_c = 400 Hz, 12 db/octave) subit une compression des amplitudes de ses enveloppes de crêtes positives et négatives. Un circuit spécialisé permet alors d'amplifier et de détecter les crêtes dont l'amplitude vient de croître brusquement alors que les amplitudes des crêtes précédentes allaient en décroissant : ce sont les crêtes significatives de la périodicité du signal.

Un algorithme câblé traite le train de crêtes ainsi sélectionnées et en déduit la période du fondamental quand elle existe.

* Puis la mesure de l'intervalle de temps qui sépare deux crêtes d'amplitude maximum donne la valeur de la période.

Ensuite, l'ensemble des informations relatives au spectre et à la source est codée sous forme numérique.

- la période du fondamental (pitch) est quantifiée sur 256 niveaux (soit 8 éléments binaires), chaque niveau correspond à une durée de 64 µs. Si le son est non voisé, cette période alors inexistante est codée Ø .

.../...

La cadence d'échantillonnage est la même que celle du spectre d'amplitude du signal : elle a lieu toutes les 13,3 ms. On considère que cette fréquence d'échantillonnage réalise un bon compromis entre l'élimination de la redondance du signal et la volonté d'obtenir une bonne définition des sons.

L'ensemble des informations numériques relatives au spectre d'amplitude et à la structure fine du spectre à chaque instant d'échantillonnage constitue ce que l'on appelle un "échantillon vocodeur" (fig.12)

Au total, les informations numériques nécessaires pour coder un échantillon vocodeur sont de 64 éléments binaires qui se répartissent comme suit :

- concernant le spectre : 14 canaux x 4 eb par canal = 56 eb
- concernant le pitch (période codée) : 8 eb.

le débit correspondant est de 64 eb x 75 échantillons/seconde = 4 800 eb/seconde

Une fois le codage effectué, les informations numériques sont transmises et stockées dans la mémoire du calculateur si l'on désire une reconstitution différée du signal de parole.

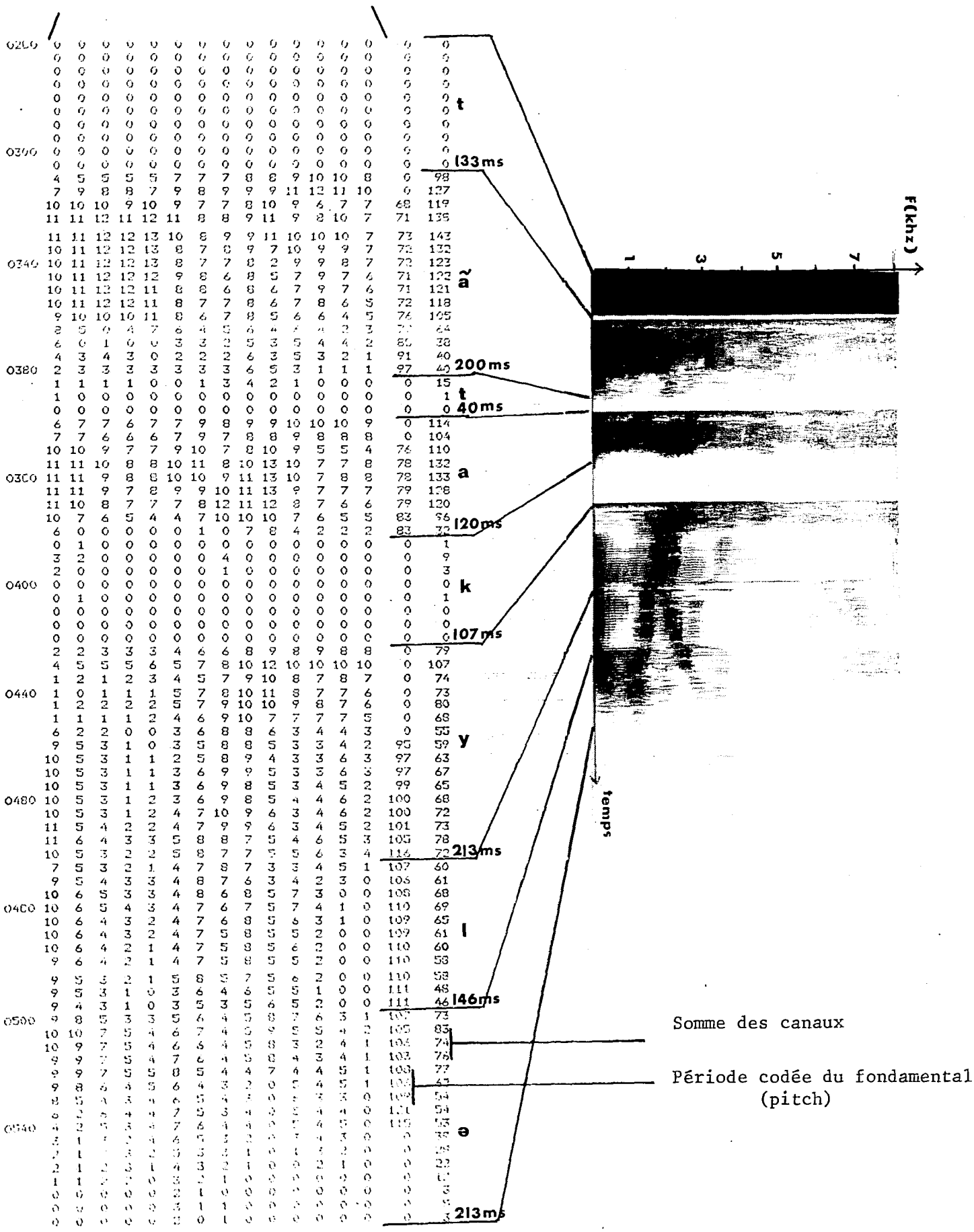
2/ Le synthétiseur du vocodeur : Restitution du signal simplifié. Il s'agit de reconstituer un signal de parole qui approche au mieux le signal original. La structure du synthétiseur est parallèle à celle de l'analyseur :

- une source de puissance (la fonction d'excitation) reconstitue l'information relative à l'excitation : restitution du spectre de raies (sons voisés) ou restitution du spectre de bruit (sons non voisés),
- un banc de 14 filtres dont les bandes de fréquences correspondent à celles de l'analyse.

Le dispositif d'excitation recrée ou non une excitation vocale (présence ou absence de fondamental) qui présente une densité spectrale moyenne d'énergie constante et qui attaque les 14 filtres du synthétiseur ; le signal d'excitation est modulé en amplitude proportionnellement aux niveaux d'énergie détectés par l'analyseur avant d'attaquer le filtre et on effectue l'addition des résultats obtenus dans chacun des canaux. En réalité, nous avons effectivement stocké les informations numériques des éléments de parole dont nous avions besoin en utilisant les valeurs délivrées par les 14 canaux de l'analyseur afin d'avoir une bonne définition spectrale des sons - en particulier des constrictives sourdes[f, s, ʃ], dont l'énergie dans les hautes fréquences est significative, mais

.../...

FIG. 12-Représentation numérique et analogique du mot /tentacule/.



CODAGE VOCODEUR

nous avons délaissé les canaux 13 et 14 du synthétiseur puisque à terme, la parole de synthèse est destinée à la transmission par voie téléphonique, et que la largeur des 12 premiers canaux recouvre la bande passante téléphonique.

De la même façon, nous avons analysé et stocké les données de parole à 4 800 eb /s afin de connaître les variations très rapides et souvent importantes de la fréquence fondamentale, mais à la synthèse, nous avons utilisé un codage numérique à plus faible débit : 2 240 eb/s qui se répartissent comme suit :

en ce qui concerne le spectre : 12 canaux x 4 eb = 48 eb

en ce qui concerne le pitch : 8 eb.

soit 56 eb x 40 échantillons par seconde = 2 240 eb/s.

En définitive, plusieurs arguments justifient notre choix d'un vocodeur à canaux comme matériel de synthèse :

* la bonne qualité de la parole de synthèse obtenue si l'appareil fonctionne à travers un poste téléphonique (300 - 3400 Hz)

* Ces appareils ne nécessitent pas une analyse préalable : les vocodeurs à canaux permettent une restitution de la parole en temps réel, alors que les synthétiseurs à formants nécessitent une analyse préalable pour déterminer l'emplacement et les valeurs des zones formantiques.

* Le principe de détection de la période du fondamental a été sans cesse amélioré (ZURCHER et al, 1975) et l'on peut dire qu'actuellement la mesure de la période est excellente : elle nous a fourni les données indispensables pour l'observation des évolutions très rapides de la mélodie.

* Enfin, et surtout le vocodeur à canaux offre un grand intérêt pour l'étude des faits prosodiques puisqu'il est possible de faire varier sur chaque échantillon, et indépendamment, ce qui relève de la source et ce qui concerne le spectre d'amplitude. On peut également modifier le rythme.

D'ailleurs, le fait de pouvoir effectuer des modifications soit sur l'enveloppe spectrale soit sur la structure fine du spectre de façons totalement indépendantes pose des problèmes dont nous aurons l'occasion de reparler plus loin.

CHAPITRE II

LA SYNTHESE PAR DIPHONES

Nous avons vu que si l'on veut réaliser de la parole synthétique, il se pose au départ un certain nombre de problèmes :

- Les méthodes de stockage de la parole sans compression, si elles permettent d'obtenir une grande qualité, présentent cependant l'inconvénient d'être limitées en vocabulaire pour des raisons de coût et d'encombrement.
- Les méthodes de codage alliées au stockage numérique de la parole ne pallient que partiellement ces inconvénients : en effet si l'on veut avoir la possibilité de composer n'importe quel message, il demeure hors de question de stocker - sous quelque forme que ce soit - l'ensemble des mots qui composent le vocabulaire d'une langue. C'est pourquoi on en est arrivé à concevoir des méthodes utilisant des éléments beaucoup plus petits : c'est dans cette optique qu'a été développée la synthèse par règles ; mais nous avons laissé entrevoir la complexité des règles que nécessite sa réalisation.

Il existe cependant un moyen terme qui repose d'une part sur le principe d'un codage et d'un stockage numérique de la parole, et d'autre part sur l'utilisation de segments de parole "microscopiques". L'avantage essentiel de cette méthode tient au fait qu'elle rend inutiles les règles de composition indispensables dans la synthèse par règles : l'assemblage des unités de parole en message est réalisée par simple juxtaposition des unités.

Ce sont les résultats des recherches entreprises sur la structure acoustique de la parole en particulier par les Laboratoires HASKINS pendant les années 1950 qui ont conduit à la conception de cette procèdure :

.../...

- La parole n'est pas constituée d'éléments discrets facilement segmentables en unités fonctionnelles, mais se présente sous la forme d'un continuum sonore évolutif.

- Les rapides changements dans les fréquences formantiques des voyelles au contact des consonnes (les transitions) ne constituent pas un phénomène accessoire mais l'un des indices essentiels pour la perception des syllabes.

Devant la multiplicité et la complexité des règles de composition des sons entre eux, des chercheurs ont mis au point une méthode astucieuse permettant d'éluder cette dificulté.

L'analyse spectrale permet d'observer dans la réalisation des séquences Consonne-Voyelle-Consonne-Voyelle... des moments de relative stabilité spectrale qui se situent dans la partie centrale des réalisations vocaliques et consonantiques (occlusives mises à part : ces réalisations étant essentiellement dynamiques), et au contraire des moments de grande instabilité dans le passage d'une réalisation phonétique à une autre.

La solution consiste à isoler et à choisir comme élément de parole pour la synthèse, une séquence allant de partie stable à partie stable et qui comprend le mouvement de passage d'une unité phonétique à la suivante (fig.13)

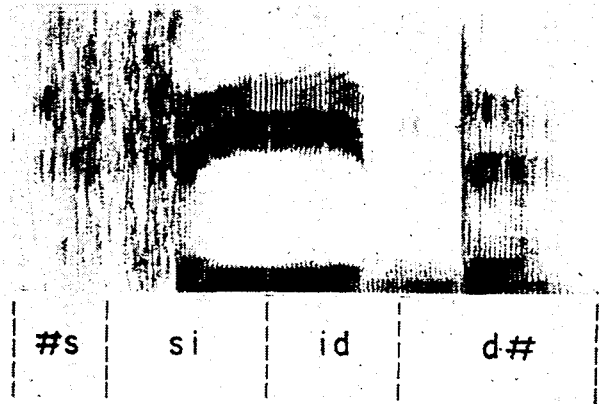


FIG. 13- Segmentation de mot [seed] en diphones -(d'après PETERSON et al, 1958)

.../...

On parle alors de synthèse par diphonèmes, ou diphones, dyads, phonatomes. Nous utiliserons pour notre partle terme anglo-américain de diphone, qui met d'avantage l'accent sur le niveau acoustique, préférant ne pas entretenir l'ambiguité liée à la notion de phonème de par son inexistence sur le plan acoustique.

ON APPELLE DIPHONE LE SEGMENT QUI S'ETEND DE LA ZONE STABLE D'UNE REALISATION PHONETIQUE A LA ZONE STABLE DE LA REALISATION SUIVANTE ET QUI PROTEGE EN SON CENTRE TOUTE LA ZONE DE TRANSITION.

La synthèse est alors réalisée par simple concaténation de ces éléments, le raccordement se faisant au niveau de deux parties spectralement identiques.

On conçoit que cette méthode soit séduisante, mais on comprend également qu'en contrepartie de sa simplicité, elle nécessite un nombre d'éléments de parole bien plus élevé que pour les méthodes de synthèse par règles : en effet, si l'on veut pouvoir composer n'importe quel message, il faut disposer de toutes les combinaisons deux à deux, c'est-à-dire que pour la synthèse d'une langue comme le français, qui compte quelque 33 réalisations phonémiques, il faut envisager de pouvoir disposer de $(33)^2$ combinaisons.

* Les premières tentatives de synthèse utilisant les diphones comme élément de parole datent de 1956 et ont été proposées par deux chercheurs allemands : KÜPFMULLER et WARNS.

Les diphones leur ont semblé être le moyen le plus astucieux pour que l'auditeur n'ait pas la sensation d'une discontinuité dans le passage entre sons adjacents.

La synthèse est réalisée par simple juxtaposition des éléments de parole; bien que la langue allemande dispose de quelque 40 réalisations phonémiques simples, il suffit cependant sur les 1 600 combinaisons possibles de disposer de 1 000 diphones pour pouvoir composer un quelconque message.

Effectivement, l'auteur cite l'exemple du mot "MUSTER" ;

.../...

celui-ci est réalisé en synthèse à l'aide de quatre diphones seulement : [MU] - [US] - [T ə] - [ə R] , tous les éléments ne présentant pas de transitions phonétiques sont éliminés du dictionnaire, ce qui permet de le réduire notablement : beaucoup de diphones composés de deux portions de consonnes cumulées ne sont pas pris en compte, c'est pourquoi le diphone [ST] n'apparaît pas.

Cependant, il semble que le nombre important de diphones indispensables auxquels on doit avoir accès en un temps minimum ait exigé pour l'époque un trop grand effort technologique ; c'est pourquoi la méthode a été abandonnée bien que, semble-t-il, l'intelligibilité de la parole obtenue ait été satisfaisante, même si la non-prise en compte des paramètres prosodiques aboutissait à une synthèse à la tonalité monotone.

* Deux années plus tard, une équipe américaine, PETERSON, WANG et SIVERTSEN (1958) reprend le principe de la synthèse à partir de "segments qui contiennent les parties de deux phones avec leur influence mutuelle au milieu du segment, et le début et la fin de la position phonétiquement la plus stable de chaque phone" ("the segments are characterized as containing parts of two phones with their mutual influence in the middle of the segment, and beginning and ending at the phonetically most stable position of each phone." - (voir fig. 13).

Ils insistent à leur tour sur la difficulté de segmenter le continuum sonore en une succession d'unités discrètes, et sur le problème posé par une juxtaposition sans discontinuités de ces segments.

Parce qu'il paraît évident aux auteurs que la prise en compte des seuls phénomènes articulatoires ne suffit pas à caractériser un segment de parole, ils envisagent de stocker autant de diphones qu'il y a de possibilités prosodiques susceptibles d'être produites lors de l'assemblage en séquences :

.../...

Certaines combinaisons d'unités sont éliminées ; en particulier les diphones [voyelle-voyelle], et une grande partie des diphones constitués de deux éléments consonantiques. Malgré tout, si un message comporte ce type d'occurrence - soit on insère un silence (noté #) entre eux : par exemple le mot "naïve" sera composé des diphones suivants :

#n - na - #i - iv - v#

- soit on passe directement au diphone suivant :

"Edna" → ɛ- ɛd - nə - ə#

On considère que le diphone [dn] est inutile.

Si l'on évalue pour l'idiolecte américain le nombre de réalisations phonémiques à 43, il reste quand même 1 418 possibilité de combinaisons en diphones. Mais là n'est pas l'essentiel de la procédure. En effet, la simple juxtaposition de ces segments ne permet pas d'obtenir une parole de bonne qualité : "on ne peut réduire aucun langage à de simples séquences unidimensionnelles d'éléments phonémiques". Pour cette raison, une solution est proposée qui consiste à imposer pour la plupart des diphones différentes versions prosodiques : certains diphones connaissent jusqu'à 9 configurations intonatives. On en arrive ainsi tout naturellement au total de 8 500 diphones (Tableau 1).

.../...

2e position / 1e position	Voyelle	glide	nasale	fricative	plosive	silence
Voyelle	0	60 x 9	45 x 9	135 x 9	165 x 9	15 x 9
Glide	60 x 9	0	0	0	0	0
Nasale	45 x 9	0	9 x 3	27 x 3	0	3 x 3
Fricative	135 x 9	0	27 x 3	81 x 3	99 x 3	9 x 3
Plosive	165 x 9	0	0	99 x 3	0	11 x 3
Silence	15 x 9	0	3 x 3	3 x 3	0	1 x 1

(y) x (x) y = nombre de combinaisons par catégorie.
x = nombre des patrons intonatifs par catégorie.

TABLEAU 1 - NOMBRE ET REPARTITION DES DIPHONES.

Il est évident qu'une telle démarche représente une somme de travail gigantesque et suppose un dictionnaire de données d'une taille impressionnante. C'est pourquoi les auteurs n'ont testé qu'un petit nombre de phrases comportant quelques dizaines de diphones, avant d'abandonner assez rapidement l'entreprise.

On leur doit cependant d'avoir mis l'accent sur l'importance des paramètres prosodiques et sur leur insertion dans un programme de synthèse. On leur doit également d'avoir noté les difficultés méthodologiques inhérentes à la synthèse par diphones : ils soulignent que si certaines catégories d'unités phonémiques ne posent aucun problème particulier de concaténation - par exemple les constrictives à cause du bruit qui est déjà présent pendant leur réalisation - il n'en va pas de même en ce qui concerne les voyelles et les glides. En effet, ce type de réalisation nécessite pour une synthèse par simple juxtaposition que l'on sache maîtriser parfaitement - sous peine de discontinuités sensibles à l'audition - :

- l'alignement de la structure formantique,
- la correspondance dans la disposition des harmoniques,
- l'égalité de l'énergie globale,

quand on passe d'un diphone au suivant lors de l'assemblage en message.

* ESTES et al. (1964) reprennent l'idée d'une synthèse à partir d'un nombre limité de courts segments de parole - les diphones - mais utilisent une bibliothèque de données aux dimensions plus raisonnables (800 diphones). Ils tirent de l'expérience de PETERSON et al. un certain nombre d'enseignements : ils constatent l'échec des méthodes fondées sur la juxtaposition de diphones extraits de la parole naturelle qui semble provenir des discontinuités de l'excitation vocale et de l'enveloppe spectrale.

Et c'est bien parce que les diphones sont issus de <u>la parole naturelle</u> qu'il y a impossibilité à assurer un continuum acoustique entre eux. C'est pourquoi, plutôt que d'utiliser des segments liés par les conditions de production naturelle, ils préfèrent <u>générer</u> un

.../...

dictionnaire de données, à partir de signaux de contrôle synthétiques qu'il sera assez facile de modifier commodément grâce aux commandes d'un synthétiseur paramétrique.

Les données sont "fabriquées" de façon à réunir tous les indices acoustiques nécessaires à leur identification, et on leur impose une image acoustique schématisée qui autorise un enchaînement harmonieux à leurs frontières : par exemple, les positions des trois premiers formants des voyelles sont fixées une fois pour toutes et leurs valeurs sont constantes ; le bruit de friction dans les réalisations constrictives est également contrôlé, etc...

On comprend effectivement que la maîtrise de ces paramètres acoustiques permet de restituer de la parole véritablement continue, mais les auteurs signalent à leur tour qu'il serait nécessaire d'insérer dans les programmes de synthèse les paramètres prosodiques (F_o , durée, intensité).

* DIXON et MAXEY (1968) posent à nouveau le problème essentiel relatif à la méthode de synthèse par diphones : est-ce qu'un nombre fini de segments synthétiques peut remplir les rôles multi-allophoniques qui sont les leurs dans la parole continue ?

Leur démarche est intéressante, leurs programmes de synthèse incluent la génération des phénomènes prosodiques : dans un premier temps, le message est écrit sous un code phonétique assez complexe qui attribue deux caractères pour chaque son. A mesure que l'on transcrit le message, sont insérés d'une part des chiffres qui précèdent le nom des diphones et d'autre part des chiffres à l'intérieur des diphones (fig.14).

Les premiers spécifient le nombre d'échantillons qu'on peut ajouter au diphone pour l'allonger (l'allongement est assuré par répétition du premier échantillon) ; les seconds font référence à la durée du diphone tel qu'il est stocké. L'allongement des éléments consonantiques et vocaliques est fixé de façon assez empirique à partir des observations de l'analyse et des tests de perception.

.../...

	XXAA	(90-100)		AA15IX	(100-125)		IXMX	(125-130)	3	MXSX	(130- 90)	I'm
5	SXAA	(90- 80)	3	AARX	(80- 95)		RXEE	(95-110)		EEXX	(110-115)	sorry,
						25	XXJX	(90-105)		JXUU	(105-115)	you've
5	UUVX	(115-130)	3	VXRX	(130-110)		RXEE	(110- 90)		EETX	(90- 75)	reached
6	CHTX	(90- 90)		XXDH	(90- 95)	1	DHIX	(95-105)		IXSX	(105-125)	this
5	SXAW	(90- 90)	5	AWFX	(90- 90)	2	FXIX	(100-105)		IXSX	(105-110)	office
5	SXPX	(110- 90)	5	BXAA	(90- 90)		AA10IX	(90- 90)		IXMX	(90- 90)	by
3	MXIX	(90- 85)		IXSX	(85- 90)	5	SXTX	(90-110)	4	DXEH	(110-110)	mis-
	EH15IX	(110- 75)		IXKX	(75- 65)	7	KXQX	(65- 65)				take,
			75	PXLXEE	(100-135)	10	EEZX	(135-135)	6	ZXKX	(135-100)	please
8	KXNX2	(100-100)	3	NXSX	(100-100)	6	SXUH	(100-110)	1	UHLX	(110-110)	consult
	LXTX	(110-110)	6	XXJX	(100-105)		JXER	(105-100)		ERDX	(100- 95)	your
4	DXER	(85- 85)		ERRX	(85- 85)	3	RXEH	(85-115)	1	EHKX	(115-115)	direct-
10	TXR3EE	(75- 75)		EEXX	(75-105)							ory,
10	XXEH	(95- 95)		EHNX	(95- 95)	3	NXDX	(95- 95)	6	DXAA	(110-105)	and
	AA15IX	(105-100)		IXLX	(100-100)	2	LXUH	(100- 90)		UHGX	(90- 90)	dial
2	GXEH	(90-100)	4	EHNX	(100- 80)	3	NXXX	(80- 65)				again.

FIG. 14 - Code phonétique et prosodique pour la synthèse d'une phrase.

Quant à l'intonation, elle est réalisée grâce à l'insertion de marques (chiffres entre parenthèses) qui indiquent pour chaque diphone les valeurs de la fréquence fondamentale de départ et d'arrivée, les valeurs intermédiaires étant déterminées par interpolation. Un programme réalise l'ensemble des opérations et transmet les signaux de contrôle à un synthétiseur à formants : l'IBM 7094.

La méthode est astucieuse mais il se dégage de l'ensemble une impression de grande complexité ; cependant les auteurs concluent en répondant à la question posée au départ : une bibliothèque d'environ 1 000 diphones suffit pour générer de la parole continue intelligible ; mais si l'on veut donner à cette parole synthétique les caractéristiques qui font l'agrément d'une langue, alors une bibliothèque de bien plus grande dimension est indispensable.

* A la même époque, une équipe japonaise - SAÏTO et HASHIMOTO (1968) - applique au japonais une méthode de synthèse par concaténation de séquences du type voyelle-consonne-voyelle pour éviter un assemblage délicat au niveau des réalisations consonantiques. Nous savons peu de choses sur la méthode utilisée si ce n'est qu'elle s'apparente légèrement à la synthèse par règles par un procédé de

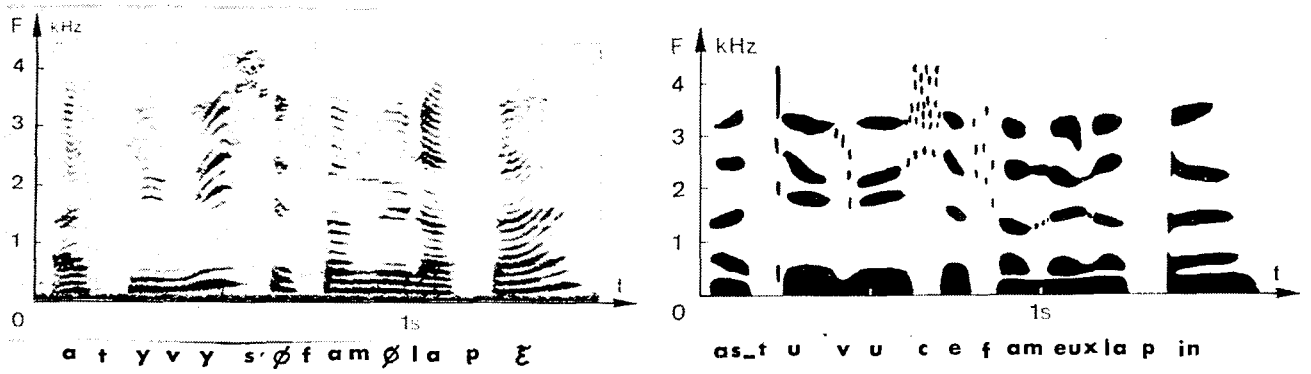
.../...

modification des frontières de séquences pour assurer un continuum acoustique entre séquences successives. Le programme de synthèse prévoit également la génération de la prosodie.

* Mais sans doute, dans le domaine de la synthèse par diphones, la réalisation la plus complète et la plus élaborée revient au Groupe de Recherche sur la Parole du Laboratoire d'Acoustique de la Faculté des Sciences de PARIS, qui a montré la validité de la méthode appliquée au français (LEIPP et al,1968 ; TEIL,1969 ; QUINIO et TEIL,1970 ; LIENARD,1972 ; TEIL et al,1974 ; CHOPPY et al,1975).

Dans un premier temps, les auteurs ont établi un catalogue de diphones (on parle ici de phonatomes ou de digrammes phonétiques) à partir de la voix chuchotée qui leur semblait la plus apte pour ce genre d'élaboration. Si l'ensemble des combinaisons de sons deux à deux était prise en compte, on disposerait de 900 diphones. Mais les auteurs remarquent qu'un certain nombre d'entre eux sont reversibles (AE et EA, RL et LR...) et que l'on peut en définitive se contenter de quelque 400 éléments.

Ces diphones sont réalisés à partir d'une analyse préalable : le sonagraphe délivre une image spectrale des éléments de parole qui permet d'identifier les diphones et de les redessiner d'une façon schématisée pour ne conserver que le "squelette sémantique"(fig. 15).



Sonagramme "Squelette sémantique"

FIG. 15.

.../...

L'Icophone à commande optique permet d'effectuer la vérification des hypothèses sur la structure acoustique de la parole et l'évaluation perceptive des diphones schématisés ; sa conception se situe dans la ligne du Relecteur de Sonagramme (COOPER et al.,1952) : les diphones dessinés à l'encre opaque sur un support transparent défilent devant un alignement de 44 cellules photoélectriques qui commandent 44 générateurs sinusoïdaux entre 100 et 4 400 Hz ; le signal ainsi produit possède un fondamental fixe à 100 Hz. Par la suite les auteurs ont présenté un synthétiseur à commande de fondamental pour générer la prosodie (CHOPPY et al,1975), mais qui entraîne aussi le déplacement des formants.

Les diphones subissent une succession de tests de perception avant d'être définitivement acceptés et mémorisés sur cartes perforées . Ils sont ensuite stocké sur disque d'un calculateur selon un classement qui tient compte de leur fréquence d'occurrence dans la parole (LIENARD, 1966).

Dans une seconde étape, il s'agit de reconstituer de la parole à partir d'un texte écrit sous une forme orthographique classique ; un programme réalise la traduction orthographique-phonétique et appelle les diphones correspondants (grâce au classement effectué, plus le diphone est fréquent, plus son temps d'accès est faible). En même temps, un traitement assez sommaire du rythme est prévu ; on double la durée de la voyelle qui précède une ponctuation, et des pauses sont associées aux signes de ponctuation pour améliorer l'intelligibilité des phrases. Ensuite les diphones sont édités sur l'Icophone à commande numérique conçu comme périphérique d'ordinateur, et le message est énoncé sous forme vocale presque instantanément.

Le résultat (95 % d'intelligibilité) prouve que la forme acoustique choisie est suffisante pour véhiculer l'information.

Cependant la voix obtenue surprend l'auditeur non averti par son timbre particulièrement métallique dû à l'excitation et

.../...

par l'absence de modulation intonative. Mais peut-être l'habitude fera-t-elle que - comme le font remarquer les auteurs - "la parole synthétique, toujours semblable à elle-même, deviendra dans un proche avenir plus intelligible que la parole humaine, de même qu'un texte imprimé est plus lisible qu'un texte manuscrit car les caractères en sont normalisés" ?

* Signalons enfin qu'une équipe de GRENOBLE (Ecole Nationale Supérieure d'Electronique et de Radio électricité ; Institut de Phonétique - BOE et al., 1974) a entrepris la réalisation d'un système de synthèse avec des diphones et grâce à un synthétiseur à formants. Mais le dictionnaire des éléments de parole n'est pas encore achevé ; son élaboration présente une réelle difficulté dans la mesure où le synthétiseur à formants travaille avec peu de paramètres (faible redondance) et qu'une erreur sur le choix des indices entraîne irrémédiablement une mauvaise intelligibilité. Un programme permet de commander la fréquence d'excitation : chaque diphone est stocké dans le dictionnaire avec des valeurs de F_0 qui sont représentatives de leurs caractéristiques intrinsèques. A celles-ci, se superpose à chaque instant un contour mélodique de phrase élaborée à l'aide de marques qui signalent dans la séquence des maxima et des minima intonatifs. D'autres marques sont prévues pour permettre une augmentation ou une diminution de durée des diphones.

CHAPITRE III

L'ETUDE DES FAITS PROSODIQUES.

Quand nous avons décidé de réaliser un système de synthèse de la parole à partir de diphones et par l'intermédiaire d'un synthétiseur à canaux, notre but était surtout d'arriver, à partir de l'analyse instrumentale d'un corpus parlé, à dégager un nombre limité de patrons prosodiques pertinents susceptibles d'être introduits en synthèse par une programmation relativement simple et de s'approcher ainsi un peu mieux des réalisations naturelles de la parole.

Mais cette option nous a conduit à nous poser un certain nombre de questions :

- Sur le plan théorique d'abord, quelle est la définition des faits prosodiques et quels en sont les paramètres pertinents?; il s'agit également de préciser quelle est la place réservée aux phénomènes prosodiques dans le système de la langue, et de déterminer ce qu'ils apportent à la parole de synthèse c'est-à-dire quelles en sont les fonctions.

- Sur le plan pratique, il faut se demander s'il est possible de dégager de l'analyse instrumentale des invariants prosodiques, et si oui quelles sont les difficultés procédurales que va induire leur prise en compte.

I - La définition de leur <u>contenu</u> semble ne pas poser de problèmes théoriques ; les auteurs s'entendent généralement pour inclure dans la prosodie l'étude des faits accentuels, du rythme (structuration de l'énoncé par les pauses) et de l'intonation dont on attribue la réalisation et la perception à l'étroite interaction de trois variables : l'évolution de l'intensité et de la fréquence laryngienne en fonction du temps ; superposées à ces évolutions se situent les caractéristiques intrinsèques, mais auxquelles on n'attribue pas de valeur linguistique puisqu'elles sont liées inévitablement - c'est-à-dire sans intervention possible du locuteur - aux caractéristiques du mode de phonation.

Mais les controverses sont nombreuses qui opposent les linguistes sur le point de savoir , au niveau de la réalisation et au

.../...

niveau acoustique, dans quel cadre se situent les phénomènes prosodiques. Parmi les tendances qui se dégagent, il est intéressant de noter l'attitude de l'Ecole de Prague.

* Pour MARTINET (1962 à 1967), la langue est un outil de communication doublement articulé :

La première articulation se compose d'unités linguistiques comprenant à la fois un signifié et un signifiant : le signifié, c'est le concept, c'est l'unité de sens minimale ; le signifiant représente quant à lui la forme, la représentation vocale du signifié, il est décomposable en unités distinctives successives : les phonèmes, qui constituent la seconde articulation du langage. Pour MARTINET, "tous les faits de parole qui n'entrent pas dans le cadre phonématique c'est-à-dire ceux qui échappent d'une façon ou d'une autre à la double articulation" ne sont pas véritablement linguistiques et c'est dans cette catégorie qu'il fait entrer tous les faits prosodiques.

Pour justifier le caractère marginal qu'il attribue à la prosodie, MARTINET prend l'exemple de l'énoncé "Il pleut ?". On peut, dit-il, effectivement décomposer ce signe en un signifié (c'est la question posée) et un signifiant (une montée de la fréquence fondamentale sur la dernière syllabe), mais pour lui, ce signifiant ne comporte pas de seconde articulation, c'est-à-dire n'est pas décomposable en une succession de phonèmes. On a affaire à des "faits supra segmentaux" qui ne sont pas segmentables en unités discrètes définies comme les unités "dont la valeur linuistique n'est affectée en rien par des variations de détail déterminées par le contexte ou diverses circonstances" : la deuxième articulation constitue un système d'opposition binaire alors que la hauteur mélodique par exemple peut, par des variations entre deux niveaux distinctifs, se trouver à des niveaux fréquentiels intermédiaires qui apportent des modifications de sens successives à l'énoncé ; il y a une continuité, un glissement, et non pas un passage discret d'une courbe à une autre. Les paramètres prosodiques introduiraient donc des différences significatives comme les monèmes, mais non distinctives comme les phonèmes ; on parle de "morphèmes d'intonation".

* D'autres linguistes pensent au contraire que l'on peut décomposer les phénomènes prosodiques en unités discrètes. MALMBERG(1962) est l'un de ceux qui ont nourri la controverse ; il conteste le refus d'introduire les traits prosodiques dans la seconde articulation du

.../...

langage et estime que ceux-ci "peuvent et doivent sans doute être réduits à un nombre limité d'invariants types (c'est-à-dire en prosodèmes, ou en phonèmes supra segmentaux) exactement comme les sons parlés ordinaires (segmentaux) sont phonémisés en unités discrètes" : l'analyse expérimentale lui a permis de mettre en évidence deux types de prosodèmes : un prosodème de continuité (2-3) et un prosodème de finalité (3-2).

* Mais l'adversaire le plus acharné de la théorie de MARTINET est sans doute FAURE (1967) pour qui la structure prosodique compose "un système tout aussi économique et tout aussi rigoureux que le système phonématique". Il estime que les phénomènes prosodiques peuvent être analysés et segmentés en unités discrètes distinctives, que l'on peut établir un nombre fini de seuils d'intonation aussi pertinents que les oppositions phonématiques et aussi parfaitement localisables dans le continuum sonore. Les faits prosodiques forment un ensemble structuré qui permet de parler d'un véritable "code de la prosodie".

- L'autre aspect qui divise les linguistes est relatif aux fonctions de la prosodie :

- Pour MARTINET et l'Ecole de PRAGUE, la prosodie appartient au domaine de la parole (conçue comme l'ensemble des possibilités concrètes découlant de l'organisation de la langue), et non à celui de la langue. De par cette conception, les fonctions attribuées à la prosodie sont fort peu linguistiques, et dans ce cas seulement significatives comme dans "Il pleut ?", et beaucoup plus ectosémantiques, expressives.

- Pour d'autres linguistes, les seules fonctions exercées par la prosodie sont d'ordre grammatical (HALLIDAY M.A.K. de 1961 à 1967).

- Mais la plupart des auteurs ont pris position de façon plus nuancée :

. BAILLY (1941) puis DANES (1959) : pour eux la prosodie structure les unités linguistiques en énoncés, établit une hiérarchie entre elles, et organise les rapports entre le thème et le propos :"l'intonation intègre , délimite ou segmente".

. PIKE (1945), WELLS (1947), TRAGER et SMITH (1951) : les manifestations prosodiques assurent le décodage de la structure syntaxique du message. FAURE (de 1962 à 1970) se situe dans la même voie : la prosodie réalise un découpage de l'énoncé qui permet de hiérarchiser

.../...

les unités syntaxiques qui le composent ; il y a une correspondance entre les manifestations prosodiques et l'agencement des syntagmes.

. Pour LIEBERMAN (1965) au contraire, la fonction de décodage de la structure syntaxique de l'énoncé n'intervient de façon indispensable que dans les cas où les unités syntaxiques présentent une ambiguïté. On connait à ce propos l'exemple cité par MALMBERG :"la belle ferme le voile".

Pour notre part, nous avons également envisagé l'étude des faits prosodiques comme l'étude de leur structure. Structure étant entendue comme "un ensemble d'un tout formé de phénomènes solidaires tel que chacun dépend des autres et ne peut être ce qu'il est, que dans sa relation avec eux" : cela signifie que l'on envisage les phénomènes prosodiques comme constitués d'un ensemble dont toutes les parties sont agencées et liées par un réseau étroit de dépendances. De la même façon que les éléments segmentaux correspondent à des possibilités et à des limites articulatoires déterminées , il faut bien voir que la fréquence fondamentale est aussi régie par le mode de fonctionnement du larynx et des paramètres physiologiques de commande qui sont les siens : pression subglottique, tension des cordes vocales, dynamique et rapidité de variation, inspiration et expiration phonatoires.

Le propre de la structure est d'être inconsciente : les schémas prosodiques que nous utilisons nous sont imposés ; ils ne résultent ni d'un hasard ni d'un libre choix ; ou plus exactement ils sont constitués à la fois d'un aspect obligatoire, imposé, et d'un aspect qui laisse place à une certaine marge de liberté. Les règles obligatoires permettent à l'intonation de remplir sa fonction proprement linguistique, les variations tolérées autour de ces règles permettent la manifestation du "MOI" dans la parole, c'est l'aspect ectosémantique, expressif de la parole, c'est la part du sujet ; on peut concevoir les manifestations acoustiques de la prosodie comme le résultat d'un hasard ou d'un choix ; nous pensons pourtant qu'au delà de l'illusion de la liberté, il y a un ordre, une nécessité, celui du code de communication.

Il est certain que dans le code linguistique - code de communication - l'intonation joue un rôle moins important que les unités segmentales : les possibilités de combinaison des schémas intonatifs sont très grandes et laissent la marge à d'importantes variantes individuelles, ce qui entraîne des difficultés d'analyse certaines auxquelles

.../...

se sont heurtés la plupart des chercheurs.

Il faut se souvenir qu'avant d'être écrite, la langue s'est manifestée d'abord sous sa forme orale. Quand on est passé à l'écrit, on a cherché à rendre visible pour l'oeil l'information véhiculée vers l'oreille par la prosodie : les espaces entre les mots permettent de les repérer immédiatement, les signes de ponctuation (virgule, point virgule, point, point d'exclamation, point d'interrogation...) aident à découper le message en ses unités constituantes. Nous pensons que là réside la fonction linguistique de la prosodie : elle réalise un décodage, elle aide à déchiffrer le message par une organisation temporelle de la chaîne parlée et la mise en évidence d'unités de groupes de sens. On s'aperçoit que les procédés d'aide au décodage du message écrit sont peu nombreux par rapport aux manifestations de la prosodie ; on peut penser que le caractère essentiellement fugitif de la parole (scripta manent, verba volant) est l'un des aspects qui rend nécessaire la multiplicité des indices acoustiques facilitant l'appréhension du sens dans le continuum sonore.

Dans la phase d'apprentissage de la lecture par l'enfant, toute son atttention est accaparée par un travail de décodage du message ; par contre, si le même message, au lieu d'être lu par l'enfant, lui est transmis oralement, le cerveau déchargé de la tâche d'agencement cohérent des unités linguistiques entre elles, peut se concentrer sur un travail de compréhension sémantique, et le message sera compris par l'enfant.

On peut penser que la prosodie n'est pas indispensable au décodage, la synthèse de la parole réalisée par les Laboratoires HASKINS avec une fréquence fondamentale fixe à 120 Hz est intelligible, mais il est connu que le fait d'augmenter la durée de certaines syllabes et d'insérer des pauses augmentent l'intelligibilité. On peut penser en définitive :

- d'une part que la prosodie est indispensable à établir le sens véritable de l'énoncé dans le cas où la structure syntaxique est ambigüe.

- d'autre part que, effectivement, on peut toujours arriver à retrouver la sémantique d'un énoncé émis sans prosodie, comme on peut retrouver le sens d'un message écrit sans espace entre les mots et sans ponctuation. Mais le but de la parole, c'est la transmission, la communication, la compréhension mutuelle, la préhension du sens : si l'on

.../...

parle, c'est pour se faire comprendre spontanément ; cette nécessité d'appréhender vite le sens rend évidente la nécessité de l'existence de la prosodie. Elle permet de décharger le cerveau d'une tâche de premier niveau qui consiste à déchiffrer des signifiants (apprentissage de la lecture par l'enfant) pour ne lui laisser qu'une tâche de second niveau c'est-à-dire d'appréhension du signifié.

On peut concevoir que si les phénomènes prosodiques jouent dans la parole naturelle un rôle de meilleure intelligibilité de l'énoncé par l'indication de sa structure linguistique, dans la parole de synthèse, il est absolument nécessaire de les introduire comme il est nécessaire de prendre en compte tous les paramètres acoustiques pertinents.

Pour ce faire, arriver à trouver l'ordre caché dans les manifestations prosodiques et leur réduction à un nombre fini d'invariants représentait une entreprise séduisante. Mais la mise en oeuvre des phénomènes prosodiques (et plus particulièrement des phénomènes intonatifs liés à la modulation laryngienne) dans un système de synthèse pose des problèmes particulièrement délicats du fait du matériel choisi : le synthétiseur à canaux, et des éléments de parole retenus : les diphones.

II - Pour réaliser l'insertion des faits prosodiques dans le système de synthèse par diphones, plusieurs voies nous étaient ouvertes :

* On aurait pu, par exemple, comme pour la synthèse réalisée aux Laboratoires HASKINS, négliger la variable fréquence laryngienne en la figeant une fois pour toute à une valeur fréquentielle fixe et unique, et ne s'intéresser qu'aux deux autres paramètres plus faciles à contrôler : la durée et l'intensité. Mais dans ce cas, la voix de synthèse obtenue possède un timbre particulièrement métallique qui surprend et qui demande un effort d'adaptation de la part de l'auditeur. Or, notre but était d'essayer d'approcher au mieux dans la parole de synthèse les caractéristiques de la production naturelle de la parole, ou tout au moins de faire en sorte que ses attributs permettent à un auditeur naïf de l'appréhender immédiatement et sans effort. C'est pourquoi nous n'avons pas adopté ce choix.

* On sait que le vocodeur à canaux délivre deux sortes d'informations : une information relative au spectre et une information concernant la source (source périodique ou source de bruit). Cela signifie qu'à chaque période d'échantillonnage, on a connaissance de

.../...

l'évolution des mouvements rapides du régime de vibration des cordes vocales, par l'intermédiaire d'une valeur représentative de la période du fondamental $\frac{1}{F_o}$. Par conséquent chaque diphone contenu dans le dictionnaire et stocké en mémoire possède l'information concernant l'évolution de $\frac{1}{F_o}$ pendant sa réalisation.

On aurait donc pu envisager de garder ces valeurs de période et les utiliser pour la synthèse. Mais les mots qui ont servi à la constitution du dictionnaire de diphones ont été enregistré volontairement (nous nous en expliquerons plus loin) sur un ton neutre. Le résultat de l'assemblage est une voix tout à fait monotone. Pour pallier à ce défaut, on aurait pu concevoir d'utiliser le traitement prosodique élaboré par PONCIN (1970) pour l'Unité à Réponse Vocale, c'est-à-dire garder toutes les valeurs de période délivrées par l'analyseur et introduire seulement pour certaines syllabes situées en des points importants de l'énoncé une modification en prévoyant un programme qui permette d'insérer par exemple un schéma mélodique montant ou descendant pour certaines syllabes finales de mot.

Mais ce qui est possible pour la synthèse par mots devient inutilisable dans la synthèse par diphones parce que cette dernière impose en plus la reconstitution de la continuité mélodique à l'intérieur du mot.

Or, quand on procède à l'enregistrement des mots à partir desquels les diphones sont extraits, il est impossible au locuteur de maîtriser parfaitement la hauteur de sa voix pour qu'elle demeure à un niveau stable pendant toute la durée de l'enregistrement. Cela signifie que l'information relative à la structure fine du spectre va présenter une différence sensible d'un diphone à un autre : malgré les précautions prises au moment de l'enregistrement, il n'est pas rare de noter un écart d'une dizaine de Hertz dans les valeurs de F_o de deux diphones différents.

Si, malgré ces différences, l'on effectue une juxtaposition, il en résulte une discontinuité dans le spectre de raies à la frontière de deux diphones, c'est-à-dire justement à un moment considéré comme relativement stable dans la réalisation ; cette discontinuité, tout en provoquant un effet désagréable à l'écoute, entraîne une dégradation de l'intelligibilité (fig. 16).

.../...

FIG. 16 - Assemblage de la voyelle [y] à partir du mot [structure] prononcé à deux niveaux de hauteur différents.

Il aurait été possible alors, avant même de stocker les diphones en mémoire, d'effectuer un alignement progressif de toutes les valeurs de période afin qu'il n'y ait pas de rupture à la concaténation. Mais cela équivaut à conserver des informations mélodiques qui ne sont même plus révélatrices des caractéristiques intrinsèques des réalisations phonémiques, et un traitement intonatif qui est limité à l'insertion de quelques schémas mélodiques sur les syllabes finales de mot est en fait insuffisant pour corriger la monotonie du message et donner un aspect naturel à la parole de synthèse. Ces raisons nous ont fait éliminer également cette possibilité de traitement de la prosodie.

* Il restait une autre alternative : se débarrasser de toutes les informations relatives à F_0 , et effectuer la reconstitution de l'intonation à partir des observations tirées de l'analyse instrumentale d'un corpus. C'est l'option que nous avons choisie.

Cependant les résultats que l'on peut attendre de ce choix ont certaines limites que nous allons préciser :
- mais il faut signaler d'abord que cette méthode permet, nous le verrons, d'éliminer les risques de discontinuités de F_0 au centre d'une réalisation vocalique puisque le but du traitement est justement de permettre un continuum mélodique durant la réalisation des voyelles par la continuité dans la structure de raies.

.../...

- la seule discontinuité est celle qui peut provenir de l'amplitude globale entre deux diphones : les mots servant de base aux diphones ont pu être enregistrés à des niveaux d'énergie relativement différents. Il ne s'agit que d'un inconvénient mineur, des corrections manuelles ont permis d'harmoniser les niveaux aux frontières de diphones .

- mais il existe un problème bien plus aigu :

Les diphones sont stockés avec pour chacun une enveloppe spectrale standard et une structure harmonique qui lui correspond. En modifiant cette structure harmonique par l'introduction des valeurs correspondant à la mélodie désirée, on introduit un bouleversement artificiel : d'une part la structure harmonique va être en disharmonie avec l'enveloppe spectrale, d'autre part on ne tient pas compte de l'interaction source laryngienne - conduit vocal.

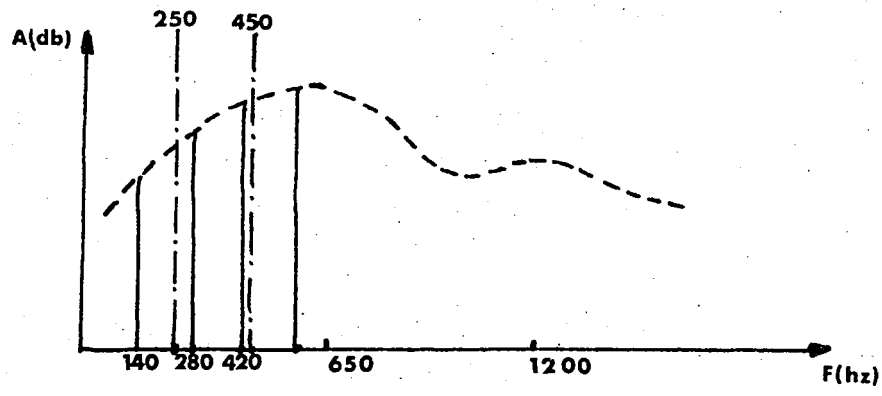
Pour tenter d'expliquer plus clairement ces difficultés, nous allons revenir sur le processus de fonctionnement du vocodeur.

1/ Quand le signal de parole est codé à 4 800 eb/seconde, cela signifie que le vocodeur délivre chaque 13,3 ms à la fois une information sur l'enveloppe spectrale et une information sur les sources. Chacune des deux informations est délivrée de façon indépendante ; cela signifie qu'à la synthèse, on peut manipuler très facilement par exemple les données qui ne concernent que l'excitation : on sépare deux types d'information qui, dans le signal de parole, sont relativement interdépendantes. Il faut dans le même temps reconnaître que le fait d'avoir un accès indépendant à l'une ou l'autre information fait du vocodeur à canaux un outil d'une souplesse particulièrement bien adaptée pour les études d'analyse et de synthèse de la parole dans le domaine de la prosodie.

Supposons maintenant la réalisation [a] prononcée par exemple avec un fondamental de 140 Hz, ce qui donne les valeurs 280, 420, 560, 700 Hz... pour les harmoniques et considérons les valeurs suivantes pour les trois premiers formants : F_1 = 650 Hz ; F_2 = 1 200 Hz ; F_3 = 2 200 Hz.

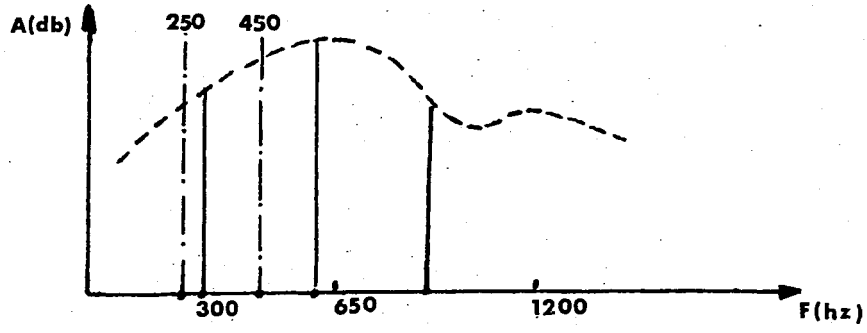
On a alors approximativement la représentation spectrale suivante :

.../...



Dans ce cas, avec un fondamental de 140 Hz, le second et le troisième harmonique sont regroupés dans la bande passante du premier filtre de l'analyseur (250-450 Hz) ; leur "moyenne" donne une certaine valeur d'énergie.

Supposons ensuite que cette configuration [a] ait été stockée avec un fondamental de 140 Hz et que pour les besoins de la synthèse, on porte la valeur du fondamental à 300 Hz.



Dans ce cas, c'est le premier harmonique et lui seul qui se trouve inclu dans la bande passante du premier filtre du synthétiseur, et c'est l'énergie (de valeur constante) correspondant à ce seul harmonique qui est modulée en amplitude proportionnellement au niveau détecté dans ce même filtre lors de l'analyse du signal avec un fondamental de 140 Hz.

Le résultat de cette modulation, c'est que l'amplitude du signal du premier filtre dans le cas de F_o = 300 Hz (une seule raie harmonique), est affaiblie par rapport à l'amplitude du signal dans le cas de F_o=140Hz.

Le spectre que l'on obtient en synthèse en modifiant seulement la structure harmonique par élévation de la voix (diminution du nombre de raies) est différent du spectre que l'on aurait obtenu en analysant un signal de parole produit avec un fondamental de 300 Hz. En effet, dans la parole, la fréquence fondamentale est corrélée - de par la pression subglottique - à l'intensité (mais dissociée dans le

.../...

chant) ; cela signifie que si F_0 augmente, l'intensité augmente aussi. Par conséquent, l'analyse d'un signal émis avec un fondamental de 300 Hz aurait donné sensiblement la même enveloppe spectrale que l'analyse d'un signal émis avec un fondamental de 140 Hz : l'énergie liée à l'harmonique d'un fondamental haut donne la même sensation d'intensité sonore que l'énergie liée à deux harmoniques d'un fondamental bas.

A la synthèse au contraire, les modifications introduites dans la structure harmonique par doublement du fondamental n'entraînent pas pour autant une augmentation de l'intensité comme dans les réalisations naturelles mais au contraire un affaiblissement - même par rapport à des réalisations naturelles produites avec un fondamental bas.

Les effets de cette transformation de l'enveloppe spectrale qu'un traitement de l'intonation impose ici sont en partie formulés dans les jugements portés par des auditeurs sur la qualité de la parole obtenue avec ce système de synthèse (voir Ve partie).

2/ La seconde difficulté du traitement de la fréquence fondamentale en synthèse est liée aux conséquences du couplage source laryngienne - conduit vocal.

En effet, chacun des diphones contenu dans le dictionnaire est destiné à intervenir dans différents contextes intonatifs. Il peut par exemple apparaître dans la dernière syllabe d'un mot terminant une phrase interrogative ou bien en finale de phrase énonciative, et l'on se retrouve alors dans les cas énoncés ci-dessus : le locuteur féminin qui a été choisi pour l'enregistrement du vocabulaire de base de la synthèse a une tessiture comprise entre 134 et 356 Hz, c'est-à-dire que F_0 évolue dans 98 % du temps entre ces deux limites (ces valeurs correspondent aux valeurs moyennes relevées en particulier pour le français : BOE et al., 1975). Or, tous les mots ont été enregistrés avec un fondamental voisin de 180 Hz ; c'est donc une modification de une octave que l'on introduit pour la synthèse de l'interrogation. Or, on sait que si la fréquence fondamentale varie, le larynx varie également, modifiant les dimensions du pharynx et donc l'enveloppe spectrale qui correspond à la fonction de transfert du conduit vocal : PETERSON et BARNEY (1952) ont montré que chez un même locuteur prononçant plusieurs fois une même voyelle, les fréquences formantiques de cette voyelle varient. Puis ARNOLD (1957) signale que l'augmentation de F_0 entraîne l'élévation du larynx, et MEYER-EPPLER (1957) explique l'élévation des formants constatée dans

.../...

le cas d'une réalisation de voyelle à mélodie plus haute par l'élévation du larynx.

Mais il résulte d'autres travaux (FANT, 1959 ; CARRE, 1971) que, de par l'élévation du larynx liée à une augmentation de la fréquence de vibration des cordes vocales, il intervient une modification dans la forme et le volume du conduit vocal qui provoque également un changement dans les fréquences de résonance de ce conduit et par là même dans les fréquences des zones formantiques : ces études menées au moyen de la simulation électrique du conduit vocal montrent que l'élévation du larynx réduit la longueur du conduit vocal par diminution de la longueur de la cavité du pharynx ; un conduit vocal d'homme tend à devenir un conduit vocal de femme.

De la même façon, on peut penser que le conduit vocal d'une femme - c'est le cas en l'espèce - est réduit quand la fréquence de vibration des cordes vocales augmente, provoquant en même temps une élévation des fréquences des zones formantiques. Nous n'avons pas calculé l'amplitude du déplacement des zones formantiques quand le fondamental est modifié d'une octave, d'une part parce que tous les phénomènes ci-dessus énoncés s'interpénêtrent et ne nous sont pas accessibles directement, et d'autre part parce que même bien connues, ces difficultés ne sont pas pour l'instant résolvables avec le vocodeur à canaux.

Cependant des tests de perception ont montré que les défauts énoncés pour la parole de synthèse obtenue par modification du fondamental de près d'une octave - sans adaptation de l'enveloppe spectrale - disparaissent quand la parole est le produit de la juxtaposition de segments analysés au départ avec un fondamental élevé, c'est-à-dire quand l'interaction source laryngienne - conduit vocal existe naturellement. Nous pensons que l'amélioration de la qualité de synthèse passe par la résolution de ces problèmes : réaliser les interactions entre les phénomènes.

3ème PARTIE

ANALYSE INSTRUMENTALE

DES

FAITS PROSODIQUES

Le but de cette analyse est de dégager les composantes acoustiques qui, dans la parole continue, ont pour fonction d'assurer la réalisation des faits prosodiques et plus particulièrement des faits intonatifs, ainsi que le découpage du message en groupes syntaxiques pour aider au décodage sémantique de l'énoncé.

Par prosodie nous entendons phénomènes accentués, phénomènes qualifiés globalement de rythmique et phénomènes intonatifs, qui se manifestent par l'intermédiaire de trois paramètres : la durée, l'intensité et la fréquence fondamentale.

Pour notre part, nous nous sommes surtout intéressée à l'étude du découpage temporel de l'énoncé - durée segmentale d'une part, répartition et durée des pauses d'autre part - ainsi qu'aux évolutions temporelles de F_o . Nous avons eu quelques difficultés dans l'étude du paramètre d'intensité parce qu'avec le matériel utilisé - le vocodeur à canaux - il ne nous est pas accessible directement et que par voie de conséquence (la synthèse est symétrique de l'analyse) on ne peut pas modifier de façon rigoureuse en synthèse les données qui le concernent.

1 - LES PROBLEMES DE L'ANALYSE SUR LE PLAN PRATIQUE:

Pour obtenir une prosodie correcte à la synthèse, il nous a semblé nécessaire de passer par les conditions suivantes :

a/ - l'analyse d'un corpus :

Comme le souligne UMEDA (1976) le meilleur moyen pour élaborer les instructions à fournir à une machine - dans le domaine de la synthèse de la parole - est d'étudier la façon dont les gens parlent et de dériver des règles à partir de ces informations.

.../...

b/ choix pour ce corpus de la parole continue : l'ensemble du corpus représente environ 400 phrases.

c/ il s'agit d'un corpus lu.

Si nous avions suivi les conseils de DELATTRE (1966) : "Pour découvrir les caractéristiques d'intonation d'une langue, il faut les saisir sur le vif, dans l'énonciation vraie de l'improvisation, il faut éviter le texte lu et récité", nous aurions dû procéder à l'enregistrement de nombreuses heures de parole avant de pouvoir disposer de tout le matériel linguistique nécessaire ; un corpus préparé permet au contraire d'avoir toutes les phrases et rien que les phrases que l'on désire analyser.

Une autre raison justifie le choix d'un corpus lu ; elle tient à ce que l'application visée - l'implantation d'un centre de renseignements automatique - ne laisse que peu de place à l'improvisation et à l'hésitation. On demande des opérateurs un renseignement énoncé clairement, distinctement et sans hésitation; celui-ci résulte de sa lecture sur un annuaire ou sur une console de visualisation . C'est cette situation que nous voulons reproduire avec la parole synthétique.

Les seules indications de lecture données aux locuteurs concernaient le débit : nous leur avons demandé de lire les phrases - disposées dans un ordre aléatoire - en les articulant correctement, c'est-à-dire à un débit moyen (BOE et al, 1975 : 160 mots par minute ; c'est d'ailleurs ce qui correspond au débit préconisé par FAIRBANKS, 1960).

Les phrases étaient précédées d'un texte de trois minutes destiné à éliminer une éventuelle appréhension de l'enregistrement, et chaque phrase ne devait être lue qu'après une pause d'environ deux

.../...

secondes. Toutes les dix phrases, l'enregistrement était interrompu pendant une à deux minutes pour éviter l'effet de liste.

Evidemment, les enregistrements ont été effectués en chambre sourde pour éliminer au mieux les bruits de fond. Nous avons remarqué que la présence d'un opérateur gênait véritablement certains locuteurs qui n'arrivaient pas à garder leur naturel (en particulier pour la formulation des phrases interrogatives), aussi avons-nous laissé les locuteurs opérer seuls dans la chambre sourde.

Les enregistrements ont été effectués sur un magnétophone de type REVOX A 77 ; un microphone Beyer ; bandes magnétiques Scotch ; la vitesse d'enregistrement est de 19 cms/seconde.

2 - LE CHOIX DU LOCUTEUR:

Les études menées au département E.T.A. sur le vocodeur à canaux et sur la compression des données de parole (CARTIER et GRAILLOT, 1974; GRAILLOT,1974, 1975) ont fourni des enregistrements de parole pour plusieurs locuteurs et à différents niveaux de compression. Ces enregistrements d'analyse-synthèse ont été soumis à des tests d'intelligiblité.

Dans un premier temps, nous avons éliminé des quelques vingt locuteurs, ceux dont l'élocution paraissait trop rapide, et ceux dont l'accent manifestait par trop leurs origines géographiques. Il restait alors trois locuteurs âgés de 28 à 30 ans, deux hommes et une femme. Ces trois locuteurs ont enregistré le corpus. Dans une première phase, nous avons choisi l'un des deux locuteurs masculins qui nous a permis de constituer un dictionnaire de diphones, d'élaborer les premiers schémas intonatifs, et de réaliser les premiers tests de perception. Les résultats nous ont semblé dès le départ conduire à des limites peu satisfaisantes. Aussi avons-nous décidé d'opérer sur la nature de la voix des

.../...

tests systématiques ; nous avons composé avec la voix des trois locuteurs quelques phrases de synthèse par diphones et nous avons simplement demandé à une cinquantaine d'auditeurs pris au hasard de donner leur impression générale sur les voix qu'ils entendaient. C'est à la suite de ce test assez sommaire que nous avons finalement opté pour la voix du locuteur féminin - d'origine grenobloise mais sans accent particulier - qui avait reçu 85 % des suffrages non indifférents. Nous tenons à préciser que c'est uniquement aux vues de ce critère d'intelligiblité (tests/logatomes) et de qualité que cette voix féminine a été retenue. En aucun cas notre choix n'a été influencé par des considérations retenues par BURON (1968).*

Lorsqu'il s'agit de mener une étude théorique sur les faits prosodiques, nous reconnaissons avec LEON et MARTIN (1970) et de nombreux autres auteurs qu'il est plus judicieux de travailler sur un corpus restreint mais avec de nombreux locuteurs que le contraire : "Il est évident qu'il vaut mieux étudier de courts échantillons de parole provenant de dix informateurs différents plutôt qu'un seul échantillon dix fois plus important d'un seul locuteur ; on évite ainsi d'interprêter des variantes individuelles pour des patrons généraux du langage". Mais il faut bien voir que le but de notre travail était de mettre en oeuvre un système de synthèse opérationnel, ceci nous a amené à opter pour le choix suivant : un seul locuteur et le corpus le plus vaste possible.

Le locuteur féminin choisi étant en même temps l'un des auteurs du système, on est légitimement en droit d'émettre l'une des critiques les plus vives énoncées par LEON et MARTIN (1970) : "Bien des études ne reposent que sur le parler d'un seul informateur, souvent l'auteur lui-même. Cette hyper-conscience des phénomènes qu'il étudie est un danger réel".

---

*"Une voix de femme a été choisie, comme reflétant le mieux, la volonté des futurs clients".

.../...

A cette critique, nous pouvons répondre que l'élaboration du corpus a été conçue au départ de l'étude, et la totalité de l'enregistrement a été effectuée également à ce moment là, c'est-à-dire à un moment de totale méconnaissance des composantes que nous allions pouvoir dégager du corpus. De plus, l'analyse du corpus a montré une similitude complète entre les trois locuteurs notés dans un choix final pour ce qui concerne certaines manifestations de la prosodie : signe de la pente de F_0 , répartition des pauses, allongement caractéristique du dernier segment situé avant une pause.

3 - LE CORPUS :

Nous avons choisi d'étudier trois types de phrases : les phrases de type énonciatif, impératif, et interrogatif, excluant dans un premier temps les énoncés de type exclamatif dont on peut penser qu'ils n'entrent pas à court terme dans les attributions d'une "machine à parler", et parce que nous pensons avec BLOOMFIELD (1933) que les énoncés de type exclamatif ne sont que des variantes, soit de phrases énonciatives, soit de phrases impératives ou interrogatives (CHALARON, 1972).

Les phrases ont été construites de façon à posséder, d'une part, des mots au nombre de syllabes recouvrant toutes les possibilités de la langue française, d'autre part une composition syntaxique très variée ; nous parlons de "phrase" dans le sens où l'entend BLOOMFIELD (1970) : "Forme linguistique indépendante, qui n'est pas incluse dans une forme linguistique plus large, en vertu d'une construction grammaticale quelconque", et nous utiliserons les termes de groupe nominal (GN 1) pour les éléments qui appartiennent au syntagme nominal sujet (1), de syntagme ou groupe verbal (2), et de groupe nominal (GN 2) pour les éléments qui composent le syntagme complément(3).

"la sociologie rurale | a connu | un développement relativement tardif"
(1) (2) (3)

Nous ne donnerons pas la liste de toutes les phrases du corpus, nous indiquerons simplement les constructions syntaxiques que nous avons utilisées : .../...

1 - Les phrases de type énonciatif :

1-1 - les phrases simples :

syntagme nominal sujet + verbe copule.

(verbe être joignant l'attribut au sujet) + attribut.

Syntagme nominal sujet (1) + syntagme verbal (2) + syntagme complément (3)
ou Syntagme complément (3) + syntagme verbal (2) + syntagme nominal sujet (1)
ou Syntagme complément (3) + syntagme nominal sujet (1) + syntagme verbal (2)
ou syntagme nominal sujet (1) + syntagme verbal (2)

(1) il..
Jean ..
Le chat..
Le dinosaure..
La documentation ..
Le petit chat..
Le petit chaperon rouge
Le gentil petit animal de la vieille femme..
Le chat, la belette et le petit lapin..
Le chat caché dans le grenier ..
Les exercices de rééducation de la colonne vertébrale ..
Les exemples de synthèse par diphonèmes...
L'institut et Laboratoire de Sociologie de Strasbourg..
Les maîtres de conférence et professeurs..
Le tracé de l'organigramme d'une administration..

(2) sont..
mange..
dort..
voudrait dormir..
ne veut pas se lever..
désespère..
ferme..
a voulu travailler..
se réveillera ..
s'est aperçu que ..

.../...

ira sans doute chasser..
va chercher..
n'a jamais voulu ..
est gentil ..
auront faim ..
sont appelées à ..
se soumettra ..
sépare ..

.
.
.

(3) des souris.
des souris blanches.
des souris blanches et des rats énormes.
des poissons rouges.
avec appétit.
toutes les cinq minutes.
aux aléas du nombre de la population étudiante.
les niveaux de la réalité sociale, les contextes ou les manifestations de la vie sociale.
la connaissance scientifique de la connaissance vulgaire.
les conditions d'une observation participante avouée.

1-2- Les phrases complexes :

Elles comprennent à la fois des propositions qui se coordonnent les unes aux autres sur le même plan grammatical,

"Après le diner, je suis sortie dans le jardin, puis j'ai appris mes leçons."

et des propositions qui se subordonnent, c'est-à-dire entre lesquelles existent une hiérarchie grammaticale :

* subordonnées introduites par un pronom relatif, subordonnées conjonctionnelles introduites par une conjonction de subordination, subordonnées infinitives qui ont pour base un infinitif ayant son sujet propre, subordonnées participes.

.../...

Ces subordonnées peuvent être sujet, attribut, en apposition, complément d'objet, complément circonstanciel , complément de nom ou de pronom :

" qui veut la fin veut les moyens".
" cela va faire bientôt une semaine que je ne l'ai vu".
" les oiseaux chantent quand le soleil se lève".
" je me demande pourquoi il est mort".
" les astres,qui éclairent le ciel, illuminent le visage".
" chaque fois que les chasseurs arrivaient, les lapins se cachaient en vitesse".

.
.
.

Nous avons utilisé le même processus avec les autres types de phrases.

2 - Les phrases de type impératif :

Veuillez indiquer le code postal !
Indiquez le code postal !
Indiquez-moi le code postal !
Recommencez votre expérience !

.
.
.

Veuillez indiquer votre numéro !
.... votre numéro de sécurité sociale !
.... le numéro de votre correspondant !
.... le numéro de téléphone de votre correspondant !
.... votre numéro de Sécurité Sociale et votre numéro d'allocation !

.
.
.

.../...

Veuillez appeler le 73-83 à VINAY !
le 842-17-18 à PARIS !
le 27-09-94 à PERROS GUIREC !

.
.
.

Appelez l'opératrice, ensuite posez votre question !
Fais tes devoirs et apprends tes leçons avant d'aller jouer !

.
.
.

3 - Les phrases de type interrogatif:

Le corpus comprend les mêmes possibilités d'expansions dans la composition des syntagmes et ce dans les trois types de phrases interrogatives susceptibles d'être rencontrées :

3-1- Phrases interrogatives construites selon le même modèle syntaxique que les phrases énonciatives (question totale).

Les questions totales représentent la catégorie qui arrive de loin en tête dans les phrases interrogatives : 30 % de l'ensemble des questions (TERRY, 1970).

Vous voulez manger ?
Vous voulez prendre des cachets avant de partir ?
Vous pensez pouvoir partir demain ?
Les voisins connaissent le chemin pour aller chez les DUPONT ?

.
.
.

3-2- Phrases interrogatives avec inversion du sujet et du verbe .

Pouvez-vous me dire quel est le numéro de Monsieur DUPONT ?
Préférez-vous fumer la pipe ou les cigarettes ?
Savez-vous si les inscriptions ont commencé ?

.../...

Pourriez-vous m'indiquer ce que je dois faire ?

3-3- Phrases interrogatives introduites par un mot ou un groupe de mots interrogatifs .

Quand partez-vous ?
Dans quelle ville habitez-vous ?
Qui a frappé Paul ?
Que connaissez-vous de Rome ?
Est-ce que vous pourriez m'indiquer le numéro de Monsieur BERTRAND ?
Quels espoirs placez-vous dans le changement ?
Quelles sont les modifications que nous pourrions apporter ?

CHAPITRE I

LE DECOUPAGE TEMPOREL DE L'ENONCE.

I - LA DUREE SEGMENTALE :

Afin de savoir quelle durée fixer pour chaque diphone dans le dictionnaire et quels traitements apporter selon sa position dans le message, il nous était nécessaire :

- d'une part de connaître la durée intrinsèque de chaque segment vocalique et consonantique dans la parole continue en fonction de son entourage phonétique, correspondant en première approximation à la fréquence d'occurrence la plus élevée : distribution normale.

- d'autre part d'étudier cette durée segmentale dans différentes situations syntaxiques pour déterminer toutes les conditions de modification de durée et la longueur des segments à ces différents moments.

Cette étude a été menée à partir du corpus de phrases défini. Les résultats que nous donnons ne concernent qu'un seul locuteur et nous ne prétendons donc pas faire oeuvre théorique en ce domaine. Pour une étude approfondie des phénomènes de durée, on pourra se reporter en particulier aux travaux suivants : BOË(1973), HOUSE et FAIRBANKS (1953-1961), LEHISTE (1970), LEHISTE et PETERSON (1961), LINDBLOM et RAPP (1973), LINDBLOM (1974), SIGURD et LINDBLOM (1971), LYBERG et LINDBLOM (1975).

Cependant nos observations théoriques concernant en particulier la durée des segments en fin de mots situés avant une pause ont été confirmées par l'analyse de deux autres corpus identiques.

.../...

Le mesure de la durée n'a pas été effectuée comme dans la plupart des autres études sur des spectogrammes mais sur les listings d'échantillons vocodeur obtenus après analyse du corpus à 4 800 éléments binaires par seconde.

Ce choix nous a été dicté directement par le but visé : la réalisation d'un dictionnaire de diphones avec pour chacun d'eux une adresse de départ et de fin correspondant à des échantillons vocodeur ; de plus, la succession des échantillons vocodeur nous donne une vision directe de la durée.

La seule difficulté qui se pose, mais qui existe aussi avec les méthodes spectrographiques (UMEDA, 1975) concerne la segmentation de certains sons, en particulier les liquides [r, l] et les semi-voyelles [w, j, y] qui présentent des transitions peu rapides avec les voyelles qui les entourent (fig. 17)et non pas une discontinuité spectrale comme par exemple dans les réalisations des occlusives ou des constrictives.

	14	11	10	8	7	5	4	3	4	7	5	3	4	4	82	69	
	14	11	10	9	7	5	4	3	3	6	6	4	4	4	81	90	
140	13	11	11	10	8	5	4	3	2	5	6	4	5	4	83	91	o
	13	10	10	11	8	5	3	3	2	3	6	4	6	5	88	89	
	12	10	10	11	8	4	2	3	2	0	6	5	6	5	90	84	
	11	11	10	9	8	3	2	2	0	2	5	5	4	4	94	76	
	10	11	9	8	7	3	2	1	0	0	5	4	5	4	98	69	
	8	10	7	7	5	3	2	1	1	0	6	5	6	5	99	66	
	6	6	7	6	6	2	1	2	2	2	6	6	6	6	104	64	ɹ
	7	5	6	5	7	3	2	3	1	1	5	5	6	5	106	61	
80	8	6	7	6	7	4	2	3	0	1	4	4	6	5	99	63	
	8	8	7	7	7	5	2	3	1	3	6	5	6	4	98	72	
	10	10	7	6	5	5	3	2	2	3	7	6	6	5	97	77	
	11	11	9	7	4	7	6	6	5	6	8	5	6	5	100	96	
	11	11	9	6	2	5	8	7	6	8	9	7	6	5	100	100	
	11	12	8	5	2	3	8	7	7	8	10	7	5	5	102	98	ɛ
	11	11	7	4	2	2	5	8	8	8	10	7	6	6	102	95	
	11	10	7	4	2	2	4	7	9	8	10	7	5	5	102	91	
60	11	9	6	3	2	2	4	4	8	8	9	7	4	4	102	81	
	11	8	5	2	0	1	3	3	7	9	8	7	4	4	102	72	
	11	7	4	1	0	1	3	3	7	9	7	7	4	4	102	68	
	10	6	3	1	0	0	2	3	5	8	7	6	4	3	101	58	
	10	6	3	1	0	0	0	2	3	6	5	6	4	3	101	49	
	9	5	2	0	0	0	0	1	0	5	4	5	3	3	97	37	
	8	4	2	0	0	0	0	1	1	5	4	3	2	3	96	33	
	7	3	2	0	0	0	0	1	1	5	4	4	2	2	97	31	
00	6	2	0	0	0	0	0	1	1	5	3	5	3	2	93	28	
	4	1	0	0	0	0	0	1	1	4	4	6	4	3	88	28	

FIG - 17 -Réalisation de [r] dans /0 r ɛ/.

.../...

Outre cette difficulté matérielle, il faut bien voir qu'il n'existe pas de relation biunivoque entre la suite des éléments discrets d'une suite phonétique et un continuum sonore, la transcription phonétique se situant déjà à un premier niveau d'abstraction lié à la connaissance de la langue. Il y a une différence de niveau entre la réalisation acoustique et la transcription phonétique.

Nous avons donc étudié les consonnes et les voyelles dans différentes situations phonétiques et dans des positions variables à l'intérieur des groupes syntaxiques, qui pouvaient nous permettre de dégager les indices que nous souhaitions mettre à jour. C'est pourquoi nous avons préféré étudier les segments dans un corpus de phrases énoncées naturellement et non pas inclus dans des mots enchassés dans des phrases de contexte toujours identiques, en particulier les phrases fixes du type "say... instead" (KLATT, 1973), ou inclus dans de courts groupes de mots (BARNWELL, 1971), ou dans des logatomes (LINDBLOM ,1973).

De par le choix du corpus, nos résultats sont sensiblement différents de ceux obtenus par ces auteurs, et nos conclusions se rapprochent davantage de celles de UMEDA (1972, 1974, 1975, 1976) qui étudie systématiquement la durée des segments dans la parole continue.

Les contextes dans lesquels nous avons étudié la durée des réalisations vocaliques et consonantiques font intervenir différentes variables :

- l'influence de la longueur du mot sur la durée des sons .
- l'influence des pauses sur la durée des réalisations situées en fin de mot.

Nous considérerons donc :

1/ les segments qui ne précédent pas une pause :
   - durée des voyelles et des consonnes des mots monosyllabiques,

.../...

- durée des voyelles et des consonnes situées à l'intérieur et en initiale des mots plurisyllabiques.

2/ les segments situés en fin de mot avant une pause:

- durée des consonnes finales,
- durée des voyelles finales en syllabe ouverte (Consonne-Voyelle) ,
- durée des voyelles de syllabe finale fermée (Consonne-Voyelle-Consonne).

I-1- La durée des réalisations consonantiques :

On distingue entre consonne située dans un entourage essentiellement vocalique et groupement consonantique comprenant au moins deux consonnes successives.

I-1-1- Les consonnes isolées :

I-1-1-1- La durée des consonnes isolées dans les séquences Voyelle-Consonne-Voyelle des mots plurisyllabiques. (On exclut délibérément de cette catégorie les consonnes initiales et finales de mot).

Les moyennes effectuées sur les résultats sont conformes à la plupart des études menées en ce domaine : on constate de façon générale pour les consonnes sourdes une durée plus longue que celle de leurs équivalents sonores.

Afin de pouvoir effectuer une comparaison acceptable entre les occlusives sourdes et les occlusives voisées, nous avons volontairement exclu des premières toute la zone de bruit entre l'occlusive et la voyelle subséquente.: celle-ci reflète la durée de l'explosion et du VOT (Voice Onset Time ou Durée d'Etablissement du voisement - SERNICLAES et BETSTER, 1974 ; SERNICLAES, 1974 ; SERNICLAES, 1976)

.../...

qui varie selon la nature de la voyelle. C'est la raison pour laquelle nous n'avons gardé pour mesurer la durée des occlusives sourdes que la durée de la tenue.

Nous n'avons pas observé de différence de durée selon la longueur des mots, mais seulement une différence qui tient à la nature de la consonne.

Dans un ordre de durée décroissante, on trouve :

- les constrictives sourdes	[f , s, ʃ]	Moyenne 144 ms
- les constrictives voisées	[v, z , ʒ]	Moyenne 94 ms
- les nasales	[m , n]	Moyenne 80 ms
- les liquides	[r , l]	Moyenne 77 ms
- les occlusives sourdes	[p, t, k]	Moyenne 70 ms
- les occlusives voisées	[b, d, g]	Moyenne 60,3 ms
- les semi-consonnes	[w, j,]	Moyenne 53 ms

Ces résultats établissent un rapport de durée de 1 à 2,7 entre les deux catégories extrêmes.

On peut remarquer que ces résultats établis pour la parole continue diffèrent notablement de ceux qui ont été relevés pour des logatomes français (BOË, 1973) et confirment la nécessité de notre choix concernant la parole continue, les différences portant essentiellement sur la durée des nasales et des liquides.

I-1-1-2- Si l'on considère les consonnes <u>initiales</u> de mots mono- et plurisyllabiques, il faut signaler d'abordqu'on ne constate pas de différence selon que ces consonnes apparaissent après une pause ou directement après la fin d'un mot. Une réserve toutefois : les occlusives sourdes n'ont été étudiées que dans les contextes où un autre mot les précède ; il est évident qu'après une pause, il est impossible de connaître le début de la tenue.

.../...

De façon générale, on peut relever qu'en position initiale, une consonne est légèrement plus longue qu'en position intervocalique dans le cas des occlusives (sauf [p]), mais globalement plus courte dans le cas des constrictives. Deux exceptions cependant : la consonne [v] est plus longue en initiale de mot ; et c'est en initiale de mot monosyllabique que [f] présente la plus longue durée. On peut tenter d'avancer : les monosyllabiques qui contiennent ces consonnes ne sont que très rarement des mots de fonction - articles ou prépositions - comme le sont la plupart des monosyllabiques recensés dans ce corpus, et plus souvent des adjectifs, des verbes ou des substantifs à contenu sémantique qu'il est nécessaire de transmettre à l'auditeur.

En initiale de mot plurisyllabique, seule [v] parmi les constrictives est plus longue qu'en position intervocalique.

Pour une même position initiale, il est donc difficile de tirer une loi selon que le mot est mono-ou pluri syllabique :

- certaines consonnes (p, t, d, v, ʒ, s, n, l) sont plus longues en initiale de mot plurisyllabique,

- d'autres consonnes (k, b, f, ʃ , m) sont plus longues en initiale de mot monosyllabique.

Toutefois, ces différences ne portent que sur une dizaine de millisecondes (fig. 18) ; les liquides, les nasales, et les semi-consonnes présentent une grande stabilité de durée en position initiale ou intervocalique, leur moyenne se situe autour de 80 ms.

I-1-1-3 - La durée d'une réalisation consonantique en position finale d'un mot situé avant une pause.

Dans un premier temps, nous avons étudié toutes les consonnes finales de mot quelles que soient les variations du fondamental (F_0 montant ou F_0 descendant) mais en excluant la position précédant une pause : la durée de la consonne finale (dans les cas de syllabe finale

.../...

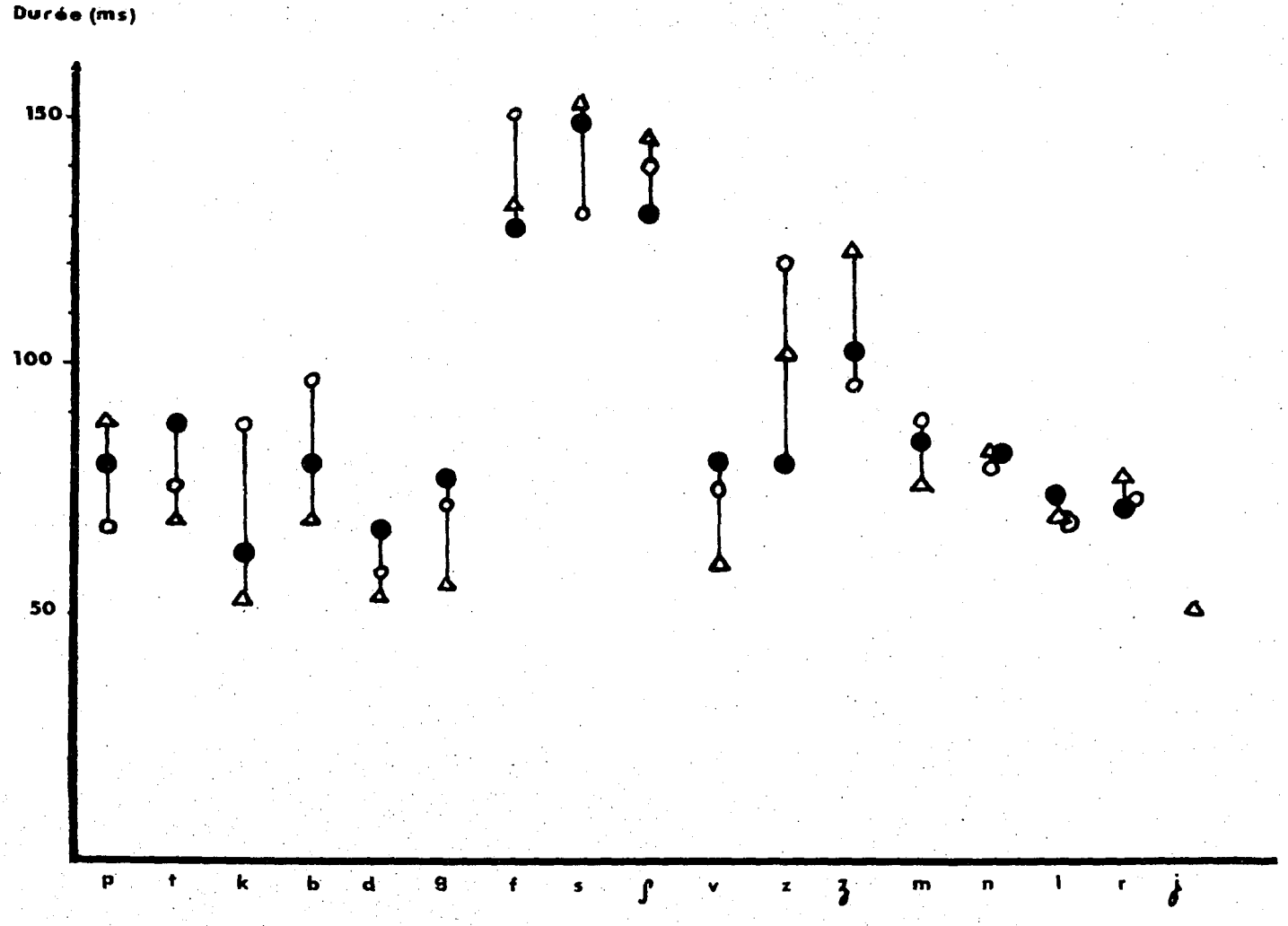


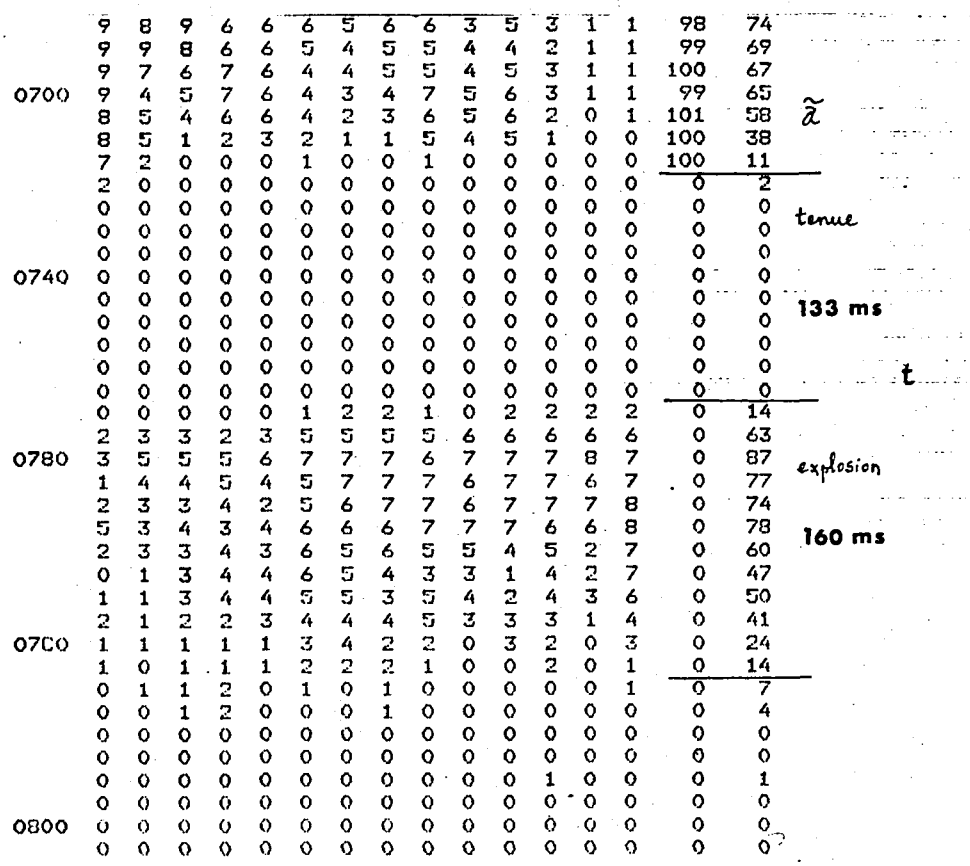
FIG. 18 - Durée des consonnes

- initiales de mots monosyllabiques ○
- initiales de mots plurisyllabiques ●
- intervocaliques (non initiales et non finales de mots) △

fermée) est identique à la durée des consonnes en position intervocalique.

Pour cette raison, nous ne considérons ici que les consonnes situées avant une pause. Mais évidemment, nous opérons une distinction selon que le schéma de F_0 est montant ou descendant sur la dernière syllabe du mot considéré.

* En ce qui concerne les occlusives sourdes, on peut décomposer leur durée dans cette position finale d'une part en la tenue de l'occlusion et d'autre part en une zone d'explosion c'est-à-dire de bruit sans vibrations laryngiennes :



Réalisation de [t] en fin de mot situé avant une pause.

.../...

* Toutes les occlusives sonores possèdent une portion voisée (tenue de l'occlusion) puis une zone qui correspond en même temps qu'à l'explosion en la réalisation d'un [ə] :

	8	6	7	8	7	9	7	5	5	4	6	8	7	7	93	94	
	7	4	5	8	6	9	8	6	5	5	6	9	7	7	91	92	
	6	5	6	6	4	8	9	5	5	5	6	8	7	7	94	87	
	3	5	6	4	5	7	8	6	5	6	5	5	6	8	92	79	ɲ
	3	6	6	5	4	6	8	6	6	5	6	5	5	7	92	78	
0200	6	7	5	6	3	4	6	6	6	6	7	6	6	4	94	78	
	8	8	7	6	4	3	5	7	7	8	9	8	5	5	87	90	
	10	8	6	5	3	3	5	7	9	9	9	9	6	6	87	95	
	11	8	7	3	0	2	4	5	8	9	9	10	6	6	91	88	
	12	8	7	4	0	1	3	3	7	10	8	9	6	5	94	83	
	12	7	5	3	1	0	1	3	5	8	8	8	7	5	95	73	
	12	7	5	1	1	0	1	3	5	5	7	7	7	5	95	66	
	11	7	4	1	0	1	1	2	5	6	6	7	7	4	96	62	i
0240	11	7	4	1	0	0	1	2	4	5	6	6	6	3	98	56	
	11	7	4	1	0	0	1	2	3	4	7	5	5	3	99	53	
	11	7	4	1	0	0	1	2	2	5	7	6	5	2	98	53	
	11	7	4	1	0	0	1	2	3	5	7	6	6	3	99	56	
	11	7	4	1	0	0	1	2	3	5	7	6	5	3	98	55	
	11	7	3	1	0	0	0	2	2	5	6	6	4	2	100	49	
	10	6	3	1	0	0	0	1	3	6	5	6	0	0	111	41	
	7	4	3	0	0	0	0	0	0	0	0	0	0	0	107	14	
0280	6	5	3	0	0	0	0	0	0	0	0	0	0	0	107	14	
	7	5	3	1	0	0	0	0	0	0	0	0	0	0	108	16	
	7	5	3	1	0	0	0	0	0	0	0	0	0	0	111	16	d
	7	4	3	1	0	0	0	0	0	0	0	0	0	0	111	15	
	6	3	2	0	0	0	0	0	0	0	0	0	0	0	114	11	
	6	2	1	0	0	0	0	0	0	0	0	0	0	0	114	9	
	7	5	5	5	5	5	4	4	6	6	5	4	5	6	110	72	
	10	8	6	5	4	4	5	5	6	5	6	5	4	5	110	78	
0200	12	10	7	4	0	2	6	7	3	1	7	6	1	2	104	68	
	12	12	8	6	2	4	7	6	5	4	8	7	4	4	105	89	
	12	12	8	6	3	3	7	5	4	4	7	7	5	6	107	89	
	11	11	8	5	3	3	7	5	2	3	7	6	2	2	107	75	ə
	9	7	7	5	2	3	7	5	2	2	5	3	1	1	106	59	
	8	7	7	4	1	3	6	4	2	1	5	5	2	3	107	58	
	6	7	7	4	1	4	7	5	3	0	5	7	2	3	103	61	
	5	4	6	4	1	4	7	4	2	0	2	6	2	3	110	50	
0300	3	2	5	2	0	1	3	2	0	0	0	2	0	2	111	22	
	2	2	4	4	0	1	2	1	0	0	0	1	0	1	0	18	
	1	2	4	3	0	1	2	0	0	0	0	1	0	0	0	14	
	2	2	4	3	0	0	0	0	0	0	0	0	0	0	0	11	

Réalisation de [d] en fin de mot situé avant une pause.

Nous avons considéré ces deux zones de façon séparée et donnerons la durée moyenne pour la réalisation du [ə]

* Les liquides-et plus particulièrement [r]-se décomposent également en deux zones : une zone voisée et une zone de bruit, cette dernière pouvant durer jusqu'à 160 ms :

.../...

14537	5	6	6	5	5	7	6	7	8	6	9	9	9	10	0	96	
14538	3	4	3	4	3	3	5	4	5	3	7	5	6	7	0	62	
14539	8	6	3	2	1	4	6	3	0	0	2	0	0	0	0	35	
14540	11	11	7	6	5	7	9	7	5	7	6	5	4	4	97	94	
14541	11	12	7	6	5	7	10	7	6	8	7	6	5	5	99	102	
14542	11	12	7	6	5	7	10	7	6	7	7	5	4	5	100	99	
14543	11	12	8	6	5	7	10	7	6	6	7	5	4	5	102	99	
14544	11	11	8	7	5	7	10	7	6	7	6	4	5	5	103	99	
14545	11	11	8	7	5	7	9	6	6	7	6	5	4	4	104	96	
14546	11	11	8	7	4	6	8	6	5	6	6	4	4	4	106	90	
14547	10	10	7	6	4	6	7	6	5	6	7	4	4	4	107	86	
14548	10	9	7	6	4	6	6	6	4	7	6	3	3	3	108	80	œ
14549	10	9	7	5	4	6	7	6	4	7	6	3	3	3	109	80	
14550	10	9	7	5	4	6	7	6	4	8	6	3	2	3	110	80	
14551	10	9	6	5	5	6	8	6	4	8	5	3	3	3	110	81	
14552	10	8	6	5	4	6	8	6	4	8	5	3	3	4	110	80	
14553	9	8	6	5	4	6	7	6	4	8	5	3	2	3	110	76	
14554	9	8	6	4	5	6	7	6	4	8	5	3	3	3	110	77	
14555	9	8	6	4	5	7	7	5	3	8	5	2	3	3	110	75	
14556	9	8	6	3	5	7	6	5	3	7	4	2	3	3	109	71	
14557	9	8	6	4	6	7	5	5	3	7	4	2	3	1	108	70	
14558	9	8	7	6	5	8	4	3	3	5	3	1	2	0	109	64	
14559	8	8	6	6	3	6	2	2	2	4	2	2	2	0	110	53	
14560	7	7	5	5	5	4	2	2	0	3	2	1	2	0	110	45	
14561	4	3	4	5	6	4	2	2	0	4	3	1	1	0	112	39	ɲ
14562	1	1	4	5	6	3	1	0	0	1	3	1	1	0	114	27	
14563	2	3	3	5	5	2	1	0	0	0	3	4	4	2	0	34	160 ms
14564	1	2	2	5	5	3	2	1	0	0	3	5	5	3	0	37	
14565	0	2	3	4	7	3	1	0	0	0	3	4	5	3	0	35	
14566	1	3	2	3	7	3	1	0	0	0	4	5	6	3	0	38	
14567	2	3	2	4	6	2	2	1	0	0	5	5	6	3	0	41	
14568	5	5	3	4	6	3	2	2	2	0	6	4	5	1	0	48	
14569	3	2	2	3	4	1	0	0	0	0	2	1	3	0	0	21	
14570	0	0	2	3	2	0	0	0	0	0	0	1	1	0	0	9	
14571	0	1	1	4	3	1	0	0	0	0	1	1	1	0	0	13	
14572	0	1	1	3	3	0	0	0	0	0	1	1	1	0	0	11	
14573	0	1	1	1	1	0	0	0	0	0	0	2	2	0	0	8	
14574	0	1	1	0	0	0	0	0	0	0	1	3	2	0	0	8	

Réalisation de [ɲ] en fin de mot situé avant une pause.

Il nous a semblé intéressant de distinguer selon que - avant la pause - la consonne finale appartient à la syllabe finale d'un mot prononcé avec une intonation montante ou au contraire avec une intonation descendante.

Nous avons calculé la durée moyenne des consonnes finales pour toutes les consonnes dont les occurrences étaient en nombre suffisant (supérieures ou égales à 20 occurrences) pour se prêter à un calcul statistique.

. Dans les phrases de type énonciatif et impératif, on peut rencontrer un mot à intonation montante (indication de continuité) situé avant une pause :

- à la fin du syntagme nominal sujet, ou bien à la fin du groupe de mots situé immédiatement avant le verbe dans le cas d'une

.../...

phrase énonciative avec inversion [syntagme complément/ syntagme verbal/syntagme sujet] :

"dans le champ + court le chien".

- à la fin du syntagme verbal :

"Veuillez indiquer + votre numéro '"

"Le petit chat mange + du crabe" .

- à la fin d'un groupe de mots situé avant le signe graphique de ponctuation noté par une virgule.

- dans le syntagme complément, entre deux groupes de sens séparés par une préposition ou une locution prépositive, ou bien avant le début d'une proposition subordonnée ou coordonnée.

. Dans les phrases de type interrogatif, il faut ajouter la fin du mot ou du groupe de mots interrogatif et/ou la fin de la phrase :

"Dans quelle ville + habitez-vous ?"

La durée des consonnes, dès lors qu'une pause leur succède, est stable dans l'une ou l'autre des situations ci-dessus énoncées.

Si l'on distingue maintenant entre les consonnes terminant un mot à F_0 montant et un mot à F_0 descendant, dans toutes les possibilités d'occurrence, on observe:

. Pour les occlusives sourdes une grande stabilité, et la tenue est toujours comprise entre 133 et 160 ms.

On se souvient que, incluse dans une séquence [voyelle-consonne-voyelle] , la tenue varie selon la consonne entre 55 et 90 ms. La durée de l'explosion est également très stable : entre 135 et 160 ms. Ces valeurs sont valables quelle que soit la consonne (p, t ou k) et quelle que soit la voyelle précédente.

En définitive, si l'on ne considère que la tenue, la durée

.../...

de l'occlusion en position finale est égale au double de la durée constatée pour une même consonne en position intervocalique.

. Les constrictives sourdes ont également une durée semblable en position finale avant une pause quel que soit le sens de la pente de F_0 , c'est-à-dire 280 ms (plus ou moins 25 ms).

18885	11	7	4	5	3	4	7	10	9	7	5	6	7	7	75	92
18886	11	8	5	5	4	4	6	7	8	7	6	6	8	8	74	93
18887	10	6	5	3	3	4	5	6	7	4	8	7	8	8	79	84
18888	10	6	5	4	4	4	4	6	7	6	9	8	7	8	75	88
18889	11	8	4	4	4	5	5	6	8	6	8	10	7	8	75	94
18890	11	8	4	4	4	4	5	6	7	6	6	10	8	8	75	91
18891	11	7	5	4	4	3	4	4	5	7	6	10	9	7	74	86
18892	11	7	5	4	3	3	4	3	6	7	6	9	8	7	73	83
18893	11	7	6	3	3	2	4	3	6	7	6	9	8	8	72	83
18894	11	7	6	3	2	2	4	4	6	8	6	9	8	7	70	83
18895	11	7	6	4	1	2	3	4	7	9	5	9	6	6	68	80
18896	11	7	6	4	1	2	3	4	6	7	5	7	5	5	66	73
18897	11	6	5	2	0	2	3	4	6	4	5	7	4	3	65	62
18898	11	5	4	2	1	1	2	4	7	5	5	8	4	4	66	63
18899	8	3	2	2	1	2	4	6	8	5	7	9	6	7	71	70
18900	0	3	2	2	2	2	4	7	9	5	7	9	7	8	0	67
18901	2	3	4	3	2	3	4	6	7	4	7	9	8	9	0	71
18902	4	3	3	4	3	4	5	7	5	5	6	8	8	10	0	75
18903	4	4	4	4	4	5	5	8	6	5	5	8	8	10	0	80
18904	6	6	6	6	6	6	6	8	7	6	7	7	8	9	0	94
18905	6	5	4	5	5	5	5	8	7	6	7	8	9	9	0	89
18906	6	5	5	4	6	6	6	8	7	6	7	8	8	9	0	91
18907	7	6	7	6	5	6	6	7	7	6	7	8	8	9	0	95
18908	7	6	5	5	5	5	6	7	7	6	6	8	8	9	0	90
18909	4	4	4	4	5	4	6	8	6	5	6	8	7	8	0	79
18910	5	4	5	4	4	5	6	7	5	5	6	7	7	8	0	78
18911	5	4	5	3	4	5	5	5	5	4	6	7	8	8	0	74
18912	4	3	3	4	3	3	4	6	5	4	7	8	7	8	0	69
18913	4	3	3	4	4	4	3	5	4	3	6	8	7	8	0	66
18914	3	2	2	2	4	4	5	4	4	3	6	7	7	8	0	61
18915	1	2	2	3	0	4	5	5	5	3	6	7	7	8	0	58
18916	1	2	2	3	2	3	6	5	5	3	5	7	6	7	0	57
18917	0	1	1	1	1	3	5	4	2	0	3	4	4	6	0	35
18918	7	5	2	2	0	1	5	0	0	0	2	1	0	1	0	26
18919	6	5	0	1	0	2	4	2	0	0	3	3	2	1	0	29
18920	0	1	0	0	2	4	6	4	3	2	4	3	3	4	0	36
18921	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2

Réalisation de [s] avant une pause.

. Par contre, dans le cas des consonnes sonores, on observe que leur durée est plus importante dans le cas de F_0 montant que dans le cas de F_0 descendant, alors que la zone correspondant au [ə], qui leur succède, est sensiblement identique dans l'une ou l'autre position :

.../...

130 à 170 ms pour la tenue dans le cas de F_0 montant,
80 à 130 ms pour la tenue dans le cas de F_0 descendant.

[ə] dure environ 135 ms (plus ou moins 25 ms).

Lors de la production des occlusives voisées, l'obstruction du canal vocal produit une diminution de la pression intra-glottique provoquant un abaissement de la fréquence fondamentale (BOË, 1973). Avant une pause, la pression intra-glottique du fait même des mécanismes des phénomènes phonatoires est elle-même diminuée (LIEBERMAN, 1967), et sur une voyelle en position finale, le maintien du voisement n'est possible qu'au prix du réajustement de la disposition des cordes vocales (variation de leur tension). On comprend que le maintien du voisement qui associe chute de la pression sub-glottique et diminution de la pression intra-glottique soit rendu encore plus difficile. Pour que le schéma mélodique soit assuré dans son intégralité, la présence du [ə] supplée vraisemblablement à la difficulté de prolongement de l'occlusive ·

La figure 19 donne les durées moyennes pour les consonnes finales à forte occurrence dans notre corpus en schéma intonatif montant et descendant dans les phrases de type énonciatif et impératif : les consonnes finales de phrases impératives (F_0 descendant) ont une durée identique à celle des consonnes finales de phrases énonciatives en schéma montant ; ce phénomène peut s'expliquer par le fait que la syllabe finale de phrase impérative a une évolution de F_0 moins importante que celle d'une phrase énonciative. Il est donc possible de faire durer plus longtemps la réalisation consonantique avant d'arriver au minimum fréquentiel de F_0 lié à des caractéristiques physiologiques. Malheureusement nous n'avons pas pu faire de comparaisons avec les consonnes finales des phrases interrogatives parce que nous n'avons pas relevé un nombre d'occurrences suffisant pour l'autoriser. Les seuls exemples dont nous disposons montrent une durée inférieure des consonnes dans ce type de phrase. Les valeurs qui apparaissent dans le tableau représentent la durée de [ə] qui prolonge les occlusives voisées, et la durée du bruit en fin de réalisation des liquides.

.../...

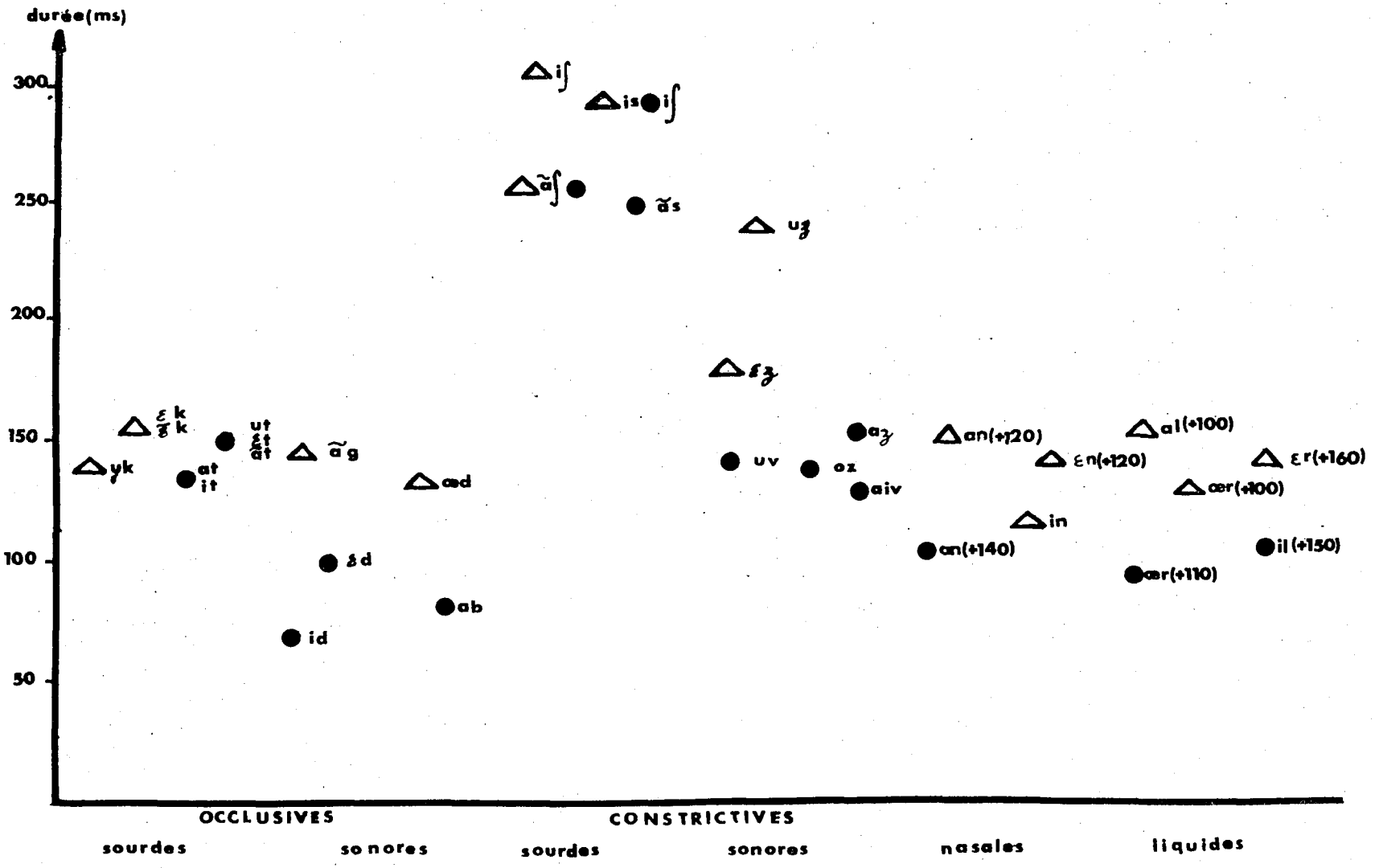


FIG. 19 - Durée des consonnes situées en fin de mot avant une pause

- avec un schéma de F_o montant △
- avec un schéma de F_o descendant ●

(les valeurs indiquées à la suite des séquences - VC représentent la durée en ms de la zone de bruit)

I-1-2- Les groupements consonantiques.

On s'intéresse ici aux suites VCCV ou VCCCV, c'est-à-dire dans une position non initiale.

Nous avons groupé dans cette catégorie à la fois les suites consonantiques appartenant à la même syllabe

. par exemple [p] + [l] dans déplaisant,

et les suites appartenant à deux syllabes séparées

. par exemple [r] + [d] dans le jardin.

Il nous était indispensable de connaître la durée de tels groupements consonantiques parce que contrairement à d'autres équipes (KÜPFMÜLLER et WARNS, 1956 ; PETERSON, WANG et SIVERTSEN, 1958) qui ont délibérément exclu le stockage de ces segments pour la synthèse, nous avons inscrit en bibliothèque tous les diphones qui représentent le passage d'une consonne à une autre, et il était important que de tels groupements connaissent une durée convenable soit pour ne pas donner par une trop grande durée l'impression auditive d'un découpage syllabique, soit pour ne pas provoquer la sensation d'une difficulté de prononciation lors de tels groupements.

Ces groupements de consonnes ont été étudiés simultanément dans les trois types de phrases énonciatives, impératives et interrogatives - parce qu'une analyse préliminaire a montré qu'il n'y avait pas lieu d'opérer une distinction entre elles.

Les groupements les plus fréquents recontrés dans le corpus comportent le segment [r] comme premier, deuxième ou troisième élément, associé à des occlusives sourdes ou sonores, ainsi que le segment [s] comme premier élément associé à une occlusive sourde.

Ces constatations confirment bien des observations systématiques sur les contraintes statistiques des groupements consonantiques : [tr] et [pr] apparaissant comme les premiers (TUBACH, 1969).

.../...

Nous n'avons pas étudié la durée de ces groupements en fonction des voyelles adjacentes parce que, à la synthèse, les diphones qui les représenteront sont destinés à servir dans différents contextes sans référence aux voyelles qui les entourent.

1ère catégorie : Le premier élément consonantique est une liquide :

1/ [l] est le premier élément consonantique.

Sa durée est très stable quelle que soit la consonne qui lui est accolée : entre 53 et 66 ms - sauf une exception - c'est-à-dire une légère diminution par rapport à un contexte intervocalique :

. Quand [l] est suivi d'une consonne nasale (par exemple "mal nutrition"), cette dernière présente la même durée que la liquide, c'est-à-dire qu'elle est réduite d'environ 10 ms, soit 13 % de réduction de durée par rapport à la position intervocalique.

. Quand le groupement est constitué de [l] + [k] ("alcool"), l'occlusive sourde est raccourcie d'environ 15 ms par rapport à un entourage vocalique c'est-à-dire que sa durée n'excède pas 40 ms (27 % de diminution de durée).

. Au contraire l'occlusive [t] en seconde position conserve une durée moyenne de 70 ms alors que [l] ne dépasse pas 53 ms.

Dans cette catégorie, la durée des deux consonnes est comprise entre 100 et 130 ms.

. Cependant, il existe une exception : [l] associé à une constrictive sourde constitue un groupement qui présente une grande durée : 250 ms, soit 71 % d'allongement pour la liquide,

dans "Delphine" par exemple, [l] = 120 ms,
[f] = 130 ms comme en position intervocalique.

.../...

2/ [r] est le premier élément consonantique :

Mis à part le groupement [rt] où [r] n'est pratiquement pas modifié, dans toutes les autres occurrences, [r] est raccourci d'environ 30 ms (soit 37 %) et sa durée est toujours inférieure à celle du second segment.

[f] et [n] subséquent à [r] subissent une diminution de durée d'environ 30 %.

2ème catégorie : constrictive sourde [f, s, ʃ] + occlusive sourde [p,t,k].
+ semi consonne [j].

exemples : escargot, dévasté, ration...

. Dans tous les cas, c'est le premier élément (constrictive) qui est le plus long particulièrement quand il est suivi de la semi-consonne [j] : sa durée est de l'ordre de 175 ms (21 % d'augmentation en moyenne), et la durée de la glide augmente de 40 ms (soit 75 % par rapport à une position intervocalique).

. Au contraire, les occlusives sourdes qui suivent une constrictive gardent à peu près (sauf [p] qui subit une diminution de 20 ms, soit 23 %) les durées qu'elles ont en position intervocalique.

. Quant à la constrictive, sa durée - entre 100 et 125 ms c'est-à-dire diminuée de 25 % par rapport à la durée observée dans les suites VCV - est d'autant plus courte que la tenue de l'occlusive est longue.

La durée totale [constrictive sourde + occlusive sourde] est en moyenne de 165 ms (plus ou moins 10 ms).

3ème catégorie : occlusives voisées [b,d,g] + liquides [r,l].

exemples : adresse, ablatif, aigri.

.../...

. Les occlusives [b] et [d] conservent, quelle que soit la consonne qui leur est associée, une durée voisine (- 15 %) de celle qui est la leur en position intervocalique (respectivement 65 et 55 ms) ; la seconde consonne [l,r] subit une réduction de durée de l'ordre de 10 %.

. Les groupements [g] + [liquide] présentent une augmentation de durée de 38 % sur la première consonne mais une diminution simultanée sur la seconde consonne :

[r] = - 33 %.
[l] = - 7 %. La durée moyenne de ce groupement est de 130 ms.

4ème catégorie : Le premier segment est constitué d'une occlusive sourde.

Le premier segment n'est pas profondément modifié en durée par le cumul consonantique. Par contre, il est intéressant de noter l'influence progressive d'une première consonne sourde sur une seconde consonne sonore ; ces dernières sont très notablement augmentées dans leur durée ; la consonne [r] en particulier , qui constitue le deuxième segment le plus fréquemment rencontré dans cette catégorie, présente un allongement au contact d'une occlusive sourde pouvant aller jusqu'à un doublement de sa durée en position intervocalique.

Dans "démocrate" par exemple : [k] = 65 ms.
[r] = 135 ms.

dans "acné " [k] = 80 ms.
[n] = 160 ms.

5e catégorie : [nasale] + [constrictive].

Dans ce type de groupement , on observe qu'une constrictive voisée tend à allonger la consonne qui la précède :

[n] < dans [ns] = 43 ms (au lieu de 83 ms soit une diminution de 48 %)
dans [nz] = 75 ms (au lieu de 83 ms soit une diminution de 9,6 %)

.../...

L'ensemble de ces résultats apparaissent au Tableau 2 dans lequel nous avons défini 8 zones :

. selon que la première consonne est plus ou moins influencée que la seconde (effet regressif ou progressif),

. selon que l'évolution de durée se fait ou non dans le même sens sur les deux consonnes (assimilation ou dissimilation),

. enfin, l'assimilation peut être positive ou négative selon que la durée des deux consonnes augmente ou diminue.

6e catégorie : Triple juxtaposition de consonnes

Les groupements les plus fréquents dans notre corpus sont les suivants :

[s]	+ [t]	+ [r]	(structure).
[s]	+ [t]	+ [j]	(question).
[k]	+ [s]	+ [p]	(expérience).
[p]	+ [s]	+ [j]	(inscription).

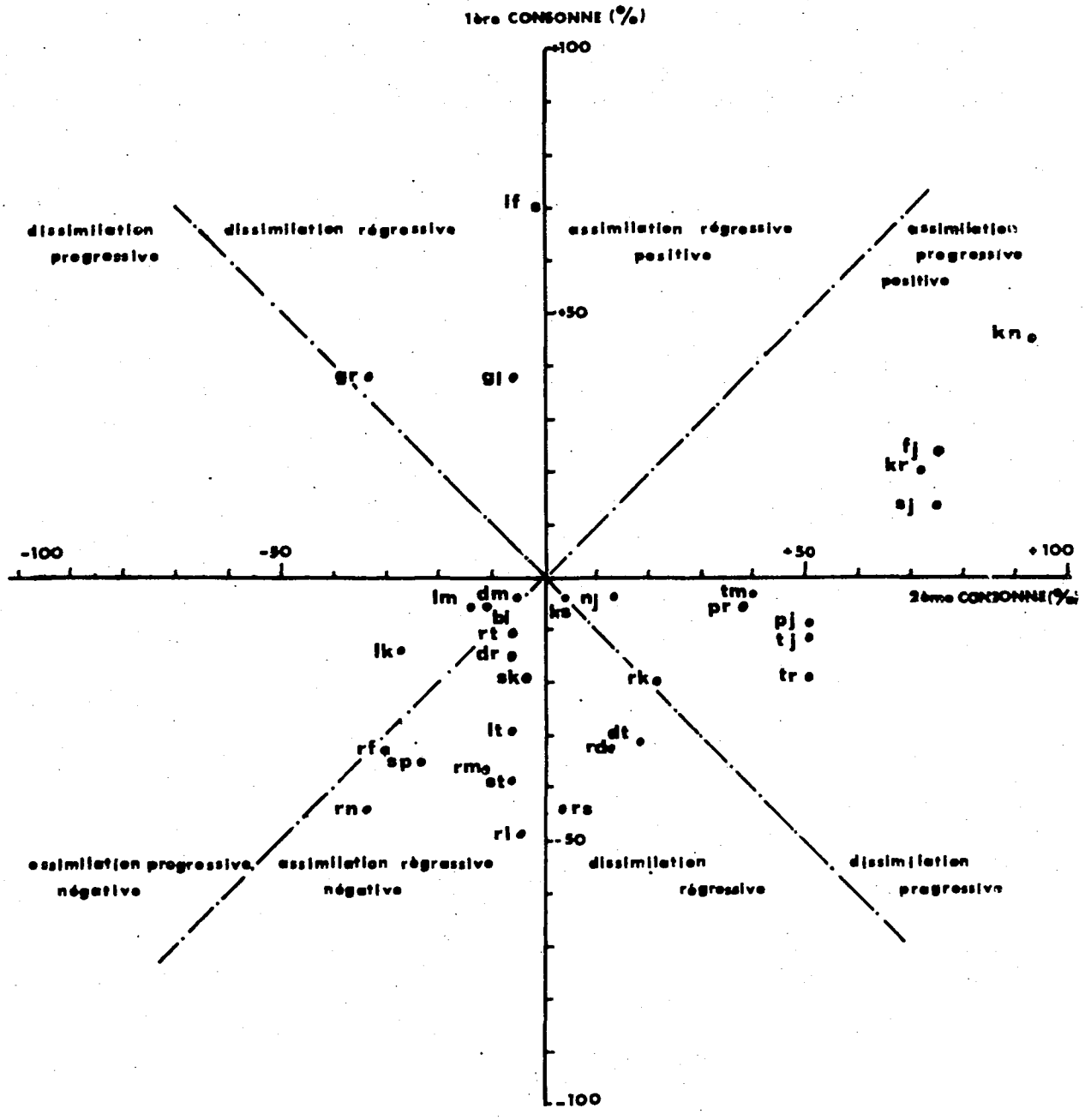
Dans tous les cas étudiés, c'est la constrictive sourde [s] qui subit la plus profonde modification : en effet, sa durée n'est plus comprise qu'entre 40 et 95 ms au lieu de 155 ms en position intervocalique, soit en moyenne une diminution de 56 %.

En définitive, il faut noter :

. Une tendance à une diminution de durée des deux segments dans les groupements consonne voisée-consonne voisée (phénomène d'assimilation).

. Une tendance à une augmentation de la durée de l'élément voisé dans les groupements associant une consonne sourde et une consonne voisée (phénomène de dissimilation).

.../...



Groupements consonantiques :

- Modifications de la durée de chaque consonne par rapport à leur durée en position intervocalique

TABLEAU 2

par exemple [t] + [r] dans /transpercé/
[k] + [n] dans /acné/
[t] + [m] dans /atmosphère/

En fait, les phénomènes que nous venons de chiffrer relèvent d'une phonétique combinatoiredéjà bien connue et pour lesquels des explications physiologiques ont été avancées (fonctionnement du larynx en particulier).

I-2- La durée des réalisation vocaliques.

Nous avons étudié les réalisations vocaliques dans trois situations susceptibles d'affecter leur durée :

. durée de la voyelle dans les mots monosyllabiques non situés avant une pause ; on n'a considéré dans cette catégorie que les mots grammaticaux :

- soit une voyelle dans une séquence consonne-voyelle :
  le, la, les, du, des, chez, son...

. dans une séquence CVC :
  car, pour, sur...

. durée de la voyelle à l'intérieur d'un mot plurisyllabique à l'exclusion de la voyelle finale de mot.

. durée de la voyelle de dernière syllabe de mot situé avant une pause. On distingue alors selon que la syllabe finale est ouverte ou fermée et selon que le mot considéré a une intonation montante ou descendante.

Enfin, on signale - quand elles existent - les différences de durée vocalique dans les phrases interrogatives par rapport à celles des phrases énonciatives et impératives.

.../...

I-2-1- Etude de la durée des voyelles dans les mots grammaticaux (auxiliaires, articles, prépositions...) monosyllabiques non suivis d'une pause (fig.20).

. Les mots grammaticaux uniquement composés d'une voyelle ont une durée moyenne légèrement supérieure aux durées vocaliques des mots composés de la suite CV ou CVC.

La durée de ces voyelles isolées est très stable quelle que soit leur position dans le message. On observe simplement que les voyelles nasales sont les plus longues (autour de 120 ms) alors que les voyelles orales varient entre 75 ms pour[a] et 85 ms pour [e] . Ces remarques concernant la différence de durée entre voyelles orales et voyelles nasales sont des phénomènes de phonétique générale bien connus.

. Les mots grammaticaux CV ou CVC présentent des durées vocaliques différentes selon leur nature et selon leur entourage consonantique. L'influence respective de chacune de ces deux variables est difficile à évaluer. Tout au plus peut-on remarquer - à la suite de nombreux auteurs - un allongement systématique de la durée des voyelles situées après une occlusive sourde [p, t, k].

I-2-2- La durée des voyelles dans les mots plurisyllabiques.

Le Tableau 3 montre la répartition des durées vocaliques en fonction de la consonne subséquente pour les réalisations à occurrence suffisante (au moins 20 réalisations).

On remarque en particulier l'allongement de durée de la voyelle [a] quand elle est suivie d'une constrictive sourde ou sonore : 117 ms en moyenne, alors que suivie d'une occlusive sourde, sa durée moyenne est de 75 ms, et de 80 ms avant une liquide ou une nasale.

.../...

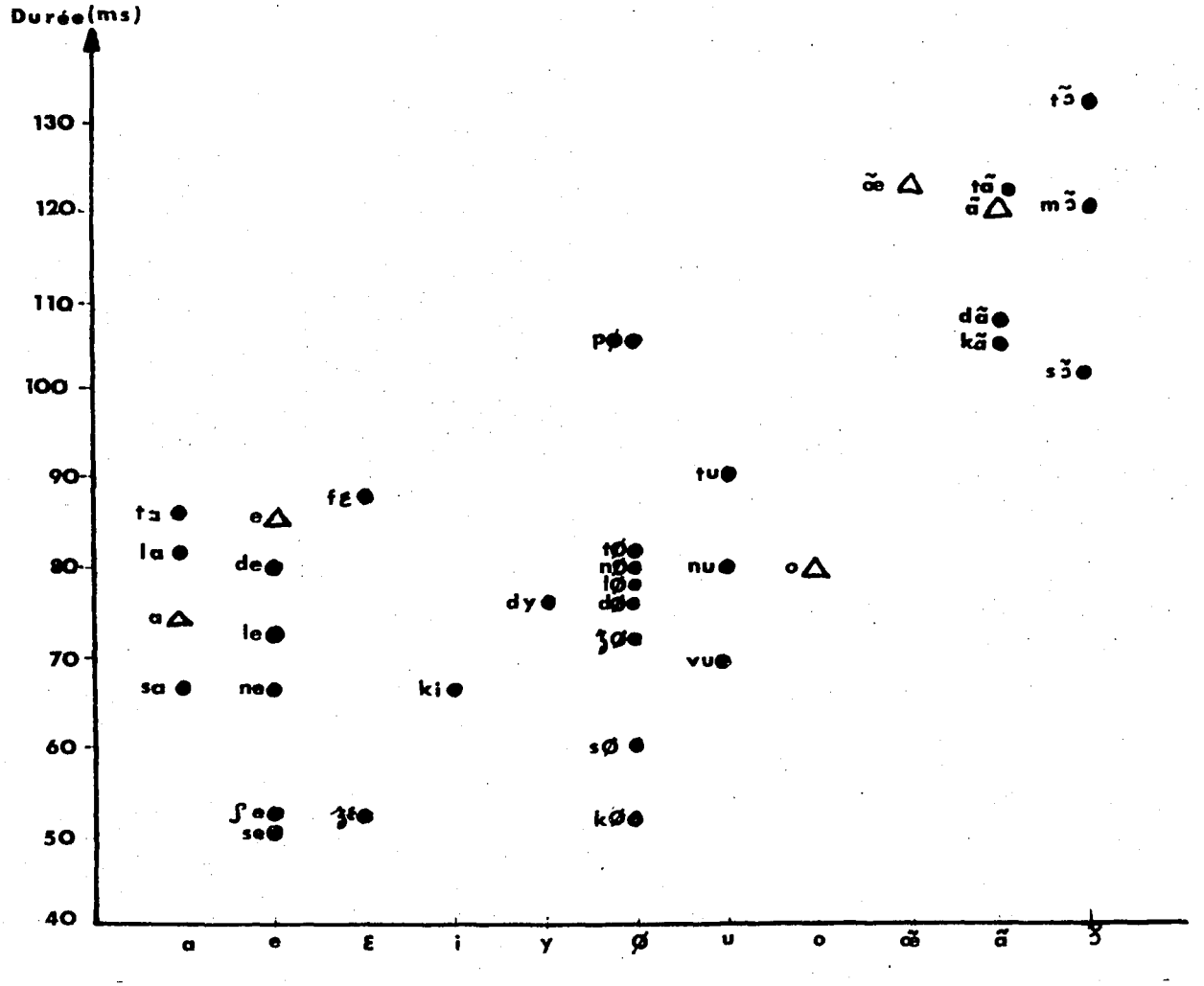


FIG. 20 - Durée des voyelles dans les mots monosyllabiques non situés avant une pause

△ voyelles isolées
● voyelles des séquences CV

	p	t	k	b	d	g	f	s	ʃ	v	z	ʒ	m	n	l	r	j
a	81 (53)	76 (53)	70	70	82 (66)	66	133	80	133 (80)	102 (66,5)	133	126	74	80	90	82 (70)	92 (40)
ɛ		76	60		66			76						66	93	90	
ã		133	100	133	113			106 (99)				147					
ɔ̃	133	120			113			124									
ɛ̃		133		133	84		118	160 (120)		118							
i	72 (53)	76	76	80	78	65	87	83 (80)	53	67	87		91	80	80	93	90
e	85 (66)	90 (60)	66		66		78	78		105	70		85	73	80	102	
ø	77	67				55		74		66		105	73	75 66,5	93	80	108
y	55	52	56					60					65	75	74	100 (80)	
o	64	73									116		80				
u	66				53			66,5		82 (53)					(53)	113	
ɔ	66,5		73,5		73,5		73,5	80					73		93	75	

A ces durées moyennes, il faut ajouter la durée de l'explosion et du VOT pour les voyelles précédées d'une occlusive sourde (Tableau 4)

TABLEAU 3 - Durée moyenne (en ms) des voyelles dans les mots plurisyllabiques pour les phrases énonciatives et impératives d'une part, et pour les phrases interrogatives d'autre part (..)

Quand elle apparaît dans un mot plurisyllabique en phrase interrogative, la voyelle[a] ne présente qu'une durée de 65 ms. Cette valeur constitue d'ailleurs la durée moyenne de toutes les voyelles orales dans les phrases de type interrogatif ; seules les voyelles nasales, réduites elles aussi, sont supérieures à 100 ms (entre 100 et 120 ms).

On remarque effectivement une différence dans la durée vocalique des monosyllabiques et des plurisyllabiques, les voyelles des plurisyllabiques ayant généralement une durée plus grande. Mais comme d'autres auteurs, en particulier UMEDA (1975), nous pensons que ces différences tiennent à la prévisibilité du mot, à sa fréquence d'occurrence dans le message.

Les prépositions et les articles sont très faciles à reconnaître, à reconstituer, grâce à des contraintes sémantiques, même s'ils ont été produits très rapidement. Au contraire un mot qui peut être important pour le sens est difficile à deviner s'il a été mal réalisé et doit entre autres attributs acoustiques posséder une certaine durée pour être identifié : "les noms forment la classe ouverte la plus large et ils sont difficiles à deviner du contexte quand ils sont oubliés dans le flux de la parole. Les prépositions au contraire sont généralement très faciles à reconstituer à partir du contexte" (UMEDA, 1975).

D'autre part, à nombre de syllabes égal, un mot qui intervient souvent dans le message possède des réalisations vocaliques plus brèves qu'un mot plus rare. UMEDA (opus cité) cite l'exemple du mot "father " qui apparaît 75 fois dans son corpus : elle note que la voyelle [a] de ce mot est considérablement réduite par rapport à la même voyelle dans d'autres situations plurisyllabiques :"les mots importants ou peu prévisibles ont beaucoup plus d'attributs acoustiques que les mots moins importants ou plus prévisibles. La durée de la voyelle peut être incluse parmi les attributs qui sont affectés par ce facteur".

.../...

Nous avons également étudié dans notre corpus la durée des voyelles dans les mots lexicaux monosyllabiques et nous n'avons pas relevé de différence de durée par rapport aux voyelles des mots pluri-syllabiques, mais simplement une plus grande durée par rapport aux mots grammaticaux monosyllabiques.

D'une façon générale, on peut noter la courte durée de la voyelle quand elle est suivie d'une occlusive sourde, et son allongement provoqué par une consonne sonore non occlusive : la voyelle [a] en particulier est très longue quand elle est suivie d'une fricative voisée (120 ms en moyenne, soit 58 % d'augmentation par rapport à sa durée quand une occlusive sourde lui succède). Il est dommage que toutes les voyelles n'aient pas présenté des fréquences d'occurrence suffisantes pour permettre une étude comparative.

Aux valeurs moyennes de durée relevées pour les réalisations vocaliques, il faut ajouter la durée de l'explosion et de l'établissement du voisement qui diffère avec chaque voyelle quand celle-ci est précédée d'une occlusive sourde. Le VOT a fait l'objet de nombreuses études, on peut se référer en particulier aux travaux de LISKER et de ABRAMSON (1964) et de SERNICLAES, opus cité.

	a	e	ã	ø	ɔ	œ	ɛ	y	u	ɔ̃	ɛ̃	i	o
p	13,3	17,29	26,6	31	20		13,3		37,24	13,5		40	26,6
t	21,28	29,26	23,94	17,29	20	40	30,59	58	53,2	26,6	26,6	60	
k	40	60	40	33,25	33,25		30,59	72	80	40		40	40

TABLEAU 4 - Durée de l'explosion et de l'établissement du voisement.

.../...

On constate que c'est la consonne [k] qui a la durée d'explosion et de VOT la plus longue, sauf une exception : l'explosion et le VOT de [t] devant [i] est de 60 ms. (Fig. 21)

L'explosion et le VOT de [p] sont les plus brefs.

Ce sont les voyelles les plus fermées [i, y, u, e] pour lesquelles la durée d'explosion et de VOT est la plus longue.

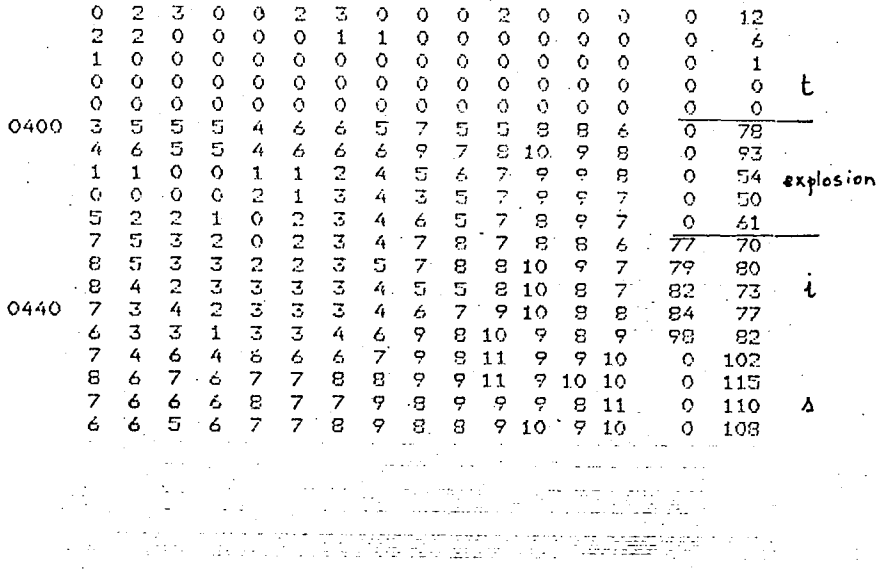


FIG. 21- Explosion de [t] devant [i].

I-2-3- La durée des voyelles en fin de mots situés avant une pause.

Les pauses, nous l'avons dit, interviennent soit en fin de phrase (pause de finalité), soit à l'intérieur d'une phrase (pause de continuité). Le dernier mot avant une pause se termine donc soit par un schéma intonatif descendant, soit par un schéma intonatif montant. Dans l'un et l'autre cas, les voyelles semblent réagir différemment quant à leur durée.

.../...

I-2-3-1- Durée de la voyelle finale en position syllabique ouverte.

. On étudie ici la situation vocalique dans les syllabes finales des mots à intonation montante suivis d'une pause, c'est-à-dire dans les positions suivantes :

- dans le dernier mot qui précède un verbe (dans les phrases énonciatives),
- dans le dernier mot du syntagme verbal,
- dans le dernier mot situé avant une virgule (indication de continuité),
- dans le syntagme complément, sur le dernier mot précédant une subordonnée ou un complément introduit par une préposition.

Nous insistons particulièrement sur le fait qu'il ne s'agit ici que de la durée des voyelles suivies d'une pause.

Il ressort de l'analyse que la durée d'une voyelle finale de mot à intonation montante est identique quelle que soit la fonction syntaxique du mot dans lequel elle est située, c'est-à-dire dans l'une ou l'autre des positions énoncées ci-dessus.

Cette durée, qu'il s'agisse d'une phrase de type énonciatif ou impératif, varie entre 190 ms [tɛ , rɛ ...] et 250 ms [bɔ̃ , rɔ̃ ...]

Par contre, la voyelle de mot à intonation descendante des fins de phrase énonciative présente systématiquement une réduction de durée par rapport à la voyelle située en schéma intonatif montant. Cette réduction est variable, elle va jusqu'à 130 ms pour la voyelle [ʏ] : ce qui correspond à une augmentation de 112 % quand F_0 est montant (FIG. 22).

.../...

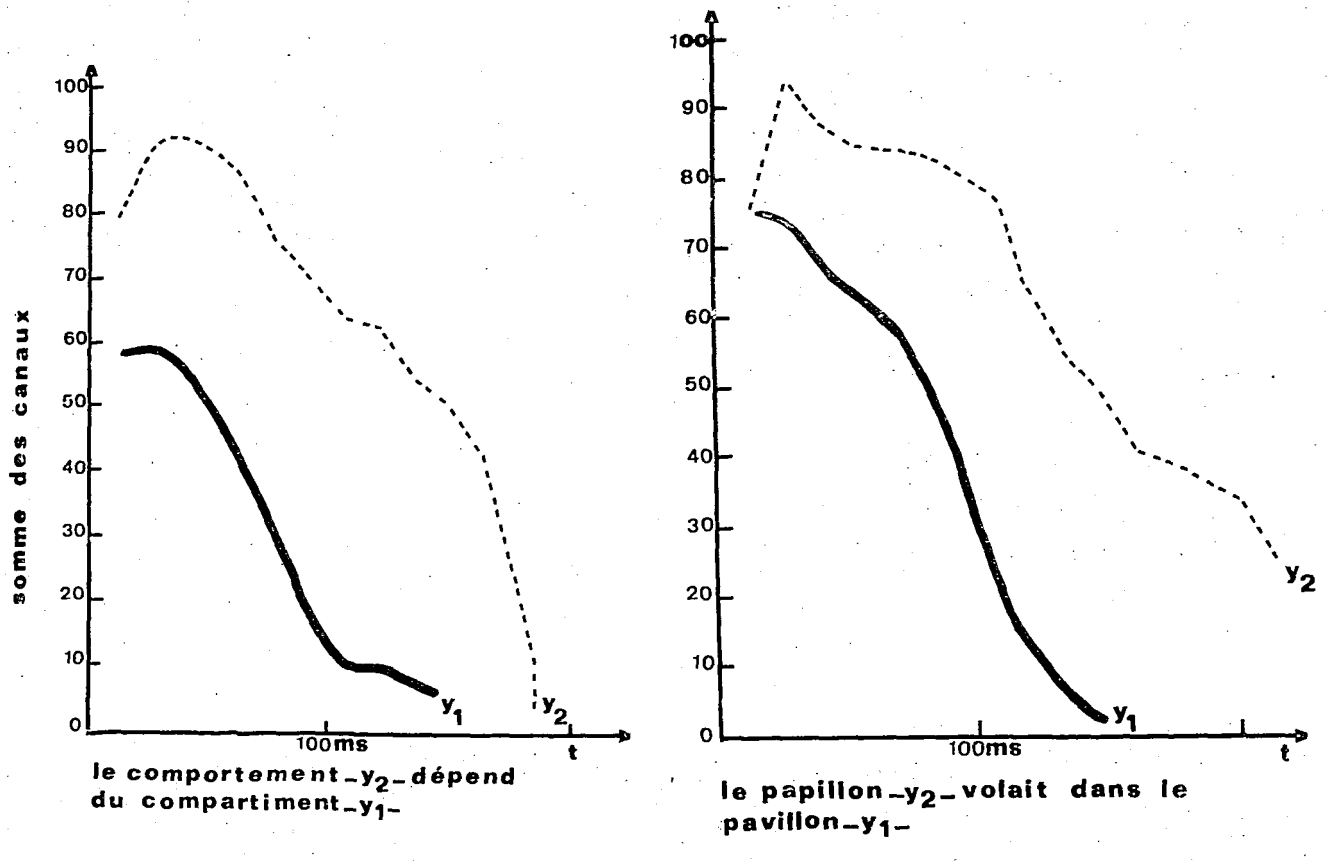


FIG: 22-Durée de la voyelle finale en schéma intonatif montant (y_2) et en schéma intonatif descendant (y_1)et évolution de "l'intensité"pour les durées correspondantes.

Les voyelles fermées [y, i, e] en schéma descendant compensent leur faible durée voisée par une très longue zone de bruit (jusqu'à 305 ms pour les voyelles les plus courtes par exemple [i] - fig. 23).

.../...

16627	5	4	3	4	7	7	5	2	0	2	4	5	4	3	87	55	
16628	2	3	3	6	6	7	5	2	0	0	0	3	4	5	97	46	ɾ
16629	4	3	3	5	4	8	5	2	0	0	2	4	3	5	99	48	
16630	5	5	4	5	5	9	6	2	1	1	3	4	3	4	96	57	
16631	5	5	3	3	4	8	7	3	2	3	5	5	3	3	99	59	
16632	5	5	3	2	0	1	6	4	3	4	5	4	3	3	98	48	
16633	9	5	2	2	0	1	5	5	4	4	6	4	3	3	91	51	
16634	10	5	2	2	1	2	2	3	5	6	7	6	4	2	94	57	
16635	9	4	2	2	0	0	1	0	2	6	5	5	2	1	97	39	
16636	8	3	1	1	0	0	0	0	0	3	1	2	3	1	103	23	
16637	7	2	1	0	0	0	0	0	0	0	1	1	3	2	106	17	i
16638	6	1	0	0	0	0	0	0	0	0	2	2	4	4	108	19	
16639	5	0	1	0	0	0	1	1	1	1	4	4	5	6	121	29	
16640	1	1	0	1	0	0	1	2	2	2	4	5	6	7	125	32	
16641	1	2	1	1	0	1	2	3	4	3	6	6	7	9	0	46	
16642	2	1	1	1	0	1	3	3	4	5	7	7	8	9	0	52	
16643	1	1	1	2	1	2	2	3	4	7	6	7	8	9	0	54	
16644	2	1	2	2	1	1	2	3	5	6	5	7	8	8	0	53	
16645	2	1	2	2	1	1	2	4	4	4	6	8	8	8	0	53	
16646	1	2	2	3	1	2	3	4	5	5	6	8	8	8	0	58	
16647	2	2	3	3	2	2	3	4	5	7	6	7	9	8	0	63	
16648	2	2	2	2	2	2	3	4	5	7	7	7	9	9	0	63	Zone
16649	1	2	1	1	2	1	2	3	4	6	7	8	9	10	0	57	de
16650	2	1	2	2	2	1	3	4	5	5	6	8	8	10	0	59	bruit
16651	3	2	2	2	1	2	3	4	5	6	7	8	8	10	0	63	
16652	2	2	1	1	1	2	3	4	6	6	6	8	9	9	0	60	
16653	3	3	2	1	2	2	3	5	5	6	7	9	9	10	0	67	
16654	3	2	2	2	3	2	3	4	6	7	7	9	8	9	0	67	
16655	2	2	2	2	2	2	4	5	7	6	7	9	9	10	0	69	
16656	2	3	2	2	1	1	3	5	6	6	7	8	9	9	0	64	
16657	3	2	1	1	0	0	3	4	5	5	7	9	8	8	0	56	
16658	3	2	3	1	0	1	3	4	5	6	7	9	9	8	0	61	
16659	4	3	3	2	2	2	3	5	5	4	6	9	9	7	0	64	
16660	5	3	2	3	2	2	3	4	6	5	5	8	7	6	0	61	
16661	3	1	2	1	0	1	2	2	4	4	4	7	4	5	0	40	
16662	1	2	1	0	0	0	0	1	3	0	4	3	1	4	0	20	
16663	1	1	0	1	0	0	0	1	1	0	1	0	1	2	0	9	
16664	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	4	
16665	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
16666	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	

FIG. 23-Réalisation de [i] en fin de mot avec F_o descendant.

On peut également comparer la durée de la voyelle portant la montée de F_o qui caractérise les phrases interrogatives avec la durée de la même voyelle en schéma intonatif montant dans les phrases énonciatives et impératives (fig. 24) : la voyelle qu'elle soit orale ou nasale n'excède jamais 205 ms dans ces énoncés.

Cette différence entre la durée des voyelles selon qu'elles présentent une évolution de F_o montante ou descendante peut peut-être s'expliquer de la même façon que la différence constatée dans la durée des réalisations consonantiques dans la même situation.

.../...

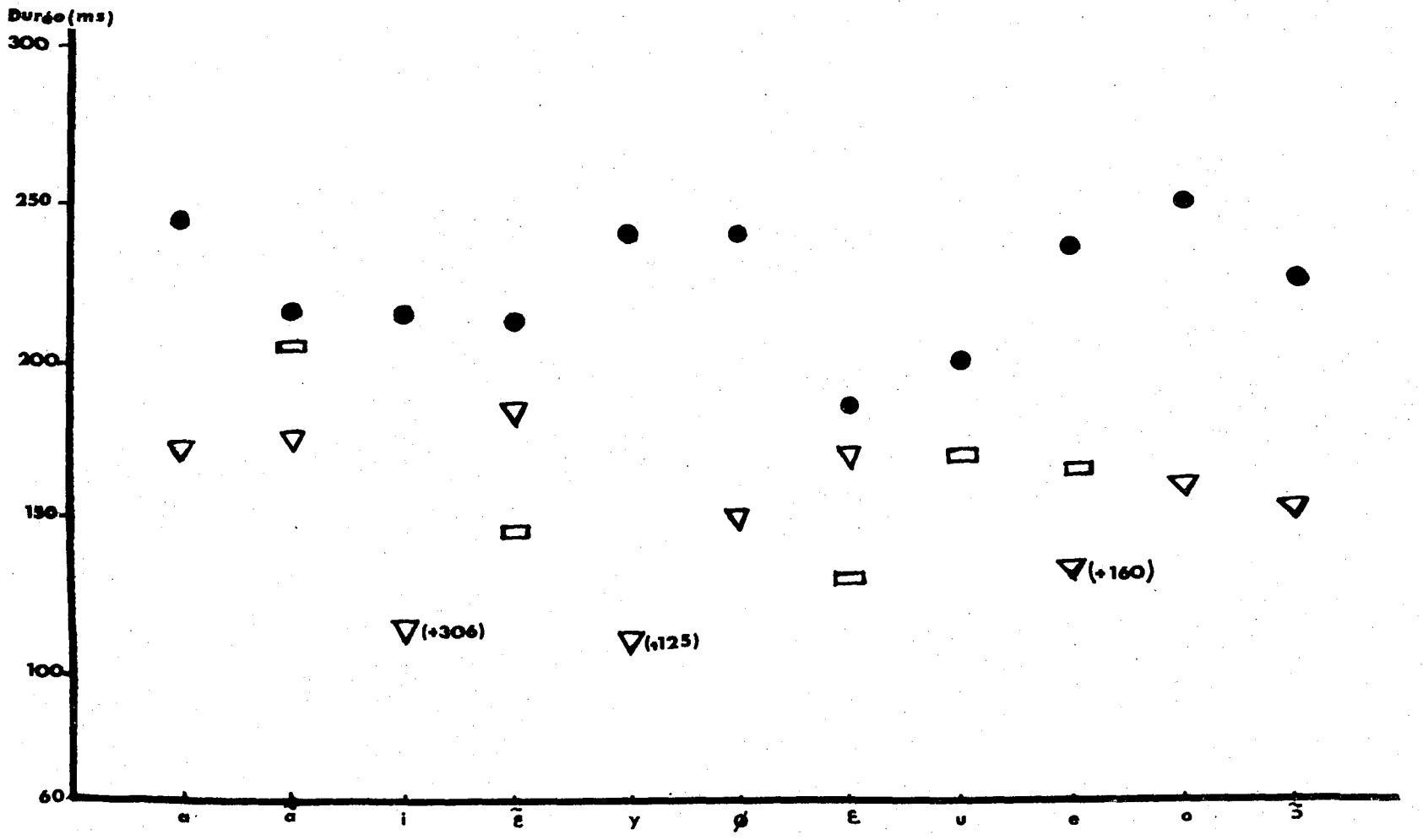


FIG. 24 - Durée des voyelles situées en fin de mot avant une pause

● avec F_o montant
△ avec F_o descendant

(la zone de bruit qui accompagne les voyelles fermées est mesurée)
▭ avec F_o montant dans les phrases interrogatives

I-2-3-2- Durée de la voyelle finale en position syllabique fermée.

On observe ici un phénomène identique : les voyelles des syllabes terminant un mot à schéma intonatif montant sont plus longues que celles des syllabes finales de mot avec F_0 descendant (ces considérations ne concernent que les phrases énonciatives et impératives, car dans les phrases interrogatives, nous n'avons pas relevé d'occurrences de mots à schéma intonatif descendant situés avant une pause ; ces occurrences sont également limitées à l'intérieur d'une phrase de type énonciatif ou impératif.

Dans ce contexte CVC fin de mot, les exemples les plus significatifs concernent les voyelles situées avant les liquides :

Quelle que soit la consonne précédant la séquence - VC, ce sont [l,r] en finale de mot qui provoquent le plus important allongement vocalique, et ce pour deux voyelles essentiellement [œ] et [ɛ] , dans le seul cas de mots à intonation montante.

La durée moyenne pour ces deux voyelles est de 280 ms alors qu'elle n'est que de 120 ms quand elles sont suivies d'une autre consonne dans cette position, et de 95 ms en schéma intonatif descendant.

Les autres voyelles ne semblent pas affectées en position finale par leur entourage consonantique.

Nous avons dégagé pour quelques voyelles les durées moyennes suivantes :

.../...

	F_o descendant	F_o montant
[i]	100 ms	135 ms
[a]	135 ms	156 ms
[ɛ]	95 ms	120 ms [ɛ] + [r,l] = 280 ms
[ã] [õ]	185 ms	200 ms

L'ensemble des résultats que nous avons obtenus pour les durées des voyelles finales de mot sont en contradiction avec la plupart des résultats d'autres études, en particulier BENGUEREL (1971). Mais les comparaisons que cet auteur a effectuées entre voyelles finales de mots à intonation montante et à intonation descendante ne font intervenir l'occurrence de la pause que pour les mots à intonation descendante, les mots à intonation montante qui servent d'élément de comparaison n'étant pas situés avant une pause.

Dans ce cas, on comprend la constatation d'une plus grande durée pour les voyelles finales de mot à intonation descendante.

Par contre, la constatation que nous avons faite d'une plus longue durée des voyelles à F_o montant situées avant une pause par rapport à la durée des voyelles à F_o descendant dans la même situation est conforme à celle de SPANG-THOMSEN (1963) qui effectue la comparaison entre syllabe de mot à intonation montante et à intonation descendante quand l'une et l'autre sont suivies d'une pause. Elle est conforme également aux conclusions de OLLER (1973).

I-2-4 - Reste à étudier la durée des séquences vocaliques .

par exemple [a-e] dans [aéroport] ,[y-a] dans [nuage], [e-a] dans [réagir].

.../...

Il est extrêmement délicat de savoir où il faut opérer la segmentation car les transitions sont réalisées de façon continue : on passe de l'une à l'autre sans discontinuité spectrale et sans discontinuité dans l'excitation.

Les échantillons vocodeur (fig. 25) qui les représentent permettent simplement de repérer la zone de transition et de déterminer la durée totale de la réalisation de la séquence vocalique. Ces informations nous suffisent pour une synthèse par diphones : il ne s'agit que d'extraire toute la zone qui représente le mouvement de passage d'un segment au suivant, et de fixer pour cette zone une durée qui permette de l'associer convenablement aux diphones environnants [CV] et [VC].

	6	7	9	8	7	9	9	7	7	8	7	6	5	6	0	101	
0640	8	10	12	11	9	11	11	8	8	10	9	7	6	7	78	127	
	9	10	12	11	10	12	12	8	8	11	10	7	6	7	76	133	
	9	11	12	11	10	12	12	9	9	11	11	7	7	8	76	139	
	9	11	12	11	10	12	12	10	9	11	10	8	7	8	76	140	a
	10	11	12	11	9	11	12	10	9	11	10	8	7	7	76	138	
	10	11	11	10	9	11	12	12	9	11	11	8	6	8	75	139	
	10	12	11	10	8	10	11	13	10	11	12	9	7	7	72	141	
	10	12	11	9	8	9	10	11	10	11	12	9	7	8	75	137	
0680	11	11	10	7	7	9	9	10	11	10	12	9	8	8	75	132	
	11	9	9	6	6	7	8	9	11	11	12	10	7	7	75	123	
	11	9	8	6	5	7	7	9	11	10	12	10	8	7	74	120	
	12	10	8	5	6	7	7	9	12	11	11	9	8	8	74	123	
	12	10	7	5	6	7	7	9	12	11	12	9	8	8	74	123	e
	12	10	8	6	6	8	8	10	12	12	11	8	8	8	75	127	
	11	10	8	6	7	8	9	10	11	11	10	8	7	8	76	124	
	11	10	9	7	8	10	10	9	8	11	9	7	7	8	77	124	
06C0	10	10	9	8	8	11	9	5	6	10	8	6	6	6	79	112	
	9	9	9	9	8	9	3	2	6	6	9	6	6	3	82	94	
	7	6	6	6	8	0	1	4	4	0	7	5	8	5	95	67	
	0	0	3	5	6	2	3	2	1	2	6	6	9	6	95	51	r
	3	3	5	6	4	4	2	0	0	0	3	4	7	4	0	45	
	5	5	6	5	3	1	0	0	0	0	6	5	7	3	83	46	
	9	8	9	7	4	3	2	1	0	1	6	5	6	3	79	64	
	11	10	11	9	6	6	3	2	3	4	6	6	7	3	77	87	
0700	12	10	11	10	6	7	4	2	4	7	6	6	8	4	76	97	o
	12	10	10	9	6	6	4	2	4	8	5	6	8	5	77	95	
	12	9	10	8	5	5	3	2	3	7	3	5	6	4	74	82	
	11	7	9	6	4	3	1	2	2	3	2	3	5	3	66	61	
	7	3	6	3	0	0	1	1	0	0	1	2	2	2	101	28	

FIG. 25 - Séquence vocalique . [a-e].

Un ensemble de deux voyelles présente une durée assez grande : de l'ordre de 265 ms quand les deux voyelles sont incluses dans le même mot. Mais en syllabe finale de mot situé avant une pause, cette situation provoque un allongement simultané des deux voyelles alors même qu'elles appartiennent à deux syllabes différentes : [y a] dans nuages dure 418 ms (fig.26).

.../...

	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	2	
01C0	6	2	1	0	0	0	0	0	0	0	1	0	0	0	0	10	
	8	5	3	3	1	2	1	1	2	3	5	2	0	0	83	36	
	8	5	3	4	1	3	2	2	3	5	6	3	0	0	81	45	
	8	5	4	5	2	3	2	3	4	6	7	4	1	0	80	54	
	9	6	4	5	3	3	4	4	4	7	7	4	1	0	79	61	n
	9	6	4	5	3	3	4	4	5	6	7	4	1	1	79	62	
	9	6	4	5	3	4	4	5	7	5	7	5	1	1	79	66	
	9	6	5	5	3	4	5	8	9	6	7	7	4	1	78	79	
0200	9	6	6	5	4	4	6	9	11	8	6	7	6	2	77	89	
	10	7	5	5	4	4	6	10	10	9	5	6	7	3	76	91	
	10	7	5	4	4	4	6	9	9	9	5	4	6	2	76	84	
	10	7	5	4	3	4	7	9	9	8	3	4	5	2	77	80	
	10	8	5	4	4	5	8	8	9	8	4	4	6	2	77	85	
	11	8	5	4	4	6	9	8	9	8	4	5	7	3	77	91	y
	11	8	6	5	5	6	10	8	9	8	5	5	6	2	77	94	
	11	8	6	5	5	7	10	8	9	8	4	5	6	2	77	94	
0240	11	8	6	5	6	7	10	8	9	8	5	6	7	3	78	99	
	11	8	7	5	6	8	10	8	10	8	5	6	7	3	78	102	
	11	9	8	6	7	9	10	8	10	9	6	5	7	4	79	109	
	10	11	9	7	8	9	11	8	10	10	6	6	6	5	80	116	
	10	12	10	8	8	10	11	8	8	9	6	6	6	5	84	117	
	10	11	11	9	9	11	10	8	8	9	7	5	7	5	83	120	
	10	10	12	9	10	11	10	8	8	10	7	6	7	6	85	124	
	10	10	12	10	10	12	10	7	8	10	7	7	7	7	86	127	
0280	9	10	12	11	10	12	10	7	8	8	6	7	7	7	88	124	
	9	10	12	11	10	12	10	7	8	9	7	6	7	7	87	125	
	9	10	12	11	11	11	10	7	8	9	7	7	7	7	87	126	
	9	10	12	11	10	11	9	7	7	9	8	7	6	7	88	123	
	9	10	12	11	10	10	9	6	6	7	8	6	6	7	89	117	a
	9	10	12	11	10	10	10	7	6	8	8	7	6	7	91	121	
	9	10	12	11	9	10	9	6	6	9	9	6	6	7	92	119	
	9	10	12	10	9	9	9	6	6	9	8	5	6	7	93	115	
02C0	9	10	12	10	8	8	10	6	5	9	9	6	6	7	94	115	
	9	10	12	9	8	8	10	6	5	8	10	7	6	7	94	115	
	9	10	11	9	8	8	10	7	5	8	10	7	5	6	95	113	
	9	10	10	8	7	8	10	7	6	8	10	7	6	5	90	111	
	9	10	9	7	6	7	9	7	5	8	10	7	6	5	96	105	
	9	10	8	6	6	7	9	7	5	8	10	7	6	5	96	103	
	10	11	8	6	5	7	9	6	5	8	10	7	6	3	96	101	
	10	11	7	6	5	7	9	5	5	8	9	7	5	3	97	97	
0300	10	10	6	5	4	6	9	5	5	8	8	8	5	2	97	91	
	9	9	6	4	2	6	8	6	6	9	8	7	4	2	97	86	
	9	8	5	3	2	6	7	5	7	10	9	6	4	2	97	83	
	9	8	5	4	3	5	7	6	8	10	8	6	5	4	97	88	
	9	7	4	4	2	6	8	6	8	11	8	8	6	5	99	92	
	9	7	4	3	2	5	8	6	9	12	9	8	7	6	102	95	
	8	7	3	4	3	5	8	6	10	11	9	9	7	7	104	97	
	8	6	3	4	3	5	7	7	10	11	9	10	8	7	100	98	
0340	8	6	4	3	2	6	7	7	10	11	10	10	9	8	99	101	ʒ
	8	6	4	3	3	5	6	7	10	12	11	10	9	8	98	102	
	7	6	4	3	3	5	6	7	11	12	11	10	9	7	99	101	
	8	6	3	3	3	6	7	7	11	12	10	12	9	9	99	106	
	8	6	4	3	3	6	7	8	11	12	10	11	10	9	100	108	
	8	6	4	3	3	6	7	8	11	13	11	11	10	8	99	109	
	7	6	4	3	3	6	7	7	10	13	11	11	9	7	98	104	
	8	6	3	3	3	5	6	7	11	12	10	10	7	7	99	98	
0380	8	7	3	3	2	5	5	5	9	9	8	6	7	6	100	83	
	8	7	3	3	1	5	5	3	6	8	7	6	4	3	100	69	
	6	5	3	3	2	5	6	3	4	9	6	6	3	3	98	64	
	3	3	3	3	2	4	6	4	6	9	5	7	3	2	100	60	
	3	5	4	3	2	5	6	4	7	9	6	7	4	2	0	67	
	3	4	2	2	1	5	6	3	5	5	4	5	2	0	0	47	
	1	0	3	2	0	3	4	1	0	0	3	3	2	0	0	22	
	0	0	2	2	0	1	2	1	0	0	3	4	2	0	0	17	
03C0	0	1	1	1	0	2	2	1	0	0	4	3	1	0	0	16	
	0	0	0	0	0	2	2	0	0	0	2	2	0	0	0	8	
	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	2	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

FIG. 26 - Séquence vocalique [y a] dans /nuage/ situé avant une pause.

.../...

Le plus souvent, les deux voyelles constituent les frontières de deux mots différents :

[e-œ̃] dans "vous prenez un sucre ?"

[ɛ-ã] dans "elle fait entrer..."

[a-a] dans "il a arrêté..."

Si ces deux voyelles appartiennent à deux syntagmes ou à deux éléments de syntagme différents, on opère bien plus facilement une segmentation entre elles, parce qu'on observe une rupture dans le schéma intonatif. La durée de chacune des voyelles est fonction du mot dans lequel elle est insérée, mot grammatical ou non, ainsi que de sa durée intrinsèque (différence voyelles orales/voyelles nasales par exemple).

En définitive, les différences qui apparaissent entre nos résultats et ceux d'autres études sur la durée segmentale sont vraisemblablement dues :

- à la nature et au choix du corpus,

- à ce que dans les procédures de dépouillement, nous avons tenu compte de la position des mots par rapport aux pauses et par rapport à l'évolution du schéma intonatif,

- à la fonction grammaticale du mot.

.../...

II - ETUDE DES PAUSES :

L'analyse de la parole continue laisse apparaître sa fragmentation et son interruption par des intervalles de silence, les pauses.

De nombreux auteurs ont étudié leur importance quantitative, et les pourcentages du temps de pause par rapport au temps total de locution : en particulier GOLDMAN-EISLER (1968), GROSJEAN et DESCHAMPS (1972) dans la conversation et l'interview, LUCCI (1973-1974) dans des études comparatives du français parlé et du français lu, BOE et al (1975), BOË (1976).

Mais des divergences entre auteurs apparaissent quand il s'agit de définir la nature et la fonction des pauses dans la parole :

- pour CARREL et TIFFANY (1960), les pauses dans la production de la parole servent de ponctuation orale ; elles reflètent et signalent la structure grammaticale de la phrase.

- pour BOOMER et DITTMAN (1962), MARTIN et STRANGE (1968), les pauses signalent les frontières grammaticales les plus importantes : "... les pauses à l'intérieur des constituants grammaticaux les plus grands représentent le procédé de sélection des mots : les pauses qui apparaissent entre des constituants majeurs indiquent que l'on veut sélectionner des structures plus larges que le mot".

D'autres auteurs au contraire pensent qu'il n'existe pas de relation étroite entre les pauses et la structure syntaxique de l'énoncé. Il semble que ces divergences tiennent entre autres à l'analyse de corpus de styles différents :

.../...

-par exemple,MAC CLAY et OSGOOD (1959), qui ont étudié les pauses dans un corpus spontané, trouvent peu de correspondances entre les pauses d'hésitation et les frontières syntaxiques.

Ces remarques ont été également formulées par LEON et MARTIN (1969) pour lesquels dans les corpus spontanés, le groupe de souffle - étudié en particulier par GRAMMONT (1948) et FOUCHE (1959) et qui correspond au "breath group" chez LIEBERMAN (1967) - ne coïncide pas forcément avec un syntagme grammatical : il ne suit pas le découpage linguistique du message. Ils proposent comme exemple, celui d'une interview de Jean-Paul SARTRE :

"J'ai/euh/bien sûr/été/ à la fois/ attiré et repoussé/
par l'oeuvre de FLAUBERT/ telle qu'elle se présente
dans ses lettres".

(les barres obliques indiquent les groupes de souffle successifs).

- GOLDMAN EISLER (1968) estime que la structure syntaxique, aussi complexe soit-elle, n'est pas reflétée par la durée des pauses ; mais que celles-ci existent pour précéder les mots à information sémantique importante.

- LIEBERMAN (1967) attribue aux pauses un rôle qui complète celui donné à la fréquence fondamentale : elles servent d'indice de démarcation qui reflète la structure constituante du message au moins pour ce qui concerne les cas d'ambiguité (problèmes de jonction). Une pause démarque par exemple "les petites roues" de "les petits trous". Mis à part ce rôle démarcatif, il ne conçoit les pauses que comme moyen de reprendre son souffle à certaines frontières syntaxiques importantes : il définit le breath group comme les éléments de parole qui apparaissent entre les pauses de respiration.

Cependant, la plupart des auteurs reconnaissent quand même qu'une des fonctions des pauses est liée à la structure syntaxique de la phrase, même si cette fonction n'existe pas toujours et même si d'autres rôles lui sont dévolus, en particulier la nécessité physiologique d'inspiration phonatoire ou bien, dans un corpus spontané, la

.../...

pause d'hésitation - sans corrélation avec la structure syntaxique - liée à la difficulté du choix d'un terme ou à un problème de mémoire.

Dans le corpus que nous avons enregistré et qui consiste en des phrases lues, il est évident que les pauses d'hésitation sont inexistantes et que ne demeurent que des pauses de reprise de souffle - qui sont en coïncidence avec la structure syntaxique - et des pauses syntaxiques qui délimitent certains syntagmes ; d'autre part, la ponctuation présente dans le texte est elle aussi une indication graphique de marque syntaxique et il faut s'attendre à une certaine coïncidence entre les pauses et ces repères graphiques.

Les deux types de pauses que nous venons de citer connaissent des durées différentes, mais succèdent l'une et l'autre à un allongement caractéristique de la durée du dernier segment les précédant. Cet allongement avant une pause a été observé également par GAITENBY (1965), MATTINGLY (1968) ainsi que par KLATT (1971) qui constate que les pauses ne correspondent pas seulement à des groupes de souffle, mais qu'elles apparaissent règulièrement et de façon prévisible à des frontières syntaxiques spécifiques : il distingue entre les pauses de longue durée (environ une seconde), pour l'inspiration, et les pauses plus brèves (inférieures à 200 ms) qui reflètent la structure syntaxique.

Notre étude porte sur l'analyse de la répartition et de la durée d'environ 500 pauses.

Le problème de la perception des pauses n'est pas simple : en effet, une pause de même durée selon sa position par rapport à l'agencement syntaxique ne sera pas perçue de la même façon ; aussi, pour éviter ce problème qui dépasse le cadre de notre travail, nous sommes-nous limité à l'étude de la durée objective des pauses.

Nous distinguerons entre trois types de pause :

- les pauses de reprise de souffle qui
  . ne peuvent apparaître qu'en un nombre de points précis limités,

.../...

. sont liées à la structure du message.
Nous les noterons /P/.

- les pauses syntaxiques situées en des endroits spécifiques de l'énoncé, nous les noterons /p/ dans les exemples.

- les pauses de démarcation de mots
  . dans les cas d'ambiguité sémantique,
  . dans les cas où les frontières communes de deux mots sont vocaliques.

II-1- Les pauses de reprise de souffle :

Quand elles existent, ces pauses qui correspondent à une nécessité physiologique se situent en des endroits privilégiés de l'énoncé :

* Elles manifestent la ponctuation orale signifiée graphiquement par les signes conventionnels suivants :

- point virgule, point, point d'exclamation, point d'interrogation pour participer à l'indication de fin de message,
- la virgule, indication de continuité du message.

Mais cette pause peut n'être que "virtuelle" (MARTINET, 1961) si le débit est rapide ; l'exemple suivant peut être réalisé sans silence malgré la présence graphique de la virgule :"l'enfant mange des pommes, des poires et des oranges".

* Ces pauses de respiration phonatoire peuvent se rencontrer également "à l'intérieur" d'une proposition - sans qu'un signe de ponctuation leur corresponde à l'écrit - mais cette possibilité n'apparaît qu'en un seul point de l'énoncé :à la fin du syntagme qui précède immédiatement le verbe.Cette pause ne semble pas devoir être liée à la longueur de ce syntagme mais semble résulter d'une anticipation visuelle des éléments linguis-

.../...

tiques qui lui sont postérieurs : la pause représente une reprise de souffle quand ces éléments sont considérés comme suffisamment longs et complexes pour justifier une inspiration :

Exemple 1 : De gigantesques inondations /P/ recommencent à s'abattre sur le pays.

Exemple 2 : Le numéro de Monsieur DUPONT/P/ est le 35-28-73 à PERROS GUIREC.

Exemple 3 : Le numéro de Monsieur DUPONT/p/est modifié.

Dans le premier exemple, l'unique pause située en fin de syntagme nominal sujet est de 592 ms.

Dans le second exemple, cette pause dure 732 ms.

Par contre, dans le troisième exemple, où le syntagme nominal sujet est identique à celui de l'exemple 2, la pause n'est que de 120 ms.

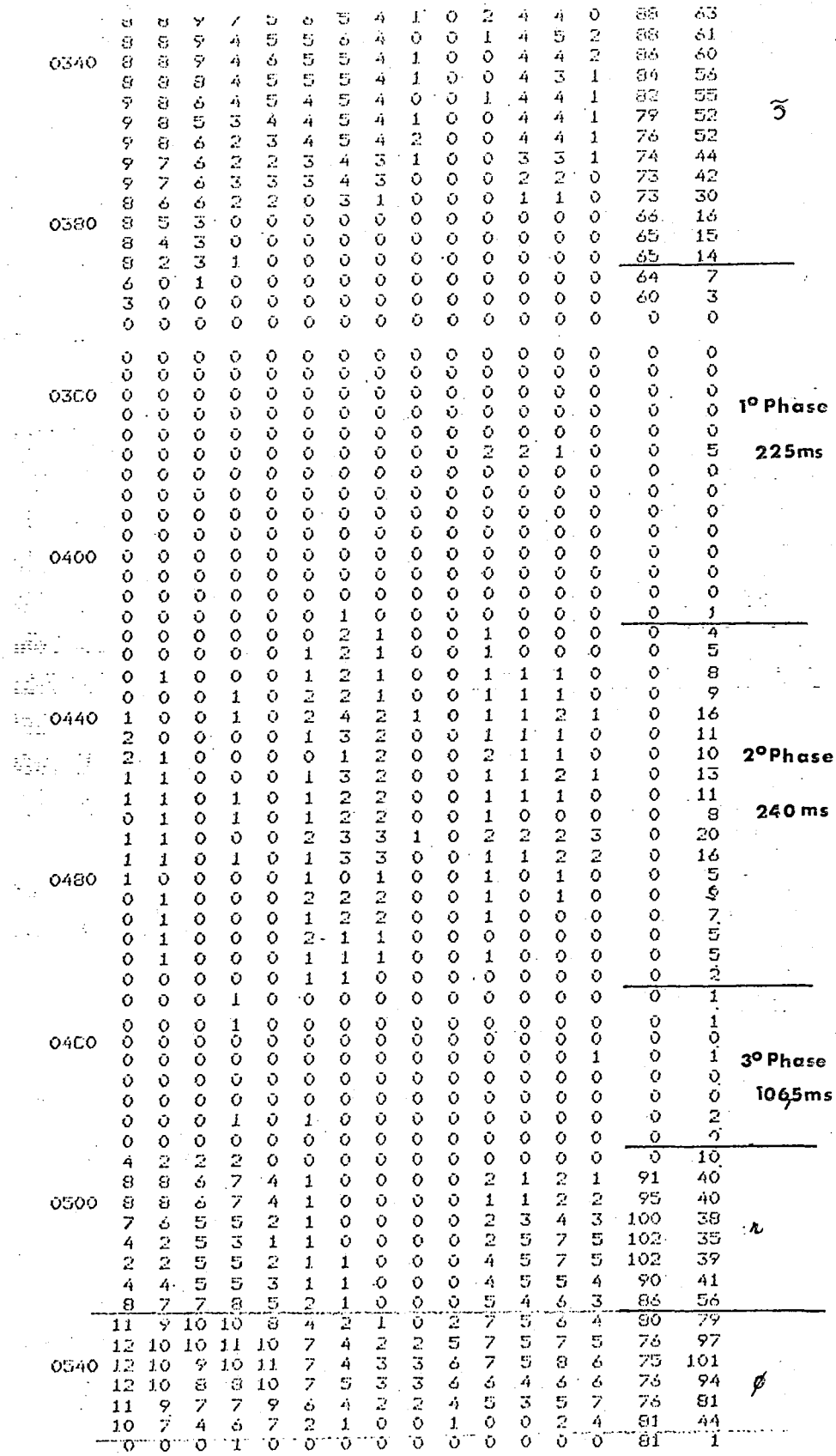
Il semble donc bien que c'est par un phénomène d'évaluation anticipatrice que la durée de la pause est déterminée.

Dans les deux premiers exemples, les pauses de 592 ms et de 732 ms sont effectivement des "pauses de reprise de souffle" ; en effet, on observe très bien sur les listings d'échantillon vocodeur (fig. 27) une décomposition en trois phases dans la réalisation de ces pauses :

Dans l'exemple 2, on observe :

1°/ Aux termes du dernier mot qui précède la pause, le début du silence est noté ∅ (aucune énergie acoustique n'est détectée) pendant environ 250 ms ; ce segment correspond à la fin de l'expiration phonatoire.

.../...



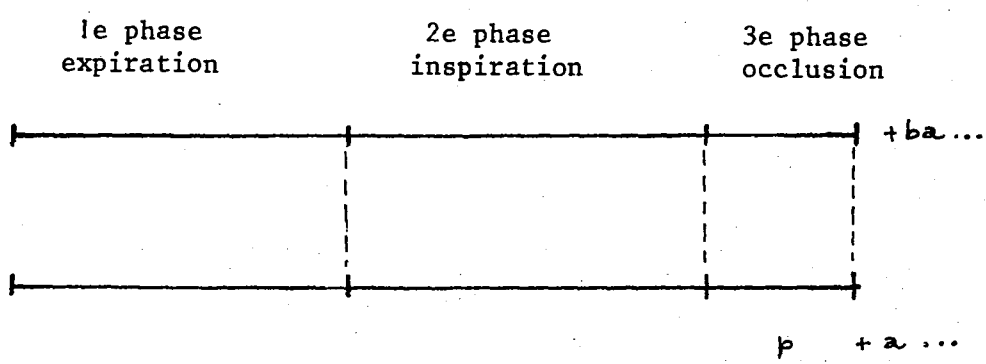
FIG·27 - Exemple 1 - PAUSE DE REPRISE DE SOUFFLE. .../...

(échantillon vocodeur)

2°/ Une zone de bruit (les canaux du vocodeur délivrent une légère énergie) qui dure également environ 250 ms ; elle correspond à une phase d'inspiration et au bruit de l'air quand il traverse les cavités phonatoires.

3°/ Une seconde zone de silence - notée Ø - d'environ 100 à 130 ms représente l'occlusion momentanée du conduit vocal avant la réalisation du premier son. SORON et LIEBERMAN (1963) constatent la même durée - Cette zone correspond à la mise en place des cordes vocales pour leur mise en vibration : les arythénoïdes pivotent vers l'intérieur, et ferment la partie postérieure de la glotte.

Mais on observe, quand le premier mot après la pause comprend une occlusive sourde [p] , [t] , [k] comme première réalisation, que cette dernière zone de la pause tient lieu de tenue de l'occlusion :



Dans l'exemple 1, on constate :

1e phase = 226 ms
2e phase = 240 ms
3e phase = 106,5 ms.

.../...

Les pauses de cette catégorie sont les plus longues, jamais inférieures dans notre corpus à 500 ms, mais leur durée maxima est très variable selon les messages, en fonction du débit de l'ensemble, et en fonction du nombre et de la longueur des pauses "intérieures" - c'est-à-dire le plus souvent syntaxiques - de la phrase, elles peuvent durer jusqu'à une seconde environ.

Quand le message comporte dans sa forme écrite plusieurs virgules successives, le locuteur a tendance à organiser, dans le passage à la lecture, la durée des pauses d'une façon qui tient compte réellement de la nécessité de respiration :

"Ce matin là, /p1/ à Domrémy,/P1/ Jeanne,/p2/ la bonne Lorraine,/P2/ gardant les troupeaux de ses parents,/p3/ fumait en cachette"*

La durée de ces cinq pauses est organisée comme suit :

/p1/ = 265 ms.
/P1/ = pause de reprise de souffle : 332 ms d'expiration + 300 ms d'inspiration + 105 ms d'occlusion glottale.
/p2/ = 80 ms.
/P2/ = pause de reprise de souffle : 332 ms d'expiration + 320 ms d'inspiration + 140 ms d'occlusion.
/p3/ = 386 ms.

On constate ici une alternance parfaite entre pauses brèves - notées/p/- pour réaliser une simple ponctuation orale, et pauses qui permettent, en même temps que la démarcation, la respiration.

II-2-Les pauses de démarcation syntaxique :

Leur présence correspond à l'intérieur d'une proposition au découpage en groupes syntaxiques auquel participent, comme nous allons le voir par la suite, les évolutions de la fréquence fondamentale et certaines durées segmentales.

---

* Gilles VIGNEAULT, Contes du coin de l'oeil, Editions de l'Arc, 1966.

.../...

Ces pauses syntaxiques font suite presque toujours à un mot dont la dernière syllabe possède un schéma intonatif montant. Les seuls cas où une pause succède à une syllabe finale de mot à schéma intonatif descendant se situent dans le syntagme qui précède le verbe:

a/ quand celui-ci est particulièrement long et découpé en plusieurs groupes de sens :

"les exemples de synthèse par diphonèmes /p1/ présentés ici /P1/ sont réalisés /p2/ de façon entièrement /p3/ automatique"

Dans cet exemple, la répartition des pauses est la suivante :

- la pause /p1/ correspond au terme d'un groupe de sens du syntagme nominal sujet terminé par un schéma intonatif descendant ; sa durée est de 250 ms.

- /P1/ est une pause de respiration en fin de syntagme nominal sujet ; elle se décompose en trois phases :

  expiration : 250 ms
  inspiration: 332 ms
  occlusion : 70 ms

- /p2/, pause syntaxique de fin de syntagme verbal, succède à une syllabe finale de mot à intonation montante et dure 80 ms.

- /p3/ pause de 40 ms qui sert à démarquer deux mots dont les frontières communes sont vocaliques.

b/ quand dans le syntagme nominal sujet, le locuteur désire insister sur certains mots qui lui semblent importants, il fait précéder ceux-ci d'une pause - le schéma intonatif de la dernière syllabe de ces mots étant descendant. Ce type de pause destiné à précéder les mots à haute information a été également signalé par GOLDMAN-EISLER (1961). Cette pause peut exister même si le syntagme nominal sujet et/ ou

.../...

l'ensemble de la phrase est court :

Le département /pl/ E.T.A. est spécialisé en acoustique
L'ordinateur /pl/ SCT va vous répondre .

Dans le premier exemple, on observe une pause de 95 ms et dans le second cas une pause de 135 ms. Dans l'une et l'autre phrase, on trouve aussi une pause située immédiatement avant le verbe, mais celle-ci correspond davantage au rôle de découpage de l'énoncé en groupes syntaxiques ; par contre les deux pauses /pl/ dont nous parlons sont des pauses d'insistance, de mise en relief du sens ; en l'occurrence, il fallait transmettre à l'auditeur le sigle énoncé ; pour ce faire, en plus de la pause préparatoire, chacun des termes composant le sigle est bien détaché, bien articulé : l'énergie est forte et bien répartie sur toute la durée des réalisations.

Ces exceptions mises à part, nous n'avons pas observé d'autres pauses faisant suite à un mot en schéma intonatif descendant, à l'intérieur d'une phrase.

Toutes les autres pauses syntaxiques répertoriées dans le corpus sont associées systématiquement à un schéma intonatif montant et à un allongement de durée sur le segment final de la dernière syllabe. Ces occurrences simultanées seront utilisées par la suite pour la formulation des règles de traitement prosodique.

Dans une proposition simple, c'est-à-dire ne comportant pas de virgule avant la pause finale de phrase, les pauses susceptibles d'être rencontrées pour opérer une démarcation syntaxique des éléments du discours ne peuvent se manifester qu'en un nombre de points finis dans l'énoncé :

Ces points sont essentiellement :

- la fin du syntagme nominal sujet dans les phrases énonciatives de construction [sujet + verbe],

.../...

- la fin du groupe de mots qui précède le verbe dans les constructions énonciatives de type [complément circonstanciel + verbe + sujet], - tous les groupes de mots dont le terme est signalé à l'écrit par une virgule,

- la fin du syntagme verbal,

- la fin d'un groupe de sens dans le syntagme complément si celui-là est suivi d'un complément circonstanciel, ou d'une proposition subordonnée.

Mais la présence de ces pauses n'est pas systématique. Leur existence et leur longueur dépend d'abord du débit du locuteur : l'importance du temps de pause est en rapport inverse de celui de la vitesse de la parole - nombre de syllabes par seconde - (LUCCI, 1973 ; BOË,1976) ; elles dépendent ensuite de la longueur (nombre de mots) du message : dans un corpus lu, le locuteur semble anticiper la longueur, la complexité syntaxique du message pour prévoir et organiser avant le début de l'énonciation orale, la répartition globale des pauses. Ces pauses n'existent qu'avec la réalisation simultanée d'une montée intonative et d'un allongement de durée du segment final de mot. Si en ces points de l'énoncé, une montée intonative de faible amplitude est seule réalisée, sans allongement de durée concomittent, on n'observe pas de pause. Enfin, il ne peut exister de pause après le verbe que s'il en existe une auparavant dans le syntagme nominal sujet. On observe également que lorsque le locuteur accélère son débit, les pauses les plus rebelles à la suppression sont celles qui marquent le terme du syntagme situé immédiatement avant le verbe.

Notons enfin l'impossibilité d'existence pour les pauses dans quelques constructions syntaxiques :

- entre le syntagme nominal sujet et le syntagme verbal quand le syntagme nominal sujet est un pronom personnel (je, tu, il, nous...). Par contre, un sujet monosyllabique mais qui n'est pas un mot grammatical peut être séparé du verbe par une pause :

Il est venu ce matin.
Jean /pl/ est venu ce matin.

.../...

La première phrase est prononcée sans pause alors que dans la seconde, le sujet est séparé du verbe par une courte interruption de 66 ms.

- entre le verbe et un attribut, ou entre le verbe et l'adverbe qui lui succède immédiatement - sauf manifestation particulière d'insistance.

En définitive, nous avons cerné les tendances suivantes pour ce qui concerne les répartitions des pauses et leur durée dans un corpus lu :

1/ Les pauses dans les phrases énonciatives :

1-1- Pauses qui précèdent immédiatement le syntagme verbal :

Leur durée semble varier selon la longueur et la complexité des syntagmes qui lui sont subséquents ; elles peuvent être soit une pause de démarcation syntaxique /p/, soit une pause de reprise de souffle /P/.

La durée de la pause syntaxique en cette position est difficile à déterminer puisqu'il nous semble que ce n'est pas la longueur du syntagme nominal sujet qui l'influence mais plutôt l'estimation de ce qui va suivre, ou encore le débit du locuteur.

En effet, dans les phrases :

le petit chat /p1/ boit du lait.
son numéro /p'1/ est modifié.

/p1/ est de 55 ms,
/p'1/ est de 40 ms.

.../...

et dans les phrases

le gentil petit chat de la voisine /pl/ boit du lait.
le numéro de Monsieur DUPONT /p'l/ est modifié.

/pl/ est de 55 ms,
/p'l/ est de 120 ms.

Mais cette pause de démarcation syntaxique peut durer jusqu'à 380 ms quand la phrase est plus longue, par exemple dans la phrase suivante :

Les abonnés /pl/ pourront consulter le fichier de la documentation automatique.

Ces pauses n'existent - nous le répétons - que simultanément à une montée intonative de grande amplitude et à un allongement de durée sur le dernier segment qui les précède.

1-2- Pauses situées dans le syntagme post-verbal.

Plusieurs conditions doivent être remplies pour qu'aucune pause de nature syntaxique existe:

* le syntagme complément doit être composé d'un groupe de sens

. complété par un complément circonstanciel introduit par une préposition,
. complété par un autre groupe de sens introduit par /et/,
. complété par une proposition subordonnée.

* dans ces cas, à la montée intonative qui marque le terme du premier groupe de sens, doit correspondre un allongement de durée sur le dernier segment vocalique ou consonantique du mot.

.../...

le journaliste /p1/ travaille dans son bureau /p2/ avec le directeur

l'université /p'1/ va fermer ses portes /p'2/ pendant le mois de juillet.

Les pauses de type /p2/, /p'2/ répertoriées dans le corpus ont une durée moyenne de 80 ms.

Il convient cependant d'introduire une distinction :

- le groupe de sens est suivi d'un complément déterminatif se subordonnant au nom pour en limiter le sens:

un cor de chasse,
une tasse de lait,
une statue de bronze,

Ce complément du nom est très souvent introduit par [de].

- le groupe de sens est suivi d'une proposition ou d'une locution prépositive qui sert à introduire un complément qu'elle unit au mot complété par un rapport déterminé (rapport de lieu, de temps, de moyen...) ; la préposition peut également introduire un rapport d'appartenance entre le mot complété et le complément :

le jardin de mon père,
le CNET de Lannion,

La préposition [de] - en particulier - pose des problèmes : en effet, quand elle sert d'introduction à un complément déterminatif, elle n'est jamais précédée d'une pause (sauf insistance particulière), on observe simplement en fin de mot complété une montée de F_0 de faible amplitude et une durée sans allongement sur le dernier segment de mot ; par contre quand cette préposition sertà lier deux groupes de mot par un rapport privilégié, une pause peut exister. Les frontières de démarcation entre les deux fonctions de cette préposition étant extrêmement

.../...

floues, on peut imaginer les difficultés que son existence provoquera dans un système automatique de Reconnaissance - Synthèse au moment de l'analyse syntaxique du message et du positionnement des marqueurs prosodiques (cf. IVe partie).

Si le débit du locuteur est accéléré, la pause du syntagme complément disparaît, il ne subsiste qu'une montée intonative de faible amplitude sans l'allongement de durée caractéristique du segment de dernière syllabe.

1-3- Pauses situées en fin de syntagme verbal.

Leur existence dépend de la suite du message :

a/ si l'on a [syntagme verbal + syntagme complément simple] du type "le chat mange des souris blanches",

on observe toujours une montée intonative sur la dernière syllabe du verbe

. soit de faible amplitude et sans allongement de durée donc non suivie de pause.

. soit de forte amplitude et avec allongement de durée du dernier segment ; dans ce cas on constate la présence d'une pause.

L'existence de l'une ou de l'autre de ces situations dépend du débit du locuteur.

b/ si l'on a [syntagme verbal + groupe de sens dans le syntagme complément + complément introduit par une préposition ou proposition subordonnée ...],

.../...

- on n'observe plus de pause en fin de syntagme verbal ni de montée intonative et d'allongement de durée sur la syllabe finale de mot de ce syntagme. Nous verrons (IIIe partie, Chapitre II) que la présence d'un complément introduit par une préposition ou une subordonnée bouleverse les traits prosodiques du syntagme verbal qui le précède.

2/ Les pauses dans les phrases impératives:

Les pauses qui correspondent au découpage de la phrase en groupes syntaxiques répondent aux mêmes normes que celles observées dans les phrases de type énonciatif : elles succèdent à un mot à intonation montante dont le dernier segment connaît un allongement de durée caractéristique. On les rencontre essentiellement en fin de syntagme verbal.

Veuillez indiquer/p1/ le code postal !
Répétez lentement /p'1/ votre question !

Cette pause est brève : environ 40 ms quand le complément qui suit est court.

Elle est plus longue (en moyenne 160 ms) quand le complément subséquent est long, par exemple :

Veuillez appeler/p/ le 842-17-18 à PARIS !

Il faut bien voir que cette pause résulte d'un <u>choix</u> ; elle peut ne pas exister, auquel cas on constate une montée intonative de plus faible amplitude et une durée de dernier segment non modifiée. Comme dans les phrases énonciatives, si le complément d'objet qui suit le syntagme verbal est suivi d'un autre complément ou d'une subordonnée, la pause qui intervient entre les deux groupes de sens dans ce syntagme élimine celle du syntagme verbal ; il se produit en même temps et de ce fait une modification du sens de la pente de F_0 sur la dernière syllabe du syntagme verbal (le schéma intonatif en fin de syntagme verbal devient descendant) :

.../...

"Veuillez expédier votre réponse /p/ avant de partir!"

Un phénomène identique est provoqué quand une virgule intervient pour séparer les éléments linguistiques dans le syntagme complément :

Enoncez vos noms, /p1/ qualité /p2/ et profession !
Appelez l'opératrice, /p'1/ ensuite posez votre question !

Dans le premier exemple, la pause est de 230 ms, elle est de 95 ms dans le second exemple.

3/ Les pauses dans les phrases interogatives:

Les pauses, quand elles existent, n'apparaissent qu'après un mot dont la syllabe finale possède un schéma intonatif montant.

3-1- Dans les phrases avec inversion [verbe + sujet] , les pauses sont situées en fin de syntagme verbal (on considère comme faisant partie du syntagme verbal, le pronom personnel sujet qui le suit immédiatement:"Recommencerez-vous"...)

Leur durée est en moyenne de 80 ms:

Voulez-vous relire /p1/ le dernier paragraphe de votre lettre ?
Avez-vous essayé /p1/ d'appeler un autre numéro ?

3-2- Dans les phrases introduites par un mot ou un groupe de mots interrogatifs, ce sont ces derniers qui sont suivis d'une pause :

quand /p/ comptez-vous partir ?
qui /p/ a frappé Paul ?
dans quelle ville /p/ habitez-vous ?

La pause en ce cas est inférieure à 70 ms, elle est liée là encore à la durée du dernier segment qui la précède.

.../...

3-3- Enfin, dans les phrases construites selon le même schéma syntaxique que les phrases énonciatives du type [sujet + verbe + complément] , les pauses sont situées à la même place mais connaissent des durées inférieures. Pour pouvoir mesurer cette différence de durée entre les deux types de phrase, il aurait fallu procéder à l'enregistrement d'une phrase identique , l'une produite de façon énonciative, l'autre de façon interrogative. Nous ne l'avons pas fait, et ce serait fausser les résultats que de faire cet enregistrement maintenant puisque le locuteur dont il s'agit ici a conscience de ce qu'il désire obtenir ("l'hyperconscience des phénomènes étudiés" dont parlent LEON et MARTIN - 1969).

De façon générale, cette différence de durée s'explique par le débit : nous avons vu que dans ce type de phrase la durée des sons élémentaires est inférieure à celle relevée dans les phrases énonciatives ; d'autre part les phrases interrogatives sont plus courtes.

II-3- Les pauses de démarcation des mots à frontières vocaliques.

Ces pauses sont systématiques entre deux mots lexicaux quand leur frontière commune est constituée de deux éléments vocaliques , par exemple dans la phrase "les abonnés pourront consulter le fichier de la documentation /p/ automatique". On observe une pause de 40 ms entre [documentation] et [automatique] alors même que ces deux mots appartiennent au même groupe de sens.

Par contre ce phénomène ne se produit pas quand l'un des deux mots est un mot grammatical :

la police a␣arrêté des manifestants,

Il n'y a aucune pause entre [a] et [arrêté].

On constate l'existence de cette pause surtout à la fin du syntagme verbal quand celui-ci est suivi d'un complément simple (c'est-à-dire sans complément introduit par une préposition et sans subordonnée). Pourtant, nous avons dit que la fin de ce syntagme pouvait

.../...

se présenter dans cette construction syntaxique selon deux modèles prosodiques :

- soit avec un schéma intonatif montant de faible amplitude, sans allongement de durée, et sans pause.

- soit avec un schéma intonatif montant de grande amplitude, avec allongement de durée et avec pause.

Tous les exemples dans lesquels le syntagme verbal s'achève par une réalisation vocalique et est suivi d'un syntagme débutant par une voyelle, laissent apparaître une pause de courte durée : 55 ms en moyenne.

les médicaments /p1/ sont vendus /p2/ en pharmacie	/p2/ =	40 ms
l'association de cyclotourisme /p1/ comprend /p2/ énormément d'adhérents	/p2/ =	66 ms
le département /p1/ ETA /p2/ est spécialisé /p3/ en acoustique	/p1/ =	95 ms
	/p3/ =	66 ms
les passagerspour Philadelphie /p1/ sont attendus /p2/ à la passerelle	/p2/ =	53 ms

Le schéma intonatif en fin de syntagme verbal étant montant cette pause permet de réaliser un passage intonatif sans brusque rupture avec la voyelle du premier mot du syntagme suivant dont le niveau fréquentiel de départ est toujours plus bas.

Quand dans cette position, l'une des frontières de mot est consonantique, la pause est moins nécessaire puisque la consonne autorise, par ses caractéristiques intrinsèques, le passage sans discontinuités de F_0 entre les voyelles.

.../...

En définitive, il faut retenir que l'existence et la durée des pauses - dans un corpus lu - résultent d'un choix - conscient ou non:

* soit le locuteur sépare de façon très nette chaque groupe syntaxique et dans ce cas, il réalise simultanément en fin de groupe :

. une montée intonative de grande amplitude,

. un allongement de durée sur le dernier segment,

. une pause :

qui peut n'être qu'une pause syntaxique brève.

qui peut servir en même temps de pause syntaxique /p/ et de pause de respiration /P/ et ce seulement à certains endroits bien précis de l'énoncé : - en fin de message pour indiquer la finalité,

- dans la phrase, immédiatement avant le syntagme verbal, ou aux points signalés graphiquement par une virgule.

* Soit le locuteur ne réalise pas une séparation très marquée des groupes syntaxiques. Dans ce cas :

. la montée intonative en fin de groupe syntaxique connaît une variation moindre,

. la durée du dernier segment du groupe ne présente pas d'allongement,

. aucune pause ne conclut la fin du groupe.

C'est ce choix que nous avons dû faire en synthèse. Nous verrons que la première option permet d'appréhender plus facilement le sens d'un message synthétique ;c'est pourquoi nous l'avons adoptée.

.../...

Il semble donc que les pauses font partie intégrante des phénomènes prosodiques pour servir d'indice supplémentaire à l'actualisation de la structure syntaxique. C'est - énoncée en d'autres termes - la conclusion implicite qui se dégage des travaux de BUTCHER (1973).

CHAPITRE II

L'ANALYSE DE LA FREQUENCE FONDAMENTALE

I - LES CARACTERISTIQUES INTRINSEQUES DES SONS :

* Les consonnes voisées présentent durant leur réalisation une évolution de la fréquence laryngienne caractéristique. Le tracé de la variation de Fo trouve son origine dans la constriction ou l'occlusion momentanée du conduit vocal pendant la production de ces sons qui offre une résistance au passage de l'air, ce qui augmente la pression supraglottique et diminue l'effet Bernouilli.

Liée à un phénomène physiologique d'articulation, l'évolution particulière de Fo est sans valeur distinctive contrairement aux phénomènes supra-segmentaux (prosodiques) qui, eux, appartiennent au domaine de la langue (dans la mesure où ils sont structurés).

C'est CRANDALL (1925) qui, le premier, a constaté que la fréquence laryngienne des consonnes est plus basse que celle des voyelles. HOUSE et FAIRBANKS (1953) ont montré dans une étude sur les caractéristiques acoustiques des voyelles (environnement consonantique voisé CVC) que la première et la deuxième consonne ont en moyenne une fréquence fondamentale plus basse respectivement de 3% et 10% par rapport à celle de la voyelle; ils indiquent que Fo dans cette séquence présente un tracé circonflexe.

Toujours à propos des consonnes voisées, KIM (1968) et MOHR (1968, 1971) concluent, dans une étude comparative du russe, du coréen, de l'allemand et du chinois, que leurs caractéristiques n'appartiennent pas au code linguistique puisqu'elles existent systéma-

tiquement dans toutes les langues étudiées; d'autre part ces variations de Fo ne sont pas toujours significatives du point de vue perceptuel.

Des études systématiques ont été menées pour le français en particulier par LARREUR et BOË (1973), et BOË (1973) qui précisent l'importance théorique et pratique des caractéristiques secondaires par rapport à une analyse des faits prosodiques. Une double analyse est conduite : l'une concerne l'évolution de Fo dans la production des consonnes voisées dans un contexte vocalique CVCVC, l'autre s'attache à l'étude de ces variations dans la parole continue.

Les résultats montrent que la différence des tracés entre occlusives et constrictives sur des logatomes a tendance à disparaître dans la parole continue : les tracés se rapprochent de l'évolution relevée pour les constrictives, soit une amplitude de variation de l'ordre de 5 1/4 de tons sur 100 ms.

Ces valeurs sont en concordance avec celles que nous avons dégagées dans l'analyse du corpus. Nous avons surtout noté la très faible variation de la fréquence laryngienne durant la réalisation des liquides et des nasales.

Si l'on peut considérer que ces caractéristiques participent à la reconnaissance des sons (pour CHISTOVICH, 1969, elles constituent un indice pour éviter la confusion entre [b] et [m]), à la perception du voisement pour les occlusives (LARREUR et BOË, 1973) et à la qualité des constrictives (LARREUR et BOË, 1973), des tests que nous avons menés avec le vocodeur ont montré que les occlusives et les constrictives ne sont pas dégradées du point de vue de l'intelligibilité si l'on ne respecte pas l'ampleur de la concavité observée à l'analyse : il suffit de reproduire la forme globale du tracé (de type descendant-montant) et de faire en sorte que les valeurs de Fo de la consonne soient toujours inférieures, d'une part à la valeur la plus haute de la voyelle précédente, et d'autre part à la valeur la plus basse de la voyelle subséquente.

Seule l'intelligibilité des liquides et des nasales est nota-

blement dégradée si l'on s'avise d'introduire une rupture de la fréquence fondamentale égale à celle des autres consonnes pendant leur réalisation. C'est pourquoi à la synthèse, nous nous sommes attachée, sans chercher à reproduire exactement le tracé observé à l'analyse, à éviter toute possibilité de discontinuité de Fo au milieu d'une réalisation consonantique voisée lors de l'assemblage des diphones. D'ailleurs, la plupart des auteurs utilisent ou envisagent d'utiliser ces caractéristiques pour l'élaboration de règles prosodiques à la synthèse :ÖHMAN et LINDQVIST (1965), MATTINGLY (1966), NEMETH (1970, 1971), PAILLE (1971), VAISSIERE (1971).

* De la même façon, des études ont été menées pour isoler les caractéristiques secondaires des voyelles : BLACK (1949), PETERSON et BARNEY (1951), HOUSE et FAIRBANKS (1953), LEHISTE et PETERSON (1961), MOHR (1971), BOË et LARREUR (1974).

On constate en général pour les voyelles une fréquence laryngienne qui augmente en même temps que leur aperture diminue : [a] est la voyelle grave alors que [i] et [u] sont les voyelles les plus aiguës.

Les écarts relevés dans différentes langues sont compris en moyenne entre 3% et 5% selon la nature de la voyelle, mais peuvent atteindre 10% - en anglais notamment (LEHISTE et PETERSON). D'autre part, HOUSE et FAIRBANKS notent que les voyelles situées dans un environnement consonantique sourd présentent toujours une fréquence fondamentale plus haute que celles qui sont entourées de deux consonnes voisées : cette augmentation est d'au moins 5%. Pour le russe, POTAPOVA et BLOXINA, (1970 , cité par LEON, 1971) observent une différence de 4% entre la fréquence laryngienne de [a] et celle de [u] et [i].

Dans l'élaboration des règles prosodiques, nous n'avons pas tenu compte de cette fréquence intrinsèque; il faudrait procéder à des tests perceptuels, et, si ceux-ci révèlent une amélioration de la qualité de la parole synthétique, prévoir - ce serait extrêmement simple - cette correction.

.../...

II - L'ANALYSE DE LA STRUCTURE INTONATIVE DE LA PHRASE :

II-1- L'organisation générale des variations de Fo :

Nous utiliserons un système de notation proche de celui employé par MARTIN (1975) mais plus simple en ce sens que, au lieu de visualiser à la fois le signe de la pente, la longueur et l'amplitude de Fo, il ne donne une indication que sur le signe de la pente et sur l'amplitude :

- ⇧ et ⇩ caractérisent la syllabe finale d'un mot à schéma intonatif montant ou descendant suivi d'une pause.
- ↑ et ↓ caractérisent la syllabe finale d'un mot à schéma intonatif montant ou descendant mais non suivi d'une pause, ce qui donne implicitement une indication : l'amplitude de Fo est moins importante et la durée du dernier segment de mot n'est pas allongée.
- ⇑ et ⇓ signifient que la syllabe qui est caractérisée par ce marqueur pourrait posséder l'un ou l'autre schéma :
  - ↓↑ → faible amplitude de Fo, faible durée, pas de pause.
  - ⇩⇧ → forte amplitude de Fo, allongement du dernier segment, pause.

Pour établir des règles d'évolution de Fo, nous allons, pour les trois types de phrases étudiées, partir d'une structure syntaxique simple puis progresser vers des structures plus complexes. Nous appellerons règles des variations systématiques de Fo que nous avons retenues pour élaborer le traitement automatique de la prosodie. Les schémas intonatifs types sont présentés en annexe sous la forme de tracés réalisés à partir d'une analyse effectuée par vocodeur à 2400 eb/s.

II-1-1- Les phrases de type énonciatif :

Nous appellerons GN_1 (Groupe Nominal 1) les mots qui composent le syntagme situé avant le verbe et GN_2 (Groupe Nominal 2) les mots qui succèdent au verbe; nous regrouperons sous le terme groupe ver-

bal tous les éléments (auxiliaire, participe passé ...) qui définissent le verbe mais sans y inclure les infinitifs précédés d'une préposition : dans l'exemple "la bicyclette permet de se déplacer rapidement", on fixe le terme du syntagme verbal après [permet].

Le GN_1 et le GN_2 pourront se décomposer en ce que nous nommerons groupe de sens c'est-à-dire en un groupe de mots unis très directement et très profondément par le sens, les limites entre deux groupes de sens pouvant résulter de la présence d'une préposition ou d'une locution prépositive, de la présence d'une virgule, d'un pronom relatif, d'une conjonction de subordination ou de coordination.

En définitive, dans la phrase :
"le gentil petit garçon de la vieille dame est allé chercher son chien dans la montagne",
on distingue les groupes suivants :

le gentil petit garçon	: 1° groupe de sens du GN_1
de la vieille dame	: 2° groupe de sens du GN_1
est allé chercher	: syntagme verbal
son chien	: 1° groupe de sens dans GN_2
dans la montagne	: 2° groupe de sens dans GN_2

Nous indiquerons du signe (*) les points de l'énoncé qui dans le corpus analysé présentent des pauses.

A - Phrases GN_1 + verbe.

1) je mange. ↓ je vous écoute. ↓ j'en mange ↓

2) le chat ↑ * mange. ↓ Jean ↑ * mange. ↓

On constate dans les trois premiers exemples où le sujet est constitué d'un pronom un schéma intonatif que nous qualifierons de neutre sur le pronom sujet et une descente intonative sur la dernière syllabe de la phrase.

.../...

Dans les deux derniers exemples par contre, on observe une montée intonative sur la voyelle finale du sujet - nom commun ou nom propre -, une descente intonative sur la dernière syllabe de la phrase.

Les règles qui se dégagent de l'étude de ce type de phrase sont les suivantes :

a) Dans les phrases GN_1 + verbe où GN_1 est un pronom :
- règles obligatoires:
  . descente intonative en fin de phrase.
  . schéma intonatif neutre pour les pronoms précédant le verbe.
  . jamais de pause entre le pronom sujet et le verbe.
  . schéma intonatif descendant neutre sur tous les mots outils (articles, prépositions, pronoms ...). (par neutre, nous entendons la plus faible amplitude observable en fin de mot).
- il n'existe pas de règles facultatives qui dépendraient du débit du locuteur.

b) Dans les phrases GN_1 + verbe où GN_1 est un mot lexical :
- règles obligatoires:
  . montée intonative sur la syllabe finale de GN_1.
  . descente intonative en fin d'énoncé.
- règles facultatives:
  . allongement du dernier segment du mot composant GN_1 et pause entre GN_1 et le verbe (nous avons dit que la pause ne peut exister que si un allongement de durée est réalisé sur le segment final d'un mot).
  Cette différence systématique dans le schéma intonatif de dernière syllabe de GN_1 selon qu'il s'agit d'un mot grammatical ou d'un mot lexical, permet d'éviter les ambiguïtés homophoniques, par exemple l'évolution de Fo permet de distinguer entre [j'en mange] et [Jean mange].

B - Phrases GN_1 + verbe copule + attribut.

1) la luminosité * est épouvantable.
↑ ↓

.../...

votre réponse * n'est pas correcte.

2) il n'est pas intelligent.

nous sommes stupéfaits.

Dans les phrases de la première catégorie, on observe une montée intonative sur la syllabe finale du GN_1, ainsi qu'un allongement de sa durée, suivie ici d'une pause. La fin de la phrase est signalée par une descente de Fo.

Pour la seconde catégorie, on n'observe qu'une descente intonative en fin de phrase : le sujet et l'auxiliaire ont des schémas de Fo descendant neutre.

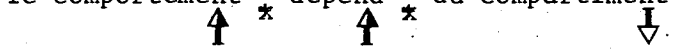
a) Dans les phrases du type 1 :
   - règles obligatoires:
     . montée intonative sur la syllabe finale de GN_1.
     . descente intonative en fin de phrase.
     . schéma descendant neutre sur le verbe copule, ou sur l'auxiliaire et sa négation.
     . absence de pause entre le verbe (auquel est associée la négation quand elle existe) et l'attribut.
   - règles facultatives, en fonction du débit du locuteur:
     . la pause après le dernier mot de GN_1 et l'allongement de durée de son dernier segment.

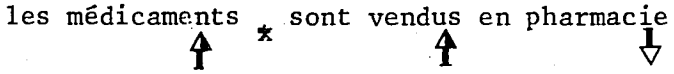
b) Dans les phrases de type 2 :
   - règles obligatoires:
     . descente intonative en fin de phrase.
     . schéma intonatif neutre sur les syllabes finales de tous les mots qui ne sont pas attribut.
     . absence de pause entre le sujet pronom et le verbe.
     . absence de pause entre le verbe copule et l'attribut.
     . la négation du verbe est incluse prosodiquement dans ce syntagme.

.../...

- Il n'existe pas de règles facultatives dans ce type de phrase.

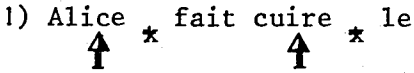
C - Phrases $GN_1 + GV + GN_2$.





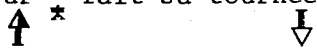
Dans tous les exemples de phrases constituées d'un syntagme sujet simple (sans épithètes ni groupes de sens), d'un groupe verbal, d'un syntagme complément simple (sans épithète ni groupe de sens), on observe une montée intontative sur la syllabe finale du sujet, un schéma de Fo descendant en fin de phrase, et une montée intonative sur la syllabe finale du groupe verbal.

Cependant il existe des exceptions à cette règle : nous avons vu déjà que les verbes copules ne présentent aucune montée intonative pendant leur réalisation, mais surtout tous les verbes transitifs à sens très général ne se dissocient jamais de l'élément linguistique qui les suit et qui leur apporte un supplément de sens. Ces verbes (particulièrement les verbes "faire", "aller", "pouvoir", "vouloir") sont totalement dépendants des éléments linguistiques subséquents (verbaux ou nominaux). De ce fait, le syntagme verbal ne connaît son terme qu'après un complément; tous les verbes transitifs qui ne peuvent être dissociés de leur complément se comportent de la sorte :

1) 

.../...

2) le facteur * fait sa tournée.



3) à Paris, j'ai rencontré Pierre.

Dans les énoncés comportant cette catégorie de verbe, la montée intonative observée en fin de groupe verbal n'existe qu'après le complément verbal ou nominal qui succède au verbe à sens très général; le verbe transitif indissociable de son complément (3), connaît quant à lui un schéma intonatif descendant neutre sur sa syllabe finale.

On peut énoncer les règles suivantes :

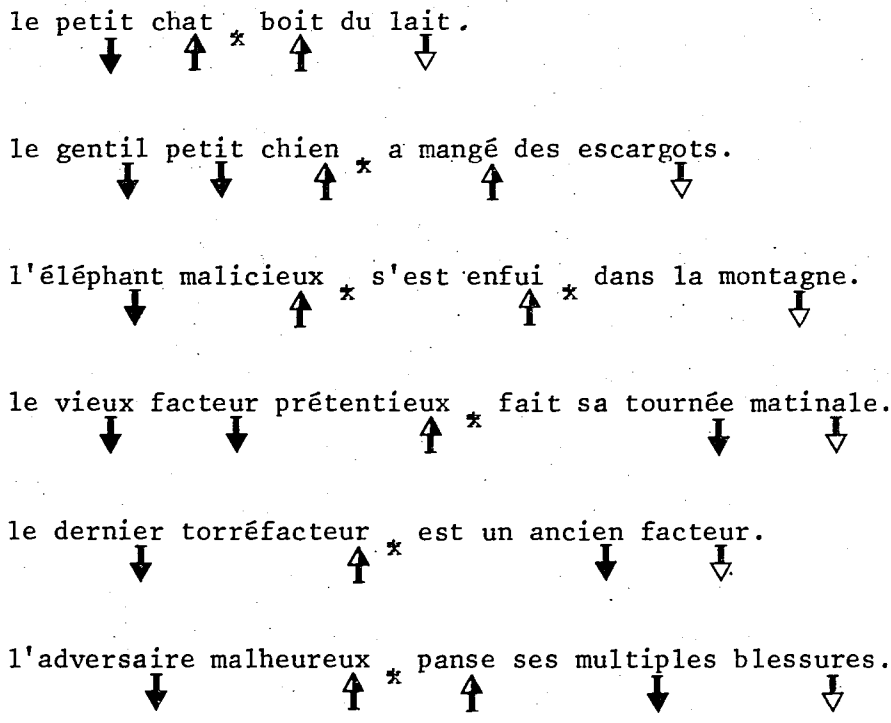
- règles obligatoires:
  . montée intonative sur la syllabe finale du dernier mot de GN_1.
  . descente intonative en fin de phrase.
  . montée intonative en fin de groupe verbal sauf dans les cas où celui-ci est composé d'un verbe transitif à sens très général ou d'un verbe transitif lié obligatoirement à un complément d'objet : dans ce cas, la syllabe finale du groupe verbal possède un schéma intonatif descendant neutre.
  . jamais de pause après le groupe verbal lié à son complément.
- règles facultatives:
  . le locuteur peut, selon son débit, réaliser ou non une démarcation du GN_1 et du groupe verbal par la production combinée d'un allongement de durée sur la dernière syllabe du GN_1 et d'une pause; il peut également - par la production des mêmes paramètres - réaliser ou non une démarcation entre le groupe verbal et le groupe nominal 2; mais il est bien évident qu'aucune démarcation n'est réalisable si le groupe verbal - de par sa nature grammaticale - connaît un schéma intonatif descendant.

D - Phrases $GN_1 + GV + GN_2$ dans lesquelles le GN_1 et le GN_2 comportent des épithètes.

Comme précédemment, nous n'indiquerons pour l'instant que le

.../...

signe de la pente de Fo pour la syllabe finale des mots; nous essaierons plus loin de procéder à l'analyse quantitative de ces variations.



Ces exemples montrent d'abord que les traits intonatifs déjà définis (montée intonative en fin de GN_1, descente de Fo en fin de phrase, montée intonative en fin de groupe verbal, sauf exceptions verbales catégorielles) ne sont pas modifiés par la présence d'épithètes successifs. Mais ces exemples montrent surtout que tous les épithètes, qu'ils soient situés dans le GN_1 ou dans le GN_2, qu'ils soient un ou plusieurs, ont un schéma descendant neutre et possèdent une durée de dernier segment égale à la durée d'un segment inclus dans un mot. On constate d'autre part que ce n'est pas la nature du mot (épithète, substantif ...) qui conditionne son évolution prosodique mais sa position dans l'énoncé (la même conclusion se dégage des travaux de OLIVE, 1975) : dans la phrase "le vieux facteur fait sa tournée", le substantif [facteur] parcequ'il est situé avant le groupe verbal, prend le schéma intonatif propre aux mots situés dans ce contexte; au contraire, dans la phrase "le vieux facteur prétentieux fait sa tournée matinale", c'est l'épithète [prétentieux] qui occupe la position propre à lui octroyer un schéma intonatif montant; le substantif [facteur] incorporé dans le syntagme nominal sujet n'a plus de situation intonative privilégiée, et la pente

.../...

mélodique de sa dernière syllabe a même signe que celle des autres mots inclus dans le syntagme. De la même façon, les épithètes du GN_2, s'ils sont situés à l'intérieur de ce groupe nominal ont tous un schéma intonatif de pente négative; ils prennent par contre en fin de phrase, le schéma intonatif spécifique aux mots situés en fin d'énoncé de type énonciatif : l'élément "matinal" dans la seconde phrase présente un schéma de Fo dans sa syllabe finale semblable au schéma final de l'élément "tournée" dans la première phrase.

- règles obligatoires:
  c'est la position dans l'énoncé, et non la nature grammaticale d'un élément linguistique qui détermine son schéma intonatif. Tous les mots intégrés dans un même groupe nominal (qui ne possède qu'un groupe de sens) ont un schéma de Fo de pente négative quel que soit le groupe nominal auxquels ils appartiennent (GN_1 ou GN_2).
- il n'y a pas de règles facultatives : il n'est pas possible d'insérer une pause entre les éléments du groupe, sauf manifestation d'une insistance particulière sur un des termes de l'énoncé : dans ce cas, ce phénomène se réalise par l'action conjuguée des trois paramètres de durée, d'intensité et de fréquence fondamentale.

E - Phrases GN_1 (composé de plusieurs groupes de sens) + GV + GN_2.

1) la magnanimité de l'éléphant * est reposante.

2) le gentil petit chat de la voisine * boit du lait.

3) le numéro des renseignements * est en dérangement.

4) le numéro de téléphone * de Monsieur Dupont * a été modifié.

5) les voyageurs pour Calcutta * sont attendus * à la passerelle.

.../...

6) les dernières informations de la justice * seront exposées *

à la bibliothèque.

7) la plupart des correspondants * en mission à l'étranger *

seront rapatriés par avion.

On a pu constater que, dans la plupart des exemples de phrases énonciatives dans lesquels le groupe nominal 1 connaît différentes expansions, toutes les syllabes finales de mot dans ce groupe possèdent un schéma intonatif de type descendant, et que les segments de fin de mot ont une durée identique qu'ils soient situés à l'intérieur du groupe de sens, ou en finale de ce groupe de sens, c'est-à-dire qu'ils ne possèdent pas d'allongement particulier dans le segment final de mot sauf quelques rares exceptions rencontrées dans notre corpus, par exemple dans la phrase suivante :
"les exemples de synthèse par diphonèmes présentés ici sont réalisés de façon entièrement automatique" : si l'on a effectivement observé des schémas intonatifs descendants sur toutes les syllabes situées en fin de mot dans le groupe nominal 1, on a relevé également la présence d'une pause de reprise de souffle dans ce GN_1 après l'élément [diphonèmes] ; cette pause est précédée de l'allongement caractéristique du segment de fin de mot en cette position, en l'occurrence, l'élément vocalique [ɛ].

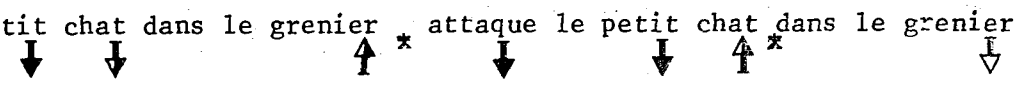
D'autre part, dans le 7° exemple, on relève l'existence d'une pause entre [correspondant] et [en mission] c'est-à-dire à la frontière des deux groupes de sens. Cette pause est courte (40 ms) et nous pensons que sa présence est due autant à l'existence d'une frontière vocalique entre les deux groupes [ã] + [ã] qu'à une volonté d'opérer la démarcation entre deux groupes de sens d'un même groupe nominal.

Mais nous le répétons, les pauses qui font suite à un schéma intonatif descendant non situé en fin de phrase, sont extrêmement rares -

.../...

tout au moins pour ce qui concerne ce locuteur.

Le corpus comprenait quelques phrases dans lesquelles le GN_1 comportait les mêmes éléments linguistiques que le GN_2, par exemple :



Dans de telles phrases, alors que l'on ne constate pas la présence de pauses dans le GN_1, on relève l'existence fréquente d'une pause subséquente à un schéma intonatif montant dans le GN_2.

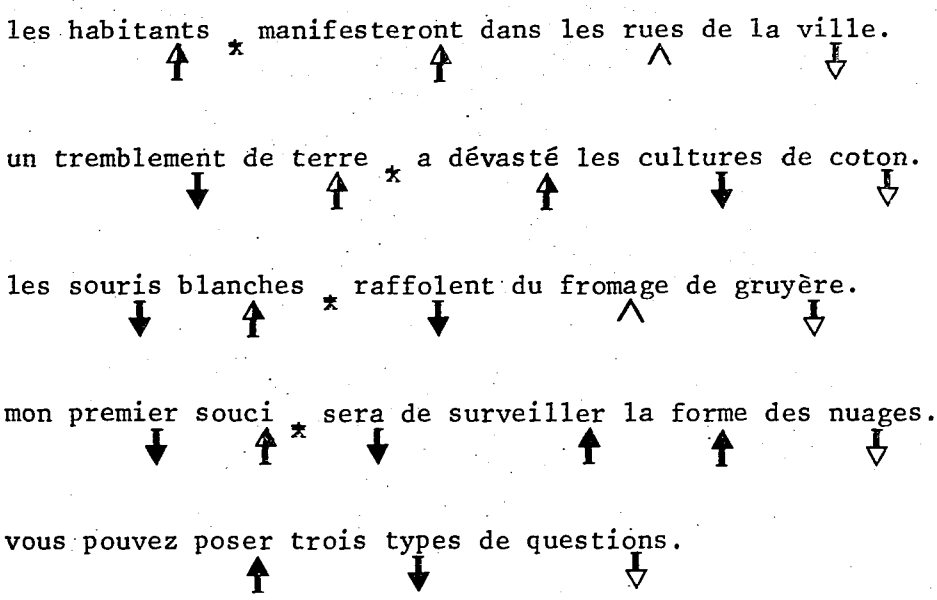
A partir de ces exemples, on peut tirer les conclusions suivantes :

- règles obligatoires:
  - . tous les mots appartenant au GN_1, qu'ils soient situés en fin de groupe de sens (non situé à la fin du GN_1) ou à l'intérieur d'un groupe de sens, possèdent un schéma intonatif descendant; si le signe de la pente est identique nous verrons que les niveaux de Fo et leur amplitude sont différents dans l'un et l'autre cas.
  - . il n'y a jamais de pause, sauf insistance particulière, entre deux mots quand le second est subordonné au premier pour en limiter le sens (exemple : un cor de chasse, un verre de vin).
- règles facultatives:
  - . des pauses peuvent apparaître dans le GN_1 entre deux groupes de sens quand le second est introduit par une préposition ou une locution prépositive (dans, depuis, chez, avec ...) ou quand il constitue une proposition subordonnée. Dans ce cas, les pauses succèdent naturellement à un schéma intonatif de dernière syllabe de type descendant et à un allongement de durée sur le dernier segment du mot.

F - Phrases GN_1 + GV + GN_2 (composé de plusieurs groupes de sens ou composé d'une subordonnée).

.../...

1) Le GN_2 se compose de deux groupes de sens dont le dernier constitue un complément déterminatif, en général [de].



Le signe ∧ indique que le schéma intonatif est d'abord montant puis descendant.

Ces exemples montrent les difficultés que présente la décomposition en mots intonatifs du GN_2 quand il présente ce type de structure.

Le premier trait constant qui se dégage c'est l'inexistence de pause entre la première partie du GN_2 et le complément de nom : le lien syntaxique et sémantique créé par un complément de nom est très profond et ne permet pas de parler de deux groupes de sens.
Pourtant les évolutions de Fo manifestent d'une certaine façon une démarcation entre un mot ou groupe de mots et le complément de nom.

On constate :

- soit une montée intonative de faible amplitude sur la syllabe finale du mot qui précède le complément de nom (la forme ↑ des nuages).
- soit un schéma intonatif de type montant-descendant sur la syllabe finale.
- soit un schéma descendant dont nous verrons cependant que les niveaux fréquentiels donnent la sensation auditive d'une montée intonative (exemple : les cultures ↓ de coton; trois types ↓ de questions).

.../...

On remarque dans certaines des phrases données en exemple un schéma descendant sur la syllabe finale du verbe :

les souris blanches raffolent ↓ du ...

Ce schéma descendant s'explique ici par la nature du verbe : on raffole de quelque chose; le verbe est par conséquent lié à son complément et n'a pas d'existence autonome.

Par contre l'explication est différente dans le phrase : "mon premier souci sera ↓ de surveiller ↑ la forme des nuages".
Dans cet énoncé, le GN_2 est composé de deux groupes de sens :

1° groupe : [de surveiller]
2° groupe : [la forme des nuages]

Nous verrons plus loin que lorsque le groupe nominal 2 est composé de deux groupes de sens, la dernière syllabe du groupe verbal subit une transformation du signe de sa pente mélodique.

En d'autres termes, dans une phrase de type "mon premier souci sera la forme des nuages", on observe un schéma montant sur la dernière syllabe du groupe verbal; par contre dans l'énoncé :"mon premier souci sera de surveiller la forme des nuages", le schéma montant de fin de groupe verbal est transformé en schéma descendant de par la présence de deux groupes de sens (en effet, la présence de plusieurs groupes de sens successifs modifient à nouveau ce schéma).

En définitive, les règles obligatoires qui se dégagent le plus aisément dans les énoncés de ce type concernent :

- la production d'un schéma de Fo qui donne la sensation sonore d'une montée intonative avant le complément déterminatif.
- l'impossibilité d'une pause quand le second groupe de mots est réellement subordonné au premier pour en limiter le sens.
- la durée du segment dans le dernier mot du premier groupe qui ne subit pas d'allongement.

.../...

Une certaine liberté semble quand même exister pour réaliser la démarcation syntaxique des deux groupes de mots puisque l'on observe soit un schéma intonatif montant, soit un schéma de type montant-descendant, soit une descente de Fo mais à des niveaux fréquentiels supérieurs à ceux des syllabes précédentes du mot, ce qui laisse supposer une équivalence à la perception avec les autres types d'évolution de Fo.

2) Le GN_2 comporte deux groupes de sens dont l'un est un complément introduit par une préposition ou par une locution prépositive.

j'ai rencontré Pierre * à Paris.

le facteur * fait sa tournée en bicyclette.

Nous avons déjà noté que dans ce type de phrase, on ne peut pas dissocier le groupe verbal de son complément. On comprend donc que le verbe seul connaisse dans sa syllabe finale un schéma intonatif descendant puisque le schéma intonatif montant qui lui est normalement attribué n'apparaît qu'à la fin de son complément (premier groupe de sens de GN_2).

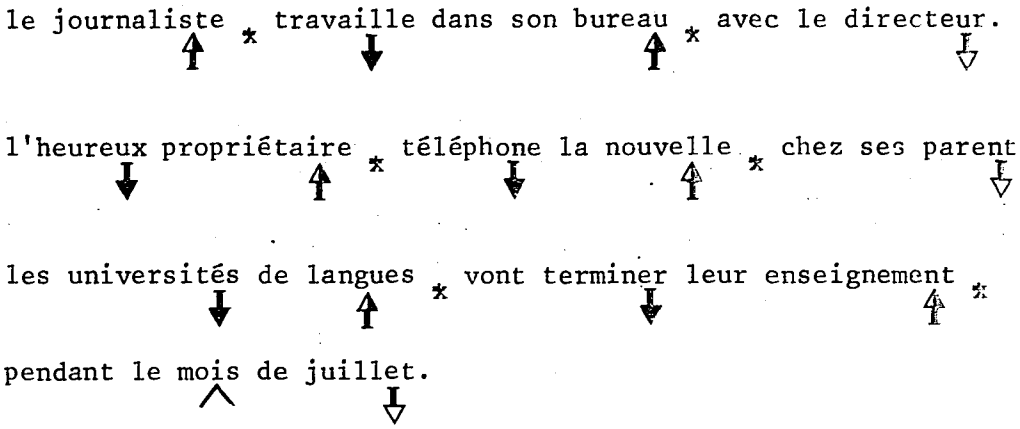
La pause entre les deux groupes de sens de GN_2 est facultative.

Les phrases suivantes comportent également un complément introduit par une préposition :

de gigantesques inondations * recommencent à s'abattre sur le pays.

le vilain corbeau fatigué * tenait dans son bec * un camembert usagé.

.../...



Dans toutes les phrases du corpus dont nous donnons ici quelques exemples, on observe une montée intonative sur la dernière syllabe du mot qui marque le terme du premier groupe de sens. Nous insistons sur le fait que cette observation ne concerne que les phrases dont le GN_2 possède deux groupes de sens.

D'autre part, alors que les phrases composées d'un seul groupe de sens dans GN_2 présentent un schéma intonatif montant sur la dernière syllabe du groupe verbal, les énoncés de cette catégorie (deux groupes de sens) font constater une modification du signe de la pente intonative sur la syllabe finale du groupe verbal : dans tous les cas, cette pente est négative.

Pour ce type de phrases, on peut énoncer les règles suivantes :

- règles obligatoires:
  - la montée intonative en fin de GN_1 demeure, ainsi que la descente intonative de fin de phrase énonciative.
  - la syllabe finale du premier groupe de sens présente toujours une montée de Fo.
  - le signe de la pente intonative pour la syllabe finale du groupe verbal est transformée : la pente est toujours négative.
  - il n'y a pas de pause entre le groupe verbal et le GN_2.
- règles facultatives:
  - on observe souvent dans un corpus lu une pause qui assure le découpage de la structure syntaxique entre les deux groupes

.../...

de sens. La pause n'existe que si un allongement de durée caractérise le segment final du premier groupe de sens.

3) Le GN_2 est composé de plus de deux groupes de sens:

a) les moyens modernes de communication * permettent de se déplacer avec rapidité * et facilement.

b) le génie des alpages * a envoyé les moutons chez des amis * au bord de la mer.

Dans les deux exemples le GN_2 comporte trois groupes de sens :

Phrase 1
- de se déplacer
- avec rapidité
- et facilement

Phrase 2
- les moutons
- chez des amis
- au bord de la mer

Cette construction fait revenir au schéma intonatif d'origine, quand le groupe verbal n'est suivi que d'un seul groupe de sens : la syllabe finale du groupe verbal présente un schéma intonatif de type montant.

En fait, on peut résumer de façon très simple les contours finals des groupes de sens dans le GN_2, puisque l'on observe, quel que soit le nombre des groupes de sens dans ce syntagme, une alternance parfaite entre schémas montants et schémas descendants depuis le syntagme verbal jusqu'à la fin de la phrase.

On a les contours suivants :

.../...

	groupe verbal	GN_2
1 groupe de sens	↑	↓
2 groupes de sens	↓	↑ 1° groupe ; ↓ 2° groupe
3 groupes de sens	↑	↓ 1° ; ↑ 2° ; ↓ 3°
4 groupes de sens	↓	↑ 1° ; ↓ 2° ; ↑ 3° ; ↓ 4°

Signalons enfin que ces groupes de sens peuvent très bien être suivis de propositions subordonnées; chacune des propositions réalisant :

. la décomposition en groupes nominaux et groupes verbaux,
. et l'alternance dans le signe de la pente de Fo,

pour les éléments qui la composent :

1) je t'ai amené mon pays par la main * pour le planter dans ton jardin.

[je t'ai amené mon pays par la main] = proposition principale non située en fin de phrase ⟶ schéma de Fo indiquant la continuité ↑ ou ↑ .

[pour le planter dans ton jardin] = proposition subordonnée composée de deux groupes de sens : le schéma descendant en fin de phrase impose un schéma montant sur le premier groupe de sens de la subordonnée.

2) le petit chat * a mangé des souris qui se cachaient dans les greniers * et qui mangeaient tous les papiers.

.../...

Les frontières essentielles de l'énoncé sont représentées par les pauses :

- pause qui met un terme à une proposition finale : dans ce cas la pause fait suite à un schéma intonatif descendant et à un allongement de durée sur le dernier segment de mot.
- pause qui conclut une proposition non finale; elle succède à un schéma intonatif montant et à un allongement de durée sur le dernier segment de mot.

On voit dans le premier exemple cité que la première proposition composée d'un groupe verbal et de deux groupes de sens dans GN_2 :

. [mon pays]
. [par la main]

présente des schémas de Fo différents de ceux énoncés plus haut pour la même structure syntaxique; c'est qu'ici, le second groupe de sens n'est pas en position terminale, il est suivi d'une autre proposition.

Par conséquent, dans le cas d'une proposition non située en fin d'énoncé, on trouve toujours sur la syllabe finale du dernier groupe de mots, un schéma intonatif de type montant; et pour les groupes qui précèdent - mais seulement à partir du groupe verbal car le GN_1 n'est pas modifié - on observe des schémas inverses de ceux constatés dans une proposition de fin d'énoncé.

C'est le même principe dans le second exemple :
on est en présence :

. d'une propositon principale non terminale,
. d'une proposition relative non finale,
. d'une proposition relative finale.

Il n'est pas nécessaire de procéder à une analyse syntaxique de l'énoncé pour connaître les schémas de Fo : toutes les propositions non finales présentent les mêmes évolutions de Fo qu'elles soient principales ou subordonnées et toutes les propositions finales ont un schéma unique. Encore une fois, ce n'est pas la nature syntaxique d'un mot ou d'un groupe de mots qui détermine ses patrons prosodiques, c'est sa

.../...

position dans l'énoncé.

Pour une proposition non située en fin de phrase, les schémas intonatifs se succèdent comme suit :

GN_1	G. verbal	GN_2	
↑ 1) le voyageur 2) les voyages	↓ a amené attristent	+ 1 groupe de sens	↑ son pays parce que ... → 2) les individus qui ...
↑ 1) le voyageur 2) ses amis	↑ a amené voient	+ 2 groupes de sens	↓ ↑ son pays par la main pour que..2) les changements dans la politique que ...
↑ le voyageur etc...	↓ a amené	+ 3 groupes de sens	↑ ↓ ↑ avec lui son pays par la main ...

G - GN_1 (composé d'une subordonnée) + GV + GN_2.

la modestie qui se plaît à être louée * est un orgueil secret.
↓ ↓ ↑ ↓ ↓ ↓

1) Il s'agit ici avec [qui se plaît à être louée] d'une subordonnée relative complément de nom. Elle se décompose en deux groupes :

. [qui se plaît]

. [à être louée]

Proposition non finale, elle se comporte exactement comme les autres propositions de ce type que nous avons décrites par ailleurs; elle

.../...

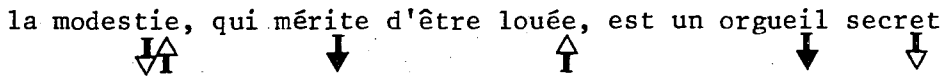
contient :

. un schéma de Fo montant sur la syllabe finale,

. et, de par l'alternance du signe de la pente de Fo, un schéma de type descendant pour la fin du premier groupe.

Quant au mot lexical [la modestie] qui appartient au GN_1, il possède obligatoirement dans ce type de phrase un schéma de type descendant, et ne peut en aucun cas être suivi d'une pause.

2) Par contre, si la subordonnée est écrite entre deux virgules (subordonnée relative détachée), le mot lexical sujet [la modestie] peut présenter un schéma de Fo soit montant soit descendant, il est obligatoirement suivi dans l'une ou l'autre réalisation de Fo d'une pause :



De la même façon, la fin de la subordonnée est suivie d'une pause. Ce double système de pause :

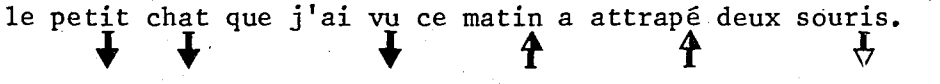
. permet de conserver à la proposition principale sa cohésion.

. permet d'éviter la confusion avec une subordonnée complément de nom.

1') C'est exactement le même processus dans une phrase de type :

"le petit chat que j'ai vu ce matin a attrapé deux souris"

on aura le schéma suivant :



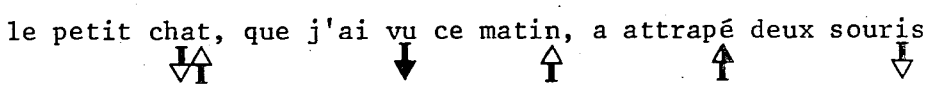
(= ce matin j'ai vu un chat; il a attrapé deux souris)

Le GN_1 possède, sauf à son terme, un schéma intonatif de type descendant parce que la subordonnée est simplement complément de nom;

.../...

le GN_1 se décompose donc simplement en deux groupes de sens; il présente naturellement un schéma de Fo montant à sa frontière avec le syntagme verbal.

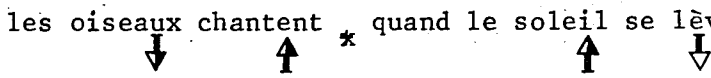
2') Par contre quand on a :

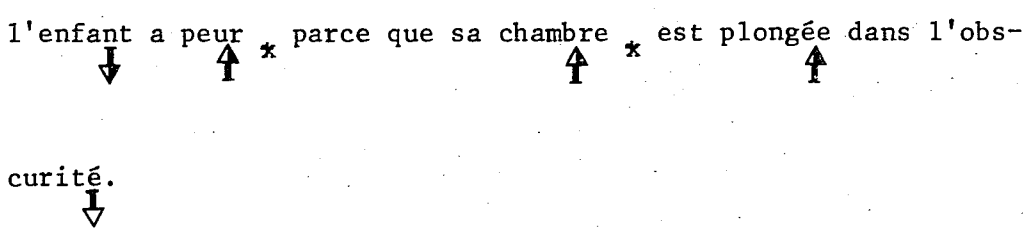


(on parle d'un chat que l'on connaît)

la subordonnée forme une entité, elle présente les schémas intonatifs de toutes les propositions non finales de phrase. Simplement et de façon obligatoire elle s'achève par une pause qui marque que cette proposition est détachée. Comme dans les exemples vus plus haut, le GN_1 [le petit chat] peut se terminer soit par un schéma montant soit par un schéma descendant mais il est suivi de toute façon d'une pause.

H - GN_1 + GV + proposition subordonnée.





Ces phrases présentent une suite de deux propositions :

- une proposition non finale,
- une proposition finale.

La première se comporte exactement comme toutes les propositions non situées en fin d'énoncé : un schéma montant termine son dernier groupe (en l'occurrence, ici, le groupe verbal) un schéma descendant achève le groupe syntaxique qui le précède immédiatement (GN_1).

.../...

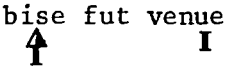
Quant à la seconde proposition, son comportement prosodique est celui d'une proposition autonome qui conclue un énoncé :

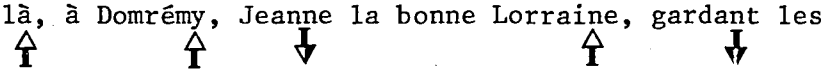
. Fo montant en fin de GN_1.

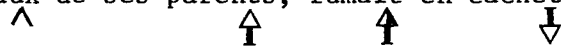
. Fo montant en fin de groupe verbal si GN_2 ne possède qu'un groupe de sens.

. Fo descendant en fin de phrase.

Les règles dégagées dans les structures $G_{2+2'}$ et H sont regroupées dans les exemples suivants :







A partir du moment où l'on a dégagé les frontières des groupes de mots formant les grandes unités de significations de la phrase, il est très facile de déduire les patrons prosodiques des mots qui les composent.

Dans la première phrase par exemple, on trouve <u>trois</u> unités non fin de phrase :

. la cigale,

. ayant chanté tout l'été,

. se trouva fort dépourvue,

et une unité de fin de phrase :

. quand la bise fut venue.

Dans les trois unités non fin de phrase, seule la première (la cigale) est susceptible de présenter deux schémas intonatifs, l'un montant

l'autre descendant. Mais l'un et l'autre étant acceptables, cela ne pose pas de problèmes pour le traitement de la prosodie à la synthèse; on décidera arbitrairement que "tous les groupes situés avant une virgule" possèderont un schéma de Fo montant sur la dernière syllabe du mot.

A l'intérieur de ces unités, l'alternance des schémas mélodiques se réalisera normalement sur les points clés déjà définis : fin de GN_1, fin de groupe de sens dans GN_1; fin de groupe verbal, fin de GN_2, fin de groupe de sens dans GN_2; l'alternance ne se réalise pas à un niveau inférieur à ceux-ci.

L'alternance des schémas existe sur les mêmes points clés que l'unité de fin de phrase, mais simplement elle est commandée par la pente négative de Fo en fin de phrase : donc le groupe situé immédiatement à sa gauche se terminera par un schéma de type montant :
[fumait ↑ en cachette].

I - Les schémas intonatifs des propositions de même nature qui sont reliéespar une conjonction de coordination.

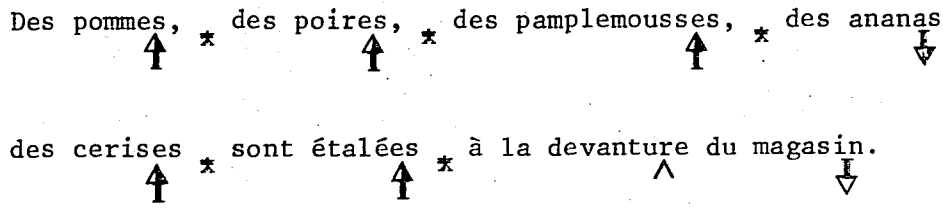
la tempête s'éloigne ↓ * ↑ et les vents sont calmés. ↑ ↓

Le principe est toujours le même : on dégage de cet énoncé deux grandes unités. L'une, non finale, se termine donc par un schéma montant qui ordonnance l'alternance pour les unités du niveau inférieur qui la composent. L'autre, finale, se termine par un schéma de Fo descendant et règle pour les unités du niveau inférieur qui sont en elles l'alternance dans le signe de la pente de Fo; il n'est pas besoin de faire référence à la nature syntaxique des mots.

J - Les schémas intonatifs des énoncés qui présentent une énumération.

Les éléments de la succession sont séparés graphiquement par une virgule, sauf les deux derniers (qui sont reliés par la copulative [et]).

.../...



Dans les contructions de ce type, on constate que les éléments du GN_1 peuvent présenter deux évolutions intonatives différentes :

- ou bien chaque unité est envisagée comme partie du tout, c'est-à-dire comme une unité d'un ensemble plus vaste, comme un groupe de sens d'un groupe de niveau supérieur non final (GN_1 en l'espèce), et dans ce cas, chaque unité (sauf la dernière) présente un schéma de Fo de type descendant, le plus souvent suivi d'une pause; la dernière unité, fin de l'unité englobante, possède un schéma de Fo de type montant.

- ou bien chacune de ces unités est conçue comme une entité autonome, comme formant à elle seule un groupe de niveau supérieur non final, et dans ce cas, chaque unité déjà démarquée par une virgule présente un schéma de type montant, sauf l'avant dernier terme de l'énumération qui lui, marque son appartenance, son lien, à la dernière unité par la conjonction de coordination [et] . Cet avant dernier terme constitue donc un groupe de sens du GN_1 et de ce fait subit l'alternance de la pente de Fo : par opposition au schéma montant de fin de groupe non final, il présente un schéma de Fo de pente négative.

Quand une telle énumération est située dans le GN_2, les possibilités de variation de Fo sont absolument identiques :

tous les termes de l'énumération non reliés par une conjonction de coordination peuvent avoir soit un schéma montant, soit un schéma descendant (soulignons que chaque terme a le même signe de la pente : on n'a jamais [... des pommes, des poires, des pamplemousses, ...]); mais l'avant dernier terme de l'énumération présentera obligatoirement un schéma de Fo montant par contraste avec l'unité finale à laquelle il est lié.

.../...

Il faut signaler cependant que le plus souvent, le schéma intonatif des termes d'une énumération dans le GN_2 est de type montant, ce qui provoque une inversion du schéma de Fo pour le groupe verbal (schéma descendant); nous rappelons que le schéma intonatif du verbe est toujours inversé par rapport au schéma du groupe de sens qui lui succède.

Il existe malgré tout des exceptions à cette règle d'inversion de la pente de Fo en fin de syntagme verbal:

"Vous pouvez appeler le 842 - 17 - 18."

dans ce type d'énoncé, on constate une montée intonative en fin de syntagme verbal, ce phénomène s'explique sans doute par une insistance particulière comme si après "vous pouvez appeler" il y avait le signe graphique [:]

"Vous pouvez appeler : le 842, dix sept *, dix huit."

dans ce cas, la fin du groupe verbal outre la montée de Fo, présente obligatoirement une pause.

Ensuite, chacune des unités qui composent le numéro de téléphone est considérée d'égale importance sémantique et possède - sauf la dernière - un schéma de Fo montant.

Par contre dans l'énoncé suivant :

"Vous pouvez appeler le 842, 17, 18 à Paris."

on peut trouver deux schémas de Fo :

- soit le GN_2 se décompose en deux groupes de sens, l'un correspondant au numéro de téléphone, l'autre relatif à l'indication de lieu; les deux informations sont considérées comme ayant le même intérêt sémantique : chacun des groupes de chiffres représentant le numéro n'est qu'un élément du premier groupe informationnel et possède alors

.../...

un schéma de Fo descendant comme tous les mots non situés à un point clé; seul le dernier élément du nombre présente un schéma intonatif montant spécifique des fins de groupes en position non fin de phrase.

- soit le GN_2 se décompose - de par une manifestation d'insistance - en quatre groupes de sens <u>successifs</u> situés au même niveau et dans ce cas, on a des schémas intonatifs qui suivent les règles dégagées pour les énoncés qui présentent une énumération.

En conclusion, ce que tous ces exemples ont montré, c'est une alternance remarquable dans le signe de la pente de Fo sur <u>trois niveaux</u> successifs de décomposition d'un énoncé.

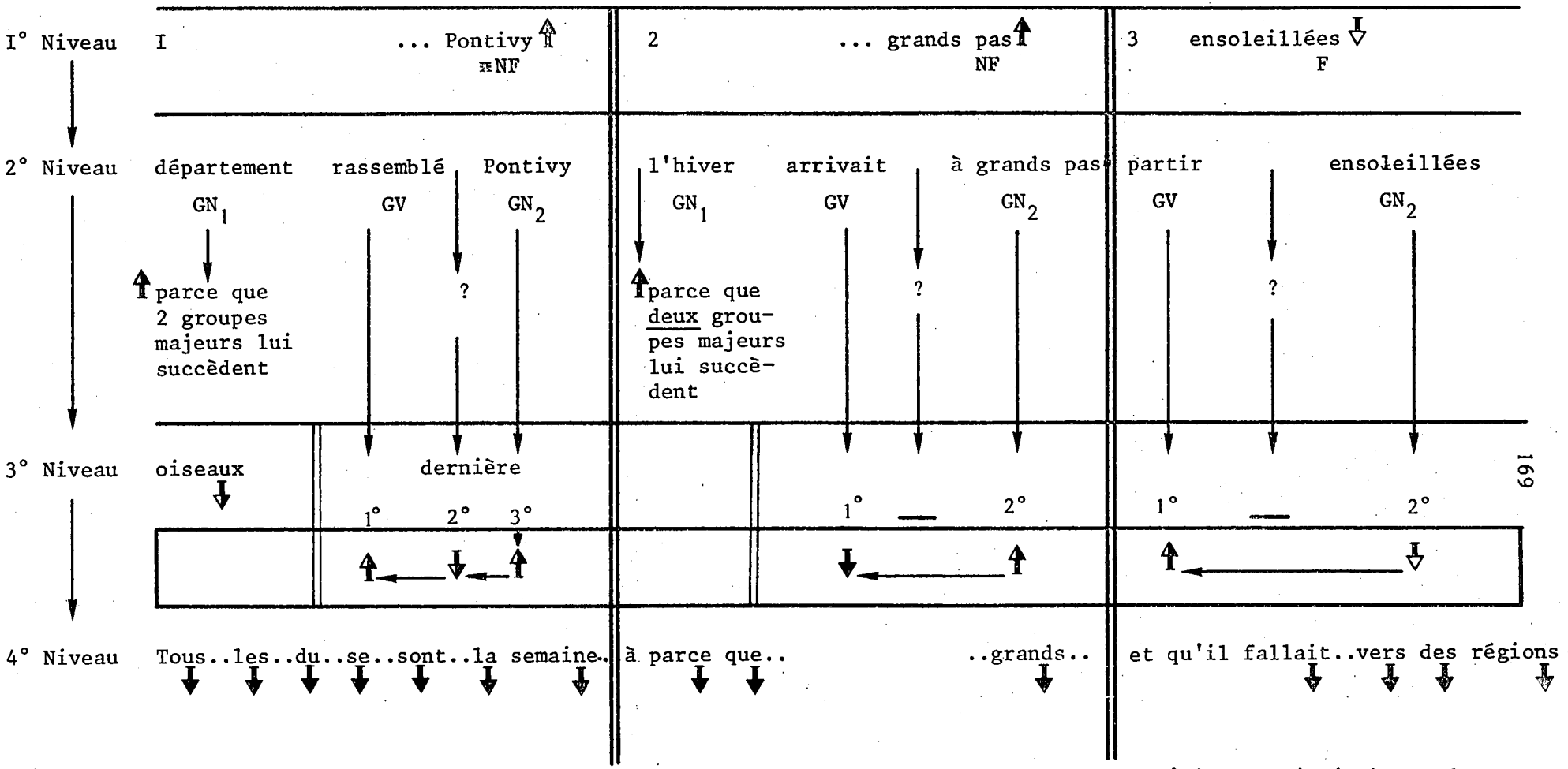
1°) niveau de la proposition : finale ↓ / non finale ↑.
2°) niveau des groupes majeurs : GN_1 / G. verbal / GN_2.
3°) niveau des groupes de sens des groupes majeurs,
[si groupe majeur = ↑ ⟶ groupe de sens = ↓]

Il existe un quatrième niveau qui se situe sur n'importe quel point de l'énoncé et qui présente une évolution identique de Fo : un schéma de type descendant.

Soit l'exemple suivant :

"Tous les oiseaux du département se sont rassemblés la semaine dernière à Pontivy parce que l'hiver arrivait à grands pas et qu'il fallait partir vers des régions ensoleillées".

.../...



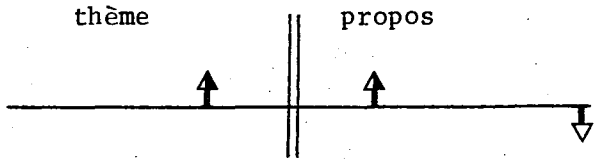
Tous les oiseaux du département se sont rassemblés la semaine dernière à Pontivy parce que l'hiver arrivait à grands pas et qu'il fallait partir vers des régions ensoleillées

* NF = non final ; F = final

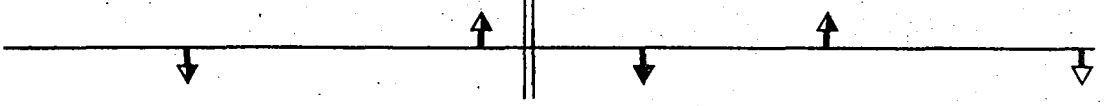
.../...

En fait, dans un énoncé où le groupe verbal est entouré de groupes nominaux, le groupe verbal joue le rôle "d'équilibreur prosodique" entre le thème et le propos :

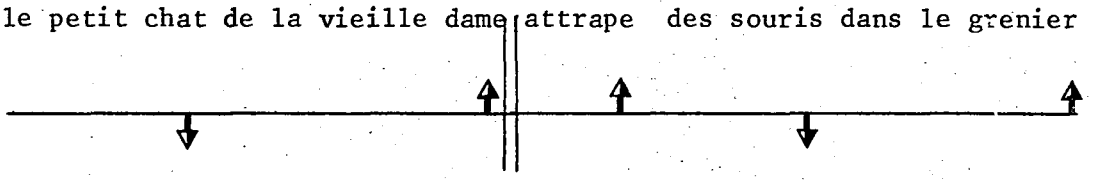
1) le petit chat attrape des souris.



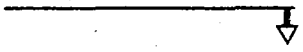
2) le petit chat de la vieille dame attrape des souris dans le grenier.



3) le petit chat de la vieille dame attrape des souris dans le grenier



avant chaque repas.



En conclusion pour fixer les règles de prosodie d'un énoncé il faut et il suffit de connaître dans l'ordre suivant :

- les frontières du groupe verbal,
- les limites des groupes de sens dans les syntagmes pré- et post-verbaux,
- et bien sûr les frontières de chaque mot.

Dans un système de "reconnaissance - synthèse" automatique l'analyse syntaxique pourra donc être limitée à ces déterminations.

II-1-2- Les phrases de type impératif :

Ces phrases suivent exactement les mêmes règles prosodiques que les énoncés de type énonciatif et présentent la structure syntaxique suivante :

Groupe verbal + groupe nominal$_2$.

La règle générale est la suivante :

. schéma intonatif montant en fin de syntagme verbal ↑.

. schéma intonatif descendant en fin de phrase.

Appelez l'opératrice!
↑ ↓

Présentez-vous au département indiqué!
↑ ↓

Veuillez indiquer * le code postal!
↑ ↓

Renouvelez votre demande!
↑ ↓

Mais dans les phrases syntaxiquement plus complexes, le décodage des réalisations intonatives devient plus difficile parce que les phrases impératives impliquent, par leur nature même, l'intervention de l'auditeur. Il ne s'agit pas seulement d'énoncer, il s'agit surtout (entre autre par l'intermédiaire de la modulation intonative) de présenter l'énoncé à l'auditeur de telle façon que celui-ci comprenne immédiatement ce qui lui est demandé pour réagir exactement à l'ordre. Pour ce faire, l'énoncé peut être découpé en groupes prosodiques différents selon que l'insistance est mise sur le groupe verbal ou sur un élément lexical du GN_2.

1) Veuillez répéter le numéro de votre correspondant!
↑ ↓ ↓

2) Veuillez répéter le numéro de votre correspondant!
↓ ↑ ↓

3) Veuillez répéter le numéro de votre correspondant!
↑ ↑ ↓

.../...

Quand l'insistance se manifeste essentiellement sur le groupe verbal, celui-ci outre la montée intonative sur la syllabe finale de son dernier élément, présente un allongement final, une augmentation d'intensité ainsi qu'une pause de courte durée (entre 40 ms et 100 ms en moyenne selon la longueur de GN_2).

Si l'insistance porte à la fois sur le groupe verbal et sur un élément de GN_2, on constate une montée intonative sur l'une et l'autre unité, chacune pouvant être ou non suivie d'une pause.

Enfin, le locuteur peut insister uniquement sur un élément de GN_2 (exemple 2). Cela signifie : ce que l'on vous demande de répéter, ce ne sont pas toutes les informations (nom, adresse ...) que vous nous avez donné sur votre correspondant, c'est uniquement son numéro de téléphone.
Dans ce cas on observe une descente intonative en fin de syntagme verbal, et un schéma intonatif montant en syllabe finale du mot sur lequel on insiste; celle-ci est généralement accompagnée d'un allongement de durée, d'une augmentation de l'intensité et est suivie d'une pause.

De ce fait, il semble délicat de proposer une règle d'évolution de Fo dans les énoncés de type impératif : le locuteur, quand il formule un énoncé de ce type, attend une réponse à une question bien précise; les manifestations prosodiques qu'il réalise lui permettent de préciser exactement à l'auditeur l'élément de réponse qu'il souhaite.

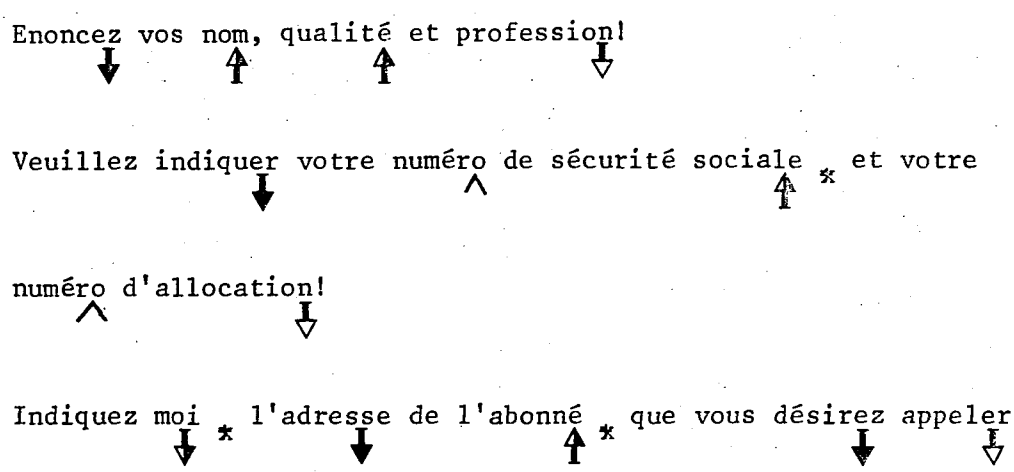
Signalons simplement que la formulation la plus neutre possible d'un énoncé impératif présente des caractéristiques prosodiques qui asssurent son découpage en groupe verbal et en groupe nominal comme dans les phrases énonciatives.

Veuillez spécifier le numéro de votre correspondant!
↑ ^ ↓

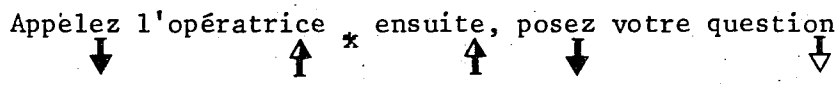
Quand la phrase impérative n'est pas située en fin d'énoncé, ou quand le groupe complément est composé de plusieurs groupes de sens, des modifications de Fo semblables à celles des phrases énonciatives

.../...

sont réalisées :



Il faut également préciser que certains verbes sont indissociablement liés à leur complément : aucune pause ne peut les séparer et ils possèdent souvent un schéma intonatif descendant, comme pour ne former qu'une seule unité avec leur complément :



II-1-3- Les phrases du type interrogatif :

La formulation d'une question peut être réalisée selon trois structures syntaxiques différentes :

- phrases construites selon le même modèle syntaxique que les phrases énonciatives:
  GN_1 + GV + GN_2 : vous voulez des cachets ?
- phrases réalisées avec inversion du sujet et du verbe:
  GV + GN_1 + GN_2 : pouvez-vous répéter votre question ?
- phrases introduites par un mot ou un groupe de mots interrogatifs:
  dans quelle ville habitez-vous ?

A - Phrase construite selon le même modèle que les phrases énonciatives :

La structure syntaxique de ce type de phrase ne permet pas de l'identifier comme phrase interrogative. Sous sa forme écrite, le point d'interrogation permet de l'appréhender comme telle; à l'oral, c'est à l'intonation qu'est dévolu ce rôle.

Vous êtes satisfait de la réponse ?

Vous pensez recommencer l'expérience ?

Vous habitez chez vos parents ?

Votre correspondant habite Paris ?

Le trait constant qui se dégage de ce type de phrase consiste en la montée intonative de fin de phrase; mais certains énoncés peuvent présenter cette montée de Fo de grande amplitude à la fin du groupe verbal, et dans ce cas la montée de Fo en fin d'énoncé est de faible amplitude :

Vous connaissez le numéro de Monsieur Dupont ?

On peut intégrer à ce type de phrase, celles qui sont introduites par la locution [est-ce que ...] puisque la suite de l'énoncé est articulée comme une phrase énonciative; on trouve d'ailleurs les mêmes caractéristiques prosodiques que dans les exemples ci-dessus :

Est-ce que vous habitez à la campagne ?

Est-ce que vous partez demain ?

Il faut également inclure dans cette catégorie les phrases qui manifestent une interrogation indirecte :

.../...

Je voulais vous demander * si vous connaissiez le numéro de

Monsieur Dupont.

Par rapport à une phrase énonciative, on note dans les schémas prosodiques les différences suivantes :

. la fin de phrase présente un schéma de Fo très légèrement montant.

. de ce fait, la fin du groupe majeur précédent (groupe verbal ici) présente un schéma de pente inversée : Fo est descendant.

Quant à la première proposition, elle possède des schémas prosodiques identiques à ceux d'une proposition non finale dans les phrases énonciatives.

B - Phrases interrogatives avec inversion du sujet et du verbe :

Pouvez-vous répéter ?

Pouvez-vous répéter ce que vous venez de dire ?

Pouvez-vous me dire * quel est le numéro de Monsieur Dupont ?

Voulez-vous relire le dernier paragraphe de votre lettre ?

Pourrez-vous m'aider * à résoudre un problème ?

Savez-vous * quel est le numéro de Monsieur Dupont ?

Dans tous les exemples de ce type nous avons observé une grande montée intonative mais une <u>seule</u> par énoncé. La position de cette montée

.../...

intonative est fonction des intentions du locuteur : elle peut se situer soit à la fin du groupe verbal, soit en fin de phrase.

Quand ce schéma montant intervient en fin de groupe verbal, le schéma intonatif de fin de phrase est très stable (□) c'est-à-dire qu'il présente une faible amplitude dans un registre fréquentiel bas (que nous fixerons plus loin). Quand au groupe verbal, à sa montée interrogative se superpose souvent un allongement de durée de dernier segment et une pause. Ce que nous appelons groupe verbal ici peut inclure soit le verbe à l'infinitif (Pouvez-vous m'aider ...) soit le pronom sujet situé après lui (Savez-vous ...).

Ce schéma intonatif peut également être situé sur une unité de GN_2 :

Préférez-vous fumer la pipe * ou les cigarettes ?
↓ ↓ ↑ □

Quand l'interrogation impose une réponse prédéfinie, la montée intonative ne se situe jamais sur le dernier terme du choix, elle intervient sur le premier des deux.

Dans les énoncés qui présentent un choix à plusieurs éléments, on peut observer une succession de schémas intonatifs montants correspondant aux choix possibles, mais le dernier élément de la question conserve un schéma de Fo avec de faibles variations fréquentielles :

Préférez-vous manger des pommes, des poires ou des abricots ?
↓ ↓ ↑ ↑ □

C - Phrases introduites par un mot ou un groupe de mots interrogatifs :

1) Qui a frappé Paul ?
◆ □

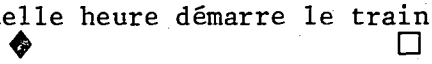
2) Que connaissez-vous de Rome ?
◆ □

.../...

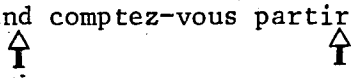
3) Quand partez-vous ?

4) Quel fichier désirez-vous ?

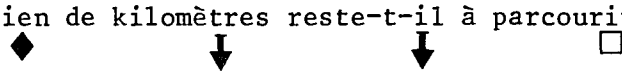
5) A quelle heure démarre le train ?



6) Quand comptez-vous partir ?

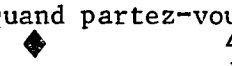


7) Combien de kilomètres reste-t-il à parcourir ?



En règle générale, la marque intonative est réalisée sur la fin du mot interrogatif; le schéma intonatif est descendant mais son attaque se situe à un niveau très élevé (ce schéma particulier est noté ◆ dans les exemples). Le schéma fin de phrase est ensuite soit faiblement montant soit faiblement descendant, soit même franchement montant.

Quand partez-vous ?



Mais d'autres énoncés présentent aussi une double montée intonative : en fin de phrase et en fin de mot interrogatif : exemples 4 et 6.

Cette variété de réalisations ne nous a pas permis de proposer une formalisation.

Nous voici parvenue au terme de cette étude sur l'organisation générale des mouvements de Fo. Il est vraisemblable que nous n'avons pas rendu compte de toutes les structures syntaxiques possibles en français ni de tous les patrons prosodiques mais il semble que les règles que nous avons induites de l'analyse du corpus doivent pouvoir s'appliquer à la plupart des énoncés du français.

.../...

Il nous reste maintenant à quantifier ces résultats mais nous voudrions tout d'abord essayer de nous situer par rapport à deux auteurs qui se sont intéressés à la structuration prosodique des énoncés français : DI CRISTO et surtout MARTIN.

II-2- Les études sur la structuration prosodique des énoncés :

II-2-1- DI CRISTO (1975) dans une étude intitulée "Recherches sur la structuration prosodique de la phrase française" analyse tous les paramètres : fréquence, durée, intensité, pour quelques types de structures syntaxiques. Il opère une décomposition en groupes prosodiques ("unité suprasegmentale délimitée par une variation perceptuelle significative d'un ou de plusieurs paramètres prosodiques") à l'intérieur desquels il étudie plus particulièrement les manifestations prosodiques de trois points clés : l'attaque, la prétonique et la tonique. Il énonce ensuite un certain nombre de règles pour les phrases définies par la modalité assertion positive, formée de deux constituants immédiats : le groupe nominal et le groupe verbal.

1ère Régle :

A - Dans une structure du type Groupe nominal (GN) + Groupe prédicat (GP), il existe une "frontière prosodique non terminale majeure" (FPnTM) qui sépare les deux constituants immédiats (à condition que le groupe nominal ne soit pas un proclitique).

B - Dans le cas où GN est un proclitique, la FPnTM se situe entre le verbe et le GN complément :

a) "les enfants - FPnTM - s'amusent"

b)"il a vendu - FPnTM - son château"

2e Règle :

Dans une structure de type GN_1 (nom propre)+ verbe + GN_2, il existe une "frontière prosodique non terminale mineure"(FPnTm) devant le GN_2 :

c) "Jean-Jacques (FPnTM) a vendu (FPnTm) son château"

.../...

La présence de cette frontière mineure n'étant pas systématique, on parle de règle facultative.

3e Régle :

Si GN_2 est suivie d'un complément circonstanciel, celui-ci est précédé d'une FPnTM :

d) "Jean-Jacques (FPnTM) a vendu (FPnTm) son château (FPnTM) en Espagne" (c'est en Espagne qu'il a vendu son château).

Ensuite sont étudiés les indices prosodiques perceptuels de ces deux types de frontières :

1) FPnTM se caractérise
   - . toujours par une rupture tonale (dans le niveau 3 : infra aigu).
   - . par une rupture bilatérale d'intensité.
   - . par un ralentissement du tempo : syllabe tonique = 200 ms.
   - . par une pause non considérée comme un indice significatif.

2) FPnTm se caractérise
   - . par une rupture tonale dans le niveau 3
   - . par une rupture d'intensité.
   - . par un ralentissement du temps moins important que pour FPnTM.
   - . par une absence de pause.

Les résultats que nous obtenons pour les phrases qui possèdent une structure syntaxique identique à celles analysées dans cette étude nous ont permis d'élaborer des règles presque semblables.

Toutefois nous pensons - à condition que les phrases soient émises sans insistance particulière sur l'un de ses termes - qu'il existe une différence fondamentale entre les deux FPnTm dans les exemples c) et d).

c) Jean-Jacques a vendu (FPnTm) son château.

GN_1 GV GN_2

.../...

d) Jean-Jacques a vendu (FPnTm) son château en Espagne.

Alors que dans le premier cas (c), le groupe verbal se terminera toujours par un schéma intonatif montant sur la dernière voyelle du mot, dans le second cas (d), le groupe verbal s'achèvera toujours (sauf insistance particulière sur l'aspect "vente", et sauf débit particulièrement lent) par un schéma de Fo de type descendant, simplement parce qu'il est suivi d'une frontière Majeure avec un schéma de Fo de type montant et que la règle d'alternance atteint le schéma intonatif du groupe verbal. Dans le premier exemple (c) au contraire, la fin du groupe verbal ne peut connaître qu'un schéma de Fo montant parce qu'il est suivi d'une frontière Majeure (la fin de phrase) qui possède un schéma de Fo descendant.

Il reste cependant qu'à ce schéma montant obligatoire en fin de verbe, il est laissé une certaine marge de liberté :
il pourra être réalisé, selon le choix du locuteur:

. soit
- par une faible montée intonative.
- sans allongement caractéristique du dernier segment de mot.
- sans pause.

. soit avec les mêmes caractéristiques que celles définies pour la FPnTM, à savoir :
- une montée intonative ample.
- une rupture bilatérale d'intensité.
- un allongement caractéristique de la syllabe tonique et une pause.

C'est un choix, un choix dans le débit.

Par contre, et c'est évident, on ne pourra jamais avoir - sauf encore une fois une volonté d'insister sur l'action du verbe - une Frontière Majeure immédiatement après le verbe, si une frontière du même type n'existe pas précédemment entre le GN_1 (composé d'un ou plusieurs mots lexicaux - par opposition aux mots grammaticaux) et le groupe verbal : la possibilité d'une Frontière Majeure après le verbe (suivi d'un simple complément) passe obligatoirement par l'existence d'une Frontière Majeure entre le dernier mot du GN_1 et le verbe. Mise à part cette

.../...

réserve, c'est au locuteur qu'il appartient de décider comment il va réaliser la montée intonative en fin de groupe verbal.

Une autre réserve concerne la pause de la FPnTM qui n'est pas considérée par DI CRISTO comme un indice significatif.

Nous pensons au contraire que la pause permet directement la réalisation des FPnTM parce qu'elle autorise en même temps que l'allongement de la syllabe tonique le passage sans difficultés à des valeurs de Fo toujours plus basses dans le mot qui lui succède (particulièrement quand cette frontière majeure est entourée d'éléments vocaliques).

Par exemple, l'absence de pause entre "vendu" et "en Espagne" interdit l'allongement caractéristique de la fin de mot [y] en même temps qu'une montée intonative vers de hauts niveaux fréquentiels (c'est-à-dire la FPnTM) puisque immédiatement après, le locuteur doit réaliser une autre voyelle [a] qui, appartenant à la classe des mots grammaticaux connaît des niveaux de Fo toujours inférieurs à ceux de mots lexicaux (même non situés en des points clés).

Par contre, nous sommes d'accord avec l'auteur quand, ayant identifié deux FPnTM dans la même phrase, il n'introduit pas de différences prosodiques entre elles selon leur position dans l'énoncé. Nous pensons également que ces frontières présentent des indices prosodiques perceptuels identiques quelle que soit leur localisation. MARTIN au contraire semble - nous le verrons - introduire de telles différences.

Enfin nous partageons l'opinion de DI CRISTO quand il conclut que "les indices acoustiques et perceptuels de la structure constituante sont toujours présents et ne se manifestent pas, contrairement à ce qu'affirmait LIEBERMAN, que dans les cas d'ambiguïté".

II-2-2- MARTIN (1973, 1974, 1975, 1976) présente une théorie générale de l'intonation; il part de l'hypothèse qu'il existe une corrélation entre la structure syntaxique d'un énoncé et les éléments suprasegmentaux (les faits prosodiques). La grammaire de l'intonation qu'il élabore vise à "dériver une séquence de contours mélodiques à

.../...

partir d'une présentation de la structure superficielle de l'énoncé. Sans entrer dans les détails de cette théorie, nous en signalerons simplement quelques éléments :
L'auteur déduit de l'analyse que les contours suceptibles d'apparaître dans la phrase française, sont au nombre de 8 et se différencient par trois traits :

- Contraste de pente (Fo montant, Fo descendant).
- Contraste dans l'amplitude.
- Contraste de longueur.

Il définit l'unité minimale de signification comme représentant le contenu de chaque couple de parenthèses.

Pour la phrase "Le vilain canard blanc buvait du lait prématurément", la décomposition en unités minimales de signification donne le parenthésage suivant :

(((A) (B)) (C)) (D) (E) (F)

Le vilain canard blanc buvait du lait prématurément

Les deux points essentiels sur lesquels nous sommes en désaccord avec MARTIN, sont les suivants :

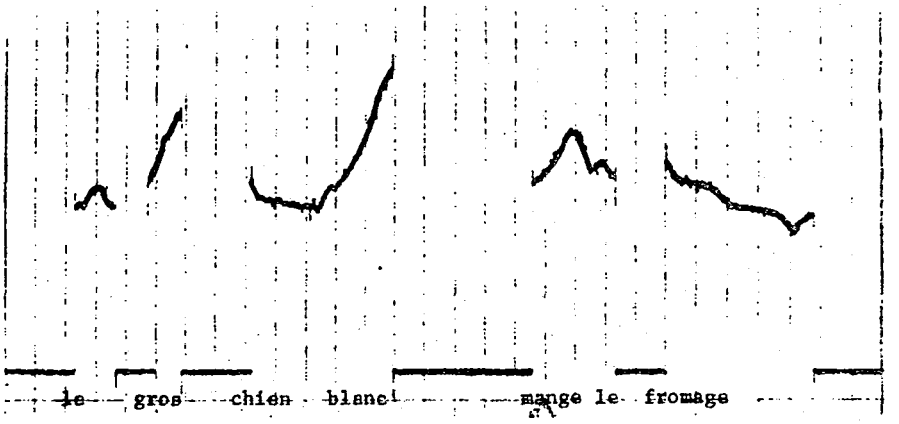
I°) Pour lui : "Un contour mélodique d'une unité ne peut prendre de valeur que <u>par contraste</u> avec le contour mélodique qui marque l'unité de niveau immédiatement supérieur et avec celui là seulement "... ..." en français cette opposition de contours d'unités composantes et d'unité composée est réalisée essentiellement par un contraste de pente mélodique (et secondairement par un contraste d'amplitude de variation mélodique".

Ainsi, par exemple la phrase "le gros chien blanc mange le fromage", donne le parenthésage suivant :
(((le gros) (chien)) (blanc)) ((mange) (le fromage)
et les contours intonatifs qui y correspondent :

.../...

le gros chien blanc mange le fromage
C_3 C_2 C_1 C_3 C_d

La courbe mélodique expérimentale qui correspond à cet énoncé est:



Toutes les phrases de même construction syntaxique que nous avons analysées, ne présentent jamais le contour C3 détecté ici sur l'adjectif (gros) et caractérisé par les traits montant, court, restreint (faible variation de Fo). Pour notre part, nous n'avons trouvé de schéma montant de Fo dans le syntagme précédant le groupe verbal que sur la première voyelle de certains mots pour réaliser ce que l'on appelle "l'accent didactique" ; un mot considéré comme sémantiquement important présente sur sa première voyelle, une montée intonative et sur sa seconde voyelle, une descente de Fo qui a pour niveau de départ une valeur fréquentielle plus haute (d'environ 1/2 ton) que la valeur d'arrivée de la première voyelle (Fig. 28- "Le numéro de Monsieur X .....).

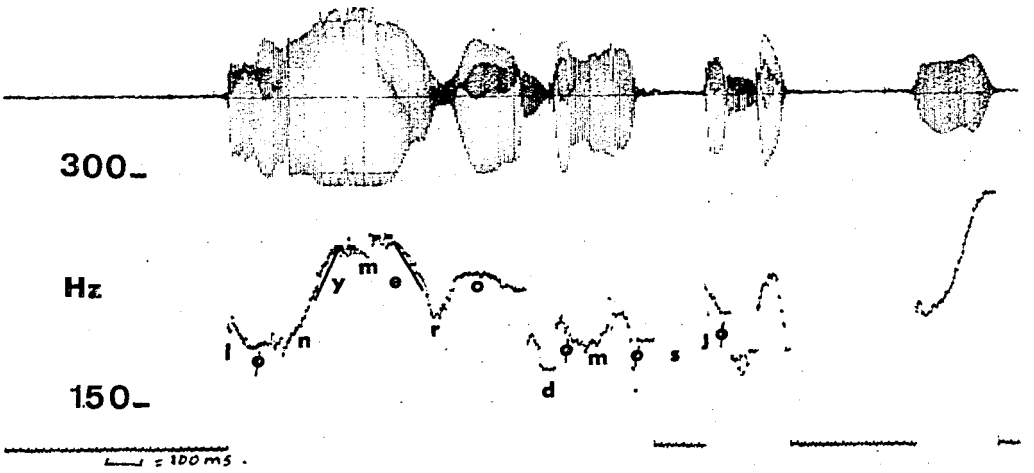


FIG. 28 - Réalisation de l'accent d'insistance.

.../...

Nous pensons que dans l'exemple "le gros chien blanc....." l'adjectif "gros" est produit avec une manifestation d'insistance qui explique (peut-être) cette montée de Fo ; nous pensons que si (gros) était remplacé par (magnifique), la montée intonative se situerait sur la première voyelle : [a] et non pas en fin de mot.

Dans notre corpus, tous les mots inclus dans le même groupe de sens du syntagme nominal situé avant le verbe, présentent à leur terme, une pente de Fo négative. Mais effectivement, nous sommes d'accord que si ce syntagme présente deux groupes de sens séparés, par exemple, par une préposition, le premier groupe de sens présentera en fin de réalisation, un contour de Fo dont la pente aura un signe différent de celui de la fin du groupe nominal ; il possèdera donc une pente négative.

C'est le même phénomène que l'on peut observer dans le syntagme nominal postérieur au groupe verbal : si une préposition, une conjonction sépare ce syntagme en deux groupes de sens, la fin du premier groupe de sens aura un contour mélodique avec un signe de pente opposé à celui de la fin de phrase. Mais, en-deçà du groupe de sens, il ne semble pas y avoir de répercussions : tous les mots regroupés dans un même groupe de sens présentent le même schéma de Fo : un schéma de type descendant ; et pour nous, "le gros chien blanc" ne forme qu'une seule unité de signification avec un seul patron prosodique significatif : celui de la fin de cette unité.

Dans un autre exemple choisi par Martin : "le vilain canard blanc buvait du lait prématurément", alors qu'il obtient par décomposition successive des unités les plus grandes, le parenthésage suivant :

(((A) (B)) (C)) (D) (E) (F)
Le vilain canard blanc buvait du lait prématurément

nous n'obtenons que quatre unités :

1. Le vilain canard blanc
2. buvait
3. du lait
4. prématurément

.../...

c'est-à-dire que si l'on applique les règles prosodiques que nous avons dégagées, nous aurons :

Le vilain canard blanc buvait du lait prématurément.

Mais nous reconnaissons que ces unités, gouvernées par le schéma descendant de fin de phrase, présentent l'une par rapport à l'autre, à partir de la fin du groupe nominal pré-verbal, un contraste de pente.

- Le premier groupe de sens du GN2 est contrasté par rapport à la fin de phrase : il présente une pente de Fo positive,

- Le groupe verbal est contrasté par rapport au premier groupe de sens de GN2 : il présente une pente négative.

Mais le contraste de pente s'arrête là, il ne continue pas pour les unités de niveau inférieur.

2°) Le second point de désaccord porte sur l'affirmation suivante : "Les contours des unités d'un même niveau sont souvent réalisés avec une amplitude correspondant à l'importance de l'unité dans la composition de l'énoncé : ces unités étant en général rangées par ordre décroissant d'importance.... les contours sont réalisés avec des amplitudes décroissantes au fur et à mesure qu'elles se rapprochent du contour de référence final".

Nous pensons, (de par les résultats de l'analyse de corpus) que si le locuteur réalise une pause après le GN1 et le groupe verbal, on constate sur la syllabe finale de l'un et l'autre groupe, un allongement dans la durée du dernier segment, et l'on observe également un schéma intonatif identique sur les syllabes situées en fin de chaque groupe, ou bien un schéma dont les différences de Fo ne sont dûes qu'aux caractéristiques intrinsèques du dernier segment vocalique : on remarque par exemple que :

- le schéma montant de Fo sur une voyelle nasale présente en général des niveaux fréquentiels d'arrivée moins élevés que ceux d'une voyelle orale.

- que dans les voyelles orales, les voyelles ouvertes

.../...

[a, ɛ, ɔ, œ] ont également des niveaux d'arrivée plus hauts que les voyelles fermées.

- que les voyelles en syllabe fermée demeurent dans une gamme fréquentielle plus basse que les voyelles en syllabe ouverte.

Mais à voyelle identique, les schémas intonatifs ne semblent pas présenter de différence significative qui tienne à leur position par rapport au "contour de référence final":

Les dernières informations de la justice $\left[\begin{matrix}214\text{Hz}\\170\text{Hz}\end{matrix}\right.$ (variation : 4 1/2 tons) seront exposées $\left[\begin{matrix}240\text{Hz}\\186\text{Hz}\end{matrix}\right.$ (variation : 4,4 1/2 tons) à la bibliothèque.

Le journaliste $\left[\begin{matrix}244\text{Hz}\\190,5\text{Hz}\end{matrix}\right.$ (variation : 4,3 1/2 tons) travaille dans son bureau $\left[\begin{matrix}237\text{Hz}\\182\text{Hz}\end{matrix}\right.$ (variation : 4,6 1/2 tons) avec le Directeur.

De la même façon, dans un énoncé présentant une succession d'unités de signification démarquée par des virgules, on ne constate pas une amplitude décroissante au fur et à mesure de leur réalisation (Voir par exemple : "ce matin là, à Domrémy ...").

Si au contraire, une pause n'apparaît pas à la fin du groupe verbal (nous avons dit que cette pause ne peut exister que si la fin du GN_1 en connaît une aussi), l'allongement du dernier segment sera moindre ainsi que la variation de Fo ; dans ce cas, mais dans ce cas seulement, on remarquera effectivement une variation de Fo moins importante en fin de G.verbal qu'en fin de groupe $nominal_1$.

Pour conclure, nous dirons qu'à notre avis, ce qui relève de l'arbitraire, du conventionnel, dans l'évolution de Fo, c'est le signe de sa pente qui permet de définir une organisation de la phrase en deux constituants majeurs répartis de part et d'autre d'une frontière située immédiatement avant le groupe verbal. On retrouve ici la notion bien connue de thème et de propos. Ce qui relève des variantes individuelles, conscientes ou non, est lié au débit du locuteur et à son

.../...

intention de bien marquer (ou non) la structure syntaxique de l'énoncé (pauses subséquentes à un allongement de durée du dernier son).

III - L'ANALYSE QUANTATIVE DES EVOLUTIONS DE Fo.

De nombreux travaux ont été menés pour tenter de dégager des niveaux intonatifs dans la phrase française et pour évaluer la pertinence de ces niveaux sur le plan perceptuel (ROSSI et CHAFCOULOFF, 1972 ; CHALARON, 1972 ; CONTINI et BOE, 1973 ; VAISSIERE 1974, 1975 ; DI CRISTO, 1975 ; BOE et CONTINI, 1976; ROSSI, 1973.

Nous présentons dans le tableau5 les fréquences moyennes et l'écart type (σ) de l'attaque et de la fin de certaines réalisations vocaliques situées en des points de l'énoncé considérés comme significatifs sur le plan prosodique.

La variation de Fo entre attaque et fin est donnée en 1/2 tons. Les mesures ont porté sur une centaine d'occurences pour chaque catégorie des phrases énonciatives, sur cinquante environ pour celles des deux autres types.

Les points analysés sont les suivants :

- Pour les phrases énonciatives :
  - (1) valeurs de Fo dans les monosyllabiques situés en initiale de phrase.
  - (2) valeurs de Fo en fin de mots plurisyllabiques à schéma intonatif descendant, c'est-à-dire les mots appelés "non marqués".
  - (3) valeurs de Fo en fin de mot à schéma intonatif descendant, situé en fin de groupe de sens du syntagme pré-verbal.
  - (4) valeurs de Fo à la fin de l'avant dernière syllabe d'un mot > 3 syllabes et valeur de Fo à l'attaque de la dernière syllabe , quand ce mot est situé avant une pause et possède un schéma intonatif montant.
  - (5) valeurs de Fo en fin de mot plurisyllabique avec un schéma intonatif montant et situé avant une pause.

.../...

(6) valeurs de Fo en fin de mot à schéma intonatif montant mais non suivi d'une pause.

(7) valeurs de Fo dans les monosyllabiques (non points clés) dans la phrase.

(8) Valeurs de Fo à la fin de l'avant dernière syllabe d'un mot $\geqslant$ 3 syllabes et valeur de Fo à l'attaque de la dernière syllabe quand ce mot est situé avant une pause et possède un schéma intonatif descendant.

(9) valeurs de Fo en fin de mot plurisyllabique avec un schéma intonatif descendant et situé avant une pause (fin de phrase).

- Pour les phrases intérrogatives :

(10) valeurs de Fo à la fin du mot qui porte la marque de l'interrogation dans les phrases introduites par un mot ou un groupe de mots interrogatifs (noté ◆).

(11) valeurs de Fo à la fin du mot qui porte la marque de l'interrogation dans les phrases construites selon le même modèle syntaxique que les phrases énonciatives ou avec inversion du sujet et du verbe (noté ⍋)

(12) valeurs de Fo en fin de phrase interrogative quand la marque de l'interrogation est déjà présente dans la phrase (□)

(13) valeurs de Fo pour les monosyllabiques.

- Pour les phrases impératives :

(14) valeurs de Fo pour la voyelle finale d'un mot à intonation montante situé avant une pause.

(15) valeurs de Fo pour la voyelle finale de phrase.

.../...

catégories	Attaque	σ (attaque)	fin	σ(fin)	variation en 1/2 ton	σ (variation)
(1)	181	13	163	11	- 1,8	1,1
(2)	208	18	186	14	- 1,9	1,0
(3)	201	10	175	11	- 2,4	1,2
(4)	179	15	188	15	0,85	1,4
(5)	189	16	257	37	5,3	2,4
(6)	198	13	219	10	1,7	1,0
(7)	182	12	164	11	- 1,8	1,0
(8)	170	12	159	8,3	- 1,1	1,2
(9)	160	9	141	6,5	- 2,2	1,0
(10)	272	24	243	13	- 1,9	1,3
(11)	210	21	311	48	6,8	2,9
(12)	179	10	178	13	- 0,11	1,2
(13)	213	14	191	13	- 1,9	- 1,2
(14)	189	18	268	50	6,0	3,3
(15)	165	10	141	8	- 2,7	0,88

TABLEAU 5.

.../...

De ce tableau se dégagent les tendances suivantes :

1 - dans les phrases énonciatives :

a) Valeurs de Fo pour les mots à schéma intonatif descendant:

- Les monosyllabiques des phrases énonciatives présentent des niveaux d'attaque et de fin identiques, qu'ils soient situés en initiale de phrase (1)ou dans la phrase (7). On peut comparer leurs valeurs à celles des syllabes finales des mots non marqués (2), c'est-à dire non situés à des points clés ; si la variation de Fo entre l'attaque et la fin, est voisine (1,8 1/2 ton), les niveaux d'attaque (et par conséquent de fin) sont différents ; les mots non marqués ont une attaque plus haute d'environ 2,4 1/2 ton. On peut également les comparer aux valeurs de Fo des monosyllabiques des phrases interrogatives (13). Ces dernières ont une attaque plus haute d'environ 2,8 1/2 ton mais même variation de Fo : - 1,8 1/2 ton.

Dans les phrases énonciatives, ce sont les monosyllabiques (1) et (7) qui, parmi tous les mots suceptibles de se terminer par un schéma descendant, exceptée la voyelle finale de phrase, présentent l'attaque et la fin les plus basses.

- Quant aux mots terminés par un schéma de Fo descendant situés en fin de groupe de sens dans le syntagme préverbal et suivis ou non d'une pause (3), si l'attaque de leur dernière voyelle est presque identique à celle des mots non marqués (2) - 205 Hz en moyenne - la variation de Fo est plus grande ;

- 1,9 1/2 ton pour les mots non marqués.
- 2,4 1/2 ton pour les mots fin groupe de sens.

- Il reste une autre possibilité de schéma descendant dans une phrase énonciative : elle se présente en fin de phrase et précède une pause (9); l'attaque de la voyelle finale se situe au même niveau que la fin des monosyllabiques (1) et (7), soit en moyenne 160 Hz ; la variation de Fo est sensiblement équivalente à celle de la voyelle (3) des fins de groupes de sens : - 2,3 1/2 ton.

.../...

b) Valeurs de Fo pour les mots à schéma intonatif montant :

- La voyelle finale de ces mots, quand ils sont suivis d'une pause (5) présente une attaque à 189 Hz et une variation de Fo importante : 5,3 1/2 ton en moyenne. Mais ce qu'il faut noter, c'est une grande instabilité de la fin dans cette position : la montée intonative semble dépendre de nombreux facteurs : voyelle orale ou voyelle nasale, voyelle ouverte ou voyelle fermée, syllabe ouverte ou syllabe fermée, présence d'une pause longue (environ 1 seconde) ou courte (inférieure à 200ms).

Nous ne sommes pas parvenue à isoler quantitativement l'influence de chacune de ces variables ; aussi, avons-nous cherché pour cette catégorie, d'autres invariants.

- La voyelle finale des mots à Fo montant mais non suivis d'une pause (6) présente une attaque plus haute que les mots suivis d'une pause (5) et une variation bien plus faible (1,7 1/2 ton seulement) ; on observe également pour cette voyelle une fin de Fo beaucoup plus stable (voir les σ correspondants) : il semble que l'allongement des segments qui précèdent une pause influence considérablement la valeur de Fo en fin de réalisation vocalique ; la pente semble donc plus significative que les valeurs finales.

c) L'évolution de Fo entre la fin de l'avant dernière voyelle et l'attaque de la dernière voyelle du même mot :

- Avec un schéma de Fo montant sur la dernière voyelle (4) : la voyelle d'avant dernière syllabe présente en fin de réalisation un niveau de 180 Hz, inférieur de 0,85 1/2 ton au niveau d'attaque de la dernière voyelle.

- Avec un schéma de Fo descendant sur la dernière voyelle (8) : si la fin de l'avant dernière voyelle présente un niveau d'arrivée assez voisin du cas précédent, la variation entre la fin et l'attaque des deux dernières voyelles est plus ample (plus de 1 1/2 ton), avec pour l'attaque de la dernière voyelle un niveau de Fo inférieur au niveau de fin d'avant

.../...

dernière voyelle, ce qui est l'inverse de ce que l'on observe pour (4).

2 - dans les phrases impératives :

Les valeurs de Fo dans ce type de phrases sont identiques à celles relevées dans les phrases énonciatives pour une même position : (15) et (9) - (14) et (5) ; on remarque la même instabilité dans la valeur de fin d'une réalisation où Fo montant est situé avant une pause - σ = 37 et 50 Hz pour (5) et (14) - et la même variation de Fo (environ 6 1/2 tons).

3 - dans les phrases interrogatives :

Ici, les niveaux de Fo sont plus hauts que ceux des autres types de phrases. Nous avons vu en particulier que les monosyllabiques (13) ont une attaque plus haute d'environ 2,8 1/2 tons.

- Le niveau de fin le plus haut est également réalisé sur les mots qui sont situés avant une pause et qui présentent un schéma de Fo montant (11), le niveau de Fo en fin de voyelle est en moyenne de 311 Hz mais avec un écart-type important (48Hz) ; la variation entre l'attaque et la fin est égale à 6,8 1/2 tons : c'est la plus forte variation relevée. Quant au niveau d'attaque, il est identique au niveau de départ des monosyllabiques (13) dans ces phrases : en moyenne 211 Hz.

- La locution interrogative ou le groupe de mots qui porte la marque intonative en début de phrase (10) présente, nous l'avons dit, un schéma de Fo descendant mais avec de hauts niveaux à l'attaque et à la fin : l'attaque (272 Hz) est plus haute que la fin d'un mot en schéma intonatif montant situé avant une pause (5) dans les phrases énonciatives (257Hz), et la variation entre l'attaque et la fin est proche de -2 1/2 tons comme pour les monosyllabiques des phrases interrogatives (13).

- La voyelle finale des mots qui concluent les phrases dans lesquelles la marque intonative n'est pas réalisée en fin de phrase (12) présente des niveaux d'attaque et de fin très voisins ; ces niveaux et ces variations sont les plus faibles relevés dans ce type de phrase en fin de mot.

.../...

En résumé, voici les points importants que nous avons retenus pour la synthèse :

- les niveaux les plus bas sont situés sur les monosyllabiques des phrases énonciatives et impératives, ils sont identiques, quelle que soit leur position dans l'énoncé,

- les niveaux de fin de réalisation en Fo montant sont très instables dans les trois types de phrase quand une pause suit, ils semblent liés à la durée de ces réalisations,

- la variation de la montée de Fo sur la voyelle qui précède une pause ne dépend pas de la position syntaxique du mot dans la phrase, mais semble corrélée à des phénomènes phonétiques (nature de la voyelle par exemple) ; elle est en moyenne et pour les trois types de phrase, comprise entre 5,3 et 6,8 1/2 tons.

- les mots qui se terminent par un schéma de Fo descendant possèdent une plus grande stabilité dans les niveaux de fin et une variation moyenne de la fréquence fondamentale comprise entre - 1,8 et - 1,9 1/2 tons pour les mots non marqués, et entre - 2,2 et - 2,7 1/2 tons pour les mots marqués (fin de groupes de sens, fins de phrase).

- la dernière voyelle d'un mot a une attaque
. plus haute que la fin de l'avant dernière voyelle dans le cas où Fo est montant (variation = 0,8 1/2 ton)
. plus basse que la fin de l'avant-dernière voyelle dans le cas où Fo est descendant (variation = -1,1 1/2 ton).

Ces résultats pour ce qui concerne les valeurs d'attaque et de fin des phrases énonciatives, sont très voisins de ceux obtenus par CONTINI et BOE (1973), et BOE et CONTINI (1976), mais pour notre part, nous observons des maxima nettement plus dispersés.

. Si nous comparons les catégories (1), (5) et (15) avec les valeurs relevées par CONTINI et BOE (1973) pour l'attaque le maximum et la fin de la phrase énonciative :

- les niveaux respectifs de ces trois points sont très voisins,
- les dispersions de l'attaque et de la fin identiques, alors

.../...

que nous obtenons pour notre part, des écarts-types deux fois plus importants pour le maximum.

. Pour la phrase interrogative (BOË et CONTINI, 1976) :

- même correspondance pour attaque et fin,
- contrairement au cas précédent, les dispersions sont importantes et concordent.

Pour la catégorie (5) (dernière voyelle de mot avec Fo montant et situé avant une pause) dans laquelle nous avons observé une grande dispersion dans la fin, nous avons développé un traitement statistique afin d'essayer de mettre en évidence des invariants.

De cette catégorie, nous avons retenu 73 observations et aux variables attaque et fin, nous avons ajouté la durée. Les corrélations R * les plus significatives apparaissent entre :

- la variation et la fin : plus la valeur terminale est élevée, plus le rapport entre début et fin est grand (R = 0,84).
- la durée et la variation : ce sont les voyelles les plus longues pour lesquelles on observe la variation la plus importante (R =0,46).

Un système de synthèse pourrait intégrer dans ses commandes, de tels résultats. Si l'on se fixe les valeurs :

- de la durée vocalique
- de la fin

il est possible de calculer la valeur d'attaque ; en effet, l'analyse des régressions nous donne pour les observations que nous avons faites :

$$\text{attaque} = \frac{\text{Fin}}{3{,}96.10^{-3}\ \text{fin} + 0{,}133.10^{-2}\ \text{durée} + 0{,}135}$$

---

* Significatives à 98% si elles sont $\geqslant$ 0,27.

.../...

En moyenne pour (5) nous avons relevé :

- durée : 160 ms
- Attaque : 189 Hz
- Fin : 257 Hz

Vérifions notre estimation :

$$\text{Attaque} = \frac{257}{3{,}96.10^{-3}.257 + 0{,}133\ .10^{-2} \times 160 + 0{,}135} = 188 \text{ Hz}$$

ce qui est la valeur mesurée effectivement.

.../...

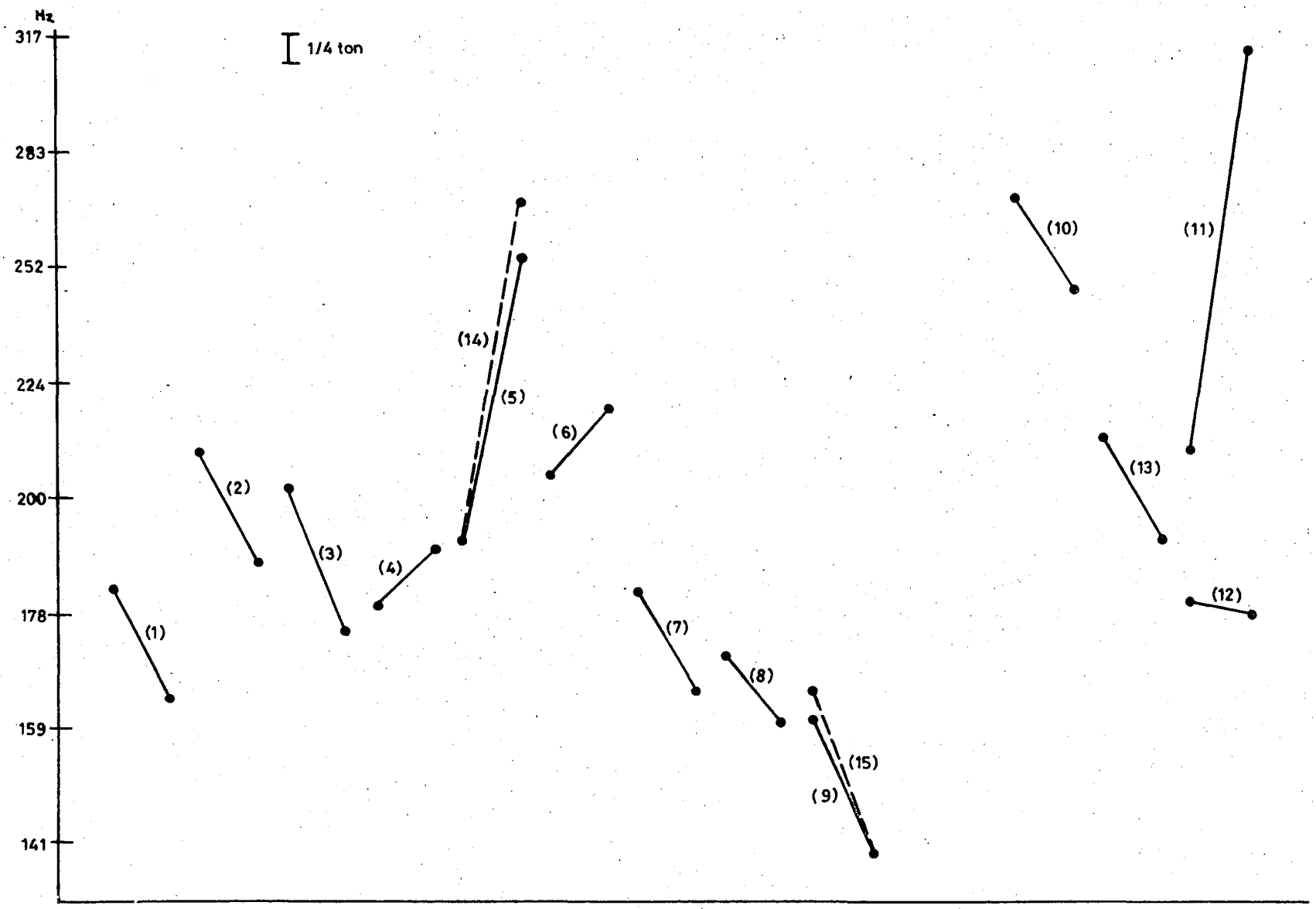


Schéma d'évolution de Fo dans les trois types de phrases (énonciatives, impératives et interrogations) et pour les voyelles des catégories étudiées (les catégories sont présentées autant que possible, dans l'ordre de leur apparition).

FIG. 29 -

4ÈME PARTIE

REALISATION

DU

SYSTEME DE SYNTHESE

Pour l'anglais-américain, de nombreux travaux ont été menés afin d'intégrer de manière plus ou moins systématique, la commande de Fo dans les paramètres de synthèse (HADDING-KOCH et STUDDERT-KENNEDY, 1964 ; OHMAN et LINDQVIST, 1965 ; HIKI, 1966 ; HIKI et OIZUMI, 1967 ; ALLEN, 1968 ; OHMAN, 1968 ; RABINER et LEVITT, 1968 ; UMEDA et al.,1968 ; RABINER et al., 1969 ; OLIVE et NAKATANI, 1974 ; OLIVE, 1975 ; ALLEN et O'SAUGHNESSY , 1976 ; KLATT, 1976).

Pour le français, les études se sont orientées dans ces mêmes directions ; elles ont été consacrées à la mise en évidence, à l'aide de la synthèse, de patrons intonatifs minimaux pour différents types de phrases (énonciatives, interrogatives, à variantes phonostylistiques) ayant une structure syntaxique plus ou moins complexe, de niveaux significatifs précisés sur le plan statistique et testés sur le plan perceptif (METTAS, 1963, 1964, 1965 ; DELATTRE, 1966 ; VAISSIERE, 1971 ; BOE et LARREUR, 1974 ; CHOPPY et al, 1975 ; BOE et CONTINI, 1975 ; EMERARD et LARREUR, 1975). Certainsde ces résultats sont exploitables mais la nature du travail qui nous était fixé, nous a imposé une connaissance quantitative très détaillée, ce qui nous a obligé à effectuer pour notre part, un travail d'analyse.

Mais il nous semble important de faire deux remarques :

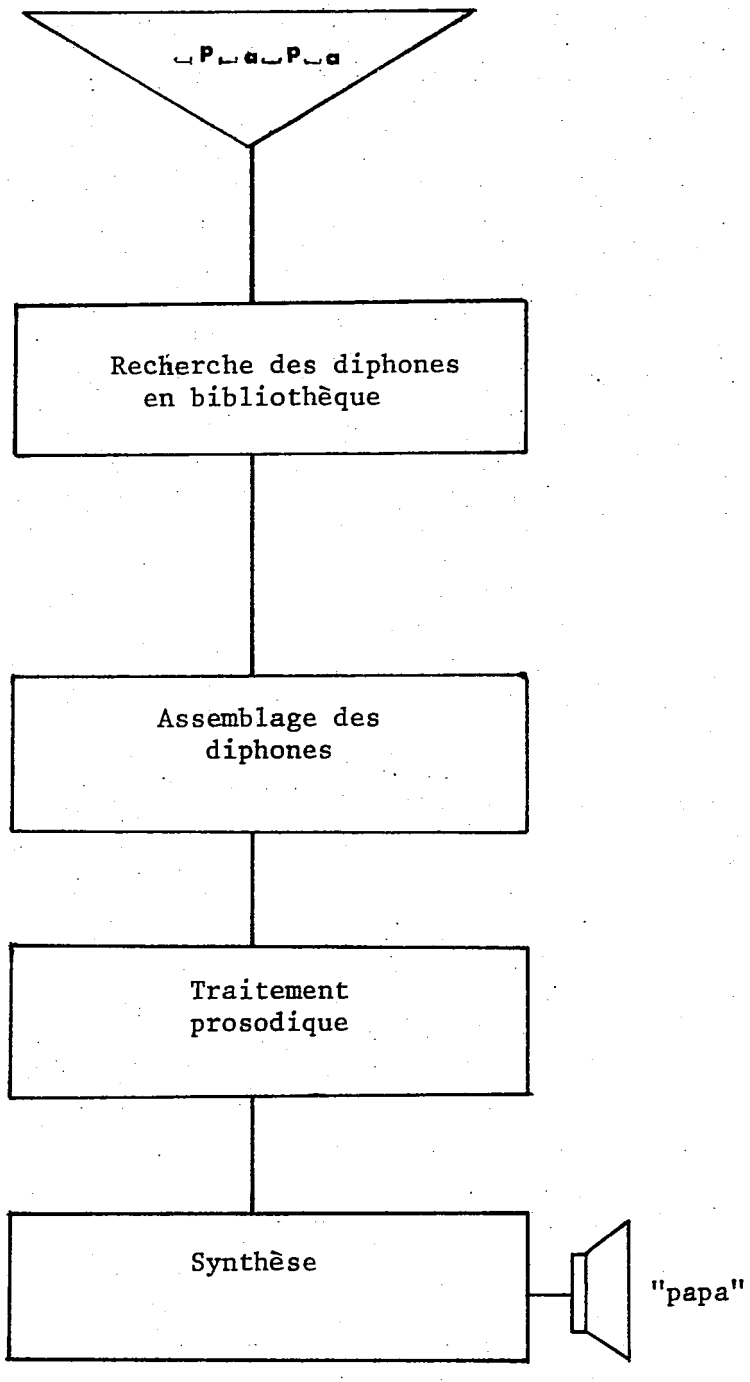
1 - Nous avons choisi pour des raisons étrangères au problème de la prosodie des éléments de parole de petite dimension puisqu'ils correspondent à la durée de réalisations d'un phonème (deux demi-réalisations de phonème). Si cette solution présente certains avantages sur le plan spectral, elle oblige à décrire l'évolution de la prosodie (Fo, durée, intensité) sur chacun de ces segments, ce qui nécessite une analyse minutieuse.

2 - D'une façon générale, l'analyse phonétique consiste à abstraire d'une multiplicité de réalisations produites par un grand nombre de locuteurs, les invariants qui sont à l'origine de tel ou tel fait de communication.

En ce qui nous concerne,le problème était radicalement différent : il s'agissait en effet de déterminer un ensemble de règles rendant compte du maximum de réalisations d'un seul et même locuteur, l'important n'étant pas de réaliser un archétype de parole mais une parole la plus ressemblante au modèle choisi et ce pour qu'elle paraisse naturelle.

.../...

Nous allons décrire ici l'ensemble des opérations que nous avons effectuées pour parvenir à la réalisation des différentes phases schématisées ci-dessous grâce auxquelles on peut passer de l'écriture d'un message sur un organe d'entrée à la réponse vocale synthétique sur un organe de sortie :



.../...

Les premières tentatives que nous avons faites en ce domaine ont consisté à réaliser une bibliothèque de diphones à partir des enregistrements de parole d'un locuteur masculin parfaitement habitué à contrôler le niveau et le débit de sa voix.

Il s'agissait de mots artificiels bisyllabiques (titi, papa, roro...) prononcés de façon isolée. Les avantages de cette procédure tenaient d'une part à ce qu'un programme de segmentation (sur Calculateur CII 10070) permettait d'extraire de façon quasi-automatique tous les diphones [consonne-voyelle] et [voyelle-consonne], et d'autre part à ce que convenablement articulés à un débit moyen, les mots présentaient des caractéristiques spectrales bien marquées.

Ces mots avaient été prononcés volontairement sur un ton monocorde ; les réalisations vocaliques présentaient entre elles une grande similitude quant aux évolutions de la fréquence fondamentale ; aussi, en réalisant une simple juxtaposition des diphones sans modification des valeurs de F_0 délivrées à l'analyse, on obtenait une parole de synthèse déjà intelligible puisque les réalisations vocaliques, responsables du message intonatif, se concaténaient sans grandes discontinuités en leur partie stable.

Mais, en contre-partie, cette procédure d'obtention des diphones présentait certains inconvénients non négligeables :

. La durée de chacun des diphones extraits des mots artificiels était bien plus importante que celle relevée dans la parole continue, les réalisations consonantiques étant les plus sensibles à cet allongement. OLIVE et NAKATANI (1974) dans une étude sur la synthèse par mots de numéros de téléphone notent la nécessité de réduire les mots prononcés de façon isolée de 20 % de leur durée quand ils sont utilisés dans la parole continue.

.../...

. Les mots, qui étaient constitués de deux syllabes ouvertes et qui étaient prononcés de façon bien articulée, engendraient des diphones dont l'assemblage produisait à la synthèse une parole au débit hâché, saccadé, la sensation auditive d'un découpage syllabique du flux de la parole : ce résultat a semblé imputable au fait que les diphones [voyelle-consonne] étaient extraits pour le premier segment (voyelle) d'une fin de première syllabe ouverte et pour le second segment d'un début de deuxième syllabe ; l'une et l'autre avaient été produites de façon trop indépendante ce qui augmentait de façon anormale la durée des transitions de la voyelle à la consonne.

. La voix de synthèse obtenue par juxtaposition simple de diphones était tout à fait dépourvue de naturel : l'ensemble du message était émis sur le même ton, aucun élément (évolution de F_0 , réduction ou augmentation de la durée des réalisations, pauses...) ne facilitait le découpage de la chaîne parlée d'abord en mots, puis en unités de sens et en groupes syntaxiques pour rendre accessible immédiatement à l'auditeur le sens de l'énoncé.

Nous avons essayé d'améliorer ce premier résultat en réalisant sur la syllabe finale de ce que nous appelons les "points clés" (endroits de la chaîne parlée qui présentent des caractéristiques prosodiques spécifiques, par exemple la fin de syntagme situé immédiatement avant le verbe ou la fin de phrase énonciative) un traitement de la prosodie : montée intonative sur la syllabe finale du mot qui précède immédiatement le syntagme verbal, descente intonative en fin de phrase énonciative ou impérative, allongement de la durée du dernier segment en ces deux points, insertion d'une pause.

Mais, il faut bien voir que la synthèse par diphones n'est pas la synthèse par mots : avant de se préoccuper des manifestations prosodiques en fin de syntagme, il est absolument nécessaire de reconstituer toute la prosodie du MOT. Aussi, ces premiers résultats n'ont-ils pas été concluants : durée trop longue des segments, absence de fluidité dans le débit, timbre de la voix trop grave lié à l'option

.../...

prise pour l'enregistrement, et mots dépourvus de leur particularisme prosodique.

Ces considérations nous ont conduits à concevoir la réalisation d'une autre bibliothèque de diphones reposant sur un corpus différent et sur une procédure d'enregistrement modifiée ; c'est également à ce moment-là que des tests comparatifs sur l'intelligibilité et la "qualité" de plusieurs voix en analyse (des tests réalisés sur des listesde logatomes indiquent que les pourcentages d'intelligibilité sont supérieurs en moyenne de 4 % avec les voix de femme analysées au vocodeur à canaux), et en synthèse nous ont fait choisir un locuteur féminin.

. Les mots qui ont servi de base aux enregistrements sont, non plus des mots artificiels mais des mots réels et le plus souvent usuels. Ce choix a permis de constater une amélioration surtout dans l'articulation des voyelles les moins tendues, en particulier [ɛ] et [œ].

. Ces mots possèdent en moyenne trois syllabes et c'est la syllabe centrale qui a été utilisée pour extraire les diphones [voyelle-consonne] et [consonne-voyelle] . Forts des enseignements tirés de l'expérience précédente, nous avons le plus souvent pris les diphones [voyelle-consonne] dans une syllabe fermée [CVC].

Exemple : /ar/ pris dans /escargot/ pour éviter la sensation auditive d'un découpage syllabique du flux de la parole, sensation due, sans doute, à une trop longue durée des transitions de la voyelle à la consonne dans les mots composés de deux syllabes ouvertes (rara, titi..) utilisés lors du premier dictionnaire.

Les résultats perceptuels donnent à penser que nous avons eu raison d'envisager cette nouvelle procédure : la parole obtenue par concaténation connaît désormais un débit moins heurté, plus fluide.

.../...

Nous avons également demandé à ce locuteur de prononcer les mots du corpus sur un ton relativement monocorde pour éviter des difficultés relatives au raccordement des spectres si F_0 est très différente d'un diphone à l'autre (déplacement de la zone des formants, et donc discontinuités spectrales à la concaténation), et parce qu'il est plus aisé de contrôler le niveau de sa voix dans ces conditions.

.L'enregistrement de ces mots a été réalisé en deux séances consécutives. L'opérateur était chargé d'arrêter le magnétophone fréquemment et sans prévenir le locuteur afin d'éviter l'effet de liste et une tendance du timbre de la voix à devenir de plus en plus grave au cours du temps. Après chaque interruption (de 2 à 3 minutes), l'opérateur faisait écouter à nouveau les derniers mots prononcés ce qui permettait au locuteur de "placer sa voix" avant de continuer.

Ces enregistrements ont porté sur environ 1 000 mots. Ceux-ci ont été alors analysés par l'intermédiaire du vocodeur à canaux à 4 800 eb/secondes. Ensuite, à partir des listings d'échantillons vocodeur, nous avons effectué "manuellement" la segmentation de tous les diphones nécessaires, et leur avons attribué une adresse symbolique de départ et d'arrivée. Cette phase de segmentation est de loin la plus longue et la plus délicate ; elle consiste également à opérer certaines corrections concernant la mesure de la période du fondamental, à contrôler si les niveaux d'énergie ne présentent pas des différences profondes de mot à mot, et à décider de la longueur qu'il est souhaitable d'attribuer à chaque diphone. Parce que chaque unité minimale a une durée fonction de son environnement phonétique et fonction de sa position dans le message, les segments sélectionnés - et, en particulier, leur zone consonantique - ont encore, malgré les soins de l'extraction et de la segmentation, une durée excessive par rapport à celle observée dans la parole continue (OLIVE et NAKATANI, opus cité).

Les diphones sont stockés dans la zone mémoire d'un disque fixe relié au calculateur , leur adresse étant calculée à partir du code ASCII de chaque caractère.

.../...

I - LA STRUCTURE DU DICTIONNAIRE DE DIPHONES

A terme, le système de synthèse que nous allons décrire n'est pas destiné à fonctionner de façon isolée comme "machine à lire et à parler", mais s'inscrit dans le chaîne "Reconnaissance et Synthèse" qui devra permettre un dialogue entre l'homme et la machine.

C'est pourquoi ce système, dans sa conception, tient compte des contraintes nées de l'application visée.

Ces contraintes se situent déjà au niveau de la structure du dictionnaire de diphones mais surtout dans la mise au point d'un traitement de la prosodie en synthèse.

Il ne faut pas oublier que dans l'application "dialogue homme-machine" on aura les séquences suivantes :

. question orale d'un correspondant → chaîne de la reconnaissance → écriture en code phonétique du message reconnu → répétition à la synthèse de la question posée → lecture de la réponse inscrite en code orthographique → transcription orthographique-phonétique et positionnement des marqueurs prosodiques → recherche et assemblage des diphones, traitement de la prosodie → réponse vocale du message.

Par conséquent, la synthèse , dernier lien de cette chaîne, dépend du bon déroulement des opérations qui la précèdent c'est-à-dire en particulier de la similitude du code phonétique utilisé et du positionnement correct des marqueurs prosodiques (ce qui implique l'élaboration de règles prosodiques simples et précises).

I-1 - Le code phonétique et les catégories de diphones.

Nous avons pris comme code phonétique, non pas le code de l'API *, mais celui utilisé par l'équipe de reconnaissance (fig. 30;31).

---

* API : Alphabet Phonétique International .../...

certains sons ayant deux caractères dans le code phonétique choisi en reconnaissance et en synthèse, on fait précéder ceux qui n'en possèdent qu'un d'un silence symbolisé par le signe [␣] , et ce afin d'harmoniser la lecture des caractères par le calculateur.

Code phonétique utilisé en synthèse	Code A.P.I.	Equivalent orthographique
1 - LES VOYELLES		
␣ I	[i]	poli
E I	[e]	café
A I	[ɛ]	procès
␣ A	[a]	papa
␣ O	[ɔ]	mode
A U	[o]	chevaux
O U	[u]	loup
␣ U	[y]	pur
E U	[ø]	creux
O E	[œ]	peur
␣ E	[ə]	particule
4 VOYELLES NASALES		
I N	[ɛ̃]	pain
A N	[ɑ̃]	espérance
O N	[ɔ̃]	honte
U N	[œ̃]	aucun

Notons surtout que la semi-consonne [ɥ] dans [huile,] [nuire] est envisagée ici comme une douzième voyelle orale, notée [ui].

FIG. 30 - CODE PHONETIQUE DES VOYELLES.

.../...

2/ LES CONSONNES		
⌴ P	[p]	pardon
⌴ T	[t]	tapis
⌴ K	[k]	cachette
⌴ B	[b]	abimé
⌴ D	[d]	adoré
⌴ G	[g]	agacé
⌴ F	[f]	enfant
⌴ S	[s]	serment
C H	[ʃ]	acheter
⌴ V	[v]	avouer
⌴ Z	[z]	azur
⌴ J	[ʒ]	jour
⌴ M	[m]	ami
⌴ N	[n]	année
⌴ L	[l]	lait
⌴ R	[r]	aride
⌴ W	[w]	oui, ouate ..
⌴ y	[j]	lieu, meunier

FIG. 31 - CODE PHONETIQUE DES CONSONNES. .../...

. la consonne nasale [ɲ] dans /agneau/ n'existe pas ici en tant que telle : on la produit en juxtaposant les deux consonnes /‿n + ‿y/

. d'autre part, nous n'avons pas réalisé la distinction entre [a] dans /papa/ et [ɑ] dans /âne/

Compte tenu du fait que la langue que l'on se propose de produire dispose de 34 réalisations phonémiques (plus le silence noté #) on aurait pu, déduction faite des juxtapositions de "phonèmes" non réalisables de par la nature physique des organes phonatoires, se contenter de stocker en mémoire à peine 900 diphones représentant toutes les combinaisons deux à deux ; mais nous allons voir que celles-ci ne donnent pas les moyens de composer n'importe quel message.

. L'ensemble des combinaisons entre les 16 voyelles et les 18 consonnes donnent :

- 288 diphones [voyelle-consonne]
- 288 diphones [consonne-voyelle]

. Mais en début de phrase, ou en début de mot subséquent à une pause, on ne peut pas commencer directement par un diphone [consonne-voyelle] ou [voyelle-consonne] car la première réalisation consonantique ou vocalique serait alors tronquée, le diphone allant de partie stable à partie stable ; il faut donc prévoir un premier segment qui représente le début d'une réalisation jusqu'à sa partie stable ; par exemple le mot /liberté/ nécessite à la synthèse la juxtaposition des diphones suivants :

/ # l / l i/ ib/ bai/...
↑

C'est pourquoi on trouve :

- 18 diphones représentant le début des réalisations consonantiques

[# - consonne]

.../...

- 15 diphones représentant le début des réalisations vocaliques :

[# - voyelle] ; le diphone [# -ə]est inexistant parce qu'irréalisable.

. De la même façon, il a fallu envisager tous les diphones susceptibles de terminer un mot avant une pause :

- 18 diphones représentent la fin des réalisation consonantiques [consonne - #]
- 16 diphones représentent la fin des réalisations vocaliques [voyelle - #] ; bien que [ɔ] par exemple soit impossible en français en fin de mot.

. Restent enfin les catégories de diphones qui représentent les séquences vocaliques et les groupements consonantiques :

- les suites [voyelle-voyelle] ne représentent que 240 diphones parce que la séquence [voyelle -ə]est irréalisable.

- les groupes [consonne-consonne] produisent 324 diphones.

On peut s'étonner que toutes les combinaisons [consonne-consonne] soient stockées bien que certaines ne soient pas réalisables.

Deux raisons justifient ce choix.

⁂ Toutes ces combinaisons peuvent effectivement ne pas exister dans une juxtaposition syllabique, mais elles peuvent apparaître côte à côte à la frontière de deux mots, et dans ce cas, il faut être capable de les reproduire.

⁂ Dans un système de Reconnaissance - Synthèse, il faut envisager le cas où l'un de ces groupements qu'on croyait impossible a été reconnu et nous est transmis. Il est indispensable alors que ce groupement possède une adresse en mémoire ainsi qu'une configuration acoustique.

.../...

Supposons en effet que la chaîne phonétique qui nous est transmise soit composée des diphones suivants :

/# a/ a t/t z/ z i/ i # /

Si le diphone /tz/ est inexistant dans le dictionnaire de diphones, la synthèse de ce mot sera quand même réalisée, mais sera le résultat de l'assemblage de [t] avant l'explosion, suivi de [z] au milieu de sa réalisation. Le résultat perceptuel - nous l'avons testé - n'est pas des plus convaincants ! Il est beaucoup plus simple d'essayer de faire prononcer ces groupements par le locuteur et d'en consigner le résultat (quel qu'il soit ; ce peut être par exemple assez proche de [dz]), en mémoire, où il correspondra - même si c'est de façon fort éloignée - à la réalisation [tz].

L'ensemble de ces combinaisons représente au total 1 207 diphones, ou plus exactement jusqu'à ce jour 1 219 diphones, parce que nous avons également stocké 12 diphones qui représentent la fin d'une réalisation consonantique située à la fin d'un mot possédant un schéma intonatif montant : l'analyse du corpus a mis en évidence la différence de spectre et d'amplitude de certains diphones situés avant une pause selon que F_0 croît ou décroît ; aussi avons-nous choisi - pour garantir une meilleure qualité de parole - de stocker les deux configurations de ces diphones plutôt que d'effectuer sur le spectre un traitement relativement complexe pour chaque message à synthétiser.

Ces diphones [consonne - #] stockés deux fois ont été enregistrés à partir de mots prononcés, d'une part, avec l'intonation descendante propre aux fins de phrase énonciative, d'autre part, avec l'intonation montante des fins de syntagmes situés par exemple avant le verbe (mais suivis d'une pause). Ce processus, qui permet de préserver la spécificité tant spectrale que temporelle de ces diphones dans l'une et l'autre position , ne concerne que les 12 consonnes voisées.

.../...

Nous voulions au départ, utiliser la même procédure en ce qui concerne les diphones [voyelle #] parce que c'est là que les différences d'évolution spectrale étaient les plus accusées selon la position de ce segment dans le message (fig.22). Cela n'a pas été possible parce que le spectre et l'amplitude de ce segment auraient présenté une trop grande discontinuité avec le diphone précédent [consonne-voyelle] qui, lui, connaît un stockage unique. De ce fait les diphones [voyelles #] sont issus de la dernière syllabe de mots prononcés avec une intonation montante ; quand F_0 possède une évolution de type descendant, on procède alors à un traitement sur les niveaux d'énergie ; ces segments contiennent en eux-mêmes, une durée qui permet de les utiliser à la fois quand F_0 croît et quand F_0 décroît, malgré la différence de l'allongement mise en évidence lors de l'analyse du corpus.

Le nombre des diphones stockés peut paraître élevé surtout quand on sait que d'autres systèmes se contentent de 400 diphones pour composer n'importe quel message (LIENARD, 1966, 1970, 1972) mais dans un premier temps, nous ne nous sommes pas soucié d'économie.

Le code phonétique que nous utilisons conduit à une écriture particulière des messages : la synthèse de la phrase "la Bretagne est magnifique" impose l'écriture phonétique suivante :

␣L␣A␣BREU␣T␣A␣N␣Y␣ EEI␣M␣A␣N␣Y␣I␣F␣ L␣K.

(le signe ␣ correspond à un espace)

Cette phrase comporte les diphones suivants :

/#L/LA/AB/BR/REU/EUT/TA/AN/NY/YE/EEI/EIM/MA/AN/NY/YI/IF/FI/IK/K#/

Ce code peut paraître complexe ; pourtant il est bien simple comparé à celui utilisé par DIXON et MAXEY (1968) ; et puis, l'habitude faisant, c'est le retour au code orthographique qui devient difficile !.

.../...

I-2- La durée des diphones en bibliothèque

Nous ne gardons en mémoire, pour une question de taille de bibliothèque, qu'une seule configuration temporelle pour chaque diphone. Il a donc fallu définir avec soin la durée qui devait être attribuée à chacun d'eux.

Les résultats obtenus à l'analyse ont servi de point de départ pour la fixer. Nous avons délibérément opté pour une durée supérieure à celle relevée dans le corpus de parole continue parce que les résultats sont bien meilleurs quand on commande par la suite une accélération de la cadence d'échantillonnage plutôt qu'un ralentissement de cette cadence.

Pour cette raison, la durée choisie pour chaque diphone correspond à la durée la plus longue relevée pour chacun des deux segments le composant dans les syllabes non situées en fin de mot. D'autre part des contraintes informatiques nous ont imposé une durée maximale de 213 ms pour chaque diphone.

Pour faciliter le traitement informatique de l'intonation nous avons fixé une durée presque identique pour tous les diphones [consonne-voyelle] : en effet le traitement de F_0 impose que la voyelle de ces diphones soit précédée, ou bien du même nombre d'échantillons qui représentent chaque consonne, ou bien de valeurs représentatives de leurs caractéristiques intrinsèques qui présentent une variation de F_0 identique.

On a donc préféré que la durée de la consonne et de la voyelle soit fixée dans le diphone [voyelle-consonne] qui suit, parce que la consonne de ce diphone ne sert pas de point de repère pour le traitement prosodique, et parce que ce processus permet de tenir compte, pour le choix de la durée vocalique, de la nature de la voyelle subséquente.

.../...

En conséquence, la durée moyenne des diphones [consonne-voyelle] est comprise entre 100 et 120 ms quels que soient les segments qui les composent ; par exemple dans les diphones /ti/, /ta/, /tan/, /tou/..., /i/ = /a/ = /an/ = /ou/.

Par contre, la durée des diphones [voyelle-consonne] est très variable :

. elle dépend de la durée de la voyelle qui varie (dans cette catégorie de diphones) entre 26 ms et 70 ms selon que cette voyelle est orale brève [i,y], ou nasale [æ̃, ɔ̃] et selon la nature de la consonne subséquente : on sait par exemple que la voyelle qui précède une consonne voisée est plus longue que la même voyelle précédant une consonne sourde.

. elle dépend de la nature de l'élément consonantique, et va de 30 ms pour une occlusive voisée à 100 ms pour une constrictive sourde.

Le fait que la durée soit presque semblable dans les diphones [consonne-voyelle] a rendu plus facile la fixation de durée des diphones [# consonne] : il a suffit de leur attribuer la durée qui leur manquait par rapport à celle relevée en initiale de mots dans le corpus de parole continue ; le procédé a été identique pour évaluer la durée des diphones [voyelle #].

Par contre, les difficultés ont concerné la fixation de la durée des diphones [# voyelle] puisque une même voyelle dans les diphones [voyelle-consonne] présente des configurations temporelles différentes selon la consonne qui lui succède. Dans ces cas, la détermination d'une durée moyenne et convenable pour l'ensemble des occurrences s'est faite par approximations successives à l'écoute. Nous avons opéré de la même façon pour calculer la durée des diphones [consonne-#].

C'est également en procédant à des tests que nous avons pu connaître la durée minimale des groupements consonantiques ; ceux-ci sont

.../...

stockés sans aucune possibilité d'allongement ou de réduction de durée : leur configuration temporelle est figée, elle a été fixée de façon à pouvoir intervenir telle quelle, dans n'importe quel contexte.

Quant aux diphones [voyelle-voyelle], ils constituent la catégorie la plus longue, comprise entre 165 et 213 ms selon la nature de chacune des voyelles ; mais un ensemble de marqueurs permet, selon le contexte, un allongement ou une réduction de leur durée.

I-3- Les difficultés de méthode liées au spectre.

La méthode de synthèse par diphones et les contraintes qui tiennent à la taille de la bibliothèque obligent souvent à des simplifications qui ne reposent sur aucune donnée de la phonétique.

De la même façon que nous ne gardons qu'une seule configuration temporelle pour chaque diphone, nous ne conservons qu'une seule image spectrale en mémoire ; les problèmes inhérents à cette contrainte sont nombreux et graves. Nous en avons déjà évoqué un certain nombre dans la seconde partie de ce mémoire (chapitre III) ; nous voudrions simplement ici en citer un autre, dont les conséquences perceptuelles sont quand même moins aiguës.

Les phénomènes d'assimilation (modifications que les sons subissent au contact d'autres sons) sont bien connues en phonétique générale. Ces influences se situent souvent au niveau de la sonorité quand deux consonnes sont juxtaposées : les liquides et les semi-consonnes qui sont voisées dans un entourage vocalique ou consonantique voisé ont tendance à s'assourdir au contact d'une consonne sourde qui les précède (assimilation progressive - fig. 32), ou qui les suit (assimilation regressive - fig. 33).

.../...

7952	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	
7953	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
7954	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	k
7955	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7956	1	1	0	0	0	2	3	2	0	0	1	1	2	2	0	15	
7957	5	6	5	6	7	7	8	6	5	5	5	5	5	7	0	82	
7958	7	8	6	7	9	9	8	5	4	2	6	6	6	9	0	92	
7959	5	6	6	6	9	9	7	5	4	3	6	6	7	9	0	88	
7960	0	0	4	5	8	8	5	6	4	3	7	7	8	8	0	73	
7961	3	3	3	5	7	6	4	4	2	3	5	7	8	7	0	67	
7962	2	3	3	4	7	5	4	3	2	2	5	6	8	7	0	61	r
7963	0	2	3	5	8	5	4	4	2	2	6	7	8	7	0	63	
7964	0	2	2	5	7	5	3	3	1	3	6	6	8	7	0	58	
7965	0	2	3	4	6	6	4	3	1	3	6	7	8	7	0	60	
7966	1	2	3	3	7	6	3	3	1	3	7	5	7	6	0	57	
7967	2	2	3	4	6	4	2	1	0	0	4	2	4	3	0	37	
7968	9	8	9	9	9	7	4	4	2	4	5	3	3	3	0	79	
7969	10	10	11	10	11	9	6	4	5	6	6	5	5	5	99	103	
7970	10	10	11	10	11	9	6	5	4	7	8	5	5	5	102	106	
7971	9	10	11	10	10	10	6	5	5	7	8	5	5	6	102	107	
7972	10	10	11	10	10	10	6	4	5	7	7	5	6	6	88	107	
7973	10	10	11	10	10	10	5	5	5	6	6	5	5	5	101	103	a
7974	10	10	11	10	9	9	5	5	4	6	7	5	4	5	102	100	
7975	10	10	10	9	9	9	5	5	3	5	6	4	5	5	107	95	
7976	10	10	9	8	8	7	4	3	2	2	5	3	4	3	114	78	
7977	9	7	0	3	5	2	0	0	0	0	2	0	0	0	114	28	
7978	5	0	0	0	0	0	0	0	0	0	0	0	0	0	109	5	
7979	6	2	1	0	0	0	0	0	0	0	0	0	0	0	109	9	
7980	6	4	1	1	0	0	0	0	0	0	0	0	0	0	109	12	
7981	6	3	1	0	0	1	0	0	0	0	0	0	0	0	110	11	b
7982	6	3	1	0	0	1	0	0	0	0	0	0	0	0	112	11	
7983	6	3	1	0	0	1	0	0	0	0	0	0	0	0	113	11	
7984	6	3	1	0	0	0	0	0	0	0	0	0	0	0	116	10	
7985	7	4	4	3	3	3	3	2	1	0	2	1	2	2	119	37	
7986	10	9	6	6	5	5	4	4	3	4	4	3	4	5	119	72	
7987	11	11	7	6	6	7	4	3	1	5	4	1	3	4	109	73	
7988	11	11	8	7	7	8	5	3	1	6	5	2	3	5	109	82	a

FIG. 32 - Assimilation progressive de [r] dans le mot/crabe/.

.../...

7430	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	
7431	4	0	0	0	0	0	0	1	1	0	0	0	0	0	0	6	
7432	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
7433	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	p
7434	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7435	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7436	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7437	5	5	4	4	4	5	5	3	1	0	4	3	3	4	0	50	
7438	10	10	9	8	8	8	7	5	5	6	6	5	6	5	138	98	
7439	11	11	12	10	10	11	8	5	5	8	8	5	7	5	71	116	
7440	10	10	12	11	10	11	7	5	5	6	8	6	7	5	72	113	a
7441	9	8	10	11	11	10	6	6	5	6	9	7	7	5	76	110	
7442	8	6	7	10	9	7	6	5	3	4	8	6	7	4	85	90	
7443	4	3	6	8	8	6	4	3	1	3	8	7	7	5	94	73	
7444	1	4	5	8	8	6	4	3	1	0	7	7	7	7	0	68	
7445	3	4	5	7	8	5	4	3	1	0	5	7	8	7	0	67	
7446	4	5	5	6	7	5	5	4	3	1	4	6	7	5	0	67	r
7447	5	6	5	6	6	6	5	5	5	3	4	5	5	6	0	72	
7448	6	7	5	4	5	7	6	6	5	5	6	5	6	7	0	80	
7449	5	6	5	4	5	6	6	6	5	5	6	6	6	7	0	78	
7450	4	6	4	3	5	5	6	6	6	6	6	6	6	7	0	76	
7451	4	5	4	4	5	4	6	5	5	4	7	7	5	7	0	72	
7452	4	3	3	4	3	4	6	5	5	4	7	6	5	7	0	66	f
7453	3	4	3	3	3	4	6	4	3	3	5	3	3	5	0	52	
7454	5	5	3	3	2	3	5	3	2	1	4	3	2	4	0	45	
7455	11	11	7	5	4	7	9	6	5	5	7	4	4	6	72	91	
7456	12	12	10	7	6	8	11	8	7	6	9	6	6	8	74	116	
7457	12	11	10	7	5	7	11	9	6	6	8	7	6	6	75	111	ɛ
7458	11	10	8	6	6	7	10	7	4	4	8	7	4	4	66	96	
7459	9	7	1	2	3	2	0	0	0	0	5	2	0	2	66	33	
7460	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7461	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	t

FIG. 33 - Assimilation régressive de [r] dans/parfaitement/.

.../...

Or, les diphones [consonne-voyelle] ont tous été pris dans des mots de type CVCVCV, ce qui signifie que la consonne de ce segment est située dans un entourage essentiellement vocalique : elle est par conséquent voisée. Au contraire et par définition, la seconde consonne d'un groupement consonantique est précédé d'une consonne : [r] dans le diphone /tr/ est précédé d'une consonne sourde : par assimilation, il devient assourdi lui-même .

Quel va être le résultat au niveau de la concaténation à la synthèse ?

Le mot/travail/, par exemple, va présenter la succession /t+r/ sourds + /r + a / sonores, c'est-à-dire qu'au centre de sa réalisation, [r] va devenir sonore, ce qui ne se produit jamais dans la parole naturelle : [r] présente un dévoisement pendant toute la durée de sa réalisation.

Pour essayer d'enrayer ce défaut, nous avons testé si les liquides et les semi-consonnes entièrement voisées sont bien tolérées même quand elles succèdent à une consonne sourde. Les résultats perceptuels ont montré que [r] entièrement voisé après une consonne sourde provoque une "résonance" désagréable responsable d'une diminution d'intelligibilité sur la syllabe à laquelle il appartient.

Comme d'autre part, l'inclusion d'un élément de programme qui viendrait étudier, lors du décodage du message à synthétiser, l'environnement de [r] , nécessiterait un temps de calcul qui risquerait de compromettre le temps réel, nous avons préféré laisser la succession entre une zone sourde et une zone sonore pour ces consonnes. Il faut d'ailleurs reconnaître que le vocodeur n'est pas vraiment sensible à cette discontinuité.

.../...

I-4-L'insertion de marqueurs prosodiques

L'assemblage simple des diphones permet une parole intelligible (après une phase d'apprentissage), mais absolument dépourvue de naturel : toutes les syllabes se succèdent à un débit assez lent, et sur un ton monocorde (désiré pour l'enregistrement), rien ne vient interrompre le flux de parole avant la fin du message ; il est par conséquent très difficile d'identifier ne serait-ce que les frontières de mots.

Le traitement de la prosodie tend à donner à cette parole "sauvage" quelques-unes des caractéristiques prosodiques qui se manifestent dans la parole naturelle.

Ce traitement comporte :

. Une face cachée : il n'était pas envisageable d'inscrire au fur et à mesure de l'écriture phonétique du message toutes les indications nécessaires à ce traitement : on serait arrivé rapidement à une trop grande complexité dans l'écriture.

Par conséquent, nous avons inscrit en bibliothèque et sur tous les diphones concernés, des marqueurs qui ont une signification spécifique pour le traitement de la prosodie.

. Une face visible : le message sous sa forme écrite doit comporter d'autres marqueurs (en nombre limité) qui renvoient à ceux positionnés en bibliothèque sur les diphones. Chacun de ces marqueurs doit être inscrit en des points très précis de l'énoncé ; cela ne pose pas de difficultés quand un opérateur écrit le message, mais en pose dans un système entièrement automatique : le positionnement correct des marqueurs suppose résolus les problèmes liés à l'analyse syntaxique automatique, il faut élaborer le processus qui permettra à une machine de passer de la lecture d'une phrase à sa décomposition en syntagmes et en groupes de sens.

.../...

I-4-1- Les marqueurs prosodiques insérés dans le dictionnaire.

a/ Les marqueurs liés au traitement de la durée.

Ces marqueurs vont être utilisés soit pour accélérer le débit, soit pour le ralentir. En particulier, une accélération s'est révélée nécessaire dans la réalisation des mots monosyllabiques non situés aux points clés définis à l'analyse, et dans la réalisation des syllabes non initiales et non finales de mots plurisyllabiques. Un ralentissement sera par contre indispensable sur certaines syllabes finales de mots, par exemple en fin de phrase ou en certains endroits qui, en même temps qu'une montée de Fo, présentent un allongement caractéristique.

Une double possibilité s'offrait à nous :

. Dans le cas d'une réduction de durée, on pouvait soit supprimer certains échantillons, soit laisser le même nombre d'échantillons, mais accélérer la cadence d'échantillonnage.

. De la même façon, dans le cas d'un allongement, on pouvait soit doubler un certain nombre d'échantillons, soit ralentir la cadence d'échantillonnage.

Nous avons choisi d'agir sur la cadence d'échantillonnage afin d'éviter des modifications dans l'enveloppe spectrale des segments.

Trois cadences sont dès lors possibles pour chaque échantillon :

vitesse normale	: 13,3	ms
vitesse accélérée	: 6	ms
vitesse ralentie	: 26	ms

. Les marqueurs inscrits dans les diphones [consonne-voyelle]:

- les marqueurs destinés à accélérer le débit :

.../...

. des monosyllabiques,

. des syllabes intérieures des mots plurisyllabiques.

* un marqueur sur le premier échantillon de la réalisation consonantique

* un marqueur sur le dernier échantillon de la voyelle, on le note X_1.

Dans l'un et l'autre cas, ce marqueur correspond à une possibilité d'accélération de l'échantillon marqué qui passera de 13,3 ms à 6 ms.

- les marqueurs destinés à ralentir le débit :

* un marqueur sur l'avant dernier échantillon de la voyelle (donc du diphone). Il indique le départ d'une zone de ralentissement à partir de laquelle la durée de chaque échantillon vocodeur passe de 13 ms à 26 ms. A compter de ce point, deux zones - que nous définirons par la suite - sont possibles : Z_1 et Z_2 .

* un marqueur sur le second échantillon de la consonne pour signaler le retour de l'horloge au rythme de 13,3 ms après l'accélération sur le premier échantillon.

6	6	8	8	7	7	4	3	1	2	7	6	7	6	128	78	0	1
6	5	7	7	7	7	4	3	3	2	7	6	7	6	254	77	0	11
7	4	6	8	7	7	4	3	3	2	6	6	7	6	254	76	0	0
7	6	8	9	8	8	5	4	4	5	7	6	7	6	127	90	0	0
9	9	10	9	8	8	6	5	5	7	9	6	6	5	1	102	0	0
11	11	11	9	7	10	9	8	7	7	7	8	8	6	1	119	4	0
11	11	12	10	9	10	8	7	7	9	8	8	9	6	1	125	0	2
11	11	12	11	10	10	8	7	7	10	7	7	9	7	1	127	0	1

diphone /ra/ - (les marqueurs de rythme sont inscrits dans la 18e colonne).

.../...

. les marqueurs inscrits dans les diphones [voyelle-consonne]:

Leur rôle est de mettre fin à une zone d'accélération ou de ralentissement qui aurait pris naissance dans le diphone précédent.

- un marqueur sur le premier échantillon de la voyelle : il indique le retour au rythme d'échantillonnage normal de 13,3 ms en fin d'accélération de X_1 .

- un marqueur sur le second échantillon de la voyelle : il marque le terme de la zone Z_1 pendant laquelle la cadence d'échantillonnage était passée à 26 ms. On le voit, cette zone ne peut excéder trois échantillons vocodeur.

- un marqueur, sur l'un quelconque des échantillons de la voyelle, qui signale la fin de ralentissement (à 26 ms) commencé sur l'avant dernier échantillon de la voyelle précédente (fin de Z_2).

Cette zone (Z_2) de ralentissement est destinée à introduire l'allongement caractéristique des voyelles situées en syllabe fermée à la fin d'un mot suivi d'une pause.

Le choix de l'échantillon marqué varie selon la nature de la voyelle et selon l'élément consonantique subséquent:

. zone plus longue pour une voyelle nasale que pour une voyelle orale.

. zone plus longue pour une voyelle située devant une constrictive que pour une voyelle précédant une occlusive sourde :

Pour une même position dans le message, [a] dans/épave/ - voyelle suivie de [v] constrictive sonore - est évidemment affectée d'une zone d'allongement sensiblement supérieure à celle de [a] dans /épate/ où la voyelle est abrégée par la présence de l'occlusive sourde [t].

.../...

On se souvient que les voyelles dans ce contexte consonantique et dans cette position présentent à l'analyse une durée différente selon que Fo est montant ou descendant.

Mais nous ne pouvions pas alourdir démesurément les diphones par un trop grand nombre de marqueurs, ni souvent allonger les voyelles autant que nous le voulions (nous ne pouvons que doubler leur durée) ; par conséquent, les voyelles présenteront une durée identique que le schéma intonatif soit montant ou descendant.

L'allongement permet pour les voyelles les plus longues ([ɛ] + [r,l])une durée maximale de 190 ms, au lieu des 280 ms relevés dans la parole continue : en effet, dans le dictionnaire [ɛ] suivi de [r] a une durée de 110 ms ; comme on préfère ne doubler la durée des échantillons qu'à partir de la zone stable de la voyelle (avant dernier échantillon de la voyelle dans les diphones [consonne-voyelle]), le ralentissement de la cadence n'intervient que sur 80 ms ; la durée finale de cette voyelle après allongement sera donc 30 ms + (80 ms x 2) = 190 ms.

										6	9	8	8	9	7	7		1	94	0	3
12	9	5	3	4	3	4	6	9	8															
12	9	6	4	5	4	5	7	11	8	9	8	7	7					1	102	0	3
11	11	8	5	6	7	9	10	8	7	10	8	6	7					1	113	0	0
12	11	8	5	6	8	9	8	4	7	9	7	5	5					1	104	4	0
11	11	7	5	7	8	8	5	5	5	7	6	5	4					126	94	0	3
10	9	5	6	7	7	8	4	4	4	7	7	5	4					1	87	0	0
9	8	6	6	7	8	8	4	2	3	6	6	5	3					1	81	0	0

Diphone /ir/

. Les marqueurs inscrits dans les diphones [voyelle-voyelle]:

- la première voyelle est considérée comme la voyelle des diphones [voyelle-consonne] , c'est-à-dire que l'on trouve un marqueur de rythme sur le premier échantillon pour signaler la fin de l'accélération (retour à l'échantillonnage à 13,3 ms) commencée en X_1 sur la voyelle du diphone précédent, et un marqueur sur le second échantillon pour indiquer la fin de la zone Z_1 (zone de ralentissement à 26 ms).

.../...

Mais évidemment on ne trouve pas ici de marqueur qui signale la fin de Z_2 puisque cette voyelle ne peut jamais intervenir comme dernière syllabe d'un mot.

- la seconde voyelle, quant à elle, contient les mêmes marqueurs de rythme que tout un diphone [consonne-voyelle] ; on trouve :

* sur le premier échantillon, un marqueur d'accélération pour cet échantillon uniquement,

* sur l'avant dernier échantillon, un marqueur de ralentissement pour indiquer le début d'une zone d'allongement (à 26 ms) Z_1 ou Z_2 ,

* sur le dernier échantillon, un marqueur indique le début de la zone d'accélération X_1 qui se terminera avec le premier échantillon de la voyelle du diphone suivant.

11	11	12	9	6	7	8	11	8	8	9	8	6	6		1	120	0	3
11	11	11	8	6	6	7	9	8	8	10	9	7	6		1	117	0	3
11	10	10		6	6	6	7	8	7	9	9	6	5		1	107	0	0
12	10	10		5	4	6	8	9	9	9	9	6	5		1	109	0	0
12	11	9	6	5	4	5	7	9	9	8	8	6	4		1	103	0	0
13	11	7	4	5	3	3	5	8	8	8	8	6	5	124		94	0	0
13	10	6	3	3	3	4	5	9	9	8	7	6	5		1	91	0	0
11	9	6	4	1	2	3	5	7	9	8	8	7	5		1	85	0	0
11	8	6	4	1	1	3	4	5	9	7	8	7	6		1	80	0	0
12	9	6	3	1	2	3	4	6	9	7	8	7	5		1	82	4	0
12	9	6	2	1	2	3	4	6	8	8	8	7	6	128		82	4	1
12	9	6	3	1	2	3	4	5	8	8	7	8	6	127		82	0	11
12	8	5	2	1	2	3	3	4	8	7	7	9	6		1	77	0	0
12	8	5	2	2	1	3	3	4	8	7	7	9	6		1	77	0	2
11	8	5	2	2	2	3	3	4	9	7	7	9	6		1	78	0	1

Marqueurs de rythme dans le diphone / a - i /.

.../...

. les marqueurs inscrits dans les diphones [# voyelle] :

Ce sont exactement les mêmes marqueurs que ceux que l'on trouve dans les diphones [consonne-voyelle] :

* un marqueur d'accélération sur le premier échantillon de la voyelle,

* un marqueur de fin d'accélération sur le second échantillon,

* un marqueur de début de zone de ralentissement sur l'avant dernier échantillon de la voyelle : cette possibilité d'allongement sera utilisée uniquement dans le cas d'un mot monosyllabique situé après une pause :

- le chat $\not{p}$ aime les souris.
  z_1

- Ah, que j'aime la campagne !
  z_2

7	8	9	8	6	6	8	5	4	4	6	5	4	5	128	85	0	1
10	11	13	11	9	8	10	7	6	7	9	7	5	7	127	120	4	11
11	11	13	11	9	9	10	8	7	7	8	7	5	7	1	123	0	0
11	11	13	11	9	9	10	8	7	7	8	7	5	7	1	123	0	2
11	12	12	11	9	9	9	9	8	8	9	8	6	6	1	127	0	1

Marqueurs de rythme dans le diphone / # a/.

. les marqueurs des diphones [# consonne].

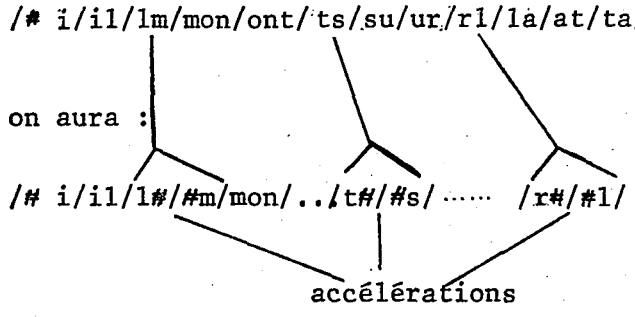
L'unique marqueur sert à indiquer la fin d'une zone d'accélération inscrite sur un échantillon fixe : sa position dépend de la longueur de ce segment consonantique.

.../...

Ce marqueur n'est évidemment pas utilisé quand ce diphone apparaît après une pause puisque dans ce cas aucun diphone ne le précède qui pourrait signaler le début de la zone d'accélération. Il est utilisé en début de mot quand le mot qui le précède se termine par une consonne. En effet, on préfère désormais, dans cette situation, plutôt que d'utiliser un diphone [consonne-consonne] pour réaliser la frontière entre les deux mots, utiliser les diphones [consonne #] et [# consonne] et procéder à une accélération pendant leur réalisation :

Dans l'exemple "il monte sur la table".

au lieu d'avoir la suite :



Pour cette raison, les diphones [consonne - #] ont un marqueur de début d'accélération (cadence d'échantillonnage à 6 ms) sur l'un de leurs échantillons, le choix de l'échantillon étant fonction de la longueur de la réalisation consonantique.

Quant aux diphones destinés à terminer un mot situé avant une pause, c'est-à-dire essentiellement [voyelle #] et [consonne #], ils sont stockés sans marqueur de ralentissement puisqu'ils réalisent déjà, dans leur forme, l'allongement nécessaire ; cet allongement est directement obtenu par simple concaténation des diphones dans le cas de syllabe finale ouverte et, dans le cas de syllabe fermée, la prise en compte de la zone de ralentissement (Z_2) sur la voyelle finale, associée au diphone [consonne #] convenablement stocké, réalise l'allongement nécessaire.

.../...

*-

5	1	0	0	0	0	0	0	0	0	0	0	0	0	125	6	0	0
7	3	1	0	0	0	0	0	0	0	0	0	0	0	5	11	0	0
7	3	1	0	0	0	0	0	0	0	0	0	0	0	1	11	0	0
7	5	2	0	0	0	0	0	0	0	0	0	0	0	1	14	0	0
7	5	3	0	0	0	0	0	0	0	0	0	0	0	1	15	0	11
7	4	2	0	0	0	0	0	0	0	0	0	0	0	1	13	0	0

Marqueur de rythme dans le diphone / #b /

La figure 34 récapitule les différentes zones d'accélération et de ralentissement inscrites selon les catégories de diphones.

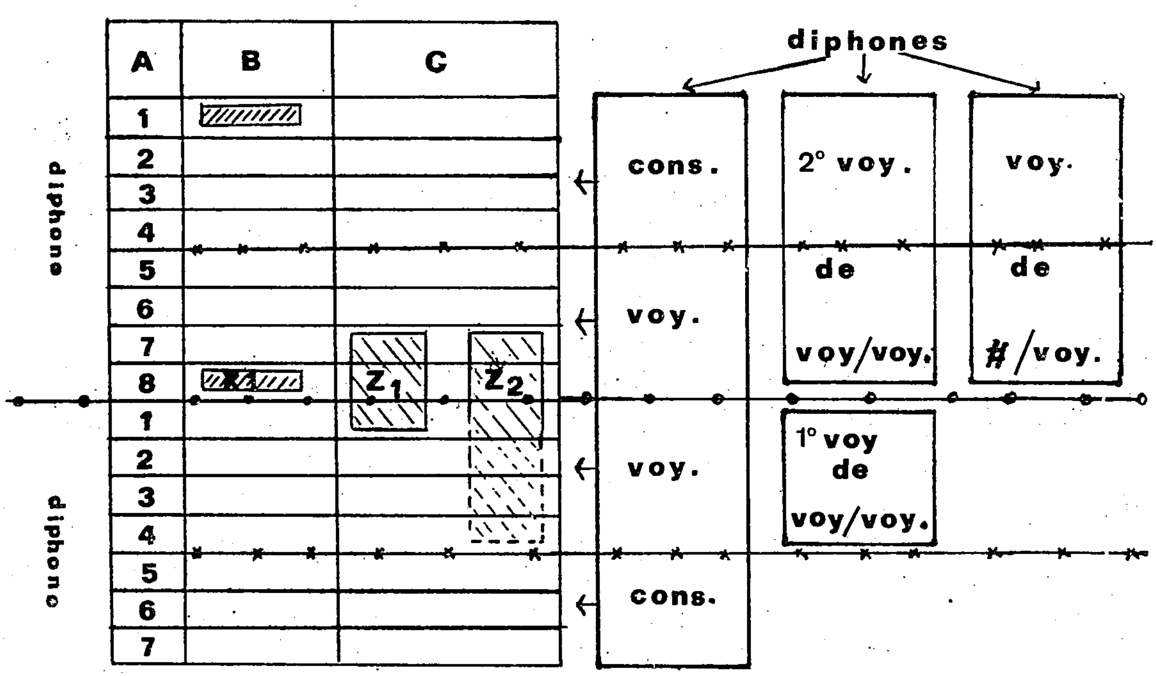


FIG. 34 - Délimitation des zones de rythme selon les diphones:

A - Numéros des échantillons pour deux diphones.

B - Différentes zones d'accélération de la cadence d'échantillonnage à 6 ms.

C - Possibilités de ralentissement à 26 ms.

.../...

b/ Les marqueurs relatifs au traitement de l'intensité .

Ces marqueurs ne concernent que les réalisations vocaliques. Parce que, comme nous l'avons dit, le vocodeur à canaux ne donne pas un accès direct au paramètre d'intensité, le traitement que nous réalisons est très sommaire et relève de l'empirisme. Il consiste à opérer une translation de - 4 db de la courbe donnant l'évolution temporelle de l'énergie des échantillons successifs du diphone.

. dans les diphones [consonne-voyelle]:

Un marqueur situé au début de la réalisation vocalique indique l'échantillon à partir duquel est réalisée cette translation (les marqueurs d'intensité apparaissent dans les listings d'échantillons vocodeur à la colonne 17).

. dans les diphones [voyelle-consonne]:

Un marqueur identique en fin de réalisation vocalique signale le terme de la zone de translation.

. dans les diphones [# voyelle]:

Le début de la voyelle est marqué pour indiquer le point de départ de l'éventuelle zone d'affaiblissement.

. dans les diphones [voyelle-voyelle]:

On a prévu un marqueur de fin de zone d'affaiblissement sur le dernier échantillon de la première voyelle et un marqueur de début de zone d'affaiblissement sur le premier échantillon de la seconde voyelle.

. dans les diphones [voyelle #]:

Il n'y a aucun marqueur : la zone d'affaiblissement commencée

.../...

au début de la réalisation vocalique du diphone précédent [consonne-voyelle] se poursuit jusqu'au terme du diphone [voyelle #].

c/ Les marqueurs relatifs au traitement de Fo.

Les mots de base utilisés pour la confection du dictionnaire ont été prononcés sur un ton volontairement monocorde et neutre par le locuteur parce que cette procédure permettait - entre autres - d'éviter les différences importantes dans l'évolution de la fréquence fondamentale et donc des discontinuités au moment de l'assemblage des diphones en message . Il n'était donc pas question de conserver les valeurs de Fo délivrées par le vocodeur à l'analyse car celles-ci ne permettent d'obtenir qu'une parole de synthèse tout à fait dépourvue de naturel et où l'ensemble du message est émis sur le même ton.

Cependant, le procédé d'enregistrement et d'obtention des diphones permet de connaître toutes les variations de la fréquence fondamentale - variations micromélodiques indépendantes de la volonté du locuteur - relatives aux caractéristiques intrinsèques des réalisations voisées.

Il nous est apparu indispensable de conserver ces variations au moins pour ce qui concerne la latérale [l] , les nasales [m] ,[n] , la vibrante [r] , et les semi-consonnes [j] et [w] , extrêmement sensibles, non pas tellement à une modification de leur tracé micromélodique, mais surtout à une brusque discontinuité de Fo durant leur réalisation. Par contre, cette possibilité de discontinuité ne semble pas avoir d'incidence sur l'intelligibilité des occlusives et des constrictives sonores.

c-1- Les valeurs représentatives des caractéristiques intrinsèques des consonnes voisées.

. Pour les raisons évoquées ci-dessus, et en ce qui concerne la fréquence fondamentale des consonnes voisées - les articulations consonantiques sont beaucoup plus évolutives que les réalisations vocaliques - nous gardons, non pas les valeurs absolues telles qu'elles sont délivrées par le vocodeur, mais des valeurs relatives, qui, par leurs

.../...

écarts, respectent grossièrement les schémas des variations micromélodiques.

. De la même façon, les valeurs absolues de Fo correspondant aux voyelles ne sont pas utilisées. On les remplace par une valeur unique - 1 - positionnée sur chaque échantillon (ces valeurs apparaissent sur les listings d'échantillons vocodeur dans la 15e colonne réservée à la mesure de la période du fondamental).

Cependant, certains diphones qui présentent une grande durée sont marqués de la valeur - 5 - sur certains de leurs échantillons ; l'échantillon ainsi marqué possèdera une répétition de la valeur de Fo attribuée à l'échantillon précédent.

. Afin de respecter le schéma spécifique des consonnes voisées, c'est-à-dire un tracé de la fréquence fondamentale de type descendant-montant, on impose à la période(*) du fondamental des consonnes dans les diphones [voyelle-consonne] , des valeurs relatives positives, et au contraire aux consonnes voisées des segments [consonne-voyelle] des valeurs relatives négatives comprises sur chaque échantillon entre [- 1] et [- 3] et dont la somme, quel que soit le nombre d'échantillons de la consonne, est égale à [- 4].

Nous verrons qu'il est extrêmement important pour le traitement de l'intonation que toutes les consonnes voisées présentent dans les diphones [consonne-voyelle] le même écart relatif entre les périodes du premier et du dernier échantillon : c'est le seul moyen dont nous disposions pour être sûre que-tout en ignorant la composition phonétique d'un diphone [consonne-voyelle] et sa configuration temporelle - la valeur de la période (x) imposée au premier échantillon de la consonne dans les diphones [consonne-voyelle] donnera (par soustraction des valeurs relatives positionnées sur la consonne) un niveau de période (y) identique sur le premier échantillon de la voyelle dans le même diphone : que cette consonne présente 6 ou 4 échantillons, la somme des valeurs relatives sera 4 et par conséquent toute voyelle aura sur son premier échantillon une valeur de période égale à la valeur de départ positionnée sur la consonne moins 5.y= x - 5.En effet, entre le dernier échantillon de la consonne et le premier de la voyelle, il y a toujours en période un niveau de différence.

(*) Pour les équivalences période codée/fréquence, voir le tableau en annexe.

.../...

. Dans les diphones [# consonne] , on a positionné sur tous les échantillons représentatifs d'une consonne sonore, des valeurs relatives positives comprises aussi entre 1 et 3 et dont la somme doit être également égale à 4.

. Les valeurs relatives positives inscrites sur chaque échantillon des consonnes voisées dans les diphones [voyelle-consonne] pour leur attribuer une micromélodie descendante sont généralement 1 quand cette consonne est longue, et 2 quand la consonne est courte.

. Dans les diphones [consonne-consonne] plusieurs cas ont été envisagés :

- groupe consonantique [consonne sourde - consonne sourde] : aucune micromélodie n'est insérée (la période, inexistante est codée Ø).

- groupe [consonne sourde-consonne sonore] : des valeurs relatives positives (Fo descendant) sont insérées sur la consonne sonore de telle façon que la somme de ces valeurs soit égale à 4 comme pour les consonnes des diphones [consonne-voyelle].

- groupe [consonne sonore - consonne sonore] : on réalise par le positionnement de valeurs relatives négatives, un schéma de Fo de type montant sur la première consonne, et au contraire de type descendant sur la seconde consonne avec des valeurs relatives positives dont la somme est égale à la somme des valeurs inscrites sur les consonnes des diphones [consonne-voyelle] c'est-à-dire 4.

8	5	3	1	0	0	0	0	0	0	0	0	0	1	255		18	0	0
8	5	2	1	0	0	1	0	0	0	0	0	0	0	255		17	0	0
9	8	7	5	6	5	4	4	5	7	6	6	5	6	254		83	0	0
9	9	9	8	6	3	4	4	2	5	7	5	5	5	125		81	0	0
8	6	6	6	7	5	4	3	1	4	7	4	5	4	1		70	0	0
6	6	6	6	7	6	4	3	1	3	7	6	6	5	1		72	0	0
6	6	7	7	7	6	4	3	1	2	7	6	6	6	1		74	0	0

Diphone /br/ - Les marqueurs d'intonation (15e colonne).

.../...

. dans les diphones [consonne sonore - ə]

On insère des valeurs relatives négatives sur chaque échantillon de la consonne pour obtenir un schéma de Fo de type montant ; sur [ə], on inscrit des valeurs relatives très petites et alternativement positives et négatives qui donneront à cette voyelle une très faible amplitude de Fo.

. dans les diphones [consonne sonore #] prononcés avec une intonation montante ou une intonation descendante, on attribue à la consonne une micromélodie de type descendant-montant en incluant d'abord des valeurs relatives positives puis des valeurs négatives.

. dans les diphones [ə - consonne sonore], les valeurs relatives insérées sur [ə] ménagent pour cette voyelle une évolution de Fo quasi-stationnaire ; sur la consonne on impose une micromélodie descendante par l'intermédiaire de valeurs positives dont la somme est égale à 4 comme pour la consonne des diphones [voyelle-consonne sonore].

La figure 35 récapitule les schémas d'évolution de la fréquence fondamentale qui seront réalisés à partir de la prise en compte des valeurs relatives insérées sur les échantillons des consonnes sonores.

c-2- Les marqueurs du dictionnaire destinés à connaître les frontières consonantiques et vocaliques dans chaque diphone.

Ces marqueurs servent de point de repère dans le traitement de la prosodie :

- pour savoir si un diphone débute par une zone consonantique,
- pour savoir sur quel échantillon commence et finit une réalisation vocalique.

. on signale le point de départ de tous les diphones [consonne-voyelle] par le marqueur 128. Il indique d'une façon générale le point de départ d'une syllabe ou plus exactement d'un "diphone syllabique", c'est à dire qui contient le début d'une réalisation vocalique :

.../...

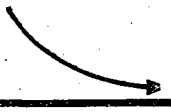
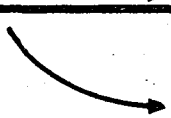
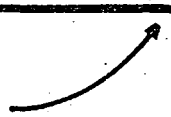
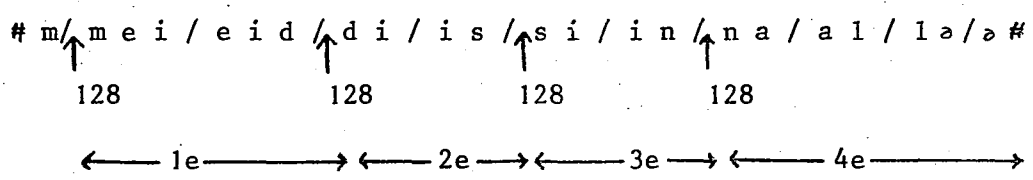
CATEGORIE DE DIPHONES	Schéma d'évolution des consonnes sonores	Valeurs relatives
[# consonne]		> 0
[consonne-voyelle] [consonne - ə]		< 0
[voyelle-consonne] [ə -consonne]		> 0
[consonne sourde-consonne sonore]		> 0
[consonne sonore-consonne sourde]		< 0
[consonne sonore-consonne sonore]		< 0 + > 0
[consonne #]		> 0 + < 0

FIG.35 - Schématisation de la micromélodie réalisée en synthèse sur les consonnes sonores par l'insertion de valeurs relatives dans le dictionnaire pour la période.

.../...

dans /médicinal/, on aura 4 diphones syllabiques :



Un signe identique est inscrit sur le premier échantillon des consonnes sourdes dans les diphones [consonne-voyelle]. Mais les consonnes des diphones [consonne - ə] ne sont pas marquées.

. le premier échantillon d'une réalisation vocalique est marqué 127. Ce marqueur ne concerne que les voyelles des diphones [consonne-voyelle] , il n'est pas inséré sur[ə]dans les diphones [consonne- ə].

. le terme de toutes les réalisations vocaliques, à l'exception de[ə]suivi d'une consonne, est noté 126.

. l'un des premiers échantillons d'une consonne dans les diphones[voyelle-consonne] est noté 124.

. le point de départ d'une consonne dans les diphones [#_consonne voisée] est marqué 125 ; les diphones [#_consonne sourde] n'ont pas de marqueurs.

. les diphones[ə - consonne] : la fin de[ə] n'est pas marquée, mais le premier échantillon de la consonne - si elle est voisée - est marqué 125 comme la consonne des diphones [#_consonne voisée].

. dans les diphones [#_voyelle], on note :
  . 128 sur le premier échantillon de la voyelle,
  . 127 sur le second échantillon.

. dans les diphones [consonne-consonne]:

- si la seconde consonne est sonore : on note 125 au début de sa réalisation.

.../...

- si la première consonne est sonore et la seconde sourde, aucun marqueur n'apparaît.

- si les deux consonnes sont sonores, la première n'est pas marquée, mais le début de la seconde réalisation est notée 125.

- si les deux consonnes sont sourdes, on n'utilise aucun marqueur.

. les diphones [consonne #]) ne sont pas marqués.
[voyelle #])
[consonne_ə])

. les diphones [voyelle-voyelle]:

- environ 5 échantillons avant la fin de la première réalisation vocalique, on inscrit le marqueur 124.

- sur le premier échantillon de la seconde voyelle, on note 128, et 127 sur le second échantillon comme dans les diphones [#_voyelle]

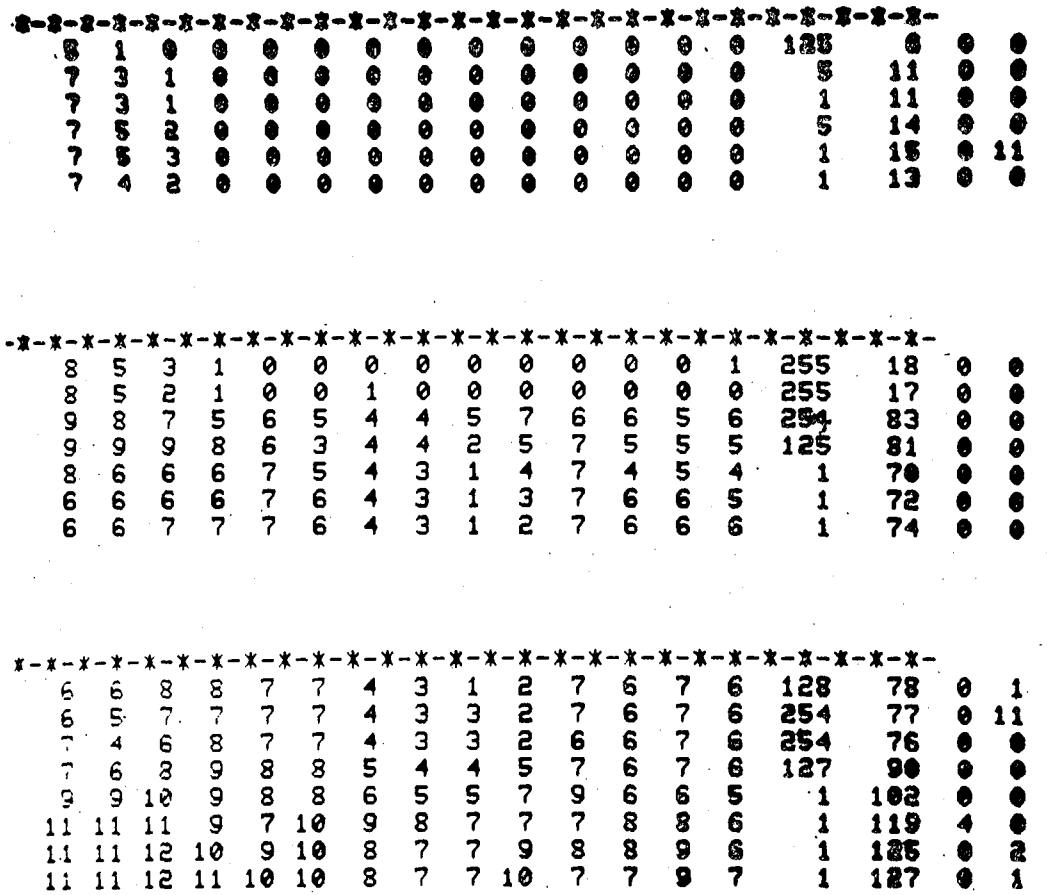
L'ensemble de ces marqueurs apparaît plus concrètement dans les figures 36 et 37.

Il est temps maintenant d'essayer d'expliquer l'utilisation que nous faisons de ces marqueurs dans le traitement de la prosodie, le plus simple étant sans doute d'opérer à partir d'exemples concrets.

Mais d'abord, et sans entrer dans les détails du traitement que nous exposerons plus loin, disons que le message présente - dans sa configuration écrite - un ensemble de marqueurs qui signalent, entre autres, les frontières des mots.

Donc, un élément de programme vient lire ces marqueurs de frontières de mots, puis, entre deux frontières, compter le nombre de voyelles, ou plus exactement le nombre de ce que nous avons appelé les "diphones syllabiques". Ce calcul est très simple, il suffit de répérer

.../...



/# b/

/br/

/ra/

← 125 = début d'une réalisation [# consonne voisée]

← 125 = début d'une réalisation consonantique voisée

← 128 = point de départ d'une consonne (sourde ou sonore) dans les diphones [consonne-voyelle]

← 127 = début d'une réalisation vocalique (sauf [ə]) précédée d'une consonne (diphones [consonne-voyelle])

FIG.36 - Marqueurs d'intonation pour trois diphones /#b/ br/ ra/.

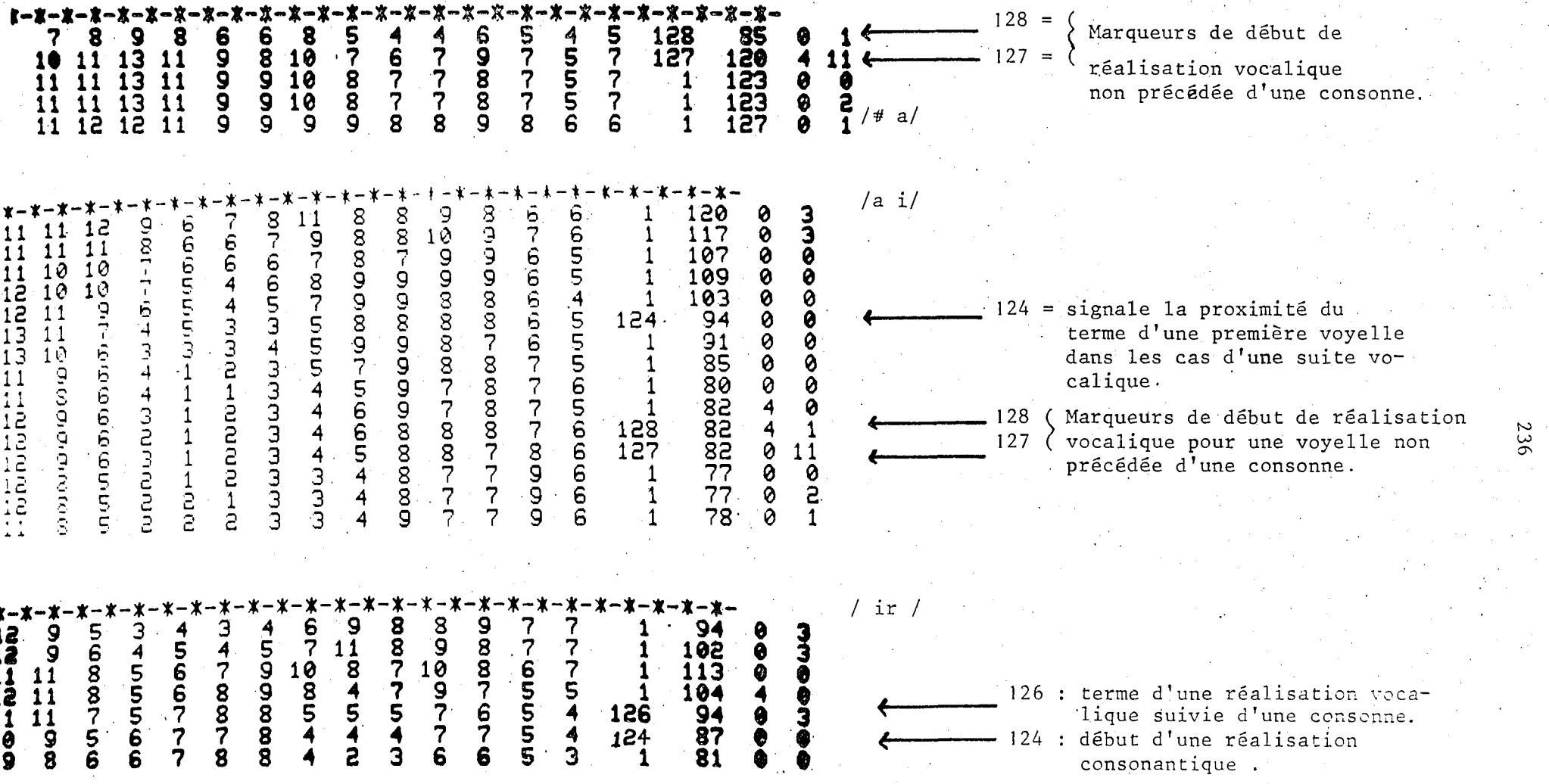


FIG.37 Marqueurs d'intonation pour trois diphones : / # a/ a i / ir /.

la présence du marqueur 128 : celui-ci n'existe que dans les diphones dont l'un des deux éléments comporte le début d'une réalisation vocalique, soit les diphones:

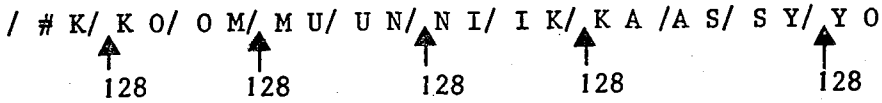
- [consonne-voyelle]
- [# - voyelle]
- [voyelle-voyelle].

Comme le traitement que nous réalisons prévoit l'attribution pour chaque voyelle d'une évolution particulière de la fréquence fondamentale, un élément de programme va consulter un tableau en mémoire qui prévoit pour chaque mot :

- en fonction de son nombre de syllabes,
- en fonction de sa position dans le message,

autant de schémas d'évolution de Fo que ce mot comporte de syllabes, c'est-à-dire de marqueurs 128:

soit l'exemple du mot / Communication/



si dans le mot testé, il a été détecté 5 fois le marqueur 128, cela signifie que ce mot est composé de cinq syllabes et qu'il lui est attribué un schéma intonatif comportant cinq évolutions de Fo spécifiques de la position que ce mot occupe dans le message. C'est donc le marqueur 128 qui sert de pilier pour l'attribution des schémas intonatifs. Or, ce marqueur, dans les diphones [consonne-voyelle] - qui représentent avec les diphones [voyelle-consonne] les catégories à plus forte occurrence - est situé sur le premier échantillon d'une réalisation consonantique dans ces diphones ; dans ce cas, quel niveau fréquentiel sera attribué au premier échantillon de la voyelle ?

Il nous faut signaler dès maintenant que le tableau qui représente l'évolution de la fréquence fondamentale syllabe par

.../...

syllabe comporte à la fois:

- une valeur absolue en période (le tableau des équivalences entre période codée et fréquence fondamentale est donnée en annexe),

- une indication sur le signe de la pente de la période,

- la progression de la période échantillon par échantillon pendant la réalisation d'une voyelle.

Supposons que le tableau donne comme première valeur d'un mot de 5 syllabes - 1.85 (le niveau 85 correspond à 183 Hz).

Cela signifie :

{ - } ⟶ que la pente est négative (Fo montant)
{ 1 } ⟶ que la période progresse de un niveau par échantillon
{85} ⟶ que cette valeur absolue sera celle de l'échantillon noté 128.

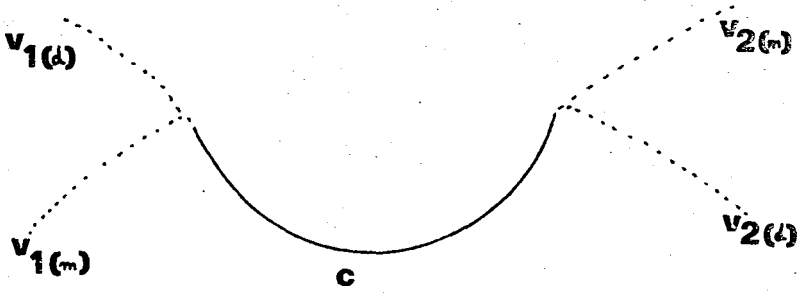
A ce stade de l'explication, il nous faut justifier l'insertion des valeurs relatives représentatives globalement de la micromélodie des consonnes voisées.

Nous avons signalé que dans les études relatives aux caractéristiques intrinsèques des consonnes voisées, on constate un schéma de F_o de type descendant-montant pour toutes ces réalisations. Nous ajoutons que les valeurs fréquentielles que l'on relève au début d'une consonne - dans une séquence [voyelle-consonne]-est toujours inférieure à la valeur de Fo observée à la fin d'une voyelle qui précède. De la même façon, la valeur de Fo en fin de consonne est toujours inférieure à celle observée au début de la réalisation vocalique subséquente.

Nous avons essayé à la synthèse, tout en respectant le schéma de Fo descendant-montant sur les consonnes voisées, de donner aux consonnes des valeurs fréquentielles très légèrement plus hautes que la valeur d'arrivée de la voyelle précédente et que la valeur de départ de la voyelle subséquente : ce schéma est très mal toléré et provoque une dégradation

.../...

de l'intelligibilité. En conséquence, on respectera toujours les schémas suivants à la synthèse dans les séquences V_1 C V_2 selon que V_1 et V_2 ont un schéma intonatif montant (m) ou descendant (d) :



Ces contraintes ont rendu nécessaire le contrôle des évolutions de Fo pendant les réalisations consonantiques ; c'est pourquoi nous avons inséré des valeurs relatives qui représentent leurs évolutions micromélodiques et qui permettent de préserver la continuité de Fo aux frontières des consonnes et des voyelles.

Reprenons l'exemple cité plus haut dans lequel une valeur d'évolution de la période (- 1.85) a été donnée pour la première syllabe d'un mot.

Nous savons déjà que le signe de la pente sera négatif et que la progression de la période se fera de 1 niveau par échantillon. Quant à la valeur absolue {85} elle correspond à la valeur attribuée soit :

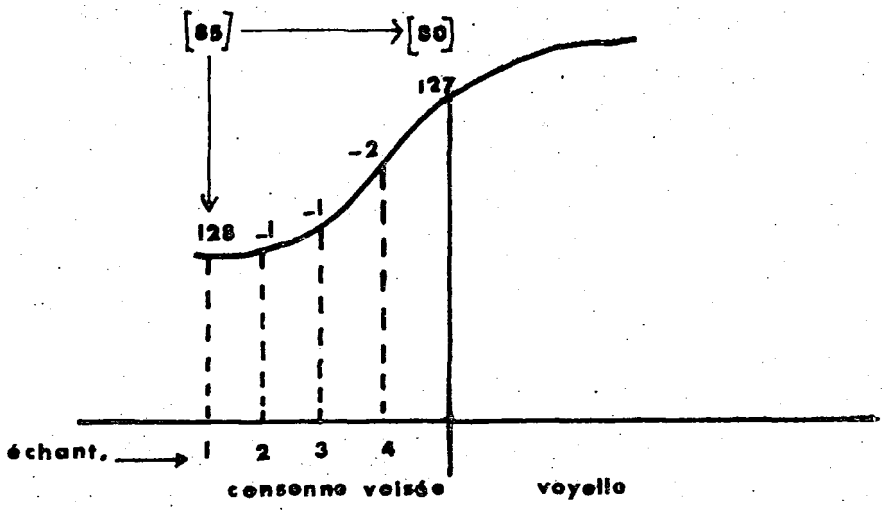
. au premier échantillon de la consonne des diphones [consonne-voyelle] qui porte le marqueur 128,

. au premier échantillon d'un diphone [# voyelle],

. au premier échantillon de la seconde voyelle des diphones [voyelle-voyelle].

Prenons le cas d'un diphone [consonne-voyelle] dans lequel la consonne voisée possède 4 échantillons : nous avons dit que sur cette consonne sont inscrites des valeurs relatives négatives qui imposent un schéma d'évolution de Fo de type montant ; nous savons également que la somme de ces valeurs doit être égale à 4. On va donc obtenir le schéma représenté ci-après.

.../...



diphone [consonne voisée-voyelle], évolution de Fo.

La valeur {85} est inscrite sur l'échantillon qui porte le marqueur 128, puis de cette valeur sont soustraites les valeurs relatives positionnées sur les échantillons subséquents. En l'occurrence, on aura pour cette consonne l'évolution suivante :

	1er échantillon = 85	⟶	= 183 Hz
[- 1]	2e échantillon = 84	⟶	= 186 Hz
[- 1]	3e échantillon = 83	⟶	= 188 Hz
[- 2]	4e échantillon = 81	⟶	= 192 Hz

et le premier échantillon de la voyelle possèdera la valeur {80}.

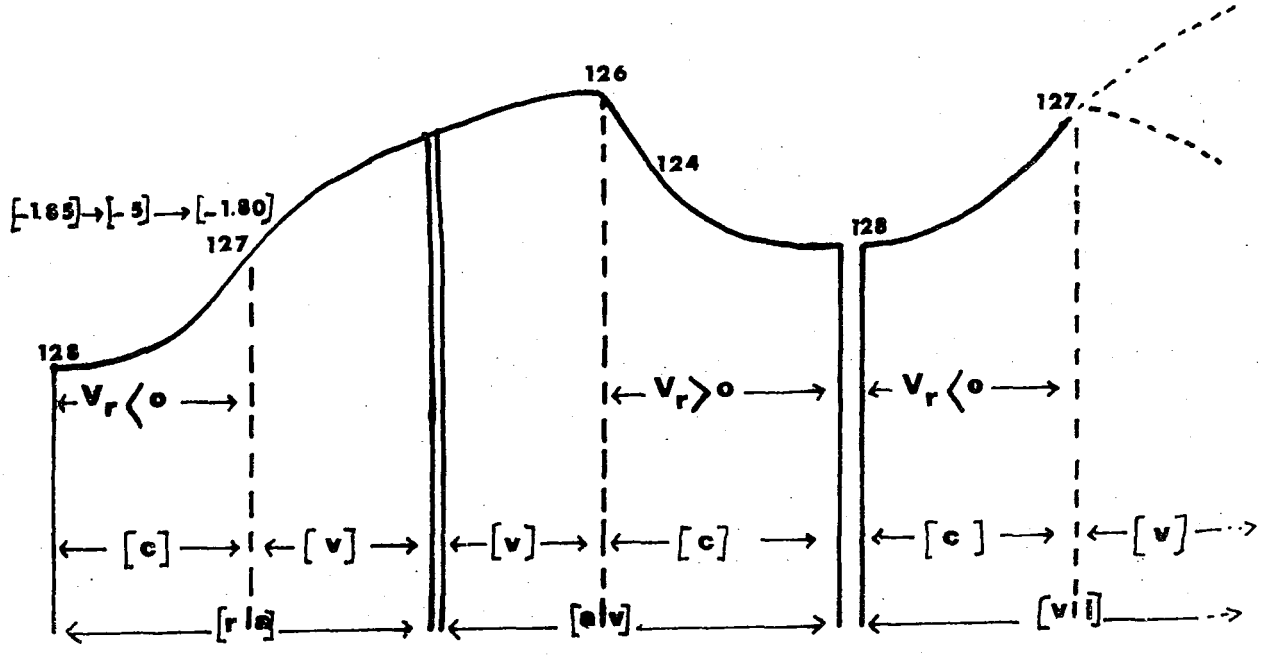
L'écart entre la valeur de départ de la consonne et la valeur d'arrivée sur le premier échantillon de la voyelle dans ce même diphone est effectivement {5} .

Ensuite, pendant la réalisation vocalique (notée 127 sur le premier échantillon), on ne prend en compte que l'indication du signe de la pente et de la progression de la période pour chaque échantillon ; cette information est ici de {- 1} , par conséquent sur le premier échantillon de la voyelle, on aura le niveau 80, puis 79, 78, 77...

.../...

Cette progression continuera jusqu'à la fin de la voyelle notée 126 (dans le diphone qui suit) ; à ce moment-là, et si une consonne lui succède, on ajoute à la valeur sur laquelle on est arrivé , les valeurs relatives positives inscrites sur la consonne, et ce jusqu'à ce que l'on trouve un autre marqueur noté 124.

Plus concrètement, ces opérations donnent le schéma suivant de Fo pour une suite CVCV :



Vr < 0 = valeurs relatives négatives.
Vr > 0 = valeurs relatives positives

Ce processus permet :

1 - d'assurer une continuité de Fo entre le premier segment consonantique et le segment vocalique d'un même diphone,

2 - d'assurer une continuité de Fo à la frontière vocalique de deux diphones successifs.

Mais d'autres juxtapositions sont possibles qui utilisent d'autres catégories de diphones ; dans ces cas aussi, il faut assurer une continuité dans le schéma de Fo ; les marqueurs que nous avons définis jouent ce rôle.

.../...

Nous allons envisager toutes les possibilités de juxtapositions à partir d'un exemple qui prend pour centre un diphone [consonne-voyelle] puisque nous venons d'expliciter son fonctionnement intonatif.

soit le diphone /ra/:

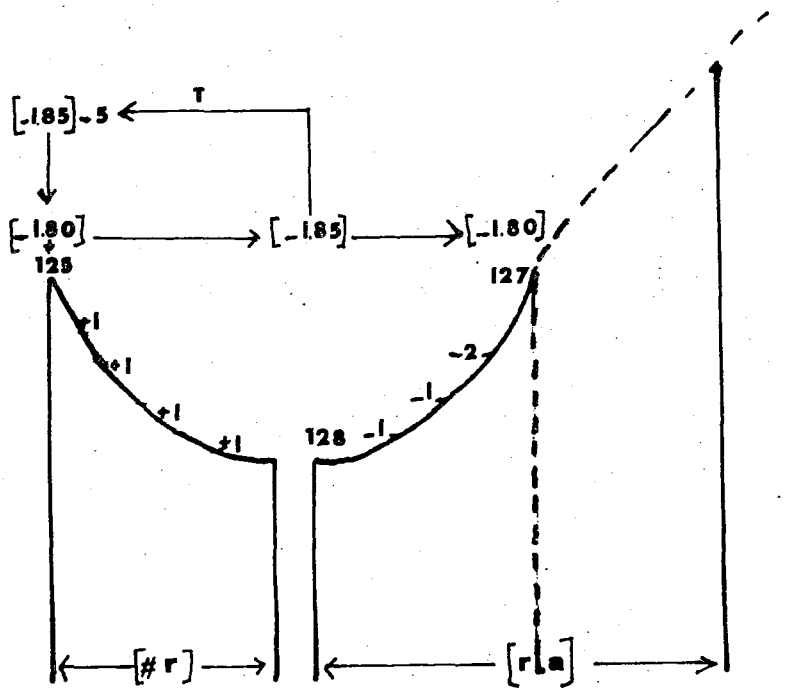
. Ce diphone, s'il appartient à la première syllabe d'un mot situé après une pause par exemple, sera précédé du diphone / # r/. La question qui se pose est de savoir comment on assure la continuité intonative entre /# r/ et /ra/.

Nous avons signalé que les diphones [# consonne voisée] sont notés 125 sur leur premier échantillon. Un élément de programme qui vient lire les marqueurs translate la valeur normalement attribuée au marqueur 128 sur le premier marqueur trouvé : 125. C'est lui qui devient prioritaire et remplace 128 pour l'attribution de la valeur de départ prévue pour la première syllabe d'un mot et affichée dans le tableau qui regroupe les schémas d'évolution de la fréquence fondamentale.

Nous avons dit également que les valeurs relatives positionnées dans les diphones [# consonne voisée] sont des valeurs positives (dont la somme est égale à 4) attribuant un schéma de Fo de type descendant.

Supposons toujours que la valeur normalement attribuée au diphone /ra/ sur son premier échantillon marqué 128 soit {- 1.85} . Il s'effectue une translation de cette valeur sur le premier échantillon de / # r/ en même temps que sa diminution de 5 niveaux de période, on obtient alors :

.../...



Après addition des valeurs relatives > 0 de /#r/ à partir de la valeur 80, et soustraction des valeurs relatives négatives positionnées sur /r/ dans /ra/, on arrive sur le premier échantillon de la voyelle /a/ au niveau {80} ; ce niveau étant identique à celui qui aurait été obtenu si le diphone /ra/ avait été précédé d'un diphone [voyelle-consonne]par exemple.

On le constate à nouveau, les marqueurs et les valeurs relatives permettent d'assurer la continuité de Fo pendant la réalisation d'une consonne voisée.

. On a la suite [# consonne] + [consonne-consonne] + [ra]:

* le premier diphone est composé d'une consonne sourde ; dans ce cas, il n'est pas marqué. Le second diphone comprend d'abord une consonne sourde puis une consonne sonore ; le début de la consonne sonore est notée 125 et sa micromélodie est réalisée par des valeurs positives : Fo descendant.

Ce cas est identique au précédent : translation sur le marqueur 125 de la valeur {85-5}⟶ addition des valeurs relatives ⟶ arrivée sur le premier échantillon de la consonne subséquente à {- 1.85}⟶niveau {80} obtenu

.../...

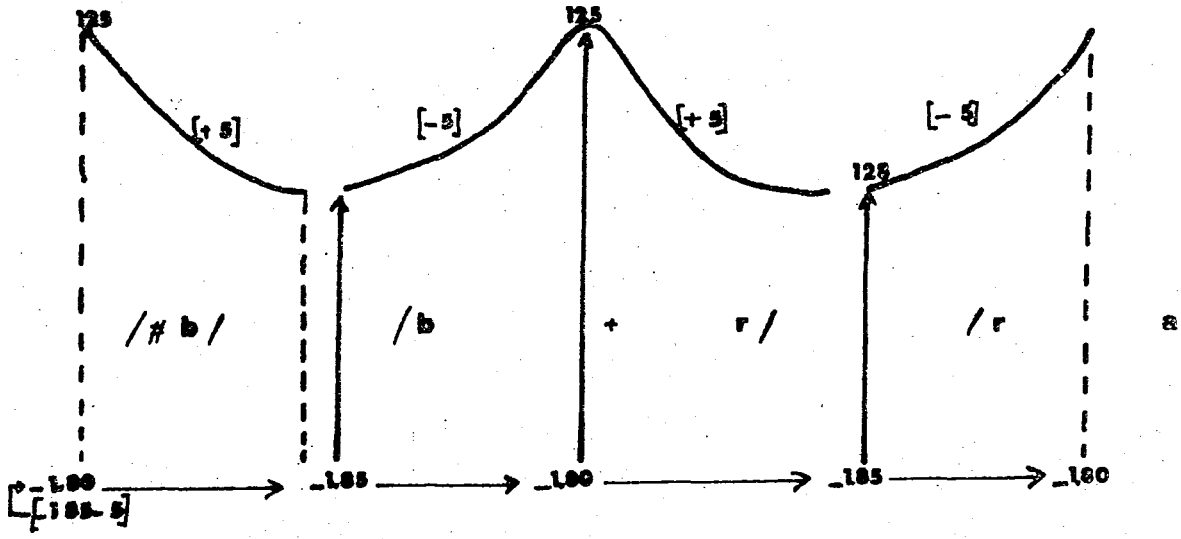
sur le premier échantillon de la voyelle, et prise en compte de la pente et de son signe jusqu'à la fin de la réalisation vocalique.

* le premier diphone se compose d'une consonne sonore ; par exemple /bras/ /# b/ b r/ r a/ a # /:

. /# b/ possède un marqueur 125,
. / r / dans /br/ possède le même marqueur 125,
. /ra / est marqué 128sur son premier échantillon.

La translation de la valeur {- 1.85} est effectuée sur le premier marqueur 125 . Cette valeur est également diminuée de {5}.

On a le schéma suivant :



Les chémas caractéristiques du groupement consonantique sont respectés, les évolutions de Fo sont continues et surtout ce traitement permet d'obtenir sur n'importe quelle voyelle d'un quelconque diphone [consonne-voyelle] une valeur de période identique - c'est-à-dire la valeur choisie pour une position donnée dans le message - et ce quels que soient les diphones qui précédent.

.../...

. Abandonnons maintenant le diphone /ra/ et choisissons un diphone [consonne sourde-voyelle]par exemple /ta/ .

Ces diphones ont également un marqueur 128 sur leur premier échantillon, mais la période inexistante est notée Ø sur les échantillons subséquents. Cette valeur de la période est testée et, si la consonne est sourde, on effectue une diminution de 5 niveaux sur la valeur absolue prévue pour le début du diphone, ce qui permet, si l'on reprend la valeur {- 1.85} utilisée dans les exemples ci-dessus, d'arriver encore à la valeur {- 1.80} sur le premier échantillon de la voyelle.

. On considère maintenant les diphones [voyelle-consonne] :

Ce diphone peut - comme dans le mot/aramis/- être précédé d'un diphone [# voyelle] .

* La réalisation [# voyelle]présente un marqueur 128 sur son premier échantillon puis un marqueur 127 sur le second.

Cela signifie que la valeur {- 1.85}va être immédiatement diminuée de 5 niveaux et que l'on se retrouve dans les cas précédents des diphones [consonne-voyelle] : la réalisation vocalique va débuter sur la valeur de période {80}.

* /ar/ est suivi de /ra/:

Dans le diphone /ar/, la fin de la réalisation vocalique est notée 126, puis la consonne /r/ sur l'un de ses échantillons (au moins 1 après la fin de la voyelle) est marquée 124.

A partir de la marque 126, la valeur atteinte en fin de voyelle évolue en fonction des valeurs micromélodiques positives insérées sur la consonne, et ce jusqu'à la marque 124 qui représente une sorte de signal d'alarme. Ce marqueur est destiné à éviter une discontinuité de Fo dans le passage de cette première partie de réalisation consonantique à la seconde partie contenue dans le diphone subséquent [consonne-voyelle] .

.../...

Nous avons signalé que le schéma de Fo durant la réalisation des liquides [r,l] , des nasales [m,n] et des semi-consonnes [w,j] ne doit pas présenter de brusque discontinuité sous peine de dégradation de l'intelligibilité ; le traitement réalisé ici vise à empêcher cette rupture à la frontière consonantique de deux diphones.

Donc, à partir de cette marque 124, un élément de programme va tester la valeur prévue pour le premier échantillon de la consonne dans le diphone [consonne-voyelle] subséquent, comparer la valeur obtenue au niveau du marqueur 124 et la valeur prévue sur le marqueur 128 et interpoler entre les deux valeurs de façon à obtenir sur les échantillons une répartition des valeurs qui ménage la plus faible discontinuité dans les niveaux de période à la frontière consonantique.

Dans le cas où le diphone /ar/ est suivi d'un groupement [consonne sonore-125 - consonne sonore], l'interpolation se fait de la même façon entre la valeur obtenue sur le marqueur 124 et celle obtenue par translation sur le marqueur 125.

. Le même processus est réalisé dans les séquences [voyelle-voyelle] :

La seconde voyelle est marquée 128 puis 127 sur ses deux premiers échantillons ; elle se comporte exactement comme la voyelle des diphones [# voyelle] .

Quant à la première, elle est dans la même situation que la voyelle des diphones [voyelle-consonne] . Les deux voyelles d'une séquence vocalique peuvent appartenir à deux syllabes différentes (en tous cas c'est comme cela que nous les avons envisagées en synthèse, parce que si on les englobe dans la même syllabe, c'est-à-dire avec un seul schéma de Fo pour les deux, leur durée cumulée étant bien supérieure à la durée de n'importe quelle autre voyelle, on arrive à la fin de leur réalisation à des niveaux fréquentiels bien trop élevés ou bien trop bas par rapport à l'amplitude de Fo autorisée pour une seule syllabe). L'analyse révèle qu'il n'existe jamais de rupture brusque de Fo dans le passage de la

.../...

première à la seconde voyelle (sauf quand l'une et l'autre se trouvent juxtaposées à la frontière de deux mots différents).

Pour cette raison, quand on arrive au marqueur 124, on compare la valeur obtenue à ce moment-là avec la valeur prévue pour la voyelle suivante, et on procède à une interpolation :

- si l'écart entre les deux valeurs est inférieur à 3 niveaux de période, on n'effectue aucune modification,
- si l'écart est compris entre 3 et 7 niveaux, suivant le signe de la pente, on incrémente ou on décrémente de 1,
- si l'écart est supérieur à 7 niveaux, on incrémente ou on décrémente de 2.

Ainsi, la continuité entre les deux voyelles est assurée .

Quant aux diphones [voyelle #] , [consonne #] et [consonne-ə] , ils ne sont pas marqués : on utilise, pour réaliser leur évolution intonative, les valeurs relatives insérées sur chacun de leur échantillon.

I-4-2 - Récapitulation des différents marqueurs :

. diphone [consonne - voyelle]
128 127

. diphone [voyelle - consonne]
126 124

. diphone [# consonne voisée]
125

. diphone [# voyelle]
128+127

.../...

. diphone [voyelle-voyelle]
124 128+127

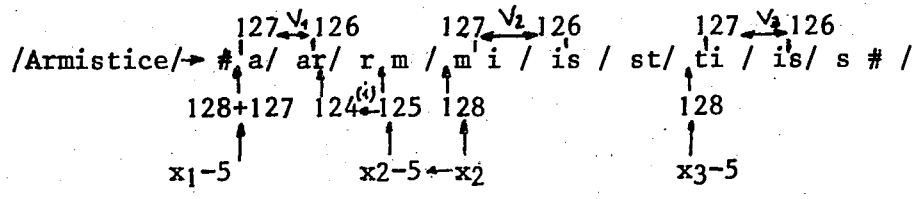
. diphone [consonne-consonne]
- [consonne sourde-consonne sourde] : aucun marqueur
- [consonne sourde-consonne voisée] 125
- [consonne voisée-consonne sourde] : aucun marqueur
- [consonne voisée-consonne voisée] 125

. diphone [ə - consonne voisée]
125

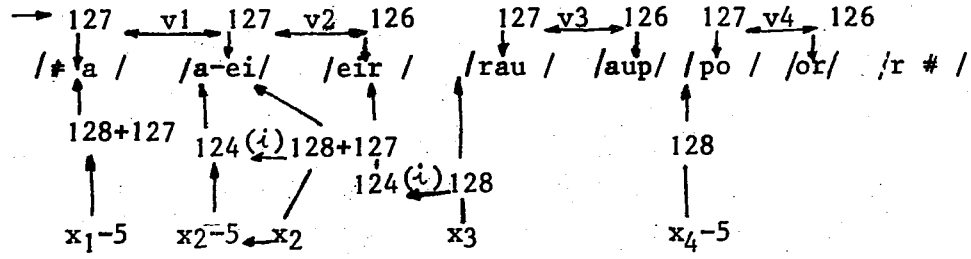
. diphones [voyelle #]
[consonne #] Aucun marqueur. On prend en compte les valeurs relatives
[consonne-ə]

- Applications :

soit (x) les valeurs absolues attribuées à chaque "diphone syllabique"
(i) l'interpolation
et (v) les voyelles

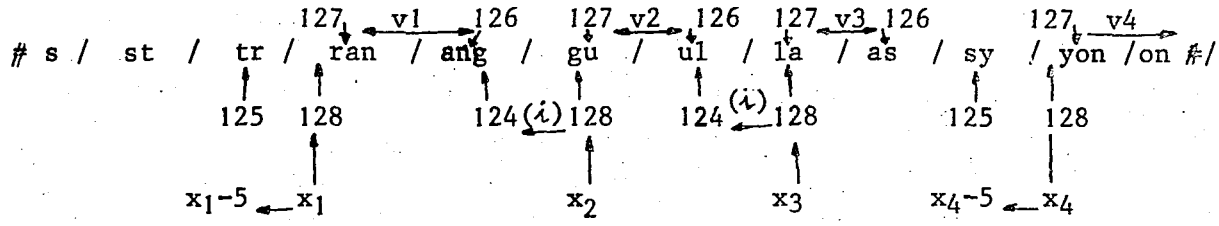


/Aéroport/



.../...

/strangulation / →



II - LE TRAITEMENT DE LA PROSODIE

Il s'agit ici d'élaborer à partir des observations de l'analyse un nombre limité de patrons prosodiques susceptibles de s'adapter à n'importe quel type et structure de phrase. L'analyse a permis de définir la tessiture du locuteur (féminin) : on peut situer les limites de son registre fréquentiel - qui doit être respecté à la synthèse - entre 134 et 356 Hz, soit une étendue d'une octave et demie. Le niveau maximal est atteint pendant la dernière syllabe du mot portant la montée intonative de l'interrogation et la valeur minimale pendant la dernière syllabe des phrases énonciatives.

Pour réaliser ce traitement prosodique, on opère une segmentation de l'énoncé en partant de l'unité linguistique la plus grande (la phrase) pour descendre jusqu'au niveau de la plus petite unité intonative : la syllabe. Bien que ces règles concernent essentiellement les voyelles, puisque les consonnes n'interviennent pas dans la perception du message mélodique, il n'empêche que la qualité et l'intelligibilité globale du message dépendent aussi d'un enchaînement harmonieux, c'est-à-dire sans brusques discontinuités de la fréquence fondamentale, entre les éléments consonantiques d'une part, entre les consonnes et les voyelles d'autre part. C'est pour tenir compte de ces phénomènes que nous avons élaboré (I-4-1 (c)) les règles de juxtaposition entre diphones.

Dans un système de Reconnaissance-Synthèse, le positionnement automatique des marqueurs prosodiques aux endroits requis suppose que soit résolu, en particulier, le problème de l'analyse syntaxique. Pour

.../...

faciliter dès maintenant leur insertion automatique, nous les avons limités en nombre, leur avons assigné une place très précise dans l'énoncé et avons implanté un programme en synthèse qui permet de les transformer si la structuration de l'énoncé rend prosodiquement incompatible leur succession.

Par exemple, un marqueur identique est prévu pour signaler chaque fin de groupe de sens dans le syntagme complément ; or, on sait que l'existence de plusieurs groupes successifs provoque des modifications dans le comportement prosodique des mots qui achèvent chacun d'eux : dans ce cas, le programme de transformation des marqueurs est destiné à rétablir une évolution correcte de la prosodie.

Nous étudierons dans un premier temps les marqueurs qui doivent apparaître sur la chaîne phonétique, nous verrons ensuite quelles sont les transformations rendues nécessaires de par la complexité de la structure syntaxique du message.

II-1- Les marqueurs inscrits sur la chaîne phonétique . On effectue une décomposition du message à plusieurs niveaux successifs :

* Connaissance du type de phrase à synthétiser :

Chaque fin de phrase est signalée par un marqueur spécifique qui est celui utilisé de façon conventionnelle comme signe de ponctuation à l'écriture.

* Ensuite, un ensemble de marqueurs - en plus de la virgule et du point virgule qui servent de marqueurs privilégiés - signale le long du message les points suivants considérés comme pertinents du point de vue prosodique :

.../...

. dans les phrases énonciatives
  - la fin du groupe de mots situé immédiatement avant le verbe.
  - la fin du syntagme verbal
  - la frontière entre deux groupes de sens séparés par une préposition dans le groupe de mots situé avant le verbe, ou entre un groupe de sens et une subordonnée.
  - la frontière entre deux groupes de sens dans le syntagme complément séparés par une préposition.., ou la fin d'un groupe de sens suivi d'une subordonnée.

. dans les phrases impératives
  - la fin du syntagme verbal.
  - la fin d'un groupe de sens.

. dans les phrases interrogatives
  - la fin du mot ou du groupe de mots qui porte l'interrogation (en début de phrase).
  - les mêmes points clés que les phrases énonciatives pour les phrases ayant même construction syntaxique.

. dans les trois catégories de phrase, un marqueur signale les frontières de chaque mot.

II-1-1- La localisation des marqueurs:

(.) - Ce marqueur se situe aux termes d'un message de type énonciatif.

(;) - Il indique à la fois la fin d'une proposition dans un message de type énonciatif, et la continuité avec le départ d'une nouvelle proposition.

(●)- Ce marqueur sert de délimitation entre deux groupes de sens dans le syntagme qui précède le syntagme verbal (c'est-à-dire soit syntagme nominal sujet, soit syntagme complément). Le second groupe de sens peut être introduit soit par une préposition, soit

.../...

par une conjonction de subordination reliant deux mots ou deux groupes de mots, soit par une conjonction de subordination :

Exemple : "le petit chat (⊖) de la vieille dame..."
"les fermes (⊖) et les villages (⊖) du département..."
"le chat (⊖) que j'ai vu ce matin..."

(#) - On utilise ce marqueur au terme du groupe de mots qui précède immédiatement le syntagme verbal .

Exemple : "le petit chat de la vieille dame (#) habite ..."
"dans cette ferme du Limousin(#)habite un paysan qui ..."

Une exception concerne tous les pronoms (je, tu, il..) monosyllabiques : ils sont rattachés directement au verbe sans marqueur prosodique.

(⋆) Ce marqueur a la même signification syntaxique que (⊖) mais il est spécifique du syntagme complément.

Exemple : "le petit chat (⊖) de la vieille dame attrape des souris (⋆) dans le grenier".

($). Ce signe marque le terme du syntagme verbal. Il est postérieur à toute forme verbale (participe passé, infinitif..) et inclut également la négation.

Exemple : "Indiquez-moi ($) l'adresse de votre abonné"
"le chat n'a jamais voulu manger ($) de sauterelles".

Mais si l'infinitif est précédé d'une préposition, il est exclu de ce syntagme et constitue un second syntagme verbal noté également ($).

"le petit chat n'a jamais voulu sauter ($) pour attraper ($) la souris".

.../...

• Ce même marqueur (**$**) est également utilisé pour indiquer la frontière entre deux groupes de sens dans le syntagme complément quand le second est un complément déterminatif introduit par /de/. Nous avons indiqué que le mot qui précède cette préposition a un comportement prosodique particulier.

(**£**) dans les phrases interrogatives introduites par un mot ou un groupe de mots interrogatif, on utilise ce marqueur :

"Quand (**£**) partez-vous ?"
"Dans quel département (**£**) travaille-t-il ?"

(!) indique la fin d'une phrase impérative.

(!.) est l'indication de la fin d'un message dont la dernière proposition était de type impératif.

(!,) marque le terme d'une proposition de type impératif, et la continuité d'un message qui peut être d'un autre type.

(?) signale la fin d'une proposition interrogative.

(?.) est l'indication de la fin d'un message dont la dernière proposition était de type interogatif.

(,) on associe logiquement à la virgule, signe de ponctuation écrite, le signe (,) comme marqueur prosodique.

Nous allons tenter de définir, dans une liste finie, la localisation de ce marqueur sur la chaîne phonétique d'un message à synthétiser.

Les éléments qui suivent ont été empruntés au "Précis de Grammaire française de HINARD (1969) et à la "Grammaire Larousse du français Contemporain " de CHEVALIER et al. (1970).

. On utilise (,) pour séparer des éléments juxtaposés : mots, groupes de mots, propositions : "le mur est gris, la tuile est rousse, l'hiver a rongé le ciment".

.../...

. Dans une énumération à plusieurs termes, on utilise (,) pour les séparer, sauf si une conjonction remplit ce rôle coordinateur (auquel cas on utilise le marqueur (*))

"je garde de mon enfance le souvenir d'années tranquilles (,) de calme(*) et de plénitude".

. Le marqueur (,) est utilisé pour isoler deux termes de fonction différente et délimiter les groupes fonctionnels:

* détacher les compléments circonstanciels placés en tête de phrase, surtout s'ils ont la forme de propositions subordonnées circonstancielles ou complément d'objet direct :

"Qu'il soit intelligent(,) cela ne fait aucun doute".
"Demain(,) dès l'aube (,) à l'heure où blanchit la campagne(,) je partirai".

* marquer une ellipse :
"Le fond du vitrail était bleu ; la bordure (,) rouge".

* encadrer les apostrophes, les appositions, les propositions incises, les propositions relatives détachées (à valeur explicative) qui interompent la proposition principale :

- "Adieu(,) muse endormeuse et douce à mon enfance,"..
- "Moi(,) Frédéric (,) seigneur du mont où je suis né".
- "je viendrai (,) lui dit-elle (,) dès demain".
- "son cocher (,) qui était ivre(,) s'assoupit tout à coup".

. Dans tous les cas où le sujet et le verbe, le verbe et le complément d'objet, le verbe et l'attribut se suivent dans la proposition, le marqueur (,) est impossible.

(-) Tous les mots qui ne correspondent pas aux points-clés ci-dessus définis sont des mots "non marqués" dont on signale les frontières par ce marqueur.

.../...

L'ensemble des marqueurs dans une phrase de type énonciatif apparaît dans l'exemple suivant :

le - lendemain (,) le - gentil - petit - chat (@) de - la - vieille - dame(#)a - failli - manger ($) tous - les - poissons - rouges (*) avant - de - mourir (.)

Les marqueurs que nous venons d'énumérer sont les seuls à devoir apparaître au fur et à mesure de l'écriture du message en code phonétique.

II-1-2- La signification prosodique des marqueurs.

a/ Traitement de la fréquence fondamentale.

Les marqueurs insérés dans le dictionnaire (I-4-1-c2) rendent possible la connaissance de la longueur des mots puisque chaque début de syllabe est signalé ; et les symboles ci-dessus définis permettent de repérer d'une part les mots-clés et d'autre part les mots non marqués (leurs frontières sont connues). Chacun de ces symboles n'entraîne de conséquences prosodiques que pour le mot qui le précède immédiatement.

Ce double système de marqueurs permet de faire correspondre à chaque mot un schéma intonatif qui tient compte :

- de sa localisation dans le message.
- de sa longueur (nombre de syllabes)
- de la nature de sa syllabe finale (ouverte ou fermée) : on voit en effet - Tableau 6 -que dans certaines positions, la syllabe finale de mot possède deux évolutions intonatives possibles selon qu'elle est ouverte ou fermée.

On réalise l'évolution de la fréquence fondamentale grâce à la mise en place d'un ensemble de niveaux de hauteur et de modèles d'évolution linéaires en période. Ainsi, pour la phrase énonciative, le nombre maximum de points de repère susceptibles d'être rencontré nécessite le stockage de 6 schémas prosodiques différents (pour un mot d'une longueur donnée) : le dernier mot situé avant la virgule, avant le

.../...

syntagme verbal et avant un nouveau groupe de sens dans le syntagme complément présente - nous le verrons - la même évolution de Fo. Comme ces schémas varient selon le nombre de syllabes que contient le mot (en général moins de six syllabes) il a été nécessaire de stocker en mémoire pour la phrase énonciative une trentaine de schémas intonatifs. Quand un mot comporte plus de cinq syllabes, on utilise l'évolution de Fo spécifique à la troisième syllabe dans un mot de cinq syllabes, et on la répète autant que nécessaire : deux fois pour les mots de six syllabes, trois fois pour les mots de sept syllabes...

Si l'on prend en compte les trois types de phrases étudiées le nombre des schémas mélodiques utilisés s'élève à 55.

Cependant, l'analyse du corpus montre que, à syllabes égales, les différences des schémas mélodiques selon les mots clés sont surtout sensibles durant la réalisation des deux dernières syllabes de ces mots. En tous cas, aux vues des premiers résultats, il apparaît que les auditeurs sont incapables d'apprécier une différence entre deux mots situés à deux points clés différents quand on ne garde que les schémas caractéristiques de leurs deux dernières syllabes et quand on intervertit les schémas des syllabes précédentes. Cette réaction laisse envisager la possibilité de réduire assez sensiblement la taille des tableaux occupée dans le calculateur pour le traitement de Fo.

Pour l'instant, l'intonation du message s'effectue en attribuant à chaque diphone [consonne-voyelle] - ou plus exactement pour chaque syllabe, après translation - une valeur absolue en période ainsi qu'une pente algébrique. Les valeurs absolues sont fixées en fonction du niveau de hauteur qui est désiré comme point de départ pour la syllabe. L'évolution de la période peut être de 1, 2 ou 3 niveaux par échantillon : le choix dépend du nombre d'échantillons de la voyelle (différent selon les points clés) et du niveau de hauteur que l'on veut atteindre en fin de réalisation vocalique: alors, connaissant les valeurs frontières (déduites de l'analyse) et la forme du schéma que l'on veut attribuer à chaque voyelle, il est très aisé, par addition

.../...

des valeurs relatives positionnées sur les réalisations consonantiques de connaître la valeur qu'il faut imposer au premier échantillon de la syllabe pour parvenir sans discontinuités de la fréquence fondamentale jusqu'à la fin du diphone [voyelle-consonne] subséquent.

L'amplitude des variations de Fo dans ces schémas varie selon les marqueurs. Les valeurs que nous fixons sont le résultat d'approximations successives entre les valeurs dégagées à l'analyse et leur adaptation après des tests de perception à la synthèse : comme le signalent LEON et MARTIN (1969) "l'oreille n'est pas toujours d'accord avec les données de l'étude expérimentale" !.

On trouvera consigné dans le Tableau 6 l'ensemble des valeurs fixées pour chaque mot en fonction de son nombre de syllabes et en fonction de sa position dans la phrase.

. En ce qui concerne plus spécialement le schéma de Fo pour la voyelle qui appartient à la dernière syllabe de mot, nous avons dégagé pour la synthèse cinq schémas caractéristiques en ce qui concerne la phrase énonciative :

* quatre schémas qui dépendent d'une position très précise du mot dans la phrase:
  - deux schémas montants que nous appellerons M_1 et M_2,
  - deux schéma descendants D_1 et D_2.

* un schéma descendant D_3 que l'on attribue à tous les mots non marqués où qu'ils se situent dans l'énoncé.

$\underline{M_1}$ se caractérise par
- une grande variation de Fo (78 Hz en moyenne).
- un niveau de départ de Fo situé pour la voyelle vers 185 Hz.
- un allongement caractéristique de la voyelle (environ 213 ms).

.../...

$\underline{M_2}$

- une amplitude de Fo de 30 Hz en moyenne.
- un niveau de départ pour Fo de la voyelle situé vers 198 Hz.
- une durée d'environ 146 ms.

$\underline{D_1}$

- une amplitude de Fo : 25 Hz en moyenne.
- un niveau moyen de départ pour Fo : 160Hz.
- une durée moyenne de la voyelle : 213 ms.

$\underline{D_2}$

- une amplitude de Fo : 20 Hz en moyenne.
- un niveau moyen de départ : 180 Hz.
- une durée moyenne : 146 ms.

$\underline{D_3}$

- une amplitude moyenne de Fo : 20Hz.
- un niveau moyen de départ : 205 Hz.
- une durée moyenne de 93 ms.

.../...

Nbre de syllabes	1 syll.	2 syll.	3 syll.	4 syll.	5 syllabes
MARQUEURS —	1.90	-1.89 ; 2.78	-1.86 ; 1.76 ; 2.84	-2.93 ; 1.72 ; 2.80 ; 1.82	-1.85 ; 1.75 ; 1.77 ; 1.80 ; 2.79
¤ ;	1.91	-1.88 ; 1.91	-1.84 ; 1.75 ; 1.91	-1.85 ; 1.72 ; 1.74 ; 1.93	-1.85 ; 1.72 ; 1.72 ; 1.74 ; 1.93
# , :*	-1.87 -2.87	-1.93 ; -1.87 -2.87 ;	-1 86 ; 1 84 ; -2 93	-1.87 ; 2.75 ; 1.85 ; -1.87 -2.87	-1.87 ; 2.75 ; 1.85 ; 1.85 ; -2.91
$	-1.87	-1. 91 ; -1.84	-1.88 ; 1.83 ; -2.86 ;	-1.87 ; 2.75 ; 1.85 ; -2.92 ;	-1.87 ; 2.75 ; 1.85 ; 1.85 ; -1.80 ;
.	1.10.3	1.88 ; 1.103	-1.86 ; 2.86 ; 1.103	-1.87 ; 2.75 ; 1.85 ; 1.103	-1.87 ; 2.75 ; 1.85 ; 1.85 ; 1.103
+	-1.86 -2.86	1.90 ; -1.89 -2.89	-1.85 ; 1.79 ; -1.81 -2.81	-1.85 ; 1.79 ; 1.84 ; -1.81 -2.81	-1.85 ; 1.79 ; 1.79 ; 1.84 ; -1.81 -2.81
=	1.90	-1.88 ; 1.91	-1.84 ; 1.75 ; 1.91	-1.85 ; 1.72 ; 1.74 ; 1.93	-1.85 ; 1.72 ; 1.72 ; 1.74 ; 1.93
!	1.104	2.80 ; 1.101 ;	-1.86 ; 1.89 ; 1.98	-1.86 ; 1.77 ; 2.84 ; 1.101	-1.85 ; 1.77 ; 1.77 ; 2.84 ; 1.101
?	-2.100 -2.95	1.85 ; -2.100 -2.95	-1.86 ; 2.86 ; -2.100 -2.95	-2 95 ; 1 79 ; 1 84 ; -2 100 -2 95	-1 87 ; 1 74 ; 1 79 ; 2 82 ; -2 100 -2 95
£	-2.80 -2.75	1.85 ; -2.80 -2.75	-1.86 ; 1.84 ; -2.80 -2.75	-2.95 ; 1.79 ; 1.84 ; -2.80 -2.75	-1.87 ; 1.74 ; 1.79 ; 2.82 ; -2.80 -2.75

TABLEAU. 6 - Evolution intonative des mots en fonction de leur longueur et de leur position en unités de pitch :

1 unité = 64.10^{-6} s.

Le schéma M_1 correspond aux marqueurs (#), (*), et (,)

le schéma M_2 correspond au marqueur ($)

le schéma D_1 correspond au marqueur (.)

le schéma D_2 correspond aux marqueurs (@) et (;)

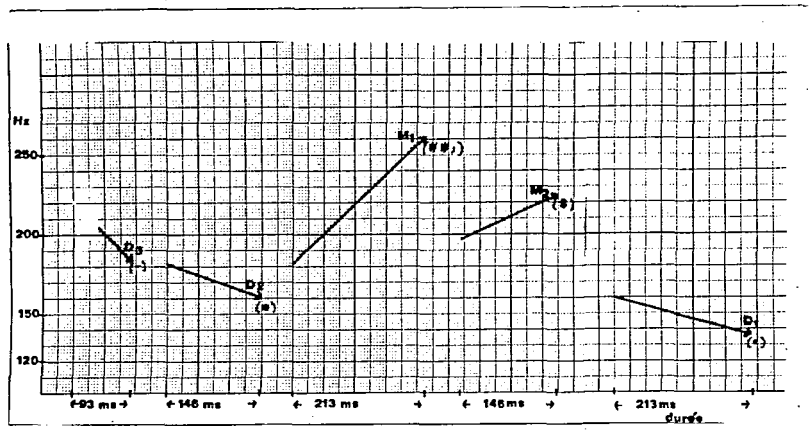


FIG. 38 - Schémas de base pour Fo en synthèse sur la dernière syllabe de mot.

. Dans les phrases impératives, le seul marqueur visible différent des phrases énonciatives concerne le point d'exclamation (!) ou (!.). En cette position, la dernière syllabe de mot est affectée d'un schéma prosodique identique à celui qui correspond au marqueur (.) c'est-à-dire un schéma de type D_1.

. Dans les phrases interrogatives, le schéma intonatif qui correspond au marqueur (?)présente sur la voyelle finale :

.../...

. une amplitude de Fo de 80 Hz en moyenne,
. un niveau de départ de 165 Hz,
. une durée que l'on a laissée pour l'instant équivalente à celle des mots situés avant une pause, soit 215 ms en moyenne.

et le schéma correspondant au marqueur (£) présente :

. la même amplitude de Fo et la même durée que le marqueur (?),
. un niveau de départ de 208 Hz.

Restent à noter quelques constantes - observées à l'analyse dans l'évolution de Fo - et reproduites à la synthèse :

- l'évolution de la fréquence fondamentale pendant une réalisation vocalique sera toujours descendante sur les syllabes intérieures d'un mot.

- nous avons observé à l'analyse, la réalisation sur la syllabe initiale de certains mots de ce que l'on appelle l'"accent didactique". Cette mise en relief se manifeste en particulier par une montée de Fo sur la première syllabe et par une descente de la fréquence fondamentale sur la voyelle de deuxième syllabe avec pour celle-ci un niveau de départ plus haut que le niveau d'arrivée de la première syllabe (en moyenne 1/2 ton)

Nous avons étendu ce schéma à la synthèse pour tous les mots qui possèdent plus de deux syllabes. Nous l'avons exclu pour les mots de deux syllabes quand cette montée de Fo risquait de rendre difficile le passage intonatif de la première à la seconde syllabe.

b/ Traitement de la durée et découpage temporel de l'énoncé.

Ce traitement concerne essentiellement l'accélération de mots ou de parties de mots, l'allongement de certaines syllabes, la répartition et la durée des pauses.

.../...

b_1/ Décision concernant la durée des mots :

* Possibilité d'accélérer la cadence d'échantillonnage, ce qui correspond à une réduction de la durée du diphone :

. Tous les monosyllabiques (-) non situés à l'un des points clés déjà cités vont subir une accélération de l'ordre de 15 ms pendant leur réalisation. Cette accélération se manifeste dans la zone stable de la voyelle : Mot de rythme R_1 (le traitement du rythme est résumé dans les tableaux 7 et 8.

. On a prévu pour tous les mots plurisyllabiques non marqués (-) (c'est-à-dire tous les mots qui ne correspondent pas à des points clés) une accélération qui porte :

- sur un échantillon (X_1) de la première voyelle : Mot de rythme R_1
- à la fois sur le premier échantillon de la consonne et sur le dernier échantillon de la voyelle dans chaque diphone [consonne-voyelle] que compte le mot : Mot de rythme R_2.

. En ce qui concerne les mots marqués , on effectue un traitement identique quel que soit le point clé auquel ils appartiennent sur toutes les syllabes sauf sur la dernière syllabe de mot :

- sur la première syllabe de mot plurisyllabique : Mot de Rythme R_1, c'est-à-dire accélération (X_1) sur un échantillon de la voyelle.

- sur toutes les syllabes qui suivent (sauf la dernière) : Mot de rythme R_2 c'est-à-dire accélération sur un échantillon de la consonne et sur un échantillon de la voyelle dans les diphones [consonne-voyelle].

. la dernière possibilité d'accélération concerne deux mots non marqués qui ont une frontière consonantique voisée commune.

* Possibilité de ralentissement de la cadence d'échantillonnage, ce qui correspond donc à un allongement de la durée. Cette décision intervient uniquement en dernière syllabe de mot, elle est prise en

.../...

fonction de la position du mot dans le message :

. En fin de syntagme situé avant le verbe (#), en fin de groupe de sens dans le syntagme situé après le verbe (*), en fin de mot situé immédiatement avant une virgule(,), en fin de proposition (signalée par la présence du marqueur (;)), et en fin de phrase (.?!), le problème de l'allongement ne se pose que dans les cas où le mot se termine par une syllabe fermée, puisque nous avons signalé le stockage dans le dictionnaire de tous les diphones [voyelle #]. Par conséquent, chaque fois que le message à synthétiser comporte l'un des points clés ci-dessus énoncés (ainsi que deux autres points-clés qui apparaissent au moment des règles de transformation (=), (+)), on affecte la dernière syllabe fermée d'une zone d'allongement (Z_2) qui varie en fonction de la nature de la voyelle et de la consonne subséquente : Mot de rythme R_4 . Le mot se termine ensuite soit par le diphone [consonne ⊖], soit par le diphone [consonne #] .

Cet allongement permet de donner à la voyelle de dernière syllabe une durée très voisine de celle observée à l'analyse, mais on ne prend pas en compte la différence qui existe entre l'allongement de la dernière syllabe d'un mot à intonation montante et l'allongement de la syllabe finale d'un mot à intonation descendante.

. En fin de groupe de sens (⊖) du syntagme qui précède le verbe et en fin de syntagme verbal ($), on a prévu une zone d'allongement (Z_1) réalisée par un ralentissement de la fréquence d'échantillonnage uniquement sur trois échantillons de la voyelle situés dans la zone stable de celle-ci c'est-à-dire à la frontière des diphones [consonne-voyelle] et [voyelle-consonne] : on utilise pour ce faire le Mot de rythme R_3 . Cet allongement permet sans utiliser de pause, de signaler la fin d'un groupe de mots important .

.../...

SYLLABES	C_n V_n	h	d	C_n V_a	h	d	C_a V_a	h	d	C_n V_{r3}	h	d	C_n $V_{r\,x}$	h	d
MARQUES DANS LE DICTIONNAIRE	128	01	1	128	01	1	128		9	128	01	1	128	01	1
	C_1	0B	11		0B	11		B	11		0B	11		0B	11
	127			127			127			127			127		
	V_1	10	2		10	2		10	2		0A	10		0A	10
		01	1			9			9		01	1		01	1
	x x	x	x	x x	x	x	x x	x	x	x x	x	x	x x	x	x
		11	3		0B	11		0B	11		11	3		11	3
		11	3		11	3		11	3		0B	11		11	3
	V_1														
	126	11	3	126	11	3	126	11	3	126	11	3	126	0B	11
	C_2														
MOTS DE RYTHME	R_0 = '0000			R_1 = '0700			R_2 = '4700			R_3 = '20C0			R_4 = '2030		

h = hexadecimal
d = décimal
Cn Vn = Consonne normale Voyelle normale
Ca Va = Consonne accélérée Voyelle accélérée

Vr_3 = Voyelle ralentie de 3 échantillons
Vr x= Voyelle ralentie de x échantillons suivant la consonne qui suit

TABLEAU. 7 - TRAITEMENT DU RYTHME.

TABLEAU.8.

Détermination des Mots de rythme en fonction du nombre de syllabes et de la position du mot.

Nombre de syllabes	—	ə $	+ = £ # ※ , ; ? ! . :
1	R_1	R_3	R_4 → Syllabe fermée R_0 → Syllabe ouverte
2	R_1 R_2	R_1 R_3	R_1 R_4 R_0
3	R_1 R_2 R_2	R_1 R_2 R_3	R_1 R_2 R_4 R_0
4	R_1 R_2 R_2 R_2	R_1 R_2 R_2 R_3	R_1 R_2 R_2 R_4 R_0
5	R_1 R_2 R_2 R_2 R_2	R_1 R_2 R_2 R_2 R_3	R_1 R_2 R_2 R_2 R_4 R_0

b2/ Décision concernant les pauses et leur répartition:

"les paroles que nous prononçons n'ont de sens que grâce au silence où elles baignent" (MAETERLINCK).

On a prévu l'insertion de pauses de différentes longueurs qui permettent de faire face à une double nécessité :

. assurer au message un flux de parole qui donne l'impression de tenir compte des nécessités physiologiques de respiration.

. effectuer un découpage de la phrase qui reflète et signale l'essentiel de sa structure, et renforce la cohésion de sens entre les mots d'un groupe.

Alors que l'analyse d'un corpus spontané révèle l'existence de pauses qui ne coïncident pas forcément avec le découpage syntaxique de la phrase, à la synthèse où il n'est question ni d'improvisation, ni d'hésitation sur le choix d'un mot, on fait toujours coïncider les pauses avec un syntagme grammatical.

Leur existence va se manifester :

- à la fin de mots à schéma de Fo montant,
- à la fin de mots à schéma de Fo descendant.

. Les pauses associées à une montée intonative apparaissent essentiellement sur trois points clés de la phrase énonciative :

- frontière entre deux groupes de sens dans le syntagme situé après le verbe : la pause prévue dure 65 ms.

- fin du syntagme situé immédiatement avant le verbe : la longueur de la pause, parce qu'elle semble dépendre de la longueur et de la complexité syntaxique des éléments qui lui succèdent, résulte d'un calcul assez complexe ; nous indiquerons simplement que sa durée minimale est de 65 ms.

.../...

- la pause associée normalement à la virgule ; elle est souvent et de loin la plus longue manifestation de silence ; elle fait suite, elle aussi, à un mot qui se termine par un schéma intonatif montant, et elle crée, nous le verrons, un bouleversement dans l'attribution des patrons intonatifs pour les mots clé qui la précédent. Sa durée a été fixée à 330 ms.

. Dans la phrase impérative, les pauses sont moins nombreuses et ont une durée faible fixée à 65 ms. On les fait intervenir soit après le syntagme verbal, soit après le complément - c'est-à-dire après le groupe qui s'achève par un schéma intonatif montant - s'il est suivi d'un autre groupe de sens.

. Les phrases interrogatives sont réalisables selon trois modèles syntaxiques :

* phrases introduites par un mot, un groupe de mot, ou une locution interrogative ; une pause de 65 ms est insérée à la fin de ce segment interrogatif.

* phrases de même construction syntaxique que la phrase énonciative ; seul le dernier mot de phrase manifeste par sa prosodie l'interrogation. Dans ce cas, les pauses ont même localisation qu'en phrase énonciative, mais leur durée à l'analyse est moindre. Cependant, les marqueurs utilisés étant identiques dans l'un et l'autre type de phrase, la durée des pauses est également semblable.

* phrases interrogatives avec inversion verbe/sujet :

une pause de 65 ms est prévue à la fin du groupe verbe/sujet ou à la fin du groupe verbe/sujet + verbe infinitif :

Habite-t-il ($)$_{p} à Paris même ?

Pouvez-vous répéter ($)$_{p} ce que vous venez de dire ?

.../...

. les pauses inscrites après un mot terminé par un schéma intonatif descendant c'est-à-dire qui correspondent aux marqueurs (;) (!) et (.) ont une durée fixée à 400 ms.

c/ Traitement sur les niveaux d'énergie:

Le français est une langue oxytonique ; les trois paramètres : durée, intensité, hauteur sont étroitement corrélés pour remplir le rôle de délimitation des fins de syntagme, mais d'une façon générale, il nous semble que pour tous les points clés de la phrase - exceptée la fin de phrase - la durée et la hauteur suffisent à réaliser la perception de l'accent.

En fin de phrase, on observe une évolution particulière du spectre de dernière syllabe en même temps qu'un allongement de sa durée et qu'une courbe descendante dans l'évolution de sa fréquence fondamentale.

Le traitement, assez sommaire (Tableau.9), ne vise à réaliser qu'une évolution correcte d'amplitude sur la dernière syllabe des mots situés avant une pause, sur tous les mots monosyllabiques du message, et sur toutes les syllabes des mots plurisyllabiques qui ont un schéma de Fo descendant.

* Tous les mots monosyllabiques non situés à l'un des points clé précédemment énoncés sont considérés arbitrairement comme des mots chevilles. On suppose donc qu'ils n'ont pas une importance linguistique primordiale et, pour cette raison, on opère sur chacun d'eux, et sur toute la longueur de leur voyelle, une translation de - 4 db par échantillon par rapport aux niveaux stockés en mémoire.

* Réalisation du traitement pour les mots plurisyllabiques:

Plusieurs cas ont dû être envisagés :

c1/ Syllabe finale d'un mot situé avant une pause qui reçoit un schéma intonatif montant :

.../...

- en syllabe ouverte :

Aucun traitement ni marquage particulier n'est nécessaire puisque tous les diphones [voyelle #] ont été stockés à partir de l'enregistrement de mots situés à cet endroit précis d'un message et prononcés avec l'intonation montante qui les caractérise.

- en syllabe fermée :

L'allongement prévu sur la voyelle finale dispense de tout traitement sur l'énergie car l'on constate pour la voyelle une faible différence entre l'évolution de l'intensité obtenue à l'analyse et celle obtenue en synthèse après ralentissement de la cadence d'échantillonnage sur la voyelle ; de plus, le dernier segment [consonne #] prévu à cet effet, vient se concaténer sans discontinuité au diphone [voyelle - consonne] qui le précède.

c2/ Syllabe finale d'un mot situé avant une pause qui reçoit un schéma intonatif descendant (essentiellement les fins de phrase):

c-2-1- en syllabe ouverte.

La configuration [voyelle #] dont on dispose (et qui correspond à celle réalisée dans un mot prononcé avec une intonation montante) a l'avantage de présenter des niveaux d'énergie relativement continus avec le diphone [consonne-voyelle] qui le précède ; malheureusement, son énergie globale ne correspond pas à celle observée à l'analyse dans cette position terminale de message. Pour cette raison, on utilise un marqueur d'intensité inscrit sur tous les diphones [consonne-voyelle] au début de l'élément vocalique, qui permet de réaliser une diminution de 4 db par échantillon sur toute la durée de la voyelle finale et d'éviter au milieu de sa réalisation une discontinuité de niveaux.

c-2-2- en syllabe fermée :

Un marqueur inscrit sur le premier échantillon vocalique diminue l'intensité de -4db sur toute la durée de la voyelle (échantillons doublés compris), la consonne dans les segments [voyelle-consonne] est laissée intacte, et le diphone final [consonne #] est utilisé tel que stocké.

.../...

TABLEAU 9

DETERMINATION DES MOTS D'INTENSITE

Marqueurs / Nbre de syllabes	— ə =	# $: ,	* £ + ?	! ; .
1	G_1	G_0	G_0	G_1
2	G_0 G_1	G_0 G_0	G_1 G_0	G_1 G_1
3	G_0 G_1 G_1	G_0 G_1 G_0	G_0 G_1 G_0	G_0 G_1 G_1
4	G_0 G_1 G_1 G_1	G_0 G_1 G_1 G_0	G_0 G_1 G_1 G_0	G_0 G_1 G_1 G_1
5	G_0 G_1 G_1 G_1 G_1	G_0 G_1 G_1 G_1 G_0	G_0 G_1 G_1 G_1 G_0	G_0 G_1 G_1 G_1 G_1

G_0 = sans variation de gain

G_1 = variation du gain = - 4 db.

* Enfin, on réalise pour toutes les voyelles incluses dans un mot et qui possèdent un schéma intonatif de type descendant une translation de - 4 db par rapport aux niveaux d'énergie stockés en bibliothèque, et ce sur toute leur longueur.

Il semble à l'écoute que ce procédé permet de mieux restituer les caractéristiques individuelles de la voix du locuteur choisi.

Le positionnement correct de ces marqueurs et leur symbolisme au niveau du traitement prosodique suffisent pour synthétiser un message court et syntaxiquement simple. Cependant, dès qu'on entre dans plus de complexité dans la composition des syntagmes, ces marqueurs apparaissent dans une succession qui ne rend plus compte des réalités prosodiques. C'est pourquoi un programme a été élaboré pour tester la suite des marqueurs et les transformer de façon à obtenir une séquence sonore satisfaisante.

II-2- Les règles de transformation

Ces règles ont nécessité l'introduction de deux nouveaux marqueurs ; ceux-ci n'apparaîtront jamais comme marqueurs prosodiques le long de la chaîne phonétique mais seulement comme marqueurs de transformation dans les programmes .

(+) est utilisé dans les phrases de type impératif, il implique pour la dernière syllabe du mot qui le précède un schéma intonatif de type montant :

- amplitude de Fo : 50 Hz.
- niveau de départ de Fo : 200 Hz.
- durée moyenne de la voyelle finale : 213 ms.

une pause de 65 ms lui correspond.

(=) ce marqueur intervient dans les trois types de phrases étudiés, il impose un schéma mélodique descendant de type D2 sur la dernière syllabe du mot dont il fixe le terme, une durée d'environ 213 ms et

.../...

une pause de 65 ms.

(L'ensemble des marqueurs utilisés dans les phrases énonciatives sont rassemblés dans le Tableau10).

Deux sortes d'évènements peuvent nécessiter l'intervention de ces marqueurs :

- la complexité des syntagmes,
- l'existence d'une virgule dans l'énoncé.

II-2-1- Complexité du syntagme situé après le verbe:

1-dans les phrases énonciatives:

La règle générale, nous l'avons dit, est de noter la fin du syntagme verbal par le signe (**$**) (schéma montant de type M2).

Cependant, l'existence de deux groupes de sens notés (*) après le verbe entraîne certaines modifications dans l'attribution des schémas prosodiques, donc des marqueurs de mots, pour respecter le contraste de pente observé dans le prédicat.

. La succession (**$**) + (*) + (.) étant impossible à cause de la succession de deux marqueurs qui imposent un schéma de Fo montant - M_2 et M_1 - on transforme cette séquence en (**ə**) + (*) + (.).

. D'autre part, la succession de plusieurs groupes de sens dans ce syntagme bouleverse l'attribution des marqueurs :

.../...

TABLEAU RECAPITULATIF DES MARQUEURS
UTILISES DANS LES PHRASES ENONCIATIVES

	LOCALISATION	SCHEMA MELODIQUE	DUREE DE DERNIER SEGMENT	PAUSES
-	frontières des mots non situés à un point-clé.	Schéma descendant D_3	Accélération sur un échantillon de la dernière consonne et de la dernière voyelle	pas de pause
ə	fin d'un groupe de sens dans le syntagme situé avant le verbe.	Schéma descendant D_2	Allongement de la voyelle finale de 40 ms	pas de pause
#	fin du groupe nominal situé immédiatement avant le verbe.	Schéma montant M_1	Durée moyenne de la voyelle finale : 215 ms	pause : durée minima = 65 ms
,	différentes possibilités d'occurrence.			pause de 330 ms
*	frontière de deux groupes de sens dans le syntagme situé après le verbe.			pause de 65 ms
$	fin du syntagme verbal.	Schéma montant M_2	Allongement de la voyelle finale de 40 ms	Pas de pause
=	. fin du premier groupe de sens quand le syntagme post-verbal en comprend deux. . fin du deuxième groupe de sens quand ce syntagme en comprend trois.	Schéma descendant D_2	Allongement de la voyelle finale. Durée d'environ 215 ms	pas de pause
.	fin d'une phrase énonciative.	Schéma descendant D_1	Durée moyenne de 215 ms	
;	fin d'une proposition énonciative.	Schéma descendant D_2	Durée moyenne de la voyelle finale : 215 ms	pause de 400 ms

TABLEAU 10

.../...

(✱) + (✱) est inacceptable, et sera transformé en (=) + (✱); dans cette situation, le groupe verbal conserve son marqueur originel (Ⓢ)

De la même façon, la succession de trois groupes de sens (✱) + (✱) + (✱) entraîne la transformation suivante : (✱) + (=) + (✱) et le groupe verbal dans ce cas redevient (ⓔ).

Exemple (a): le chat (#) a mangé (ⓔ) (D_2) les poissons rouges (✱) (M_1) avant d'aller mourir (=) (D_2) derrière le fauteuil (✱) (M_1) pour être seul (.) (D_1).

En définitive, les compatibilités de succession des marqueurs dans le propos sont les suivantes :

(Ⓢ) + (.) ... mangé (Ⓢ) les souris (.)

(ⓔ) + (✱) + (.) ... mange (ⓔ) les souris (✱) dans le grenier (.)

(Ⓢ) + (=) + (✱) + (.)... mangé (Ⓢ) les souris (=) dans les greniers (✱) et dans les caves (.)

(ⓔ) + (✱) + (=) + (✱) + (.) = exemple (a) inscrit plus haut.

2 - dans les phrases impératives :

Les marqueurs inscrits sur la chaîne phonétique sont les mêmes que ceux de la phrase énonciative mais ils ne correspondent pas aux faits prosodiques qui se manifestent dans la phrase impérative. On les a choisis pour faciliter leur positionnement automatique ; les transformations visent à rendre aux points de l'énoncé signalés une prosodie correcte.

Sur la chaîne phonétique, les marqueurs inscrits peuvent être les suivants :

.../...

Veuillez répéter ($) votre question (!.)
Répétez-moi ($) le numéro (*) de votre correspondant (!.)
Indiquez-nous ($) l'adresse (*) du correspondant (*) que vous désirez joindre (!.)

Ces marqueurs imposant tous une succession de schémas de Fo montant, les règles de transformation réalisent aussi l'alternance du signe de la pente de Fo:

la suite ($) + (!.) devient (+) + (!.) (1)

($) + (*) + (!.) ⟶ (=) + (+) + (!.) (2)

($) + (*) + (*) + (!.) ⟶ (+) + (=) + (+) + (!.)

(1) Veuillez répéter (+) votre numéro (!.)

(2) Veuillez me préciser (=) votre numéro (+) à la Sécurité Sociale (!.)

On peut formuler la règle suivante :

Quand une phrase se termine par un point d'exclamation

- et que les marqueurs de la phrase sont en nombre pairs, il y a alternance des marqueurs (+) et (=) en commençant par le marqueur (+).

- et que les marqueurs sont en nombre impairs, il y a également alternance mais en commençant par le marqueur (=).

II-2-2- Existence d'une virgule (,) :

Outre la complexité des syntagmes, c'est l'existence d'une virgule qui est susceptible de détruire les schémas qui correspondent aux marqueurs inscrits sur la chaîne phonétique. (Tableau 11).

.../...

a/ transformations dans la phrase énonciative :

. la virgule est située en fin de syntagme qui précède le verbe (#) : cette situation n'entraîne aucune modification dans le schéma intonatif du mot qui la précède : simplement le marqueur (#) est remplacé par le signe (,) :

Ex : le chat(,) qui s'est caché (,) attend la souris.

. la virgule est située à la fin du syntagme verbal noté ($) : dans ce cas, il y a transformation du marqueur précédent (#) en un marqueur de type (=), c'est-à-dire qu'on passe d'un schéma montant M_1 à un schéma descendant de type D_2 :

Ex : la souris (=) a été mangée (,) malgré ses soupçons,

par le chat .

. la virgule est située après un groupe de sens du syntagme post-verbal (*) : il y a transformation du marqueur précédent, vraisemblablement ($) en un marqueur de type (ə) ; là encore on passe d'un schéma montant M_2 à un schéma descendant D_2.

Mère Grand (#) a recontré (ə) la souris (,) la belette (*)

et le petit lapin .

. la virgule est située à la place du point (.) ou du point virgule (;)

Les conséquences sont prévues dans les exemples ci-dessus : elles dépendent du marqueur qui précède.

.../...

Ces règles de transformation font qu'on ne peut jamais trouver un schéma montant de type M_1 - comme celui des marqueurs (,) (*) ou (#) - précédé d'un autre schéma montant ; il est toujours obligatoirement précédé d'un mot clé terminé par un schéma de type descendant D_2 . Seule exception : la succession dans le message de deux ou plusieurs virgules (schéma M_1) non séparées par l'existence d'un autre point clé.

b/ transformations dans la phrase impérative :

On retrouve ici le principe d'alternance des marqueurs (+) et (=) ; la règle de transformation est la suivante :

dans une phrase simple on avait les suites :

(1) veuillez répéter (+) votre question (!.)

(2) veuillez me préciser (=) votre numéro (+) de Sécurité Sociale (!.)

. dans le premiers cas, la virgule va transformer le marqueur (+) en un marqueur de type (=):

veuillez répéter (=) votre question (,) préciser (=) votre demande (+) et attendre (+) notre réponse (!.)

. dans le second cas, la virgule va provoquer l'inversion des schémas mélodiques, donc des marqueurs qui précèdent ; on aura par exemple :

veuillez me préciser (+) votre numéro (=) de Sécurité Sociale(,) votre numéro (=) au chomage (+) et votre numéro(+) d'allocation (!.)

.../...

TABLEAU 11

RECAPITULATIF DES DIFFERENTES REGLES DE TRANSFORMATION ENTRAINEES PAR LA PRESENCE D'UNE VIRGULE.

A/ PRESENCE D'UNE VIRGULE.	MODIFICATION DES MARQUEURS.
1/ PHRASE ENONCIATIVE.	
Virgule située après le syntagme qui précède le verbe	Pas de modification des marqueurs antérieurs, simplement le marqueur (#) est remplacé par le marqueur (,)
Virgule située après le syntagme verbal (#) + (Ⓢ)	(Ⓢ) devient (,) (#) devient (=) (=) + (,)
Virgule située après un groupe de sens dans le syntagme post-verbal	
1°/ (Ⓢ) + (✱)	➡(ⓔ) + (,)
2°/ (Ⓢ) + (✱) + (✱)	➡(Ⓢ) + (=) + (,)
3°/ (Ⓢ) + (✱) + (✱) + (✱)	➡(ⓔ) + (✱) + (=) + (,)
Virgule située à la place du point final	
1°/ (Ⓢ) + (.)	➡(ⓔ) + (,)
2°/ (Ⓢ) + (✱) + (.)	➡(Ⓢ) + (=) + (,)
3°/ (Ⓢ) + (✱) + (✱) + (.)	➡(ⓔ) + (✱) + (=) + (,)
2/ PHRASE IMPERATIVE.	
(Ⓢ) + (,)	➡(=) + (,)
(Ⓢ) + (✱) + (,)	➡(+) + (=) + (,)
(Ⓢ) + (✱) + (✱) + (,)	➡(=) + (+) + (=) + (,)
ou (Ⓢ) + (ⓔ) + (#) + (,)	

.../...

D'autre part, quand une proposition impérative contient la marque (!,), cela signifie qu'elle est suivie d'une autre proposition qui peut être de n'importe quel autre type (énonciatif, interrogatif) et qui sera conclue par le marqueur (.) ou (?.)

Cette possibilité de marqueur (!,) provoque des modifications dans le positionnement des marqueurs précédents, modifications que l'on peut résumer dans la règle suivante :

Quand une proposition se termine par le marqueur (!,) et que

- les marqueurs sont en nombre pair, on réalise une alternance des marqueurs (+) et (=) en commençant par (=).

- les marqueurs sont en nombre impair, l'alternance des marqueurs commence par le signe (+).

II-2-3- Les règles de transformation dans la phrase interrogative :

Dans ce type de phrase, le rôle joué ailleurs par la virgule, est tenu ici par le point d'interrogation.

a/ Phrases introduites par un mot ou un groupe de mots interrogatif. Mis à part le point d'interrogation qui impose un schéma montant pour le dernier mot de phrase, on signale la fin de la locution interrogative par un marqueur (£) qui attribue également pour le mot qui le précède un schéma de Fo de type très montant.

Dans ce type de phrase, la transformation ne vise que l'avant dernier marqueur, sauf si celui-ci est justement (£) car il y a compabibilité de succession entre (£) et (?). Pour le reste, ($) sera transformé en (ø), et (#) ou (*) en (=) .

.../...

Ex : Quand (£) partez (...) vous ?
($) M_2 transformé en (⊛) D_2

Ex : Dans - quel - département (£) votre - correspondant
(# → =) habite-t-il ?

On attribue à tous les mots non situés sur ces points clés, un schéma mélodique identique à celui prévu pour les mêmes cas dans la phrase énonciative.

b/ Phrases construites selon le même modèle que les phrases énonciatives : seul le mot final indique l'interrogation.

(1) On a d'abord sur la chaîne phonétique :

↓

vous-habitez ($) chez - vos - parents (?.)

(2) puis, après application de la règle de transformation:

vous - habitez (£) chez - vos - parents (?.)

(1) vous - savez ($) quel - est - le - numéro (*) de - monsieur DUPONT (?.)

↓

(2) vous - savez (£) quel - est - le - numéro (=) de - Monsieur - DUPONT (?.)

c/ Phrases interrogatives avec inversion verbe/sujet .

Comme dans le premier cas, l'interrogation est déjà annoncée par la structure syntaxique de la phrase.

Dans ce type de phrase, on rattache au syntagme verbal, le pronom (tu, il, nous...) qui suit immédiatement le verbe.

.../...

(1) Pouvez-vous - répéter (**$**) votre - question (?.)

devient :

(2) Pouvez-vous - répéter (**ə**) votre - question (?.)

(1) Pouvez-vous - relire (**$**) le - dernier - paragraphe (*) de - votre - lettre (?.)
(2) Pouvez-vous - relire (**£**) le - dernier - paragraphe (=) de - votre - lettre (?.)
(1) Votre - correspondant (#) habite-t-il (**$**) à PARIS (?.)

(2) Votre - correspondant (#) habite - t - il (**ə**) à PARIS (?.)

(1) Votre - correspondant (#) habite-t-il (**$**) le - 15e - arrondissement(*) à PARIS (?.)

(2) Votre - correspondant (#) habite - t - il (**$**) le - 15e - arrondissement (=) à PARIS (?.)

Toutes les règles de transformation sont rassemblées dans le tableau12.

.../...

TABLEAU 12

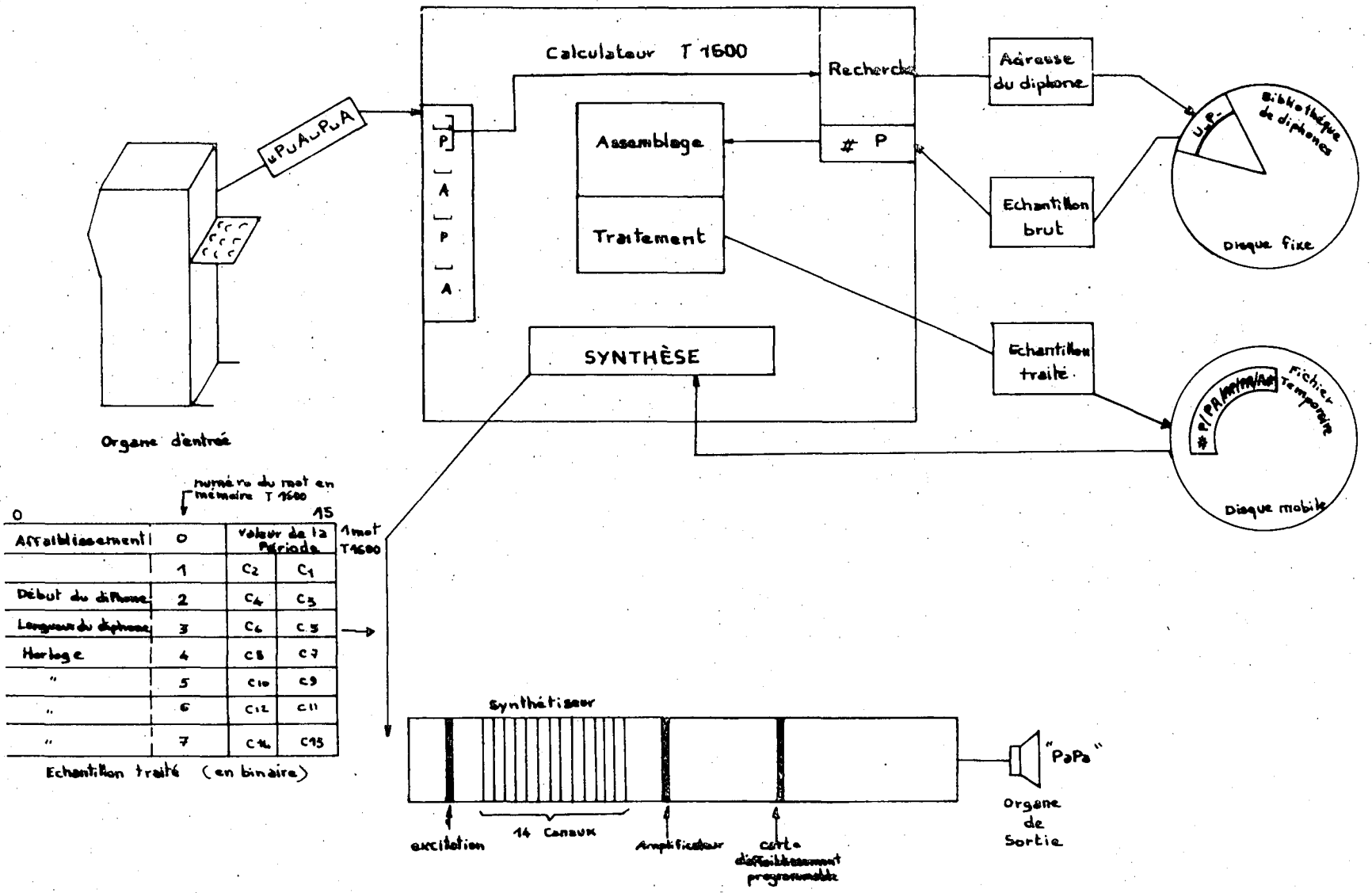
REGLES DE TRANSFORMATION DANS LES PHRASES INTERROGATIVES

1/ PHRASES INTRODUITES PAR UN MOT OU UNE LOCUTION INTERROGATIVE	
Phrase de type (£) + (?)	➡ Pas de modification. Il y a compatibilité dans la succession des marqueurs
Phrase de type (£) + ... [($) + (?)]	➡ ($) devient (ə) → (£) + ... [(ə) + (?)]
Phrase de type (£) + ... [(#) ou (✱) + (?)]	(#) ou (✱) deviennent (=) ➡ (£) + ... [(=) + (?)]
Phrase de type (£) + ($) + (✱) + (?)	➡ (£) + ($) + (=) + (?)
2/ PHRASES DE MÊME STRUCTURE SYNTAXIQUE QUE LES PHRASES ENONCIATIVES, OU CONSTRUITES AVEC INVERSION DU SUJET ET DU VERBE	
($) + (?)	➡ (ə) + (?)
($) + (✱) + (?_)	➡ (£) + (=) + (?_)
($) + (✱) + (✱) + (?_)	➡ (ə) + (✱) + (=) + (?)
($) + ($) + (✱) + (?_)	➡ (£) + ($) + (=) + (?_)
(#) + ($) + (✱) + (?_)	➡ (#) + ($) + (=) + (?_)
(#) + ($) + (?_)	➡ (#) + (ə) + (?_)
($) + ($) + (?_)	➡ (£) + (ə) + (?_)
($) + (ə) + (#) + (S) + (✱) + (?)	➡ (£) + (ə) + (#) + ($) + (=) + (?_)

.../...

Parvenue à ce point des explications, nous allons, à partir de la figure39 qui visualise chaque opération entre le moment où le message est écrit sur un organe d'entrée et le moment où l'on obtient la réponse vocale synthétique à la sortie d'un haut parleur, prendre un exemple concret pour récapituler l'ensemble des opérations d'assemblage et de traitement prosodique. Les résultats du traitement prosodique apparaissent sous forme de tracés aux figures 40, 41 et 42.

.../...


Calculateur T 1600
Recherche
Assemblage
Traitement
P
SYNTHÈSE
Adresse du diphone
Echantillon brut
Echantillon traité
Bibliothèque de diphones
U.P.
Disque fixe
Fichier temporaire
Disque mobile
Organe d'entrée
numéro du mot en mémoire T 1600
0
15
1 mot T1600
Attaque
Début du diphone
Longueur du diphone
Horloge
Valeur de la Période
Echantillon traité (en binaire)
Synthétiseur
excitation
14 canaux
Amplificateur
carte d'affaiblissement programmable
"PaPa"
Organe de Sortie

EXEMPLE : SYNTHESE DE LA PHRASE :

"le journaliste travaille dans son bureau avec le directeur

Entrée de la phrase en code phonétique avec les marqueurs

␣LEU-␣JOU␣R␣N␣A␣L␣I␣S␣T≠≠␣T␣R␣A␣Y∅␣DAN-␣SON-␣B␣U␣RAU*␣A␣VAI␣K-␣LEU-␣D␣I␣RAI␣K␣TOE␣R.

Décodage des caractères et création d'un tableau qui donnera ensuite les informations suivantes

1. Nombre de syllabes qui composent un mot
2. Nature du marqueur qui accompagne le mot : (-) ou (≠≠) ou (*) etc...
3. Nature de la syllabe qui termine le mot : ouverte ou fermée
4. Réalisation de la combinaison V($\frac{\partial}{\emptyset}$) V ou G($\frac{\partial}{\emptyset}$) à la frontière de deux mots

Application des règles de transformation des marqueurs

(≠≠) + (∅) + (*) + (.) $\xrightarrow{T}$ (≠≠) + (ə) + (*) + (.)

Introduction des pauses propres aux marqueurs

Appel des diphones du disque fixe et assemblage sur disque mobile

Représentation de la phrase sous forme d'échantillons vocodeur

Traitement de l'intonation à partir des informations données dans le tableau 1

Création d'un tableau de mot de "pitch" de la forme :
$[+ p, P_S]$ (p = pente ; P_S = valeur de période codée = pitch)
A chaque syllabe correspond un mot de pitch

Nombre de syllabes		Nature de la syllabe finale (ouverte ou fermée)	Marqueurs	Mots de pitch
1	leu	ouverte	_	+ 1.90
3	journasliste	fermée	≠≠	-1.86;1.84;-2.93
2	travaille	fermée	ə	-1.88;1.91
1	dans	ouverte	-	+ 1.86
1	son	ouverte	-	+ 1.90
2	bureau	ouverte	*	1.85 ; - 1.86
2	avec	fermée	-	-1.89 ; 2.78
1	le	ouverte	-	1.90
3	directeur	fermée	.	-1.86;2.86;1.103

Tableau 1

Traitement du rythme

Création d'un tableau de mot de rythme, rendant compte des zones à accélérer ou à ralentir suivant l'emplacement et la nature de la syllabe

Nombre de syllabes		Nature de la syllabe finale	Marqueurs	Mot de rythme
1	le	ouverte	-	LEU Consonne normale voyelle accélérée R_1
3	journaliste	fermée	≠≠	JOUR Consonne normale voyelle accélérée R_1
				NA Consonne accélérée voyelle accélérée R_2
				LIST Consonne normale voyelle ralentie car syllabe fermée R_4
2	travaille	fermée	ə	TRA R_1
	. . . etc	. . . etc		VAY R_3 . . .

Tableau 2

Traitement de l'intensité

Création d'un tableau de mot d'intensité ; ces mots ne réalisent pour l'instant qu'une sorte d'affaiblissement de 4 db ; en général l'affaiblissement suit l'intonation : il intervient quand F_0 est descendant.

nombre de syllabes		Marqueurs	mot d'intensité	
1	le	-	affaiblissement de 4 db	G 1
3	journaliste	#	affaiblissement sur la syllabe du milieu , les autres restant identiques	G 0 G 1 G 0
2	travaille	ə	affaiblissement sur la dernière syllabe	G 0

Tableau 3

A partir des 3 tableaux ainsi créés, le traitement est réalisé sur chaque échantillon vocodeur représentatif de la phrase.

le repère [128] implique l'appel d'un nouveau mot de pitch de rythme, et d'intensité

Avant passage des échantillons ainsi traités dans le synthétiseur, un système hard mémorise chaque échantillon, l'envoie vers une ligne d'affaiblissement dès la valeur de pitch affichée, et compte tenu des zones à ralentir ou à accélérer définies par les mots de rythme, positionne l'horloge sur 26 ou 6 ms avec retour à 13 ms pour les zones à débit normal.

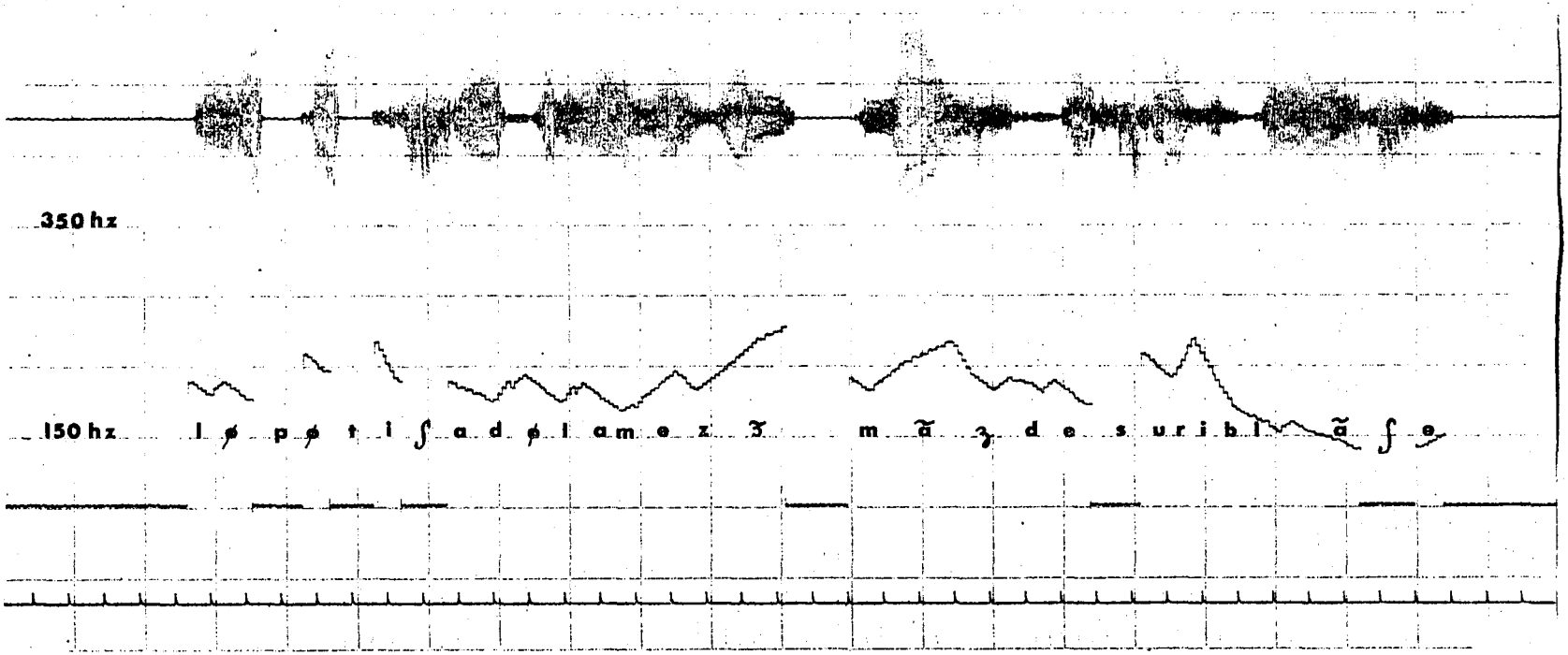


FIG 40 - Synthèse : le petit chat de la maison mange des souris blanches.

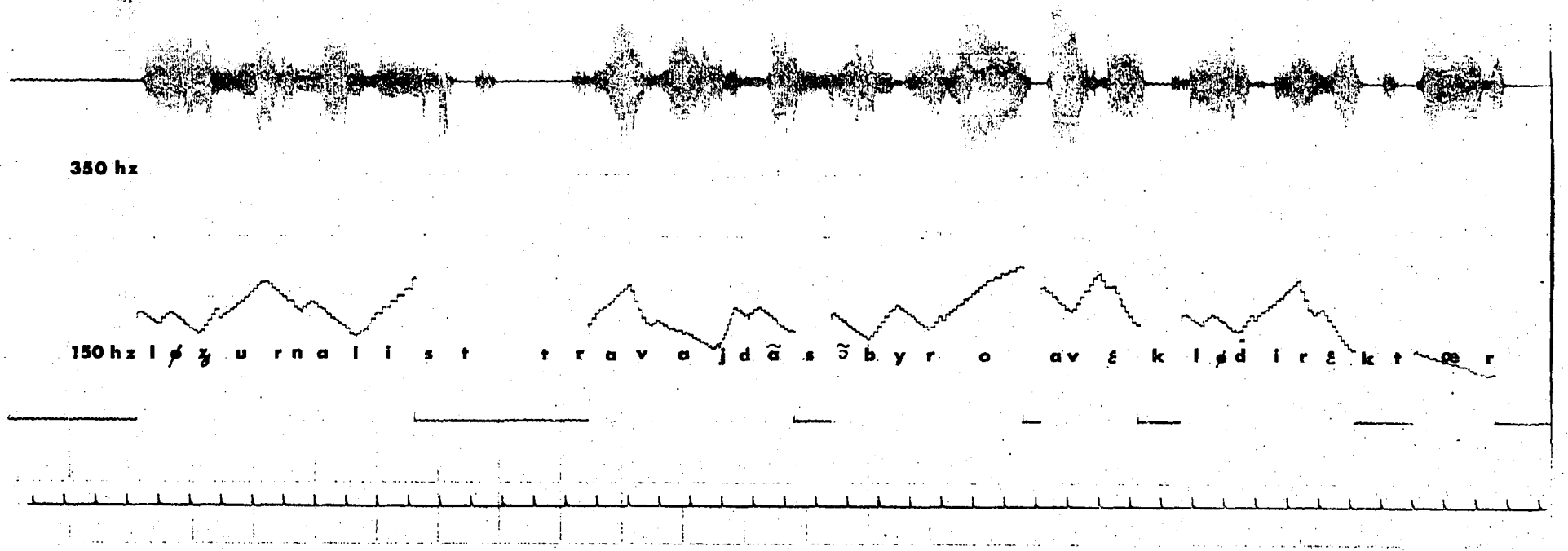


FIG 41 - Synthèse : le journaliste travaille dans son bureau avec le directeur.

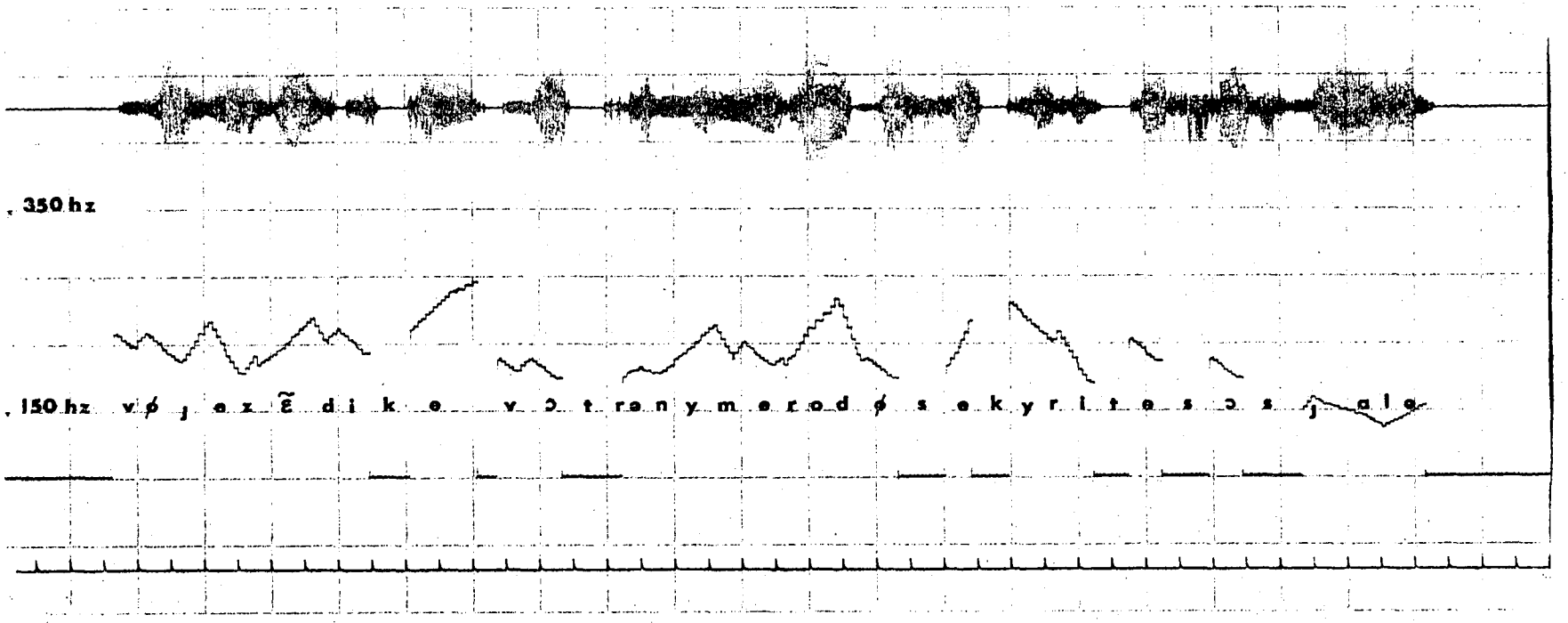


FIG 42 - Synthèse : veuillez indiquer votre numéro de Sécurité Sociale !

5ÈME PARTIE

APPROCHE

PSYCHO-LINGUISTIQUE

Nous avons signalé que l'application visée par un système de synthèse concerne l'automatisation d'un Centre de Renseignements ou plus précisément à court terme la lecture de lignes d'annuaire.

Des tests sont menés actuellement au Département ETA du CNET pour juger de leur intelligibilité et plus particulièrement de celle des noms propres ; les programmes de recherches prévoient également l'élaboration de procédures de tests qui permettent d'estimer non seulement l'intelligibilité mais aussi l'agrément de la parole synthétique. Ces points feront l'objet de rapports ultérieurs.

Pour l'instant, ce qui nous a intéressée et dont nous voudrions commencer à rendre compte ici, c'est la façon dont réagissent des auditeurs qui, après une communication téléphonique, apprennent qu'ils ont dialogué avec une "machine qui parle".

A ce sujet, deux expériences nous ont paru intéressantes, elles ont tenté de dégager le comportement tant linguistique que psychologique de personnes naïves face à la technologie moderne : une simulation de parole synthétique.

* La première expérience consiste en la simulation d'un dialogue homme-machine (QUINTON et al,1975) : on appelle quelqu'un par téléphone et on lui demande s'il accepte de tester un prototype de renseignement ; on lui précise qu'il peut formuler comme il veut une question portant sur le numéro de poste, le grade ou le numéro de bureau d'une personne d'un groupement du CNET. Le dialogue s'engage avec un opérateur qui déforme sa voix en branchant le téléphone sur l'analyseur du vocodeur et en parlant d'une façon très mécanique (monotone). Les résultats montrent que les sujets adoptent très rapidement un comportement de mimétisme tant au niveau de la formulation de leurs questions qu'au niveau de l'articulation : spontanément, ils sont prêts à faire un effort pour faciliter le dialogue (voir aussi VAISSIERE , 1976).

..../...

* La seconde expérience a été réalisée par une équipe de la Télévision française (Emission "La Caméra invisible").

Un haut-parleur est dissimulé dans un "photomaton" (cabine photographique automisée) et un opérateur - également caché - engage le dialogue (la voix est volontairement déformée et dépourvue de naturel) avec le client dès que celui-ci pénètre dans la cabine.

- "Bonjour Monsieur (ou "Madame", selon le client) : vous êtes ici dans une cabine entièrement automatique ; est-ce la première fois que vous utilisez une cabine automatique ?"

- L'opérateur obtient rarement une réponse spontanée à cette première question : le client manifeste sa surprise ("hein ?"), cherche d'où vient la voix, puis exécute les gestes habituels (introduction d'une pièce, positionnement du siège) sans se soucier de la voix.

- L'opérateur répète alors sa question : "je vous ai demandé si c'est la première fois que vous venez dans une cabine automatique ?"

- Le client finit toujours par répondre : "non !".

Opérateur : - "Bon ! alors nous allons commencer ! Prêt pour la première photo ?
(le client répond oui ou non) cinq, quatre, trois, deux, un, top !...

- Prêt pour la seconde photo ? Mettez vous de profil !...
... si le client n'obtempère pas :
- Je vous ai demandé de vous mettre de profil ! - Davantage ! - Encore ! (le client excédé finit par s'éxécuter).
Souriez s'il vous plait ! - " ... etc...

.../...

Sauf un client qui a feint d'ignorer totalement les ordres de l'opérateur, tous les autres sujets ont fini par se soumettre aux injonctions de la machine. Soumission débonnaire ou soumission excédée (quand on utilise ce type d'appareil, c'est essentiellement pour le gain d'argent, pour la rapidité d'exécution, et surtout pour la solitude avec l'appareil : nulle obligation de sourire béatement sur l'ordre d'un photographe !...). Les personnes, une fois la première surprise passée, ont exécuté assez volontiers la plupart des conseils d'ordre technique concernant la hauteur du tabouret ou l'éclairage de la cabine à modifier, mais se sont soumis avec beaucoup de mauvaise volonté à tous les ordres relatifs à l'esthétisme ("souriez , tournez-vous", etc...)

Il est en tous cas remarquable de constater l'absence totale de surprise au fait que la machine puisse non seulement parler, mais voir, comprendre et décider en fonction de la préhension de la situation. Il est regrettable toutefois qu'à la sortie de cette cabine, les clients n'aient pas été interviewés.

Nous avons élaboré notre test pour obtenir des informations concernant :

1/ L'intelligibilité globale d'un message synthétique par téléphone dans une ambiance naturelle.

2/ Les"impressions" des correspondants quand on leur signale simplement que la voix qu'ils viennent d'entendre "sort d'un ordinateur" : surprise, angoisse, amusement ...?

Pour ce faire, nous avons branché le synthétiseur sur le téléphone, appelé une personne connue ou inconnue et envoyé au fur et à mesure du dialogue, des phrases de synthèse élaborées par anticipation du dialogue. A la fin de celui-ci , nous expliquions brièvement au correspondant qu'il venait de parler avec une machine et nous lui demandions de nous donner ses impressions, nos interventions étant dans la mesure du possible volontairement très limitées.

.../...

Les entretiens que nous rapportons ici (sans aucune modification - style - ni coupure) peuvent sembler peu élaborés du côté émission (voix de synthèse). Ceci tient aux difficultés d'élaboration de ce type de test.

En effet, avant d'appeler quelqu'un, il faut envisager toutes les possibilités de dialogue.

Il faut prévoir par exemple que le correspondant peut :

1 - comprendre la question posée par la machine mais demander confirmation.

2 - ne pas comprendre et demander répétition.

3 - croire comprendre mais demander un supplément d'information, etc...

Ensuite, tous les messages que ces possibilités sous-entendent doivent être écrits sur la télé imprimante, synthétisés, et mémorisés dans des fichiers séparés du calculateur.

On ne peut en effet pour réaliser un dialogue spontané attendre la question du correspondant, taper la réponse sur l'imprimante et la synthétiser ; le temps de réponse serait trop long et le correspondant déjà vraisemblablement mis en méfiance par la voix "un peu bizarre" de la première phrase, aurait eu le temps de raccrocher.

C'est pourquoi pour chaque entretien, d'autres phrases synthétisées qui n'apparaissent pas ici avaient été prévues pour essayer de parer à toutes les questions possibles.

Par exemple, dans l'Entretien 1 avec une opératrice d'un standard téléphonique, nous avions synthétisé et rangé dans des fichiers différents les phrases suivantes :

1 - Bonjour, je voudrais le 73-83 s'il vous plait.

2 - le 73-83

.../...

3 - non. Je voudrais le 73-83 : sept, trois, huit, trois.

4 - oui, merci.

Une vingtaine d'entretiens a été réalisée selon ce modèle. Nous reproduisons ici ceux qui reflètent les tendances générales. Nous avons noté /VS/ la voix de synthèse, /CA/ les interventions du correspondant appelé, et /VL/ la voix du locuteur qui a servi pour l'enregistrement du dictionnaire de diphones . En effet, même si on n'a pas l'habitude d'écouter de la voix synthétisée par diphones, il est possible par instant de reconnaître le locuteur qui a servi pour la construction du dictionnaire. En l'occurrence, il a été intéressant de jouer sur cette ressemblance dans la mesure où l'on retrouve /VL/ à l'origine de /VS/ dans la seconde partie du dialogue.

.../...

ENTRETIEN 1 (Avril 1976)

On appelle un standard téléphonique situé dans le département de l'Isère.

/ CA / - "ST MARCELLIN, 14 ; j'écoute !"

/ VS / - "Bonjour, je voudrais le 73-83, s'il vous plait."

/ CA / - "73-83" (le numéro a été répété sans hésitation et sur un ton affirmatif)

/VS / - "oui, merci".

... Nous obtenons la communication avec la personne demandée ...

Dix minutes plus tard, nous rappelons le standard de ST MARCELLIN.

...

/ VL / - "Allo, est-ce qu'il serait possible de parler avec l'opératrice qui a le numéro 14 ?"

... on nous passe l'opératrice sans nous poser de question.

/ VL / - "Bonjour, Madame, Je crois que c'est à vous que j'ai demandé tout à l'heure le 73.83 ?"

/ CA / - "Oh, alors là, Madame, je ne m'en rappelle pas. J'ai peut-être répondu à dix appels depuis. Pourquoi donc ?" -

/ VL / - "Et bien voilà. On est dans un centre PTT en Bretagne, et c'est une petite expérience qu'on a faite avec vous : c'était une machine qui vous a demandé le numéro de téléphone... Je peux vous repasser le message si vous voulez !"...

/ CA / - (Rires) "Oui, oui, bien sûr ; oui, oui, bien sûr !" -
(on sent l'opératrice soudain détendue et manifestement contente,

.../...

amusée, de participer à cette expérience.)

... on diffuse à nouveau le même message...

/ CA / - "oui, oui ; on entend bien,hein !".

/ VL / - "Vous avez compris facilement ?"

/ CA / - "Ah oui ! oui, oui ! j'ai bien compris qu'il vous demandait ... qu'il me demandait le 73-83 !"

/ VL / - "Mais la voix vous a choquée un peu ?"

/ CA / - "... heu... oui, elle est un petit peu... heu... un peu tremblotante, un peu lente si vous voulez... mais sinon... on dirait par exemple quelqu'un qui est pas jeune, vous voyez..."

/ VL / - "Mais ça ne vous a pas surpris outre mesure ?"

/ CA / - "Non. Non, non ! bien que vous savez, on a des correspondants qui ont quelquefois des voix comme ça, alors ça ne m'a pas surpris, non. Ca ne me surprend pas, non, non".

...

.../...

ENTRETIEN 2 - (Septembre 1976)

On appelle le service de réception dans une usine ; on obtient l'hôtesse :

/ CA / - "Allo ?"

/ VS / - "Bonjour, je voudrais parler à Monsieur X ..."

/ CA / - "Oui, je vous le passe '" -

...

Monsieur X : " - Allo ?"

/ VS / - "Ne quittez pas, le calculateur du CNET vous passe Monsieur Y..."!
(Rires de Monsieur X... qui sait que Monsieur Y... travaille dans le domaine de la parole. On lui explique de quoi il s'agit. Sa réaction est la suivante : - " C'est pas mal, hein?... on dirait une voix mi-mâle, mi-femelle !" -
On lui demande cinq minutes plus tard de nous remettre en ligne avec l'hôtesse :

...

/ VL / - "Allo, Madame ? on vous a demandé tout à l'heure Monsieur X... Vous vous souvenez ?"

/ CA / - "Bien oui! je vous ai bien passé Monsieur X ... ! (affirmation impatiente)

/ VL / - "Oui ! mais je voulais vous demander : vous n'avez pas été surprise par la voix ?"

/ CA / - "par la voix de qui ?"

.../...

/ VL / - "par la voix qui vous a demandé Monsieur X...".

/ CA / - "par votre voix à vous ?"

/ VL / - "... heu... en quelque sorte..."

/ CA / - (exaspération) " - Mais comment voulez-vous que je sois surprise par votre voix, Madame ! je ne vous connais pas !..."

/ VL / - "... oui. C'est vrai ! Excusez-nous !... c'était une petite expérience !...

/ CA / - Ah, bon ! d'accord (Rires). Mais vous le voulez encore ce Monsieur ?"

.../...

ENTRETIEN 3 - (octobre 1976)

On appelle une librairie à PARIS. On obtient une première correspondante (environ 55 ans).

/ CA / - "Allo ?"

/ VS / - "Allo ? c'est la Librairie Dargaud ?"

/ CA / - "Oui !"

/ VS / - "Vous avez les bandes dessinées de Christin et Bilal ?"

/ CA / - "heu, oui... je crois qu'il nous en reste quelques uns... Quel titre voulez-vous ? - "

/ VS / - "le Vaisseau de pierre"

/ CA/ - "le vaisseau de pierre ? oui, oui Madame, nous l'avons" -

On arrête le dialogue ici ; on demande à cette personne de nous passer une dame rencontrée trois mois plus tôt et à laquelle on avait un peu expliqué ce qu'était la "synthèse de la parole" -

Cette seconde personne prend le téléphone (mais a été plus ou moins avertie de la provenance de l'appel -)

/ CA / - "Allo ?"

/ VS / - "Allo ?"

/ CA/ - "Oui, oui Madame !..."

/ VS / - "Allo, c'est la librairie DARGAUD ?"

.../...

/ CA / - (en même temps que la phrase ci-dessus énoncée)"Ah !... (compréhension) Ah ? vous faites parler l'ordinateur !?... je vous ai reconnue hein ? ... Ah, ben... il parle drôlement... hein ?... Repassez-le moi... Attendez !, ma collègue arrive... Allez-y ! faites le parler ! elle écoute !..."

/ VS / - "Allo, c'est la librairie DARGAUD ?"

/ CA / - "Oui !" (la personne se prête au jeu)

/ VS / - "Vous avez les bandes dessinées de Christin et Bilal ?"

/ CA / - "Oui, oui ! on les a ! ..."

...

/ VL / - "Qu'est-ce que vous en pensez alors ? ... La première personne que j'aieue tout à l'heure a eu l'air de comprendre ?"

/ CA / - "Oui, on comprend très bien, très très bien oui !..."

/ VL / - "Mais vous avez quelle impression quand vous entendez ça ?"

/ CA / - "... c'est une voix... morte un peu ! ... c'est ça ! ça fait une voix morte... qui viendrait des cavernes un peu ! (rires...) ... C'est vrai, ça fait un effet comme ça. Vous avez bien vu des fois des films de science fiction ? ... ça fait un peu ça !..."

/ VL / - "Mais d'après vous, c'est une voix d'homme, de femme, de rien du tout ?..."

/ CA / - "DE RIEN DU TOUT !... une voix de petit homme vert, moi je dirais"(rires...)

.../...

... elle parle ensuite d'un client de la veille qui leur a longuement et très sérieusement parlé de soucoupes volantes et des extra-terrestres et conclut :

/ CA / - "Il nous a donné le trac avec ses histoires !... Alors ça, avec la voix d'aujourd'hui !... (rires) c'est complet !..."

(Pendant tout le temps où elle parlait, nous avons eu le temps de préparer un autre message sur la téléimprimante)

/ VL / - "Ecoutez !..."

/ VS / - "Nous sommes en Bretagne et nous vous disons bonjour !..."

/ CA / - "Ah oui !... ça fait plus une voix d'homme que tout à l'heure ! Et c'est vous qui parlez là ?..."

/ VL / - "En ce moment, oui, c'est moi qui parle !..."

/ CA / - "Maintenant oui ! non, mais tout à l'heure ?... (silence)... Ah mais non ! c'est personne qui parle puisque c'est vous qui la faites parler !..."

/ VL / - "Vous pensiez que c'était moi qui parlais ?"

/ CA / - (Silence) ... "oh, je ne sais plus !"..

/ VL / - "Attendez, je recommence !"

... on diffuse à nouveau la dernière phrase synthétique...

/ CA / -(silence)- "Oh, je ne sais plus... c'est bizarre !... je n'en sais rien..."

/ VL / - "Ça vous choque ?"

.../...

/ CA / - (Eclat de rires) "Oui, on dirait Max La Menace !... (rires) Oh!... ce que c'est marrant !... oh, c'est rigolo !... (rires) oui !..."

/ VL / - "Mais vous comprenez ce qui est dit ?"

/ CA / - "Très bien ! oh oui ! très très bien !... c'est distinct !... sauf que ça surprend comme ça parce que... on sent pas la chaleur humaine !... Voilà !... Alors, mettez-lui un peu de chaleur...
parce que quand même ça surprend... la nuit on entendrait cette voix comme ça qui raconterait en plus des choses bizarres, ça donnerait une sacrée frousse !..."

.../...

ENTRETIEN 4 (novembre 1976)

Nous avons appelé une personne que nous connaissons, à qui nous avons simplement dit que nous essayions de "faire parler des machines" et qui n'avait aucune idée ni sur le système utilisé, ni sur les résultats . Nous n'avons pas simulé de dialogue, nous l'avons appelé et lui avons fait écouter un enregistrement de phrases synthétiques à sémantique variée (depuis l'énonciation de lignes d'annuaires et de "phrases téléphoniques" jusqu'à la description des développements à utiliser sur une bicyclette !...)

Voici les impressions de cet auditeur à la fin de l'écoute :

/ CA / - "Dis donc, c'est marrant ce truc là... Enfin, ça surprend, hein ? ... Alors, c'est la voix de qui, ça ? ... c'est ça une voix synthétique ?..."

/ VL / - "Oui c'est ça !... A ton avis, c'est la voix de quoi ?... Qu'est-ce que tu penses ?..."

/ CA / - "c'est marrant !... Je ne sais pas... ça surprend !... A un moment, j'ai pensé que c'était féminin, et le reste du temps ça m'a paru masculin... Mais, j'ai l'intuition que c'est féminin... et même j'ai eu l'impression pendant un moment - oh, ça a dû durer trois secondes - que ça ressemblait à ta voix. C'est idiot, hein ? mais ça doit être le téléphone qui déforme... ou alors c'est des problèmes d'intonation...

/ VL / - "Mais tu as tout compris, ou non ?"

/ CA / - "Non, il y a deux ou trois mots que je n'ai pas compris. J'ai compris... (suit la répétition par l'auditeur d'environ 80 % du texte. Le sens des messages a été parfaitement assimilé, le vocabulaire n'étant pas toujours celui utilisé dans le message synthétique. Les deux mots non compris étaient effectivement dégradés dans l'enregistrement)... mais ce qui m'étonne c'est

.../...

que ce n'est pas trop saccadé ? je pensais que ça serait beaucoup plus automatique tac - tac - tac !... tu vois ?. C'est presque un débit normal hein ?..."

/ VL / - "Mais tu ne trouves pas que c'est un peu mort ? tu crois qu'il y a de l'intonation ?

/ CA / - "Ah oui ! il y a du débit et de l'intonation !... Mais je ne sais pas comment te dire... heu... il n'y a pas d'intonation... affective, je dirais !... , mais le texte lui-même a une intonation - comment dit-on ? - linguistique ? je te dis, il y a même des intonations qui ressemblent à ta voix. C'est stupide, hein ?

/ VL / - "Finalement, ça ne t'a pas semblé bien chaleureux ?...

/ CA / - "Bien tu sais, chaleureux hein, le texte lui-même n'a rien de bien chaleureux. C'est sûr, que si tu m'avais parlé de tes états d'âme sur ce ton là, je me serais fait du souci pour ta santé !... mais quand tu téléphones pour avoir l'heure ou pour demander une adresse, ça ne m'a jamais semblé tellement plus chaleureux... je ne vois pas pourquoi on serait chaleureux pour me donner l'heure ! ? ...

.../...

ENTRETIEN 5 (décembre 1976)

/ VS / - "Allo, c'est Madame X... ?"

/ CA / - "Oui !"

/ VS / - "Ne quittez pas, on vous appelle de Bretagne".

/ CA / - "Ah bon, je vous remercie !"

... on arrête là le dialogue, et on dit simplement à notre correspondante (environ 60 ans) qu'il s'agit d'une machine qui parle...

/ CA / - "Et bien c'est très rigolo !".

/ VL / - "Vous avez compris ?"

/ CA / - " Ah oui très bien : Allo, c'est Madame X... ? on vous parle de Bretagne... j'ai cru d'abord que c'était Modane parce que j'attendais un coup de fil de ma fille."

/ VL / - "Alors qu'est-ce que vous en avez pensé, de cette voix ?"

/ CA / - "Figurez-vous, j'ai cru que c'était un canular? Je me suis demandée si on ne me faisait pas une farce... parce que la voix était un peu rigolotte. On aurait dit comme une voix de gnome. A mon idée, vous voyez ? - Amusant... qu'est-ce que vous en pensez, vous ? vous l'entendez, vous ? ... Mais c'est drôle !... je trouve ça sympa en tous cas !... je vous assure que je me suis demandée si c'était une farce."

/ VL / - "Qu'est-ce que vous pensez de la voix ?"

/ CA / - "Je pense que c'est une voix d'eunuque. hein ? Je vous dis exactement le fond de ma pensée. Moi voilà, j'ai eu l'impression qu'on me faisait une farce, que c'était une

.../...

voix un peu déguisée, et une voix d'eunuque. Une voix sans sexe en tous cas. Mais rigolotte !..."

... on lui fait écouter ensuite la synthèse du texte de "la chèvre de Monsieur Seguin"

/ VL / - "Vous avez compris ?"

/ CA / - "Oh, très bien, très très bien ! C'est une belle histoire !..."

.../...

ENTRETIEN 6 (décembre 1976)

On téléphone à un publicitaire (40 ans)

On obtient d'abord l'hôtesse d'accueil

/ CA / - "Allo, ici Agence X..., j'écoute !"

/ VS / - "Allo, je voudrais parler à Monsieur X ..."

/ CA / - "Oui, c'est de la part de qui ?"

/ VS / - "Michel(e) DARNOWSKI" (nous avons volontairement fait le choix d'un prénom qui est à la fois féminin et masculin pour ne pas influencer les réactions du correspondant sur la voix qu'il entend)

/ CA / - "Michel(e) ... ?

/ VS / - "Michel DARNOWSKI".

/ CA / - "Un petit instant s'il vous plait"

....

/ CA / - (la même personne, Monsieur X... est occupé). "Allo, c'est à quel sujet s'il vous plait ?"

/ VS / - "Bonjour ! écoutez, j'ai des problèmes. J'organise un congrès médical en juin 77, et j'aurai voulu savoir si vous pouviez m'aider ?..."

/ CA / - "Ah ! vous voulez... vous organisez quelque chose et vous voulez que Monsieur X... vienne vous aider ? ..."

/ VS / - "Oui !"

.../...

/ CA / - mais ce serait quoi exactement ?...

... on arrête là le dialogue ; on explique brièvement à l'hôtesse (25 ans) qu'il s'agit d'une machine qui parle. L'écoute téléphonique est de très mauvaise qualité (grésillements importants).

/ CA / - "Ah ? moi je croyais que c'était quelqu'un qui ne parlait pas bien le français, que c'était un homme qui parlait. C'était un homme ou une femme ? Je pensais que c'était un homme, une personne âgée qui avait la voix qui tremblait et qui ne parlait pas bien le français." (la consonance étrangère du nom a sans doute renforcé cette impression).

/ VL / - "Mais vous avez compris" ?

/ CA / - "Oui, j'ai bien compris. Mais il faut dire que j'écoutais bien, mais ça grésillait dans le téléphone..."

/ VL / - "Mais vous n'avez pas compris le nom propre quand même ?"

/ CA / - "Michel... euh... enfin je ne me souviens plus maintenant, mais je l'ai répété après, hein ?

/ VL / - "Je vous le fait écouter encore une fois !"

/ VS / - "Michel(e) DARNOWSKI"

/ CA / - "Michel DARNOWSKI, c'est ça ? mais là, on ne dirait pas une voix de femme, hein ? je l'ai bien écouté et on dirait une voix d'homme !"

/ VL / - "Mais on comprend quand même ?"

/ CA / - "Oui, on comprend "

.../...

/ VL / - "Mais vous trouvez ça comment ?... triste ?..."

/ CA / - "Oh, bien non, vous savez, non ! c'est une voix... non, ça peut être une voix normale à mon avis, hein. Je pense... qu'il y a des personnes qui parlent comme ça, il y en a d'autres qui parlent plus mal encore, hein !..."

... on lui demande ensuite s'il est vraiment impossible de parler à Monsieur X... , on lui précise qu'on le connaît. La communication téléphonique continue à être de très mauvaise qualité .

Monsieur X prend le téléphone :

/ CA / - "Allo ?"

/ VS / on lui passe le message synthétique.

/ CA /-"Qu'est-ce que c'est ?"

/ VL /- "C'est une machine qui parle".

/ CA / - "Oui, mais qu'est ce qu'elle demande ?"

/ VL / - "Elle demande si vous pouvez lui organiser un congrès "

/ CA / - "Ah bon ! oui. Et bien, écoutez, pour collaborer avec vous, d'accord ! Avec plaisir ! mais c'est quand ce congrès ?"

... Fin du dialogue parce que la liaison téléphonique est très mauvaise du côté du correspondant.

.../...

REFLEXIONS A PROPOS DE CES ENTRETIENS

Les jugements portés par les auditeurs dans ces entretiens sont de deux ordres. D'une part ils signalent le degré d'acceptabilité de la parole synthétique (intelligibilité et agrément) , d'autre part, ils portent sur la préhension du phénomène "ordinateur", situent la Machine par rapport à l'Homme et précisent ses possibilités, ses limites.

- 1/ Les jugements portés sur la parole de synthèse :

Tous les entretiens sous entendent une bonne intelligibilité globale des messages. Il n'a jamais été nécessaire de répéter les énoncés de synthèse sauf le nom propre dans l'Entretien 6 (d'ailleurs peu courant). D'autre part, nous voulons signaler ici une anecdote qui fait suite à l'entretien 1 : l'opératrice a compris immédiatement le numéro de téléphone qui était demandé. Quelques jours plus tard, le locuteur (dont la voix est à la base du dictionnaire de diphones) a redemandé le même numéro au même standard et a dû le répéter trois fois (sans grésillements sur la ligne et sans aucune volonté de déformer sa voix)!... Ce fait est à rapprocher du jugement de l'hôtesse (Entretien 6) : "je pense qu'il y a des gens qui parlent comme ça, il y en a d'autres qui parlent plus mal encore !..." et nous rappelle, si besoin était, les multiples paramètres qui interviennent dans la communications parlée.

Un autre élément qui se dégage et qui soulève un gros problème concerne le caractère masculin ou féminin de la voix. Le locuteur choisi dans ce système de synthèse était un locuteur féminin, pourtant la plupart des auditeurs hésitent quand on leur demande qui parle :

"une voix mi-mâle, mi-femelle", "une voix d'eunuque", "c'est un homme ou une femme ?" "à un moment j'ai pensé que c'était féminin, et le reste du temps ça m'a paru masculin". "On ne dirait pas une voix de femme". Nous pensons que ce caractère vient de ce que l'enregistrement des diphones a été fait sur un ton volontairement grave, et de la méthode de synthèse : la fréquence des formants

.../...

demeure inchangée quelle que soit la valeur de Fo introduite ; or, les valeurs de la fréquence fondamentale que nous avons utilisé pour le traitement intonatif sont plus élevées que celles des diphones stockés. Il serait peut être souhaitable d'envisager un nouveau dictionnaire de diphones avec la voix d'un locuteur masculin. Il n'y aurait plus alors de risque d'ambiguïté sur l'identité du locuteur. En effet, il faut ajouter que l'aspect masculin attribué à la voix de synthèse ne l'est jamais quand le locuteur à l'origine de cette voix communique naturellement par téléphone avec un auditeur inconnu. Il arrive effectivement souvent que l'auditeur trompé par le timbre de la voix commette des erreurs sur le sexe de son correspondant. Ce n'est pas le cas en l'occurrence.

Les autres défauts qui sont signalés par les auditeurs concernent surtout l'aspect"vieux" et "tremblotant" de la voix. Le premier aspect pouvant découler du second. Cette réflexion nous a permis depuis ces entretiens de remédier notablement à ce défaut.

En effet, jusque là, nous pensions que la discontinuité de Fo susceptible de se produire au centre des réalisations consonantiques était sans incidence sur l'intelligibilité. Or, il est apparu qu'un certain nombre de consonnes - en particulier les liquides, les nasales et les glides - sont extrêmement sensibles à une rupture brusque de la fréquence fondamentale pendant leur réalisation : l'incrémentation réalisée pour supprimer cette discontinuité a permis une amélioration sensible. D'autre part, il est certain que l'accélération globale du débit augmente le naturel de la voix et donne une impression de plus grande assurance. Il est évident que des tests systématiques devront être élaborés pour objectiver ces impressions.

Quant à la prosodie, elle semble apparaître satisfaisante "oui, il y a du débit et de l'intonation". Le fait le plus remarquable et qui n'est que très peu souligné dans ces entretiens concerne les réflexions des personnes qui connaissent le locuteur et qui trouvent épisodiquement des ressemblances entre sa voix et la voix synthétique tout en s'excusant immédiatement de l'ineptie de leurs réflexions :

.../...

"Et c'est vous qui parlez, là ?... mais non ! c'est personne qui parle puisque c'est vous qui la faites parler" "j'ai eu l'impression pendant un moment que ça ressemblait à ta voix... il y a des intonations qui ressemblent à ta voix"...

- 2/ Les jugements relatifs aux relations Homme-Machine :

La remarque qui s'impose d'abord concerne l'extrême gentillesse de tous les correspondants auxquels nous avons fait appel, et leur disponibilité pour parler et donner leurs impressions. Il faut noter également le plaisir manifesté par ces personnes de participer à une expérience. Nous pensons en particulier à l'opératrice du standard téléphonique qui nous a paru inquiète au début de l'entretien commencé sous la forme d'un interrogatoire ("c'est à vous que j'ai demandé le 73-83 ?" puis soudain détendue quand nous lui avons expliqué de quoi il s'agissait. Cette réaction nous laisse envisager favorablement une extansion plus élaborée de ce type de test.

Cependant leur conduite pose de réelles difficultés : il faut anticiper la forme du dialogue ; dès que le correspondant s'éloigne du schéma prévu, il n'est plus possible de continuer l'échange antérieur.

Les questions posées ensuite aux correspondants induisent les réponses : "la voix vous a choqué ?" "vous ne trouvez pas que la voix est un peu morte ?"

D'autre part, nous ne donnons que peu d'explications aux personnes appelées ("c'est une machine qui parle"... "c'est une petite expérience" "vous venez de parler avec un ordinateur"..". Et de ce fait, aucune d'entre elles ne nous a demandé plus de précision ; on ne peut donc pas tirer de conclusions quant à la préhension de la voix synthétique par ces auditeurs parce qu'il est possible que le procédé de synthèse soit pressenti comme une variante d'une technique d'enregistrement . Cependant, il faut noter que la plupart des personnes interrogées bien qu'ignorantes du processus ne semblent absolument pas effrayées, ni angoissées par cette voix, mais plutôt amusées : "c'est rigolo !" est le terme qui revient le plus souvent, "une voix sympa !".

.../...

Enfin, l'interview est trop rapide, trop immédiate, on ne laisse pas assez le temps à la réflexion. Nous pensons qu'il serait peut être plus intéressant de commencer un dialogue avec la voix de synthèse puis interrompre la communication sans intervention de notre part (voix naturelle), enfin appeler à nouveau le correspondant quelque temps plus tard et lui demander par exemple s'il n'a pas reçu une communication téléphonique auparavant, ce qu'il en a pensé, en repoussant le plus possible le moment où on lui expliquerait de quoi il s'agit.

Beaucoup d'autres tests sont envisageables qui élimineraient tous les défauts mentionnés. Nous pensons en particulier à des expériences du type "Photomaton" relatées plus haut qui pourraient être suivies d'une interview (choix ouvert, choix fermé).

Que dire des résultats obtenus ?

que des défauts demeurent
que des qualités existent
que les possibilités d'amélioration ne sont pas épuisées

1 - des défauts demeurent :

- "La voix est un peu vieille et tremblotante" - Nous avons remédié à cette impression de chevrotement causée par la voix de synthèse en nous attaquant à ses causes :

  * un débit un peu trop lent,
  * la possibilité de discontinuité de Fo au centre d'une réalisation consonantique : le traitement réalisé pour éviter ce risque de rupture a notablement amélioré l'intelligibilité des consonnes liquides et nasales, et des semi-consonnes.

Mais nous ne pensons pas que l'aspect "vieux" de la voix ait été supprimé de ce fait. La solution consiste peut-être à éliminer le locuteur qui a servi à l'enregistrement du dictionnaire de diphones !?.

.../...

- Les voyelles situées en fin de mot et dans une syllabe fermée sont modifiées quand celui-ci présente un schéma de Fo montant. Nous remarquons en effet que toutes les voyelles dans cette position tendent à prendre le timbre d'une voyelle plus fermée :

[a] est perçu comme [ɛ]
[ɛ] est perçu comme [e]
[o] est perçu comme [u] , etc...

Nous avons dit que les diphones ont été extraits de mots prononcés avec une fréquence laryngienne moyenne inférieure à celle d'un corpus lu. Le fait d'introduire artificiellement une structure harmonique indépendante de l'enveloppe spectrale est sans doute responsable de cette modification de timbre.

Il est possible également que ce phénomène provienne de l'allongement de durée que l'on introduit par un ralentissement de la cadence d'échantillonnage dans la partie stable des voyelles en syllabe fermée situées avant une pause. ENDRES (1974) utilise en synthèse le raccourcissement ou l'allongement de la section stationnaire dans le domaine temporel des voyelles pour en changer la tonalité .

Mais nous pensons que la non-corrélation de la structure harmonique et de l'enveloppe spectrale est la cause principale du phénomène observé puisque les mêmes voyelles dans le même contexte consonantique demeurent inchangées quant à leur timbre*. Ce défaut n'existe que pour les voyelles en syllabes fermées car en syllabes ouvertes (diphones [voyelle #]) elles ont été enregistrées avec un Fo montant : leur synthèse dans les mêmes conditions intonatives réalise donc une parfaite reconstitution spectrale tant au niveau de la répartition des harmoniques que de l'enveloppe. Lorsqu'elles sont générés avec un Fo descendant, bien que la répartition spectrale soit plus riche, l'enveloppe n'en est pas moins respectée.

* Si Fo est descendant.

.../...

2 - des qualités existent :

Nous n'avons pas encore effectué de tests systématiques pour juger de l'intelligibilité et de la qualité de parole de synthèse. Cependant les avis (nombreux) émis après écoute de la voix de synthèse donnent à penser que l'intelligibilité globale des messages synthétiques est bonne et que l'introduction de la prosodie est à l'origine pour une grande part des avis favorables qui se manifestent.

Enfin et surtout, la synthèse - assemblage des diphones et traitement prosodique - est réalisée en temps réel.

3 - Il reste que des améliorations sont encore possibles : particulièrement en ce qui concerne la réduction du dictionnaire (ENDRES et GROSSMAN, 1974 ; LEIPP et al, 1968) et la réduction de la mémoire utilisée pour le traitement de la prosodie sans pour autant détériorer la qualité. C'est ce à quoi nous veillerons.

Pour finir, nous laisserons "la parole" à GENIN (1976) :

..." La réponse vocale doit être agréable et intelligible, elle ne doit cependant pas chercher à tromper son auditoire... Soyons rassurés, les artisans de la synthèse de la parole, en toute modestie, reconnaissent qu'ils ont encore des progrès à accomplir avant que leurs machines ne jouent les imitateurs parfaits".

CONCLUSION

Cette étude s'insère dans le cadre général de l'automatisation des Centes de Renseignements. Elle avait pour but de réaliser un système de synthèse opérationnel, fonctionnant en temps réel, et produisant une parole non seulement intelligible, mais aussi relativement naturelle.

Pour arriver à ce résultat, nous avons effectué un certain nombre de choix qui nous ont amené à retenir comme type de synthétiseur le vocodeur à canaux, comme élément minimal de parole le diphone, et comme moyen d'amélioration de la qualité la mise au point d'un traitement systématique de la prosodie (caractéristiques intrinsèques, accent et intonation).

Deux étapes importantes ont été franchies ;

1 - la constitution d'un dictionnaire :

- A la suite de tests d'intelligibilité et de qualité, nous avons opté pour un locuteur féminin.

Les diphones ont été extraits d'un enregistrement de mots naturels effectué sur un ton volontairement grave afin :

- de permettre au locuteur de stabiliser sa voix et éviter ainsi des discontinuités spectrales à la concaténation,
- d'obtenir pour ce type de voix, la meilleure définition spectrale.

Au terme de ce travail, nous pensons que si cette solution présente des avantages certains, elle est vraisemblablement à l'origine du caractère "asexué" de la voix synthétique déplorée par de nombreux

.../...

auditeurs lors des tests de qualité.

- L'analyse a été réalisée par l'intermédiaire d'un vocodeur à 14 canaux ; les diphones ont été codés à 4800 eb/s. et stockés en mémoire d'un calculateur.
- Le nombre des diphones retenu s'élève à 1200, ce qui permet de composer n'importe quel message de la langue française.

Les premiers résultats perceptuels enregistrés avec la parole synthétique obtenue par simple concaténation de ces diphones ont conforté notre intention d'introduire un programme de régénération de la prosodie : peu naturelle, la voix n'était pas structurée dans son déroulement temporel.

2 - Le traitement de la prosodie :

- Avec le vocodeur, il n'est pas facile d'accéder directement à l'intensité globale, aussi nous sommes-nous attachée principalement au paramètre temporel (durée segmentale et organisation de l'énoncé) et à la commande de variation de Fo.

- Dans un premier temps, l'analyse d'un corpus de quelques 400 phrases, de complexité syntaxique variée, nous a permis d'isoler des invariants prosodiques et d'induire un certain nombre de règles.

- A la synthèse, nous avons utilisé comme points de repère un double système de marqueurs :

- Inscrits en mémoire sur chaque diphone pour permettre de repérer :

. le début d'une consonne dans les segments [consonne-voyelle] ,

. le début et la fin d'une réalisation vocalique, cela élimine ainsi toute discontuinité de Fo au moment de l'assemblage,

. des zones d'accélération ou de ralentissement de la cadence d'échantillonnage,

. les séquences vocaliques au cours desquelles l'intensité sera diminuée.

.../...

- Insérés, en nombre limité, dans la chaîne phonétique pour signaler les points considérés comme significatifs du point de vue prosodique, soit la fin :

. de chaque mot,

. du groupe nominal qui précède et suit le verbe, des groupes de sens dans ces syntagmes, du groupe verbal.

A chaque mot correspond donc en mémoire, un patron prosodique fonction de son nombre de syllabes et de sa position dans l'énoncé. Ces règles sont simples à mettre en oeuvre et relativement économiques (en taille mémoire).

- L'ensemble - assemblage des diphones et traitement de la prosodie - est réalisé en temps réel.

- Une série de tests a permis de mettre en évidence la bonne intelligibilité globale de cette parole de synthèse ainsi obtenue à 4800 eb/s. et transmise par téléphone. Nous disposons de nombreux éléments relatifs à l'impression qu'elle produit sur des auditeurs ignorants de sa genèse. Une série de tests sont actuellement conduits au Département ETA du CNET - LANNION pour évaluer quantitativement cette parole quand elle est synthétisée à 2400 eb/s. avec un vocodeur à 12 canaux.

- Les résultats obtenus peuvent être améliorés - réduction de la taille de la bibliothèque des diphones et des tableaux prosodiques, prise en compte systématique du paramètre intensité - et utilisés avec d'autres types de synthétiseurs (codage prédictif par exemple).

- L'analyse du corpus a porté sur un locuteur, les règles induites ont été vérifiées pour deux autres locuteurs : elles semblent facilement transposables d'un registre à un autre. Il est cependant certain que leur généralisation sur un plan phonétique exigera une vérification sur un grand nombre de sujets.

BIBLIOGRAPHIE

ABRY, C., BOË, L.J.,& ZURCHER J.F.,

La détection du voisement par les propriétés physiques résultant de l'excitation périodique du conduit vocal : comparaison statistique de trois procédés, 6e Journées d'Etude sur la Parole, GALF TOULOUSE, 228-245, 1975.

ALLEN, J.,

Machine-to-man communication by speech. Part. II : Synthesis of prosodic features of speech by rule. Proc. Sprint joint Comput. Conf. D.C., AFIPS 32, 339 - 344, 1968.

ALLEN, J., & O'SHAUGHNESSY, D.,

A comprehensive model for fundamental frequency generation, IEEE Int. Conf. on Acoustics, speech, and Signal Processing, PHILADELPHIE, 701-704, 1976.

ARNOLD, G.E.,

Morphology and Physiology of the speech organs. Manual of Phonetics, Ed. L. Kaiser, 31-64, 1957.

ATAL, B.S.,& HANAUER, S.L.,

Speech analysis and synthesis by linear prediction of the speech wave, JASA, 50,2, 637-655, 1971.

BAILLY, C.,

Intonation and Syntaxe, Cah. Saussure, 1, 33- 42, 1941.

BARNWELL, T.P.,

An algorythm for segment durations in a reading machine context, Tech. Rep. 479, M.I.T. Res. Lab. Electronics, Cambridge, MA., 1971.

BEAUVIALA, J.P., CARRE, R., LANCIA, R., PAILLE, J.,& VERMEILLE, H.,
Recherches sur l'analyse et la synthèse paramétriques de la parole entreprises à l'ENSERG, R. Acoust.,3/4 , 227-229, 1968.

BENGUEREL, A.P.,
Duration of French vowels in unemphatic stress,Language and speech, 14, 4, 383-391, 1971.

.../...

BLACK, J.W.,
Natural Frequency, Duration and Intensity of vewel in Reading, J. Speech, Dis., 216-221, 1949.

BLOOMFIELD, L.,
Language, Holt, Rinehart et Winston, NEW YORK, 1933
Trad. française - Le Langage - Payot , PARIS, 1970.

BOË, L.J.,
Etude de l'interaction source laryngienne - conduit vocal dans la détermination des caractéristiques intrinsèques des consonnes du français. Mesure de la durée. Bulletin de l'Institut de Phonétique de GRENOBLE, II, 1-24, 1973.

BOË, L.J.,
Les faits prosodiques et la fréquence laryngienne. Approche théorique et expérimentale. Bulletin d'audiophonologie, 2, 3-24, 1973.

BOË, L.J.,
Etude des vibrations des cordes vocales dans la parole : Méthodes, résultats et applications, Revue d'acoustique 2/4, 37, 105-112, 1976.

BOË, L.J., CONTINI, M.,&RAKOTOFIRINGA, M.,
Etude statistique de la fréquence laryngienne, Phonetica, 32, 1-23, 1975.

BOË, L.J., & CONTINI, M.,
Synthèse paramétrique de la phrase interrogative en français, 7e Journées d'Etude sur la parole, GALF, NANCY, 130-144, 1976.

BOË, L.J., & LARREUR, D.,
Etude de l'influence des variations de la fréquence laryngienne sur l'intelligibilité et la qualité des consonnes sonores générées par vocodeur - Bulletin de l'Institut de Phonétique de GRENOBLE, II, 103-126, 1973.

BOË, L.J., & LARREUR D.,
Synthesis by rule of Enonciative Sentence in French. Preliminary study. Speech Communication Seminar, STOCKHOLM, 1974.

BOË, L.J., & LARREUR, D.,
Les caractéristiques intrinsèques de la fréquence laryngienne : production, réalisation et perception, 5e Journées d'Etude sur la Parole, GALF, Orsay, I, 19-28, 1974.

BOOMER, D.S.,
Hesitation and grammatical encoding, Language and Speech, 8, 148-158, 1965.

.../...

BOOMER , D.S., & DITTMAN, A.,

Hesitation pauses and juncture pauses in Speech, Language and Speech, 5, 215-226, 1962.

BURON, R.,

Audio Unit Connected to a digital computer, 2ème Congrès Intern. des Techniques des Télécommunications, DIV. I-I/1, 1-13, MADRID 1965.

BURON, R.,

Le traitement de la parole et les ordinateurs . Revue d'Acoustique, 3/4, 186-190, 1968.

BUTCHER, A.,

La perception des pauses. 4e Journées d'Etude sur la Parole, GALF, BRUXELLES, 371-382, 1973.

CARCAUD, M., COURBON, J.L., GENIN, J.,&LUCAS, J.P.,

A hardware vocal source simulator, IEEE International conference on acoustics, speech and signal processing, Philadelphie,51-54, 1976.

CARRE, R.,

Contribution aux études sur l'analyse et la synthèse de la parole ; rôle et importance des formants. Thèse d'Etat, Université Scientifique et Médicale de GRENOBLE, juin 1971.

CARREL, J., & TIFFANY, W.,

Phonetics : Theory and Application to speech improvement. NEW YORK : Mc. Graw-Hill, 1960.

CARTIER, M., GENIN,J.,&LORAND, P.,

Synthèse de la parole : une unité de réponse vocale ; Echo des Recherches, 65, 43-51, 1971.

CARTIER, M., & GRAILLOT, P.,

Reduction and reconstitution of spectral data, Speech Communication Seminar, STOCKHOLM, 1974.

CHALARON, M.L.,

Contribution à l'étude des faits prosodiques dans les énoncés à caractère exclamatif. Bulletin de l'Institut de Phonétique de GRENOBLE, I, 77-91, 1972.

CHEVALIER, J.C., BLANCHE-BENVENISTE, C ., ARRIVE, M.,&PEYTARD, J.,

Grammaire Larousse du français contemporain, Larousse, PARIS , 1970.

CHIBA, T.,&KAJIYAMA, M.,

The vowel - its nature and structure, Tokyo - Kaiseikan Publishing Co Ldt, Tokyo, 1941.

.../...

CHISTOVICH, L.A.,

Variation of the fundamental voice pitch as a discriminatory cue for consonants, Soviet Physics - Acoustics, 14, 372-378, 1969.

CHOPPY, C., LIENARD, J.S.&TEIL, D.,

Un algorithme de prosodie automatique sans analyse syntaxique. 6e Journées d'Etudes sur la parole, GALF, TOULOUSE, 387-395, 1975.

COKER, C.H.,

Synthesis by rule from articulatory parameters, Proc. Conf. on Speech Comm. and Processing, MIT, CAMBRIDGE, 55-63, Mass., 1967.

CONTINI, M., & BOE, L.J.,

Contribution à l'étude quantitative de l'évolution de la fréquence laryngienne dans la phrase énonciative en français. Bulletin de l'Institut de Phonétique de GRENOBLE, II, 77-92, 1973.

CONTINI, M., & BOE, L.J.,

Etude quantitative de l'intonation en français ; premiers résultats, 8 th Int. Congr. Phonet. Sci, Leeds, 1975.

COOPER, F.S., DELATTRE, P., LIBERMAN, A.M., BORST, j.M., & GERSTMAN, L.J.,

Some experiments on the perception of synthetic speech. JASA. 24, 597-606, 1952.

CRANDALL, I.B.,

The sounds of speech, B.S.T.J., 4, 586-626, 1925.

DAVID, E.E., SCHROEDER, M.R., LOGAN, B.F., &PRESTIGIACOMO, A.J.,

Voice excited vocoders for practical speech bandwidth reduction, Int. Symp. of Inform. Theory Proc., BRUSSELS, 1962.

DELATTRE, P.,

Les dix intonations de base du français, French Review, 40, 1, 1-14, 1966.

DENES, P.,

A preliminary investigation of certains aspects of intonation, Language and speech, 2, 106-122, 1959.

DI CRISTO, A.,

Recherches sur la structuration prosodique de la phrase française (essai d'analyse phonosyntaxique), 6e Journées d'Etude sur la Parole - GALF - TOULOUSE, 94-116, 1975.

DIXON, N.R., &MAXEY, H.D.,

Terminal analog synthesis of continuous speech using the diphone method of segment assembly. IEEE - Transactions on Audio and Electroacoustics, AU-16, 1, 40-50, 1968.

.../...

DUDLEY, H.,

Remaking speech, JASA, 11, 169-177, 1939.

DUDLEY, H., & TARNOCZY, T.H.,

The speaking machine of Wolfgang von Kempelen, JASA, 22, 2, 151-166, 1950

DUNN, H.K.,

The calculation of vowel resonances and an electrical vocal tract, JASA, 22, 740, 1950.

EL MALLAWANY, I.,

Contributions aux recherches sur la communication parlée, thèse de Docteur Ingénieur, GRENOBLE, 1975.

EMERARD, F., & LARREUR, D.,

Synthèse par diphones et traitement de la prosodie, Bulletin de l'Institut de Phonétique, GRENOBLE, IV, 103-116, 1975.

ENDRES, W.,

The transitional sounds of the German Language as link elements for a speech synthesis, Acustica, 26, 33-36, 1972.

ENDRES, W.,&GROSSMANN, E.,

Manipulation of the time functions of vowels for reducing the number of elements needed for speech synthesis, Speech communication Seminar, STOCKHOLM, 267-275, 1974.

ESTES, S.E., KERBY, H.R., MAXEY, H.D.& WALKER, R.M.,

Speech synthesis from stored data. IBM Res. and Develop. Journal,8, 2-12, 1964.

FAIRBANKS, G.,

Voice and articulation Drillbook, Harper and Row, NEW YORK, 1960

FANT, G.,

Transmission properties of the vocal tract, MIT Acoustic Lab., Q.P.R., july, 20-23, 1950.

FANT, G.,

Acoustic Analysis and Synthesis of speech with applications to swedish. Ericson Technics, 1, 1-108, 1959.

FANT, G.,

Acoustic theory of speech production. Mouton and Co, the Hague, 1960.

FANT, G., RISBERG,A., & STEVENS, K.N.,

Evaluation of various analysis - synthesis speech systems, JASA, 35,5, 804(A), 1963.

.../...

FERRIEU, G., & PERSON, J.M.,

Etude d'un vocoder : le projet A.S.P.I.C., Etude CRL 2013/ETA, 1968.

FLANAGAN, J.L.,

Speech analysis, synthesis and Perception, Springer Verlag, BERLIN, 2ème édit., 1972.

FLANAGAN, J.L.,

Computers that-talk-and-listen : Man-machine Communication by voice ; Proceedings of the IEEE, 64, 4, 405-415, 1976.

FAURE, G.,

La description phonologique des systèmes prosodiques, Proc. 6 th Int. Congr. Phon. Sci. PRAGUE, 1967.

FOUCHE, P.,

Traité de prononciation française, Librairie Klincksieck, PARIS, 1959.

GENIN, J.,

Les études de synthèse de parole au CNET, l'Echo des Recherches, 85, 40-49, juillet 1976.

GLEASON, H.A.,

An introduction to descriptive linguistics, NEW YORK, Holt Rinehart and Winston, 1967.

GOLDMAN-EISLER, F.,

The distribution of pauses duration in speech, Language and speech, 4,4, 232 - 237, 1961.

GOLDMAN-EISLER, F.,

Psycholinguistics : Experiments in Spontaneous speech, NEW YORK : Academic, 1968.

GRAILLOT, P.,

Projection, compression et reconstitution de données spectrales de parole, Bulletin de l'Institut de Phonétique de GRENOBLE, III, 52-71, 1974.

GRAILLOT, P.,

Système pour la transmission à très faible débit, Recherches Acoustiques, CNET LANNION, II, 83-87, 1975.

GRAMMONT, M.,

Traité pratique de prononciation française, Delagrave, PARIS, 1947.

.../...

GROSJEAN, F. , & DESCHAMPS , A. ,

Analyse des variables temporelles du français spontané, Phonetica, 26, 129, 1972.

GUIRAUD, P.,

Problèmes et Méthodes de la statistique linguistique, PUF, 1960.

HADDING-KOCH, K., & STUDDERT-KENNEDY, M.,

Intonation contours evaluated by American and Swedish Test Subjects, Proc. 5th Int. Congr. Phonet. Sci. Münster, 326-331, 1964.

HALLIDAY, M.A.K.,

Catégories of the theory grammar, Word, 17, 241-292, 1961.

HARRIS, C.M.,

A study of the building - blocks in speech, JASA, 25,5, 962-969, 1953.

HARRIS, M.S., & UMEDA, N.,

Effect of speaking mode on temporal factors in speech : vowel duration, JASA, 56, 3, 1016-1018, 1974.

HARRIS, Z.S.,

Structural Linguistics, The University of Chicago Press, Chicago, 1951.

HATON, J.P., & LAMOTTE, M.,

Etude statistique des phonèmes et diphonèmes dans le français parlé, Revue d'Acoustique, 16, 258-262, 1971.

HIKI, S.,

On the control rules of voice pitch for sentences speech synthesis, JASP 22, 364-367, 1966.

HIKI, S., & OIZUMI, J.,

Controlling rules of prosodic features for continuous speech synthesis, Proc. Conf. on speech Comm. and processing, AFCRL paper A-4, Bedford, Mass., 1967.

HINARD, A.,

Précis de grammaire française, Edit. Magnard, Coll. J. Le Lay, 1969.

HOUSE, A.S.,

On vowel duration in English, JASA, 33, 1174- 1178, 1961.

.../...

HOUSE, A.S., & FAIRBANKS, G.,

The influence of consonant environment upon the secondary acoustical characteristics of vowels - JASA, 25, 105-113, 1953.

KELLY, J.L., & GERSTMAN, L.J.,

An artificial talker driven from a phonetic input, JASA, 33. 6, 835(A), 1961.

KIM, K.,

Fo variations according to consonantal environments. Monthly Internal Memorandum, Phonology Laboratory Univ. Calif. Berkeley, 33-43, 1968.

KLATT , D.H.,

A generative theory of segmental duration in English, 82nd Meeting of the Acoust. Soc. Amer. Denver, Co, JASA, 51, 101 (A), 1971.

KLATT, D.H.,

Discrimination of fundamental frequency contours in synthetic speech : Implication for models of pitch perception, JASA, 53, 8-16, 1973.

KLATT, D.H.,

Structure of a phonological rule component for a synthesis -by-rule program, IEEE Trans. on acoustics, speech, and signal processing, 24, 5, 391-398, 1976.

KLATT, D.H.,

Interaction between two factors that influence vowel duration , JASA, 54,4, 1102 - 1104, 1973.

KOENIG, W., DUNN, H.K., & LACY, L.Y.,

Sound spectrograph, JASA, 18, 1, 19-49, 1946.

KRATZENSTEIN, C.G.,

Cité par FLANAGAN,J.L.,1972.

KUPFMULLER, K., & WARNS, O.,

Sprachsynthese aus Lauten, Nachrichtentechn. Fachber, 28-31, 1956, cité par ENDRES, W.,1974.

LARREUR, D., & BOE, L.J.,

Les caractéristiques intrinsèques des consonnes voisées du français dans la parole continue. Bulletin de l'Institut de Phonétique de GRENOBLE, II, 25-29, 1973.

.../...

LARREUR, D., & BOE, L.J.,

Synthèse paramétrique de la phrase énonciative en français. 5ème Journées d'Etude sur la Parole, GALF, ORSAY, II, 1974.

LAWRENCE, W.,

The synthesis of speech from signals which have a low bit information rate, Communicationtheory, Butterworth Scientific publications, Edit. W. Jackson, LONDON, 460-469, 1953.

LEHISTE, I.,

Suprasegmentals, MIT Press, CAMBRIDGE, MASS, 1970.

LEHISTE, I.,

Timing of utterances and Linguistic boundaries, JASA 51, 2018, 1972.

LEHISTE, I., & PETERSON, G.E.,

Some basic considerations in the analysis of intonation, JASA, 33, 419-425, 1961.

LEIPP, E., CASTELLENGO, M.,&LIENARD, J.S.,

La synthèse de la parole à partir de digrammes phonétiques, Rep. 6th Int. Congr. Acoustics TOKYO, Paper C-5-6, 1968.

LEON, P.R.,

Où en sont les études sur l'intonation, 7ème Congrès International des Sciences Phonétiques, MONTREAL, 113-156, 1971.

LEON, P.R., & MARTIN, Ph.,

Prolégomènes à l'étude des structures intonatives, Studia Phonetica, 2, Didier, Montréal, Paris, Bruxelles, 1969.

LEVITT, H., & RABINER, L.R.,

Analysis of Fundamental pitch contours in speech, JASA 49, 569-582, 1971.

LIBERMAN, A.M.,

Some results of research on speech perception, JASA, 29, 1, 117 - 123, 1957.

LIBERMAN, A.M., COOPER, F.S., SHANKWEILLER, D.P.& STUDDERT KENNEDY, M.,

Perception of the speech code, Status Report on speech research, S.R.9,1-67, 1967.

LIEBERMAN, Ph.,

On the acoustic basis of the perception of intonation by linguists, word, 21, 40-54, 1965.

.../...

LIEBERMAN, Ph.,

Intonation, Perception and Language, MIT, CAMBRIDGE, 1967.

LIENARD, J.S.,

Le dictionnaire des éléments phonétiques et ses applications à la linguistique, Bulletin du GAM, 22 bis, 1966.

LIENARD, J.S.,

La synthèse de la parole, historique et réalisations actuelles, Revue d'Acoustique, 11, 204-213, 1970.

LIENARD, J.S.,

Analyse, Synthèse et Reconnaissance automatique de la parole, Doctorat d'Etat, PARIS VI, 1972.

LIENARD, J.S., & TEIL, D.,

Les éléments phonétiques et la traduction automatique du message écrit en message parlé. Automatisme, Tome XV, 10, 505-513, 1970.

LINDBLOM, B.,

Les mécanismes des contrôles moteurs, Bulletin de l'Institut de Phonétique de GRENOBLE, III, 1-21, 1974.

LINDBLOM, B., & RAPP, K.,

Some temporal regularities of spoken swedish, Papers from the Institute of Linguictics, University of STOCKHOLM, 21, 1973.

LISKER, L., & ABRAMSON, A.S.,

A cross-language study of voicing in initial stops : Acoustical measurements. Word, 20, 384-422, 1964.

LUCCI, V.,

Etude phonostylistique du rythme et de la variabilité de la longueur en français parlé et français lu. Bulletin de l'Institut de Phonétique de GRENOBLE, II, 139-161, 1973.

LUCCI, V.,

Rythme et longueur du message parlé ; la conversation, Bulletin de l'Institut de Phonétique de GRENOBLE, III, 139-152, 1974.

LYBERG, B., & LINDBLOM, B.,

Computational models of swedish Prosody, 8th Int. Congr. of Phonetic Sciences, Leeds, 1975.

MAC CLAY, H., & OSGOOD, C.,

Hesitation phenomena in spontaneous English speech, Word 15, 19-44, 1959.

.../...

MAC CLEAN, M.D., & TIFFANY, W.R.,

The acoustic parameters of stress in relation to syllable position, speech loudness and rate, Language and speech, 283-291, 1973.

MAEDA, S.,

A characterization of American English Intonation, Thèse, MIT , 1976.

MALMBERG, B.,

Les nouvelles tendances de la linguistique, P.U.F. PARIS, 1962.

MARTIN, J., & STRANGE, W.,

Determinants of hesitations in spontaneous speech, J. exp. Psychol, 76, 474-479, 1968.

MARTIN, Ph.,

Les problèmes de l'intonation : Recherches et Méthodes, Langue française, 19, 4 - 32, 1973.

MARTIN, Ph.,

Intonation et reconnaissance automatique de la structure syntaxique, 6e Journées d'Etude sur la parole, GALF, TOULOUSE, 51-62, 1975.

MARTIN, Ph.,

Eléments pour une théorie de l'intonation. Rapport d'activités de l'Institut de Phonétique (BRUXELLES) n° 9/1, 97-126, 1974-75.

MARTIN, Ph.,

Une grammaire de l'intonation de la phrase. Rapport de l'Institut de Phonétique (BRUXELLES) n° 9/2, 1975.

MARTIN, Ph.,

Analyse phonologique de la phrase française, Linguistics, 146, 35-69, 1975.

MARTIN, Ph .,

Synthèse par règles de l'intonation de la phrase. 7e Journées d'Etude sur la parole, GALF, NANCY, 207-213, 1976.

MARTINET, A.,

Eléments de linguistique générale, Colin, PARIS, 1967.

MARTINET, A.,

A functional view of language. London, Oxford University Press , 1962 ; Traduction française : Langue et fonction, Denoël, PARIS, 1971.

MATTINGLY, I.G.,

Synthesis by rule of prosodic features, Language and speech, 9, 1-13, 1966.

.../...

MATTINGLY, I.G.,

Synthesis by rule of General American English, S.R. 7/8, 4.1, 1966.

MEO, A.R., MEZZALAMA, M., RIVOIRA, S., & RUSCONI, E.,

A general system for synthesizing speech in any Language, 8 th Int. Congr. Acoustics, London 1, 293, 1974.

MERMELSTEIN, P.,

Determination of the vocal tract shape from measured formant frequencies, JASA, 41, 1283, 1967.

METTAS, O.,

Etude sur les facteurs ectosémantiques de l'intonation en français, Travaux de Linguistique et de Literrature, STRASBOURG, 1, 143-154, 1963.

METTAS, O.,

Etude sur l'intonation en français, Travaux de Linguistique et de Littérature, STRASBOURG, 2, 1, 99-105, 1964.

METTAS, O.,

Aperçu historique sur les appareils de synthèse de la parole, Travaux de Linguistique et de Litterature, STRASBOURG, 1, 185-200, 1965.

MEYER-EPPLER, W.,

Realization of prosodic features in whispered speech, JASA, 29, 104-106, 1957.

MOHR, B.,

Intrinsic fundamental frequencies IV : voiced consonants. Monthly Internal Memorandum, Phonology Laboratory, Univ. of Calif. Berkeley, 17-22, 1968.

MOHR, B.,

Intrinsic variations in the speech signals, Phonetica, 23, 65-93, 1971.

MÜLLER, J.,

Von der Stimme und Sprache. Handbuch ur Physiologia des Menchen. J. Holscher. COBLENZ, II, 4, 133-245, 1837.

NEMETH, A.,

La synthèse par règles de la parole, 1ère Journée d'Etudes sur la parole, GALF, GRENOBLE, 53-61, 1970.

.../...

NEMETH, A.,

Synthèse par règles de la parole à l'aide d'un vocodeur programmé avec sortie en modulation par impulsions codées, 7th Int. Congr. Acoust., Paper 24 C2, BUDAPEST, 1971.

ÖHMAN, S.,

Word and sentence intonation : a quantitative model. Quart. Progr. Status Rep. 2-3, Speech Transmission Lab. STOCKHOLM, 20-55, 1967.

ÖHMAN, S.,

A model of word and sentence intonation, STL, QPSR 2-3, 6-11, 1968.

ÖHMAN, S., & LINDQVIST, J.,

Analysis-by-synthesis of prosodic pitch contours.Quaterly Progress and Status Report, Speech Transmission Laboratory, R.I.T., STOCKHOLM, 4, 1-6, 1965.

OLIVE, J.P.,

Fundamental frequency rules for the synthesis of simple declarative English sentences, JASA, 53, 2, 476-482, 1975.

OLIVE, J.P., & NAKATANI, L.H.,

Rule-synthesis of speech by word concatenation : A first step. JASA, 55, 3, 660-666, 1974.

OLIVE, J.P., & NAKATANI, L.H.,

Speech synthesis by rule. Speech communication seminar, STOCKHOLM, 255-260, 1974.

OLLER, D.K.,

The effect of position in utterances on speech segment duration in English, JASA, 54, 5, 1235-1247, 1973.

PAILLE, J.,

Contribution aux études sur la synthèse paramétrique de la parole ; synthétiseur à formants ; Analogue de la source vocale. Thèse d'Etat, Université Scientifique et Médicale de GRENOBLE, 1971.

PETERSON, G.E., & BARNEY, H.L.,

Control methods used in a study of the vowels, JASA, 24, 175-184, 1951.

PETERSON G.E., & LEHISTE, I.,

Duration of syllable nuclei in English. JASA, 32, 693-703, 1960.

PETERSONG.E., WANG W.S.Y., and SIVERSTEN, E.,

Segmentations techniques in speech synthesis, JASA, 30, 8, 739-749, 1958.

.../...

PIKE, K.L.,

The intonation of American English, University of Michigan Publications, Linguistics, I, Univ. Mich. Press, Ann Arbor, 1945.

PONCIN, J.,

Etude d'un système de synthèse de messages vocaux, Annales des Télécommunications, 25, 11/12, 405-418, 1970.

QUINIO, J., & TEIL, D.,

La synthèse de la parole par ordinateur à partir de digrammes phonétiques. Revue d'Acoustique, 9,28-32, 1970.

QUINTON, P., VIVES,R., & GRESSER, J.Y.,

Dialogue avec un robot. 6e journées d'études sur la parole. GALF TOULOUSE, 412-421, 1975.

RABINER, L.R.,

A model for synthesizing speech by rule, IEEE Trans. on Audio and Electr. AU. 17, 7-13, 1969.

RABINER, L.R., LEVITT, H., & ROSENBERG, A.E.,

Rules for synthetizing prosodic features of speech, Preliminary investigation, JASA, 44, 1, 390,(A), 1968.

RABINER, L.R., LEVITT, H., & ROSENBERG, A.E.,

Investigation of stress pattern for speech synthesis by rule, JASA, 45, 92-101, 1969.

ROSEN, G.S.,

Dynamic Analog speech synthetizer, JASA, 30, 201-209, 1958.

ROSSI, M.,

L'intonation prédicative dans les phrases transformées par permutation, Linguistics, 103, 64-94, 1973.

ROSSI M., & CHAFCOULOFF, M.,

Les niveaux intonatifs, travaux de l'Institut de Phonétique d'AIX, 1, 167-176, 1972.

RUDER, F.K., & JENSEN, P.J.

Fluent and hesitation pauses as a function of syntactic complexity, Journal of speech and Hearing Res., 15, 1, 49-60, 1972.

SERNICLAES, W.

Perceptual processing of acoustic correlates of the voicing features. Preprints of the SCS, Speech Transmission Laboratory, STOCKHOLM, 87-94, 1974.

.../...

SERNICLAES, W.,

Prévoisement et délai d'établissement du voisement ; deux indices indépendants pour la perception des occlusives. Rapport d'activités de l'Institut de Phonétique (BRUXELLES) 10/1, 84-104, 1976.

SERNICLAES, W., & BEJSTER, P.,

Influence du contexte vocalique sur la perception du voisement des occlusives, Rapport d'Activités de l'Institut de Phonétique (BRUXELLES) 8/1-2, 101-108, 1974.

SIGURD, B., & LINDBLOM, B.,

Maximum rate and minimum duration of repeated syllables. Papers from the Institute of Linguistics, University of STOCKHOLM, Publication 3, 1971.

SORON, H.I., & LIEBERMAN, P.,

Some measurements of the glottal area waveform, JASA, 35, 1876 (A), 1963.

SPANG-THOMSEN, B.,

L'accent en français moderne. Orbis Litterarum, Supplementum 3, 181, 1963.

STEVENS, K.N., FUJISAKI, S., & HOUSE, A.S.,

Analysis of vowel spectra, M.I.T., Res. Lab. of Electronics, Quart. Prog. Rep., 60, 177, 1961.

TEIL, D.,

Etude de génération synthétique de parole à l'aide d'un ordinateur. Thèse CNAM, PARIS, 1969.

TEIL, D., CASTELLENGO, M., & SAPALY, J.,

L'unité à réponse vocale Icophone V, 5ème Journées d'Etude sur la parole, GALF, ORSAY, 1, 89-94, 1974.

TERRY, R.M.,

Contemporary French interrogative structure, Editions COSMOS, MONTREAL, 1970.

TRAGER, G.L., et SMITH, H.L.,

Outline of English structure, Studies in Linguictics, 3, Battenburg, Press, Norman, Oklahoma, 1951.

TUBACH, J.P.,

Etude des contraintes statistiques des groupements phonématiques. Colloque : l'informatique au service de l'Homme. GRENOBLE, 31-48, 1969.

.../...

TWADDELL, W.F.,

Phonemes and Allophones in speech analysis, JASA, 24, 6, 607-611, 1952.

UMEDA, N.,

Vowel duration in polysyllabic words in American English, JASA, 52, 133 (A), 1972.

UMEDA, N.,

Vowel duration in American English, JASA, 58,2, 434-445, 1975.

UMEDA, N.,

Linguistic rules for Text-to-Speech Synthesis, Proceedings of the IEEE, 64, 4, 443-446, 1976.

UMEDA, N., MATSUI, E., SUZUKI, T., & OMURA, H.,

Synthesis of Fairy Tales using an Analog Vocal Tract, Proc. 6 th Int. Congr. Acoust. TOKYO, paper B-5-3, 1968.

UMEDA, N., & COKER, C.H.,

Allophonic variation in American English. Journal of phonetics, 2, 1, 1-5, 1974.

VAISSIERE, J.,

Contribution à la synthèse par règles du français, Thèse 3e cycle, GRENOBLE, 1971.

VAISSIERE, J.,

On French prosody, Quaterly Progress Report, Res. Lab. of Electronics, MIT, 1974.

VAISSIERE, J.,

Further note on French prosody, Quaterly Progress Report, Res. Lab. of Electronics, MIT, 1975.

VAISSIERE, J.,

Caractérisation des variations du fondamental dans les phrases françaises, 6e Journées d'Etude sur la parole, GALF, TOULOUSE, 39-50,1975.

VAISSIERE, J.,

Réflexions après un stage au centre de renseignements de PARIS BRUNE : positions d'opérateur, dialogue dirigé et difficultés de la recherche. Rapport RP/CEI/CSE/7, 1976.

VON KEMPELEN, W.,

Le mécanisme de la parole suivi d'une description d'une machine parlante, J.V. Degen, VIENNE, 1792.

WAJSKOP, M.,

La perception de la parole : orientations et perspectives, 1e Journées d'étude sur la parole, GALF, GRENOBLE, 1-15, 1970.

WANG, W.S.Y., & PETERSON, G.E.,

Segment Inventory for speech synthesis, JASA, 30,8,743-746, 1958.

.../...

WARNS, O.,

Über die Synthese von Sprache aus Lauten und Lautkombinationen, Diss. Techn. Hochschule Darmstadt, 1957 (cité par ENDRES et GROSSMANN, 1974).

WELLS, R.S., REVIEW of PIKE, K.L.,

The intonation of American English in Language, 23, 255-273, 1945.

YOUNG

Natural Philosophy, 1845, (cité par DUDLEY et TARNOCZY, 1950)

ZURCHER, J.F.,

Dispositif de détection et de mesure du fondamental de la parole humaine, Analyse et synthèse de la parole, rapport d'activités ETA/CNET, I, 7-15, 1972-73.

ZURCHER, J.F., CARTIER, M., & BOE, L.J.,

Détection et mesure du fondamental, 6e journées d'étude sur la parole, GALF, TOULOUSE, 12-21, 1975.

ANNEXES

TRANSCRIPTION ORTHOGRAPHIQUE-PHONETIQUE.

1/-Texte orthographique:

Monsieur SEGUIN n'avait jamais eu de bonheur avec ses chèvres.

Il les perdait toutes de la même façon.

Un matin, elles cassaient leur corde, s'en allaient dans la montagne, et là-haut, le loup les mangeait.

Ni les caresses de leur maître, ni la peur du loup, rien ne les retenait !

C'était des chèvres indépendantes, voulant à tout prix le grand air et la liberté.

Le brave Monsieur SEGUIN, qui ne comprenait rien au caractère de ses bêtes était consterné.

Les chèvres s'ennuient chez moi ! je n'arriverai jamais à en garder une ? Pourtant, il ne se décourageait pas, et, après avoir perdu six chèvres de la même façon, il en attacha une septième .

Seulement, cette fois, il eut soin de la prendre toute jeune, pour qu'elle s'habitua mieux à demeurer chez lui.

Comme elle était jolie la petite chèvre de Monsieur SEGUIN ! comme elle était jolie avec ses yeux doux, sa barbiche de sous-officier, ses sabots noirs et ses longs poils blancs qui lui faisaient une houppelande !.

Monsieur SEGUIN avait derrière sa maison un clos entouré d'aubépines ; c'est là qu'il avait mis la nouvelle pensionnaire. Il l'avait attachée à un pieu au plus bel endroit du pré, en ayant soin de lui laisser beaucoup de corde, et de temps à autre il venait voir si elle était bien.

Alors, Blanquette, tu te plais chez moi ?

..../...

La petite chèvre broutait l'herbe de si bon coeur, que le brave Monsieur SEGUIN était ravi.

Enfin, en voilà une qui ne s'ennuiera pas !

Monsieur SEGUIN se trompait, sa chèvre s'ennuya.

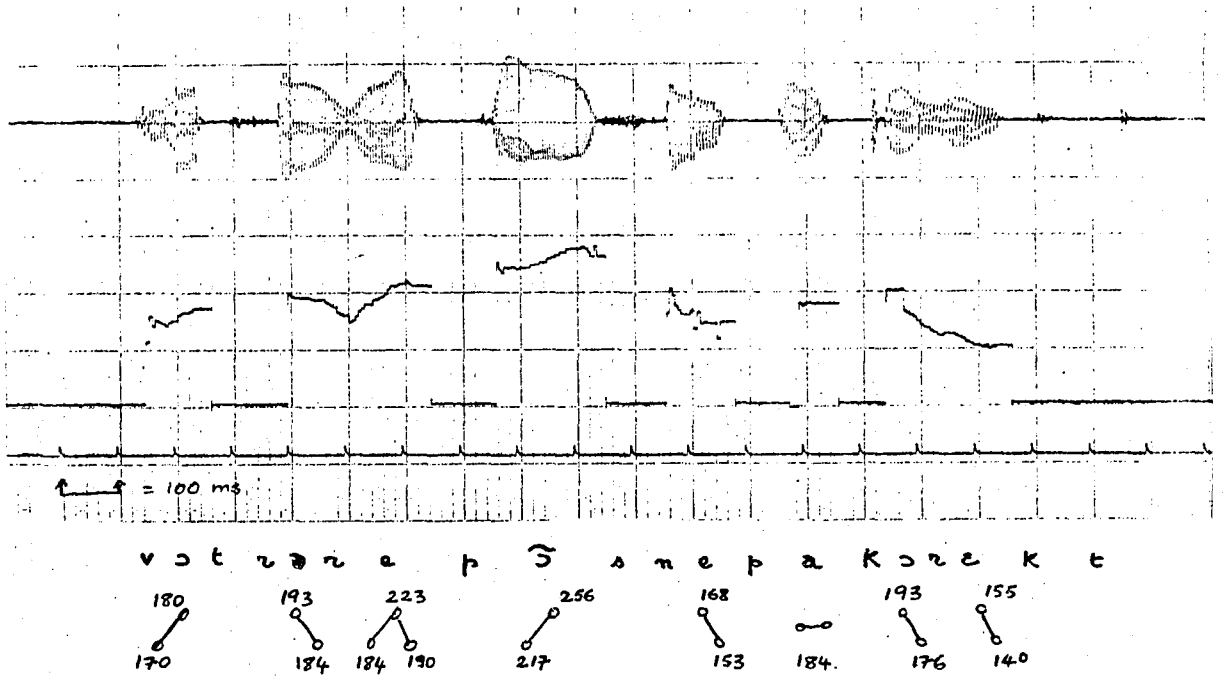
Un jour, elle se dit en regardant la montagne :

"Comme on doit être bien là-haut ! Quel plaisir de gambader dans la bruyère, sans cette maudite longe qui vous écorche le cou ! C'est bon pour l'âne ou pour le boeuf de brouter dans un clos ! les chèvres, il leur faut du large !".

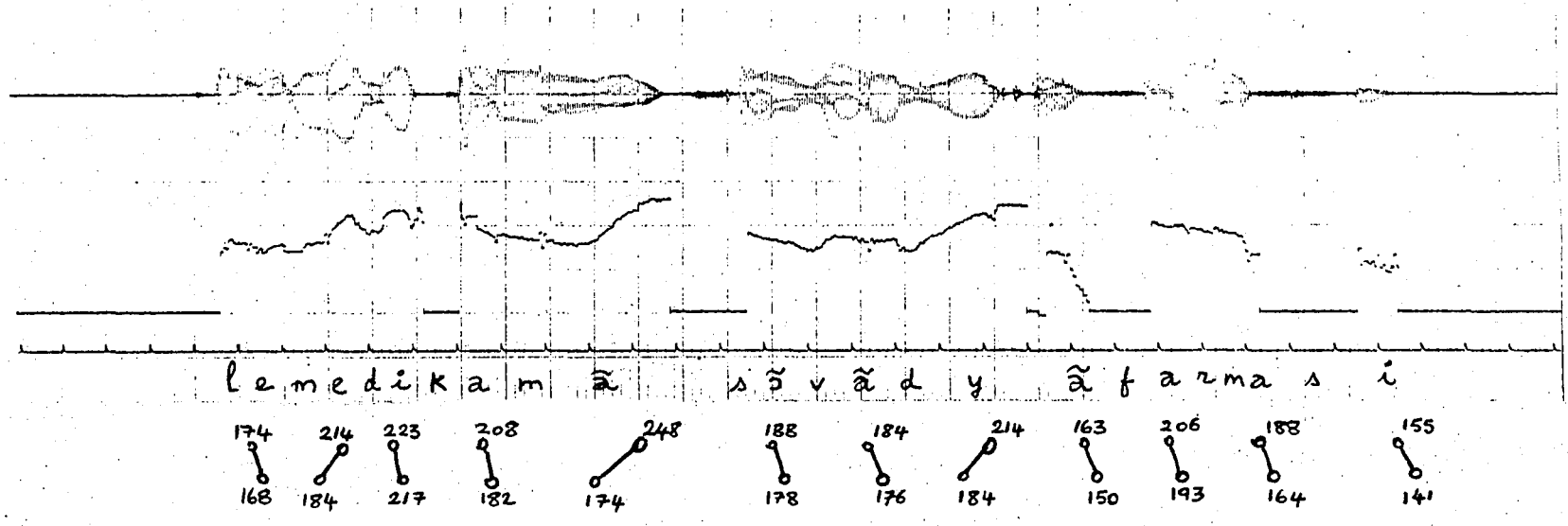
2/ - Texte phonétique :

Positionnement direct de tous les marqueurs prosodiques.

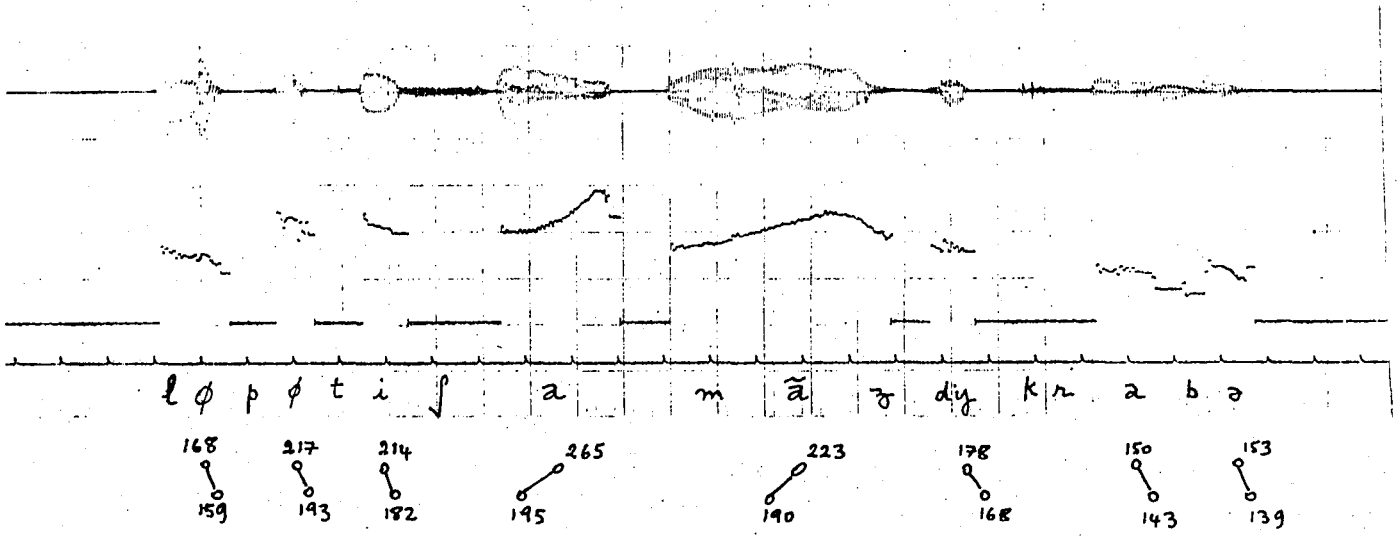
```
 1   MEU S YEU- SEU GIN= N A VAI- J A MAI- U@ DEU- B O NOE R* A VAI K- SAI
 2  -CHAI V R E.
 3   I L- LAI- PAI R DAI- TOU T$ DEU- L A- MAI M- F A SON.
 4  UN- M A TIN,AI L- K A SAI@ LOE R- K O R D E, SAN- N A LAI@ DAN- L A- M
 5  ON T A N Y,EI- L A-AU, LEU- LOU= LAI- MAN JAI.
 6   N I- LAI- K A RAI S@ DEU- LOE R- MAI T R E, N I- L A- POE R@ D U- LOU
 7  , R YIN= NEU- LAI- REU TEU NAI.
 8   SEI TAI@ DAI-CHAI V R E-IN DEI PAN DAN T, VOU LAN- A- TOU- P R I@ LEU
 9  - G RAN- TAI R*EI- L A- L I BAI R TEI.
10   LEU- B R A V- MEU S YEU- SEU GIN, K I- NEU- KON P REU NAI- R YIN$AU-
11  K A R A K TAI R@ DEU- SAI- BAI T*EI TAI- KON S TAI R NEI.
12   LAI-CHAI V R E= SAN N U I$CHEI- M W A; JEU- N A R I V RAI- J A MAI@ A
13  -AN- G A R DEI- U N? POU R TAN, I L- NEU- SEU- DEI KOU R A
14  JAI- P A,EI, A P RAI- Z A V W A R- PAI R D U$ S I-CHAI V R E= DEU- L A
15  - MAI M- F A SON, I L-AN- N ACH T A$ U N- SAI T YAI M.
16   SOE L MAN, SAI T- F W A, I L- U- S WIN@ DEU- L A- P RAN D R E- TOU T-
17   JOE N, POU R- KAI L- S A B I T U A- M YEU$ A- DEU MOE REI-
18  CHEI- L U I.
19   K O M-AI L-EI TAI- J O L I= L A- PEU T I T-CHAI V R E* DEU- MEU S YEU
20  - SEU GIN; K O M-AI L-EI TAI- J O L I= A VAI K- SAI- Z YEU-
21   DOU, S A- B A R B ICH- DEU- SOU- Z O F I S YEI, SAI- S A B O- N W A R
22  =EI- SAI- LON- P W A L- B LAN* K I- L U I- FEU ZAI$ U N-OU P
23   LAN D.
24   MEU S YEU- SEU GIN= A VAI@ DAI R YAI R- S A- MAI ZON*UN- K L O*AN TOU
25   REI- DAU BEI P I N.
26   SAI- L A, K I L- A VAI- M I$·L A- NOU VAI L- PAN S Y O NAI R.
27   I L- L A VAI- T A T ACHEI@ A-UN- P YEU*AU- P L U- BAI L-AN D R W A@ D
28   U- P REI,AN- NAI YAN- S WIN$ DEU- L U I- LAI SEI@ BAU KOU-
29   DEU- K O R D E,EI- DEU- TAN- A-AU T R E, I L- VEU NAI- V W A R$ S I-A
30  I L-EI TAI- B YIN.
31   A L O R- B LAN KAI T, T U- TEU- P LAI&CHEI- M W A?.
32   L A- PEU T I T-CHAI V R E= B ROU TAI$ LAI R B E= DEU- S I- BON- KOE R
33  ; KEU- LEU- B R A V- MEU S YEU- SEU GIN=EI TAI- R A V I.
34  AN FIN,AN- V W A L A- U N, K I- NEU- SAN N U I R A- P A.
35   MEU S YEU- SEU GIN= SEU- T RON PAI, S A-CHAI V R E= SAN N U I Y A.
36  UN- JOU R,AI L- SEU- D I@AN- REU G A R DAN- L A- MON T A N Y.
37   K O M-ON- D W A-AI T R E- B YIN- L A-AU%.
38   KAI L- P LAI Z I R* DEU- GAN B A DEI= DAN- L A- B R U I YAI R, SAN- S
39  AI T- MAU D I T- LON J* K I- VOU- ZEI K O RCH E$ LEU- KOU.
40   SAI- BON@ POU R- L A N*OU- POU R- LEU- BOE F* DEU- B ROU TEI- DAN- ZU
41  N- K L O; LAI-CHAI V R E, I L- LOE R- FAU$ D U- L A R J E.
42  FIN DE FICHIER
43
```



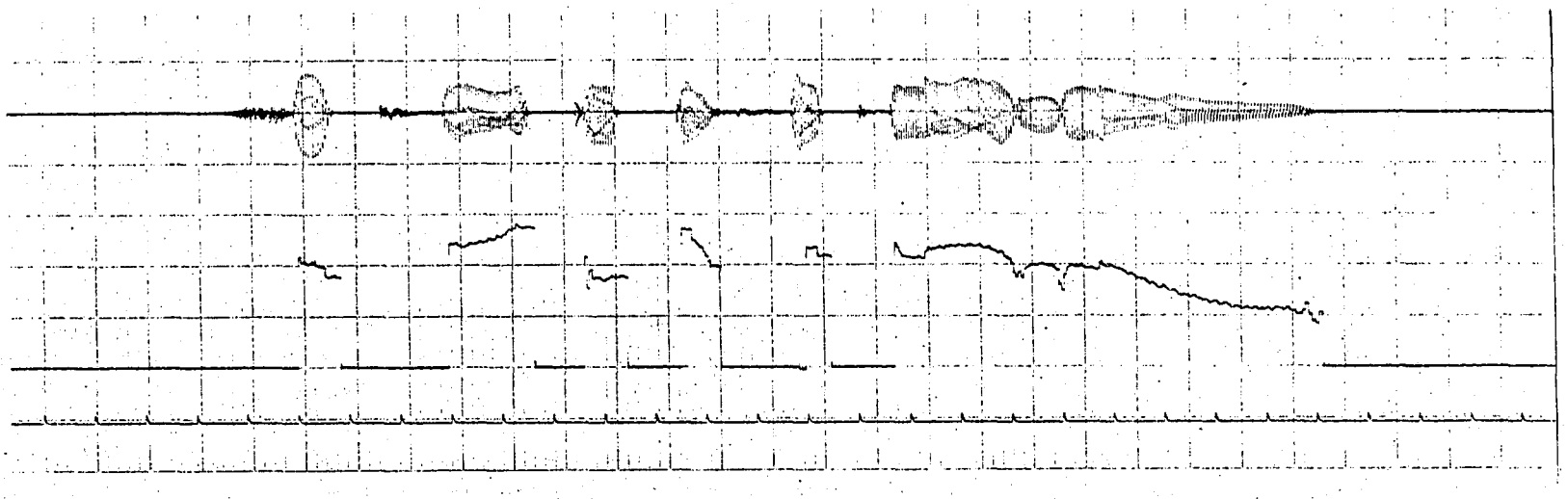
REGLE B - VOTRE REPONSE N'EST PAS CORRECTE .



REGLE C - LES MEDICAMENTS SONT VENDUS EN PHARMACIE..



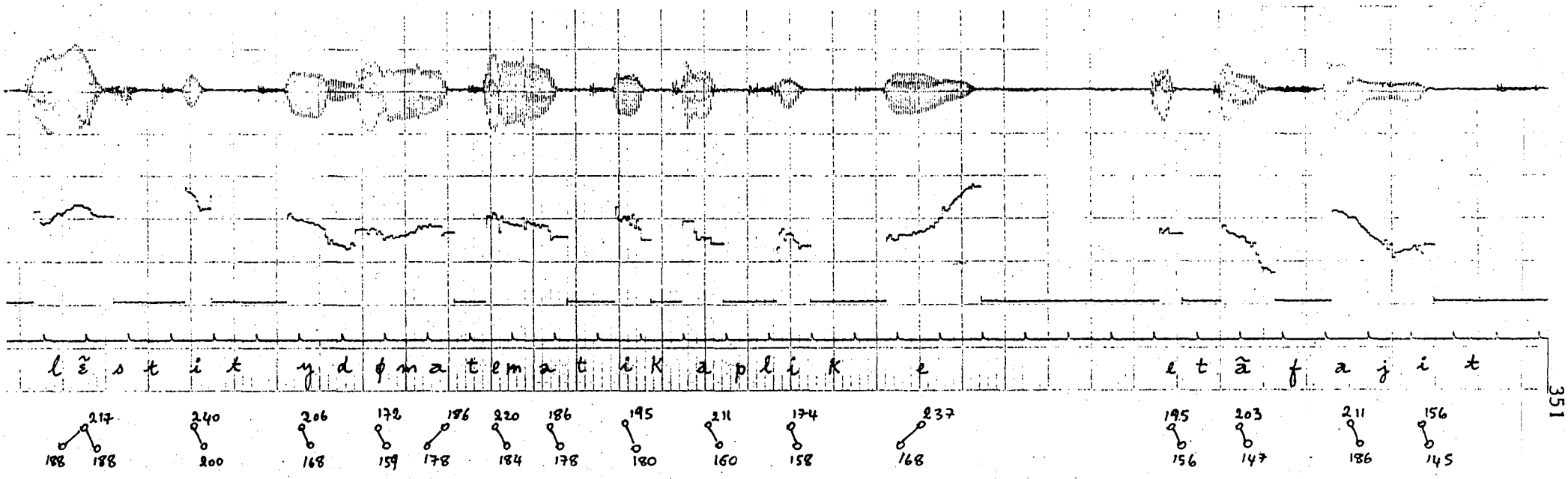
REGLE D - LE PETIT CHAT MANGE DU CRABE .



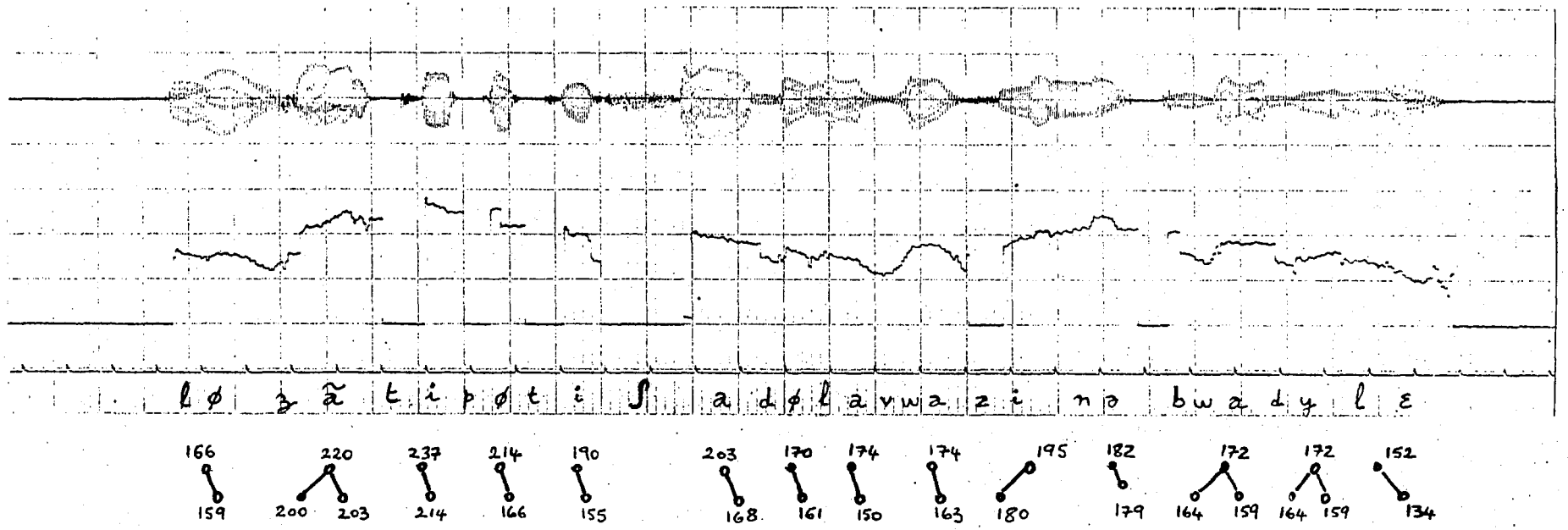
ə ɸ t ɾ ɛ e p a ɾ f ɛ t m â l i n e ɛ ɾ

211 172 230 214 175 226 166 217 206 217 211 195 186 186 166 164 145

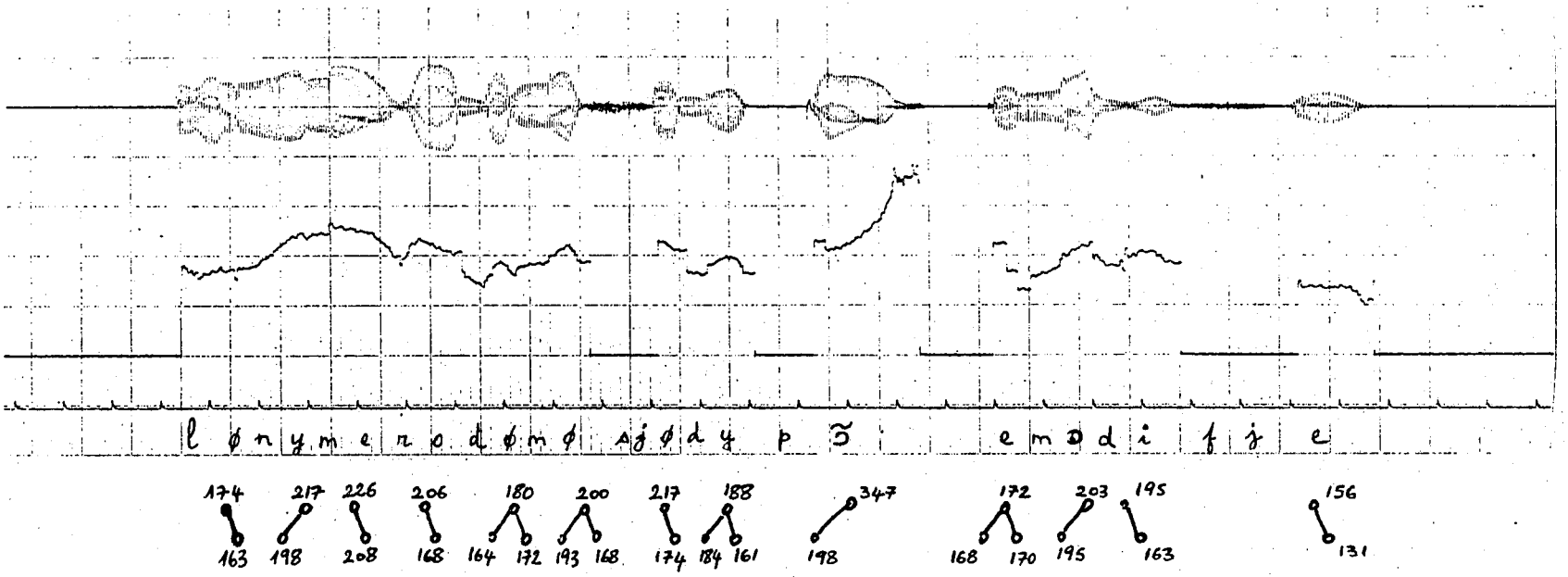
REGLE D - CE TRAIT EST PARFAITEMENT LINEAIRE .



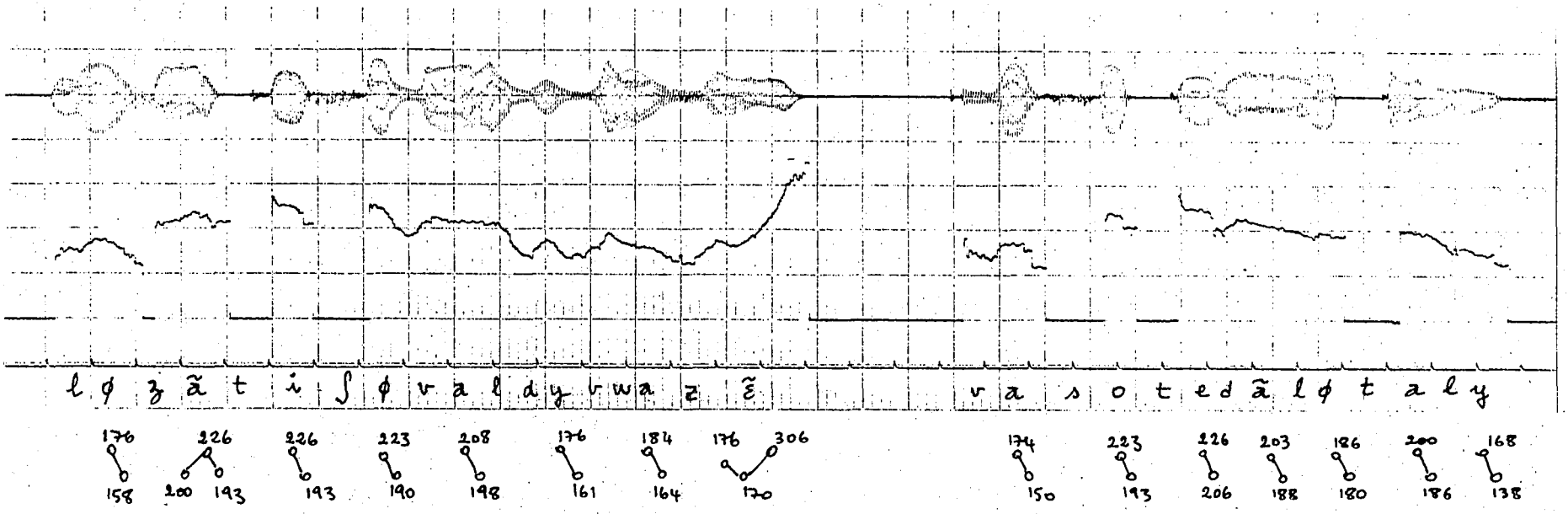
REGLE E - L'INSTITUT DE MATHEMATIQUES APPLIQUEES EST EN FAILLITE.



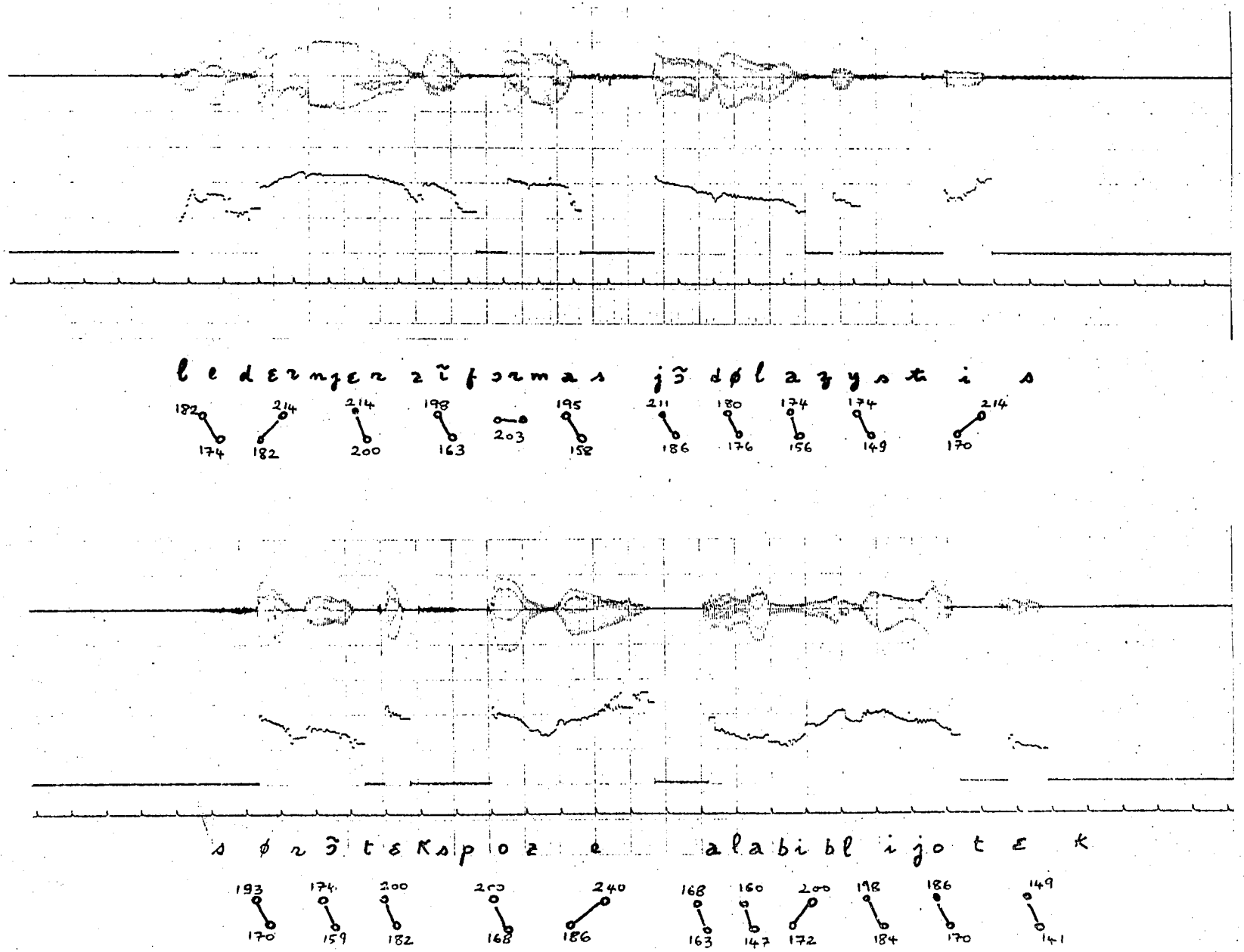
REGLE E - LE GENTIL PETIT CHAT DE LA VOISINE BOIT DU LAIT .



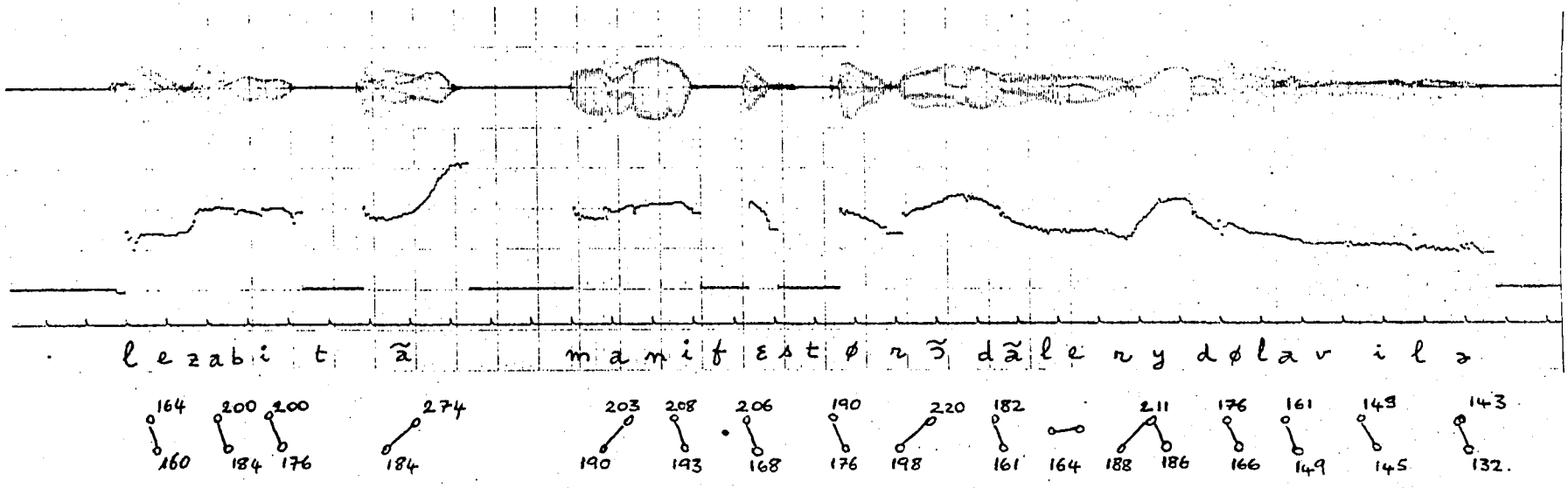
REGLE E - LE NUMERO DE MONSIEUR DUPONT EST MODIFIE.



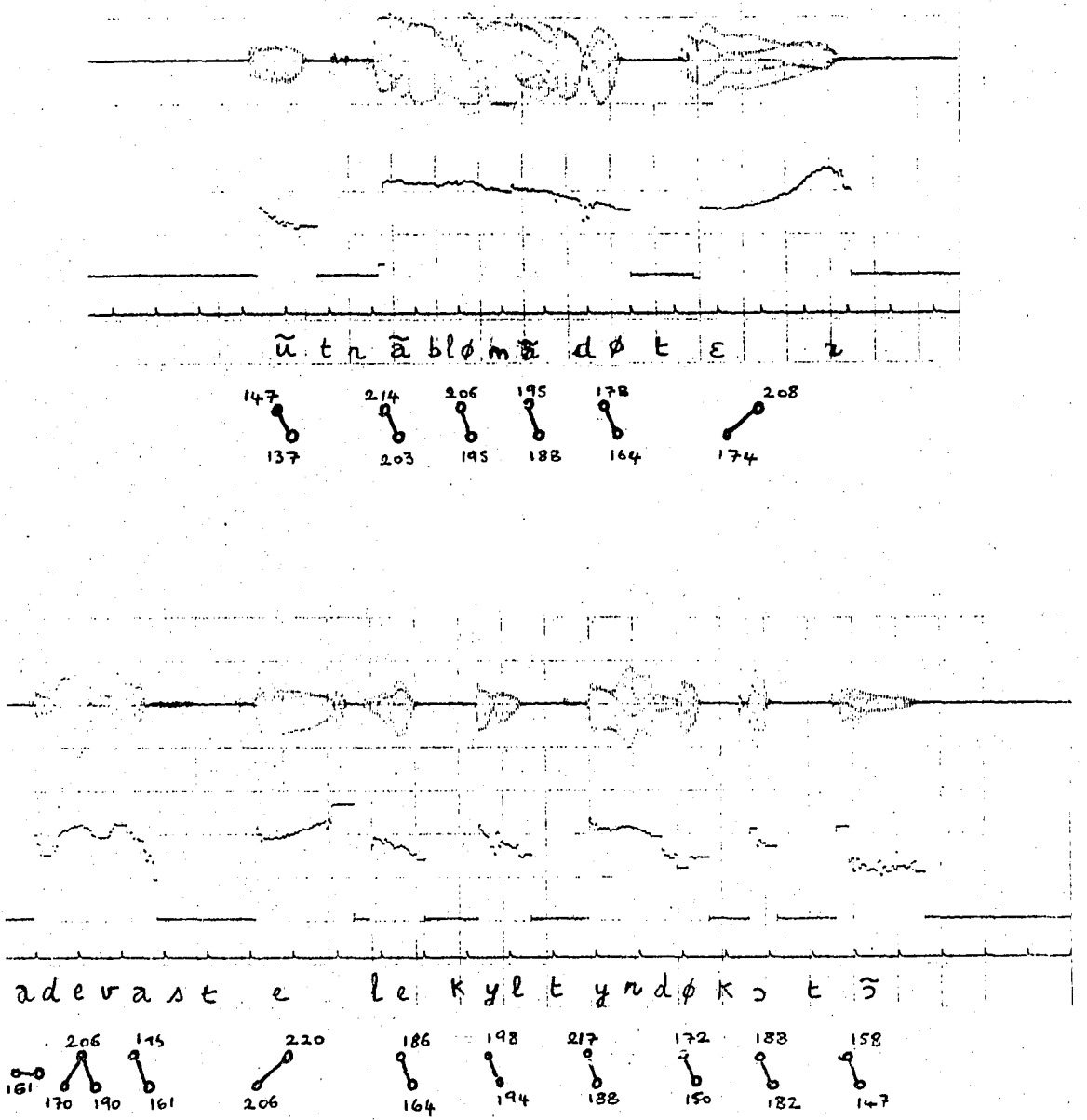
REGLE E - LE GENTIL CHEVAL DU VOISIN VA SAUTER DANS LE TALUS .



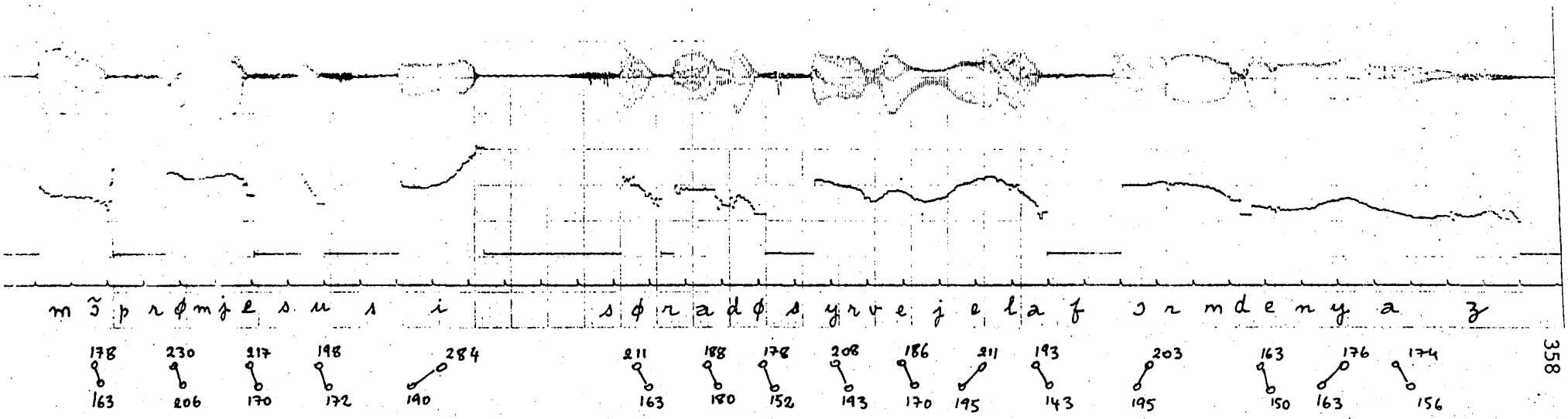
REGLE E - LES DERNIERES INFORMATIONS DE LA JUSTICE SERONT EXPOSEES A LA BIBLIOTHEQUE .



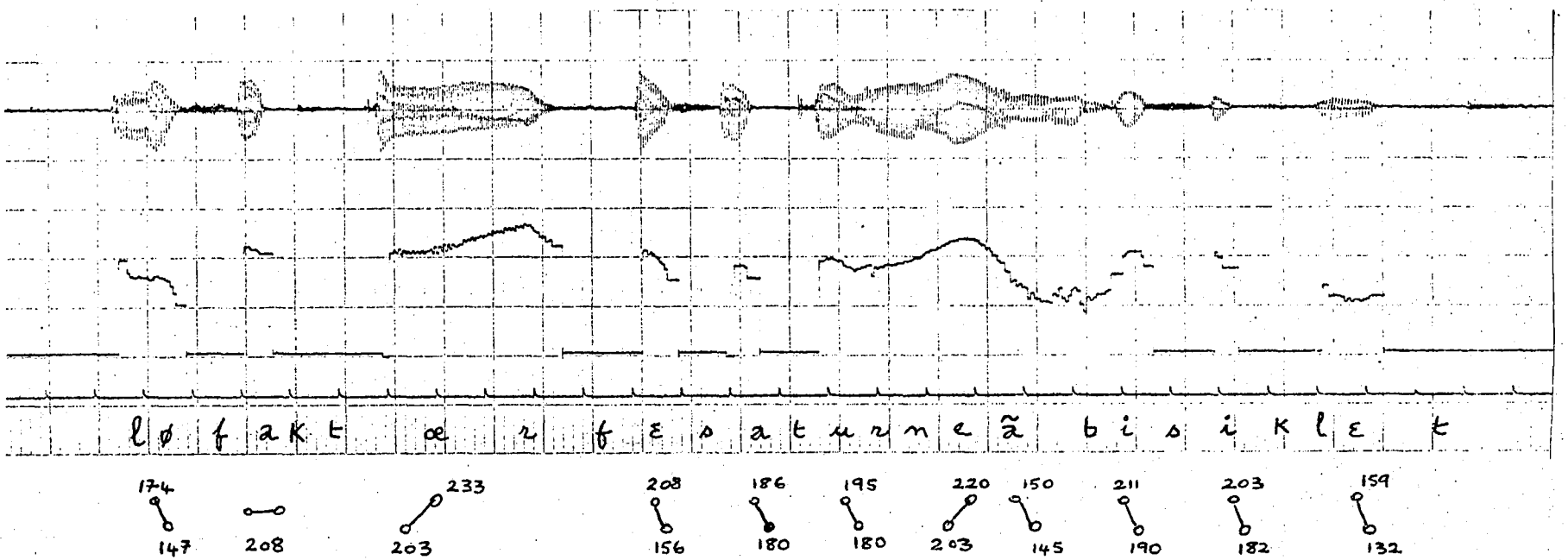
REGLE F_1 - LES HABITANTS MANIFESTERONT DANS LES RUES DE LA VILLE .



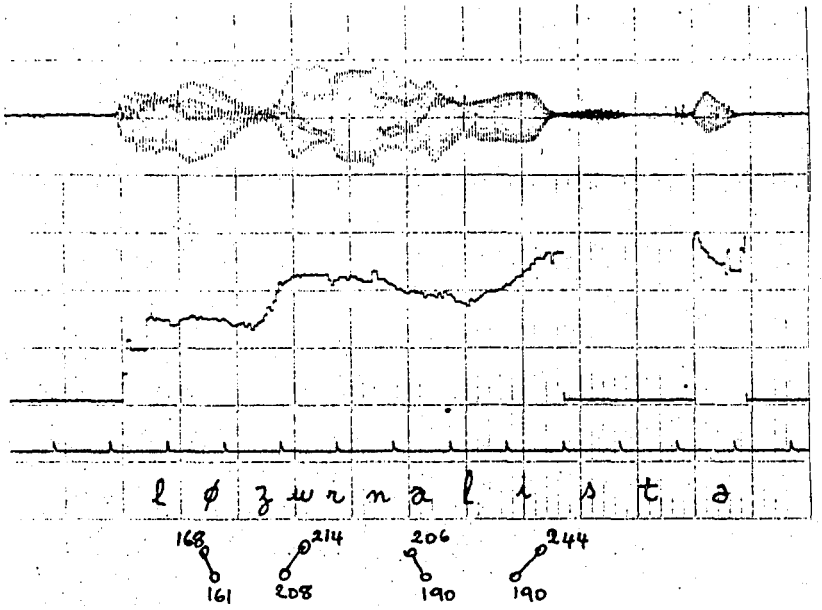
REGLE F_1 – UN TREMBLEMENT DE TERRE A DEVASTE LES CULTURES DE COTON.

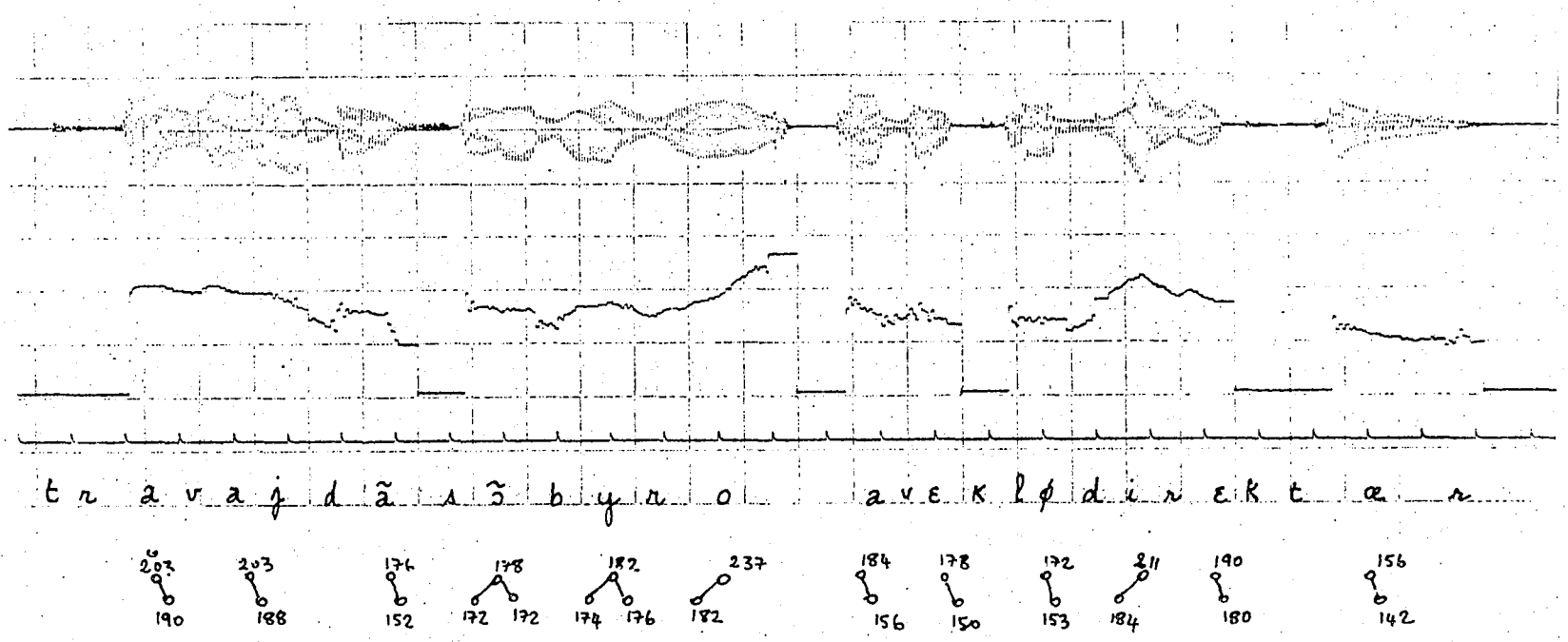


REGLE F_1 - MON PREMIER SOUCI SERA DE SURVEILLER LA FORME DES NUAGES .

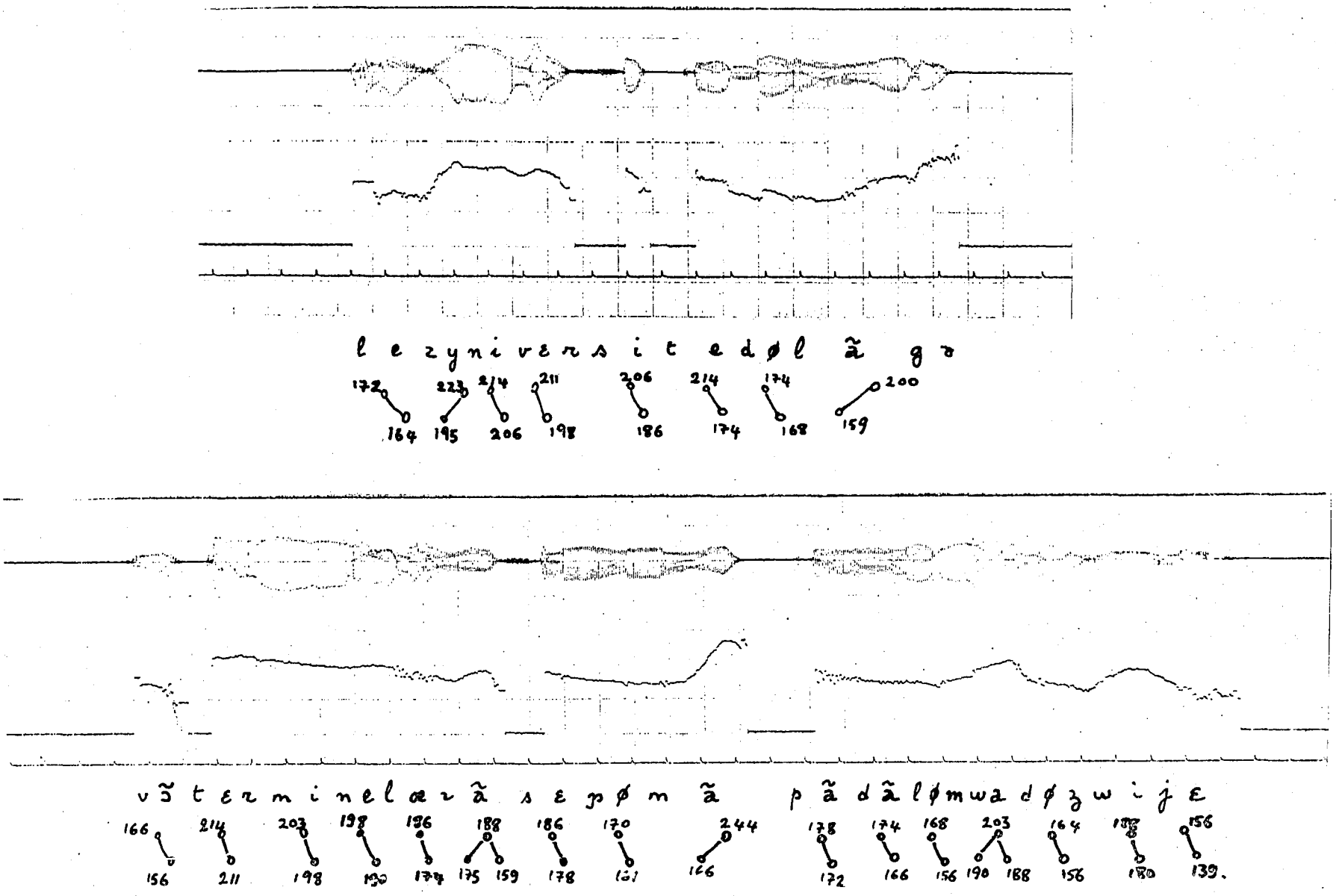


REGLE F2 - LE FACTEUR FAIT SA TOURNEE EN BICYCLETTE .

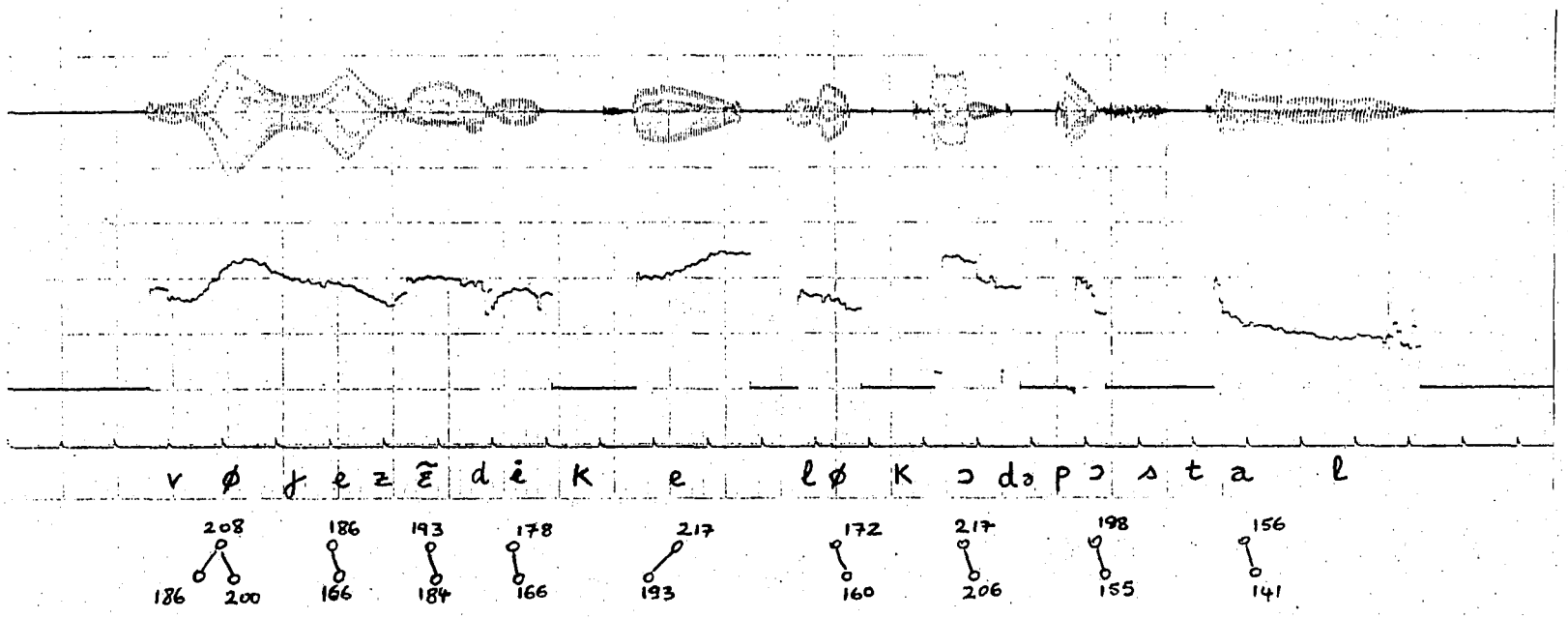




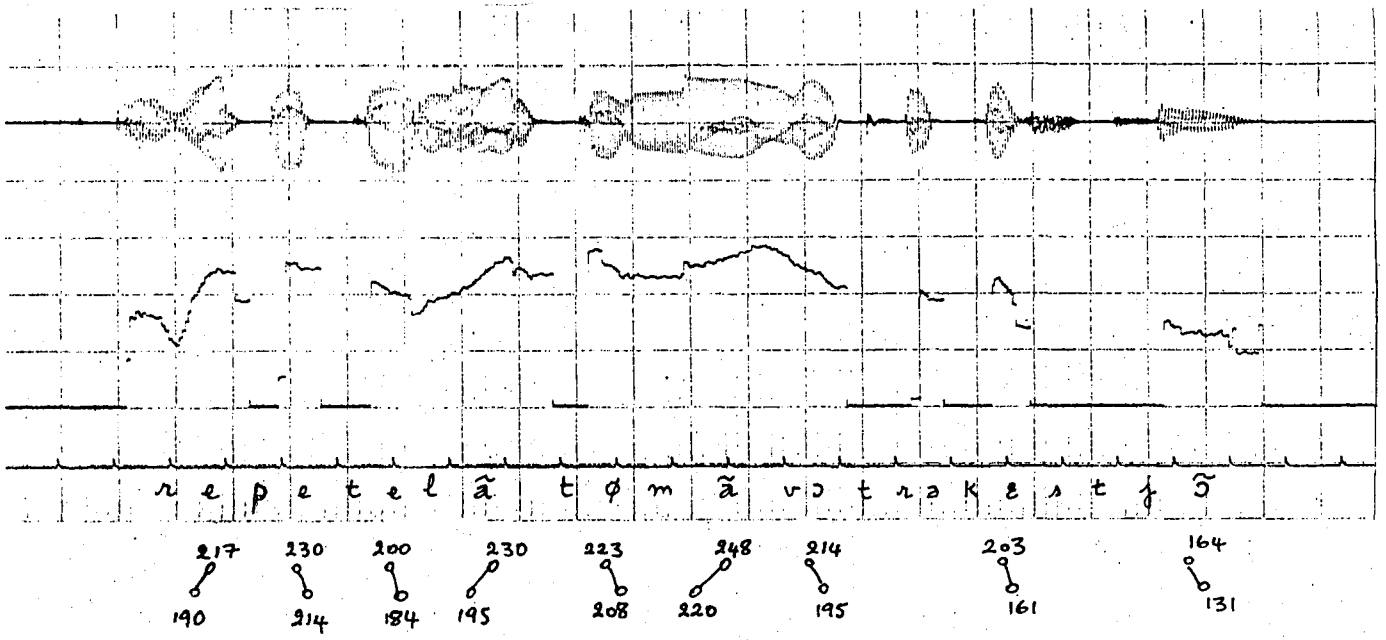
REGLE F_2 - LE JOURNALISTE TRAVAILLE DANS SON BUREAU AVEC LE DIRECTEUR .



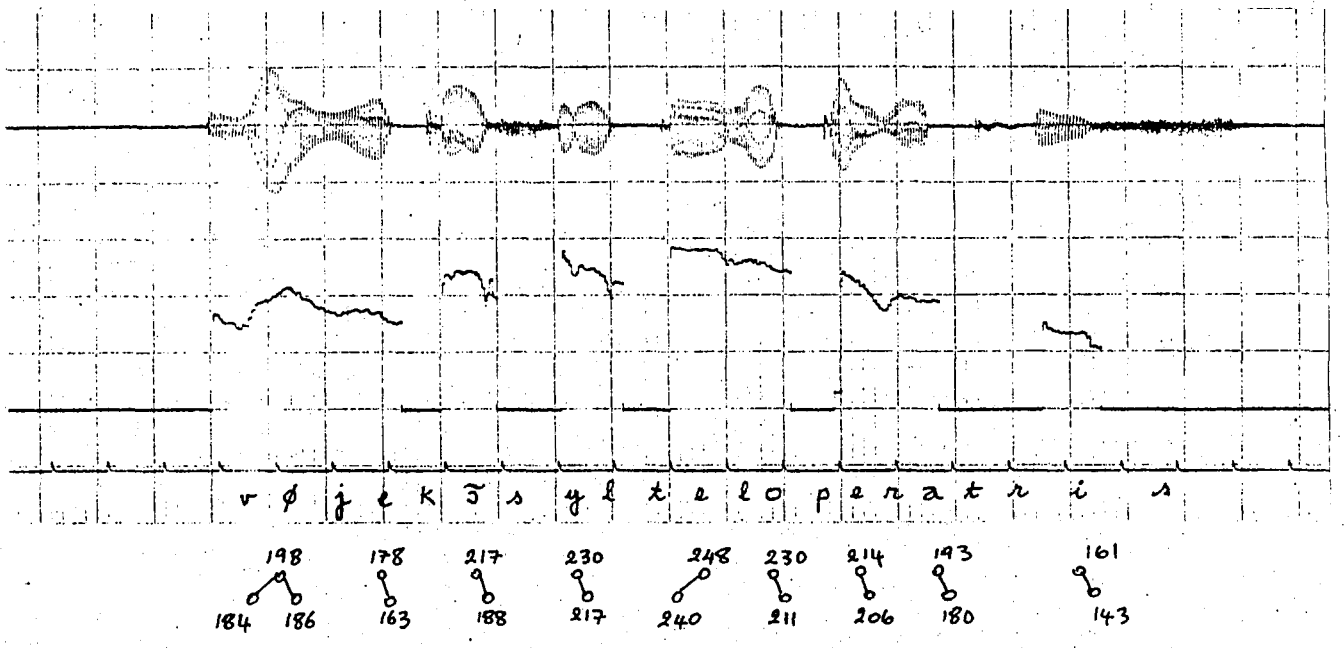
REGLE F2 - LES UNIVERSITES DE LANGUES VONT TERMINER LEUR ENSEIGNEMENT PENDANT LE MOIS DE JUILLET.



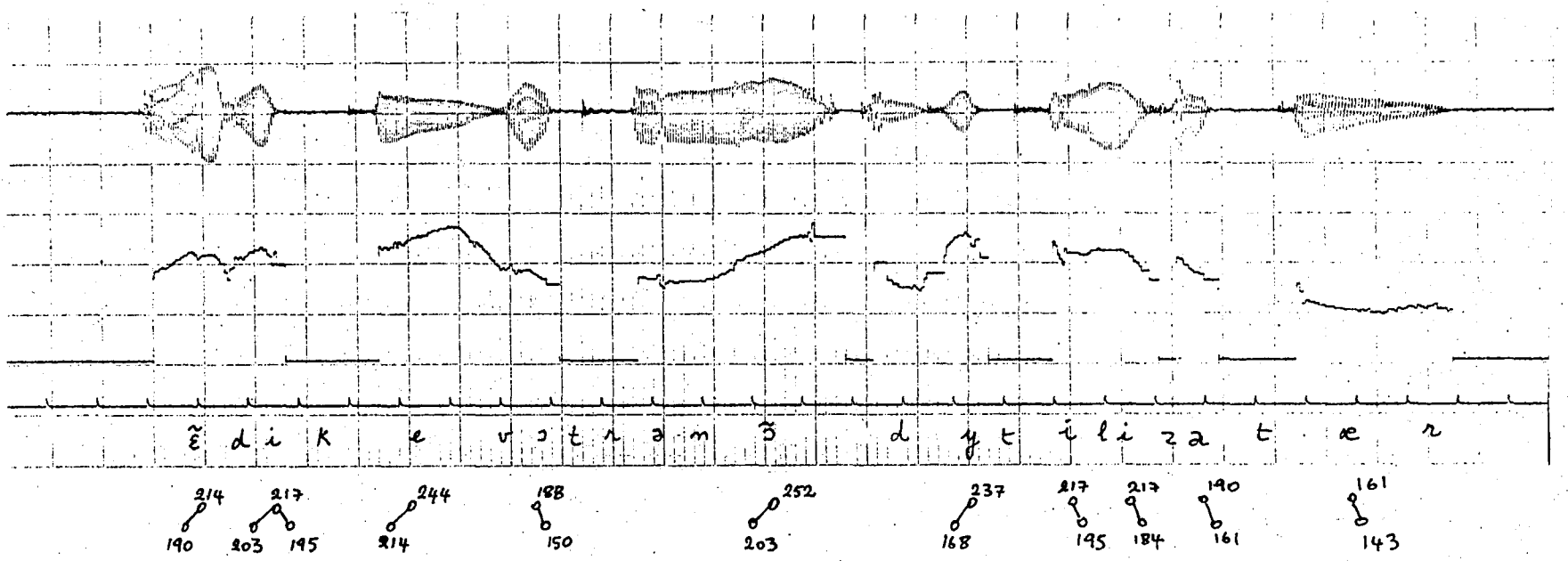
Veuillez indiquer le code postal !



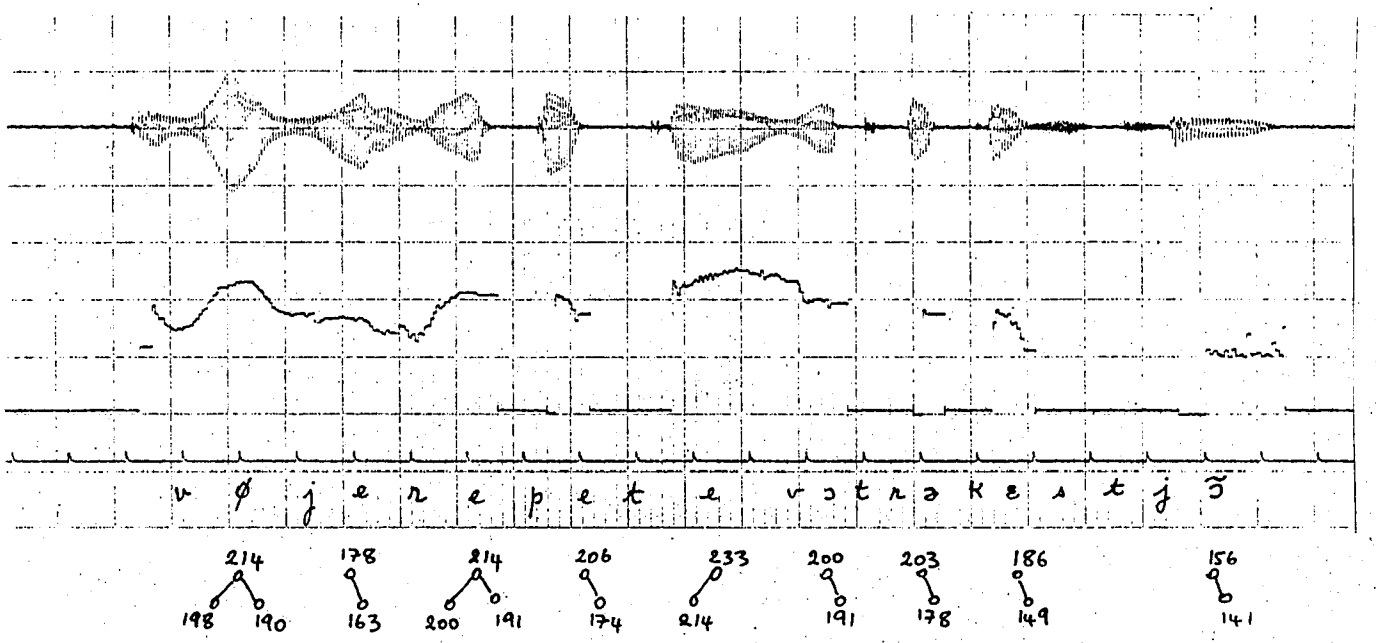
Répétez lentement votre question !



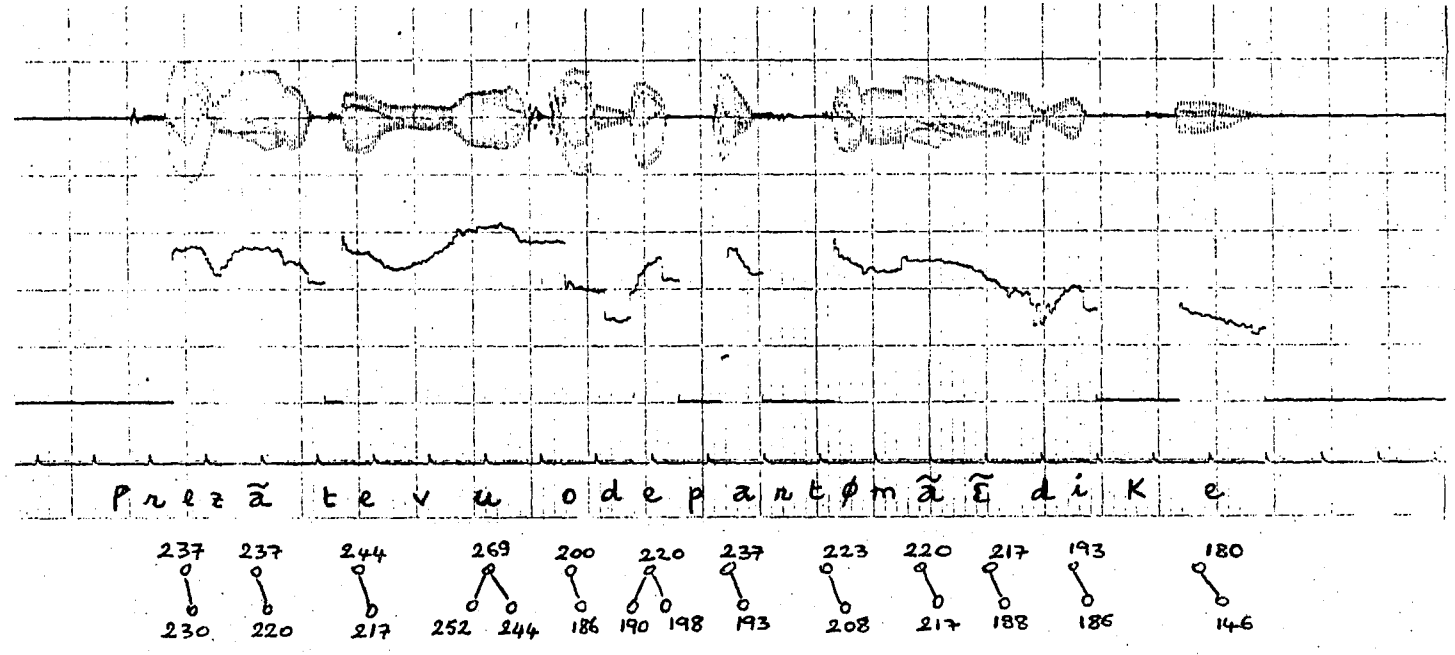
Veuillez consulter l'opératrice !



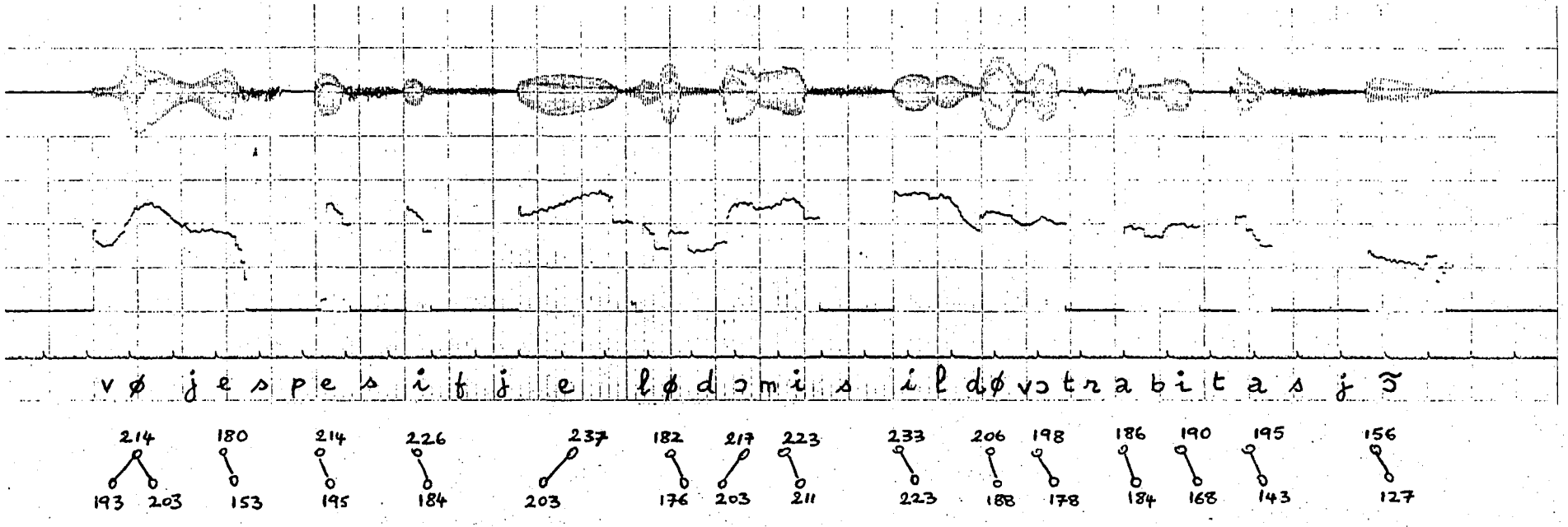
Indiquez votre nom d'utilisateur !



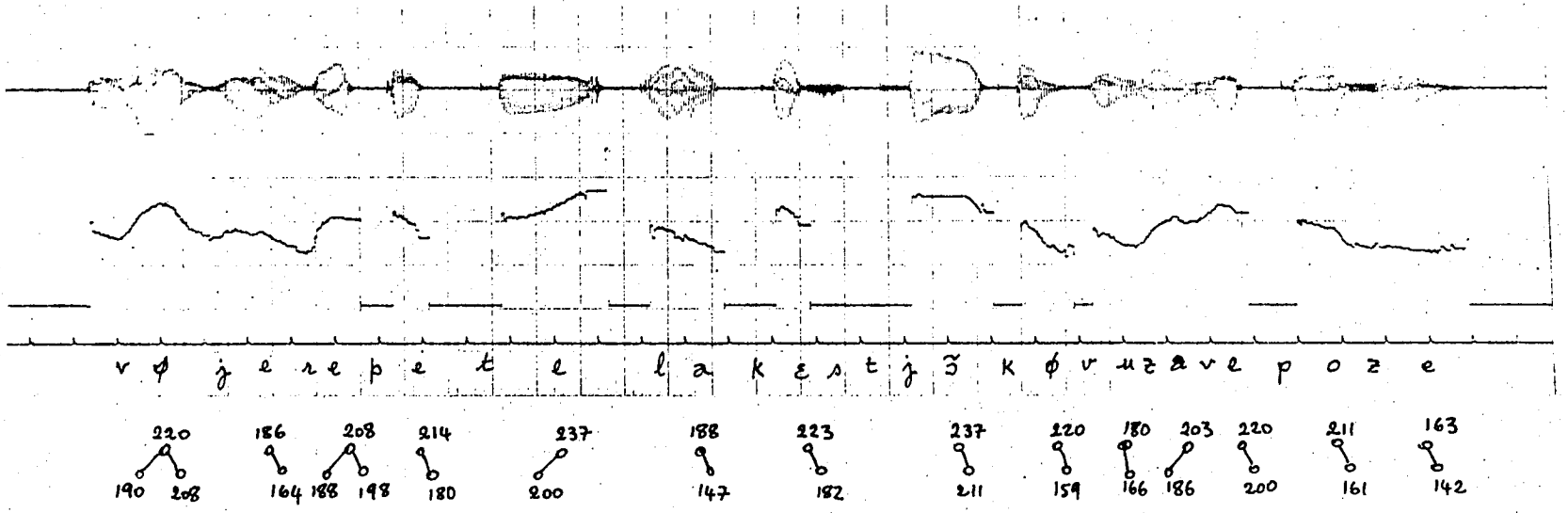
Veuillez répéter votre question !



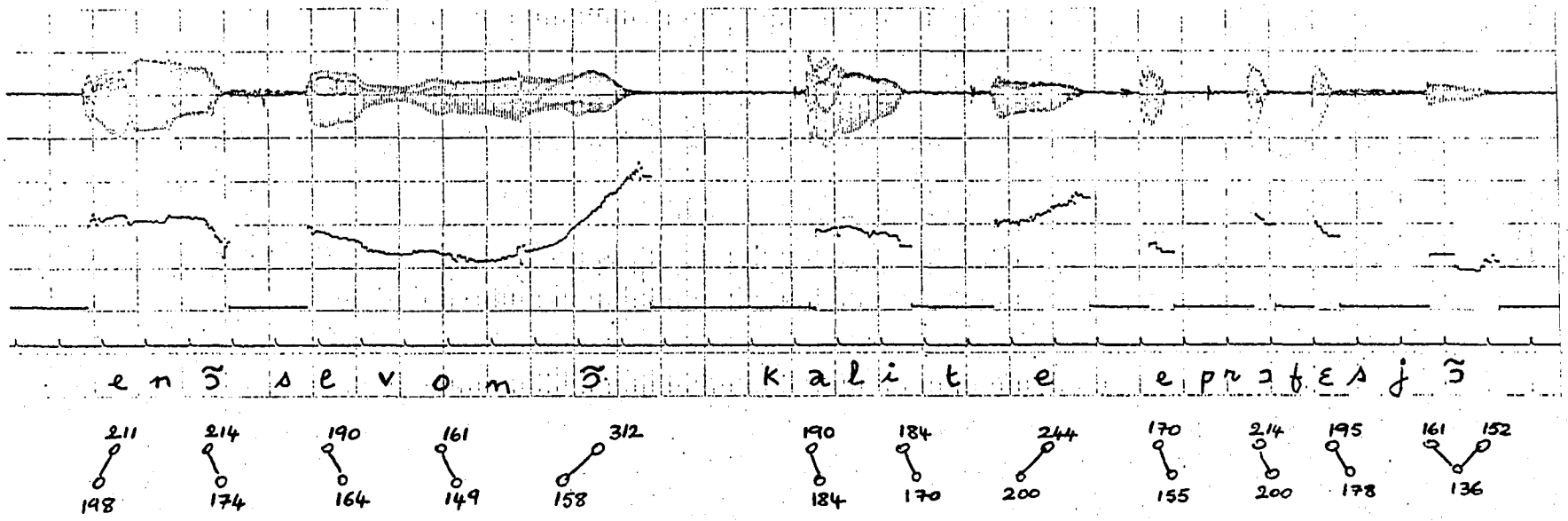
Présentez-vous au département indiqué !



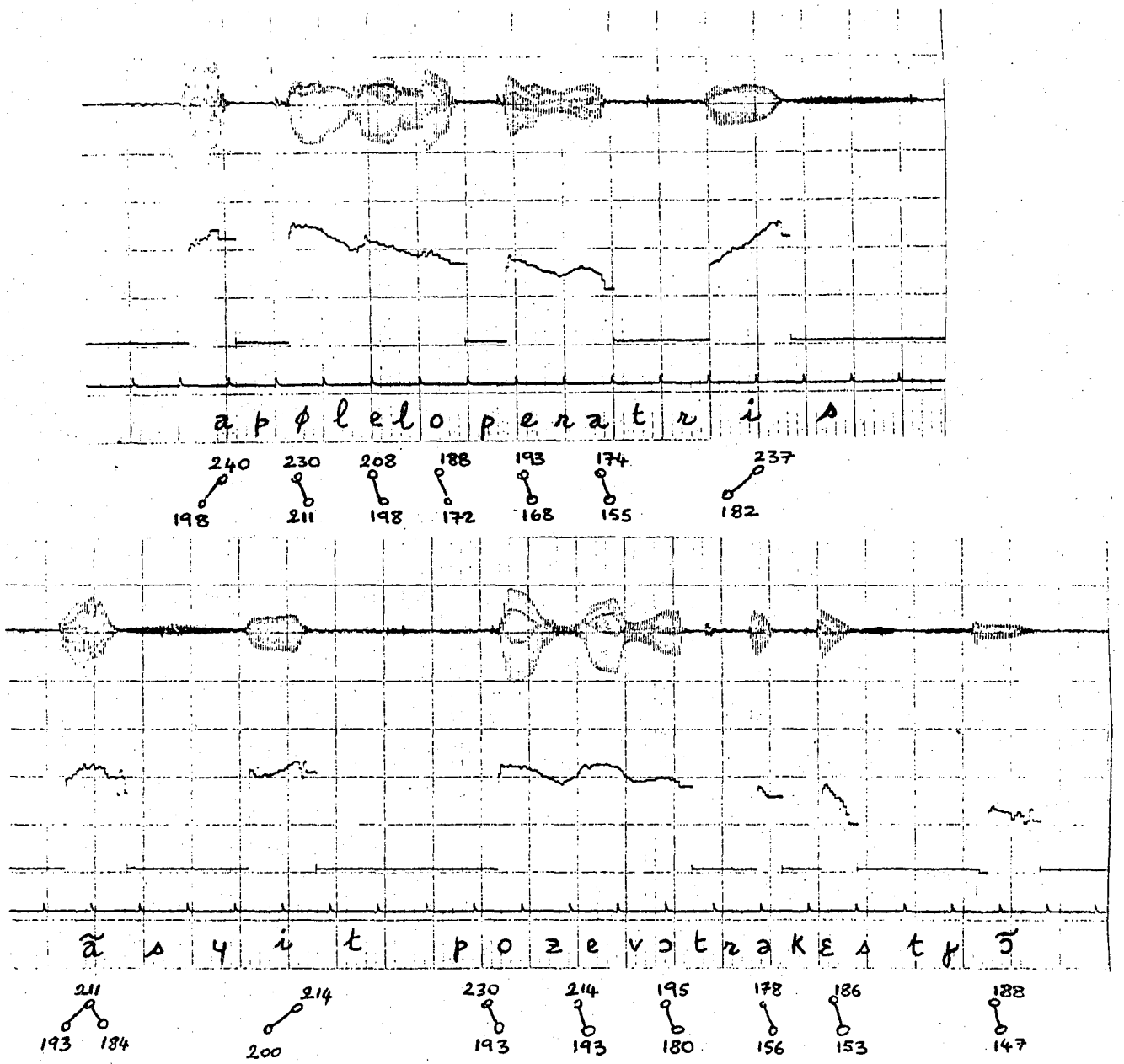
Veuillez spécifier le domicile de votre habitation !



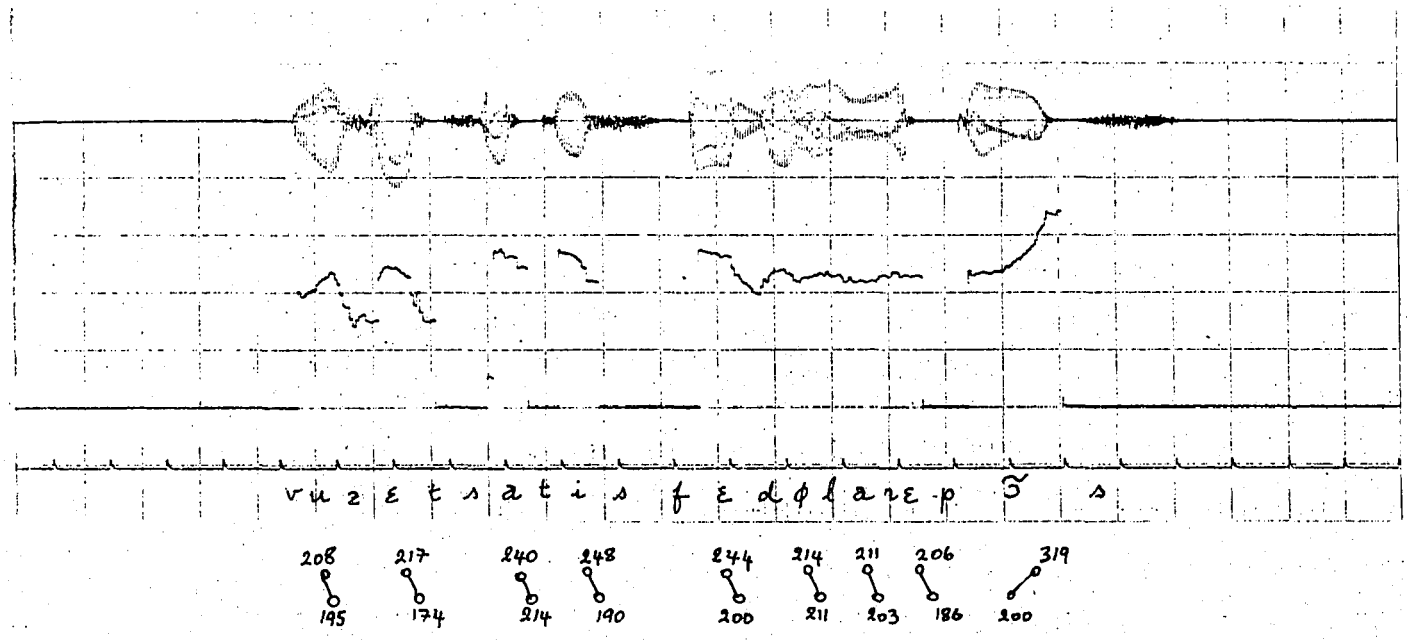
Veuillez répéter la question que vous avez posée !



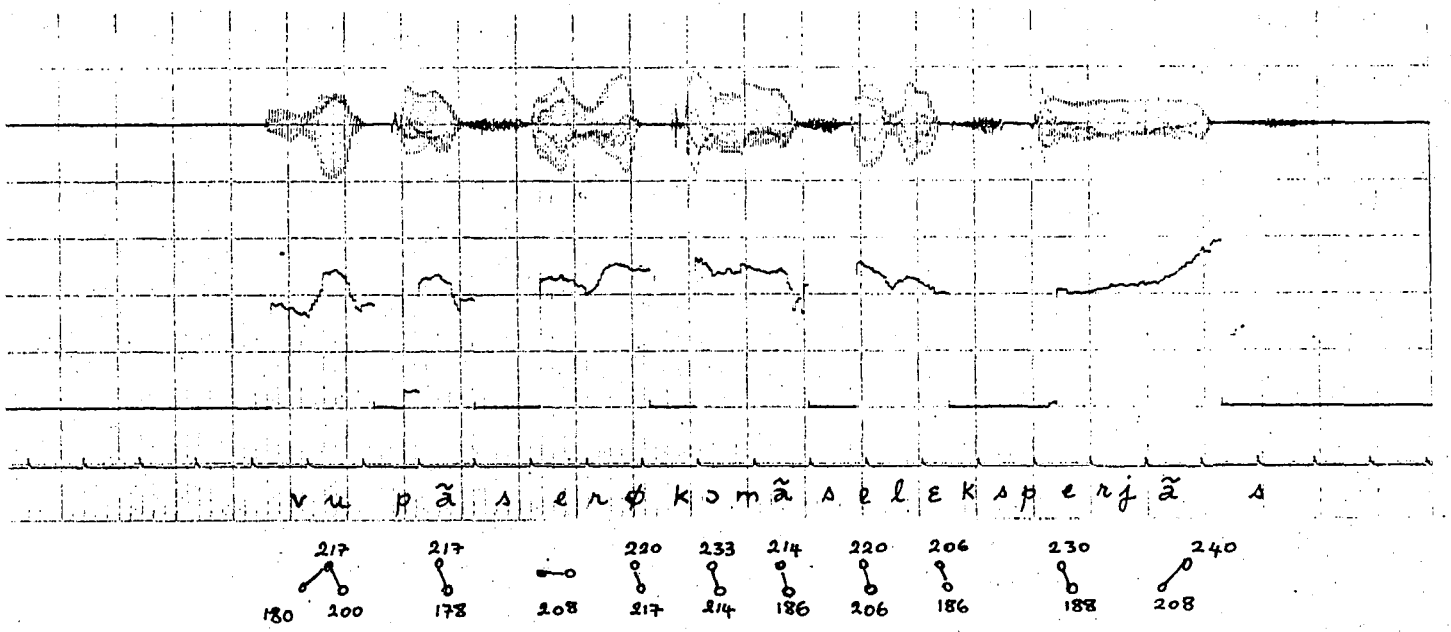
Enoncez vos nom, qualité et profession !



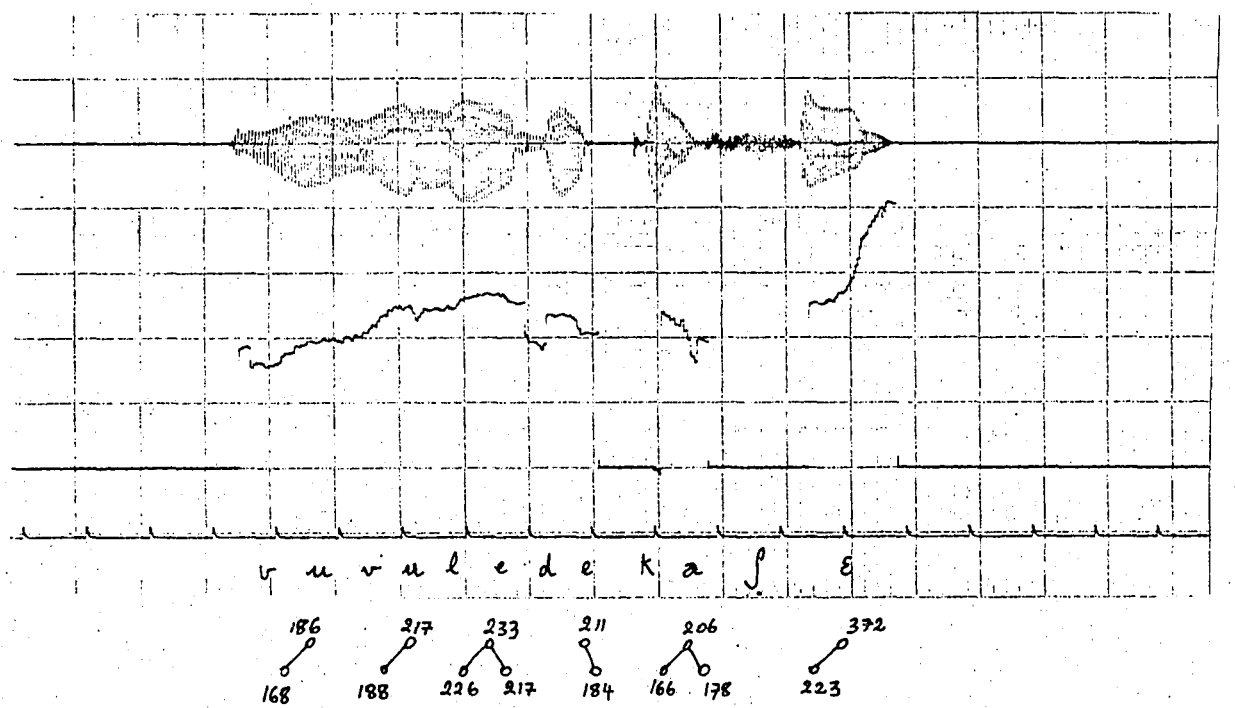
Appelez l'opératrice, ensuite posez votre question !



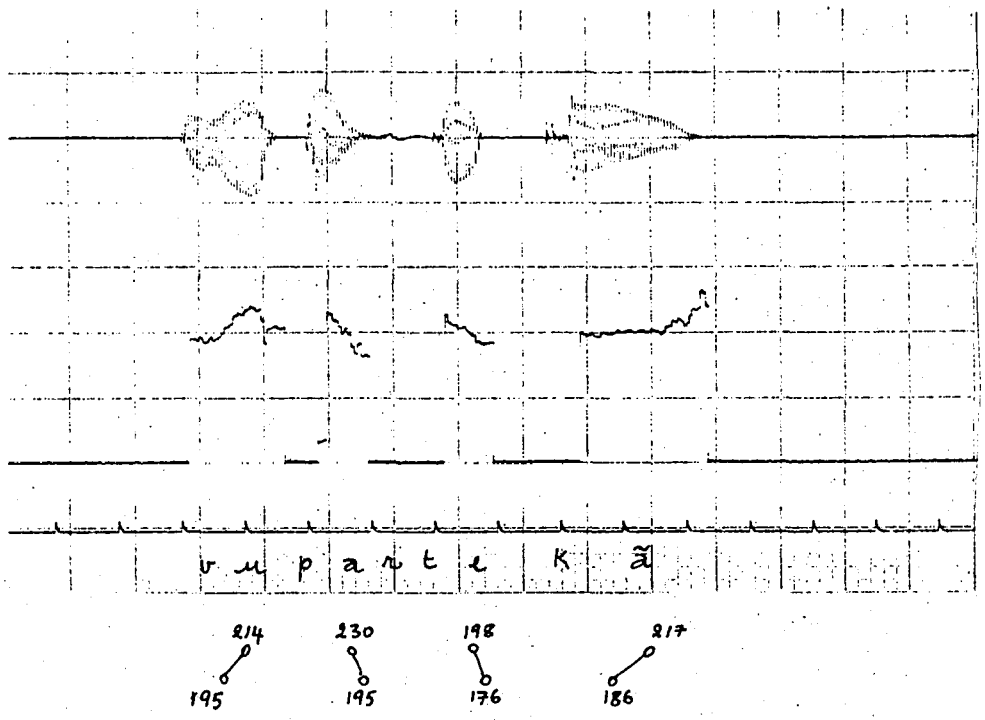
a - Vous êtes satisfait de la réponse ?



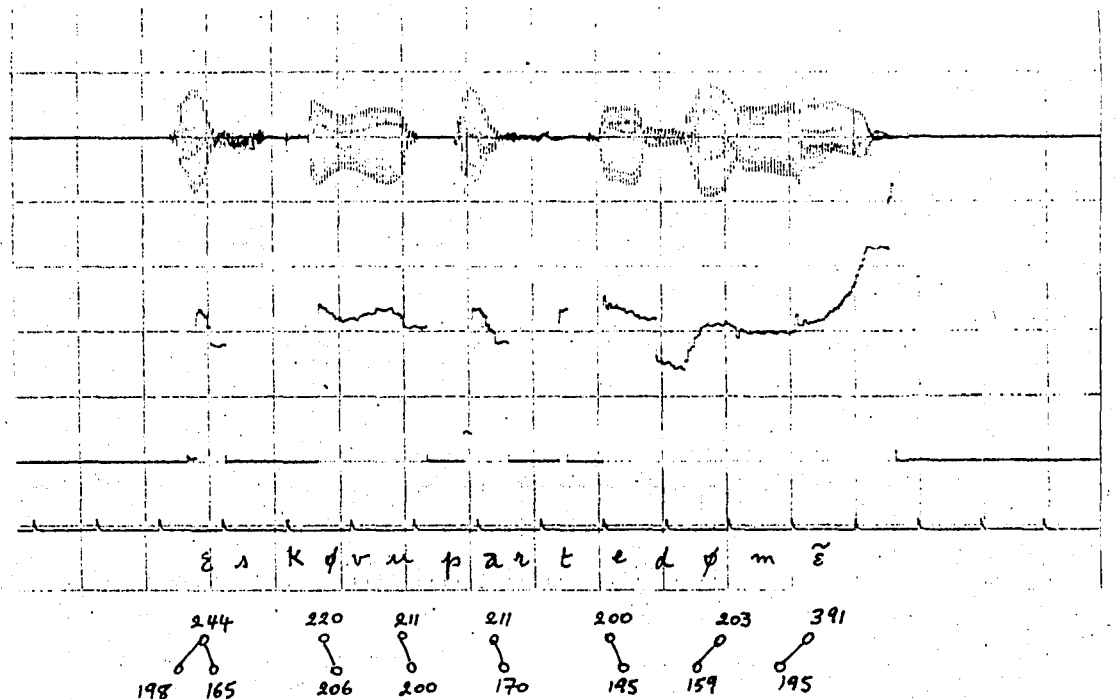
a – Vous pensez recommencer l'expérience ?



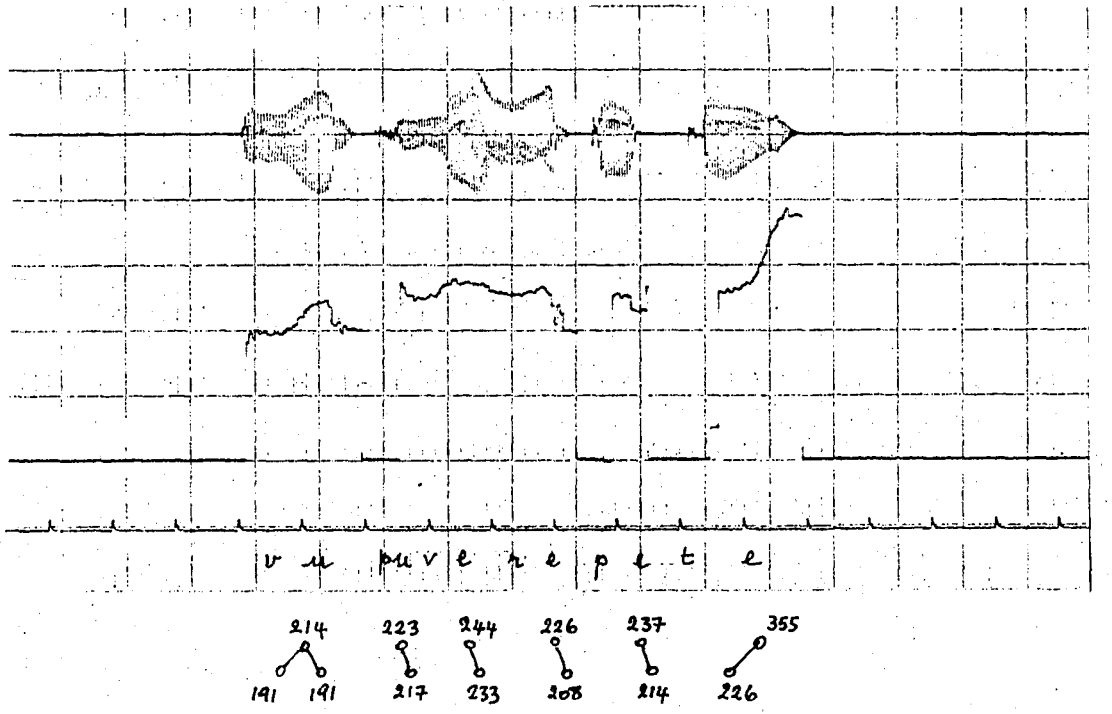
a - Vous voulez des cachets ?



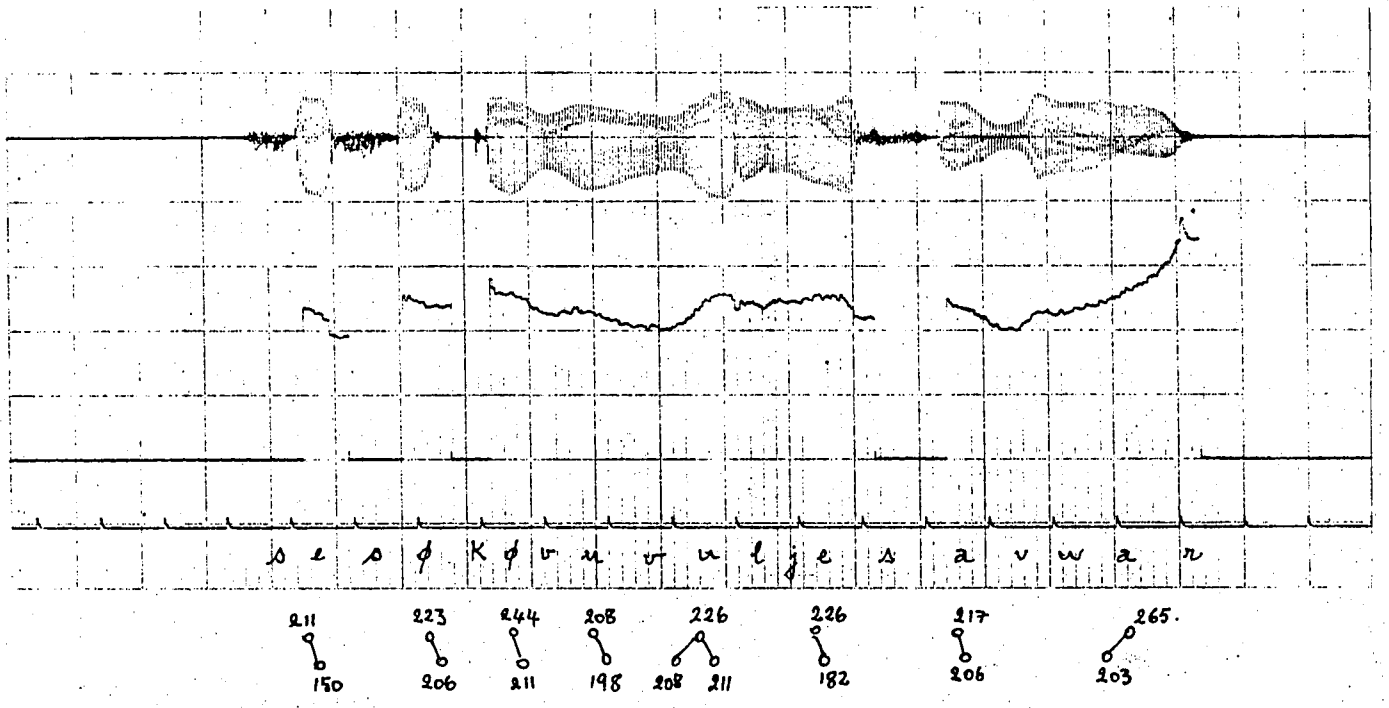
a - Vous partez quand ?



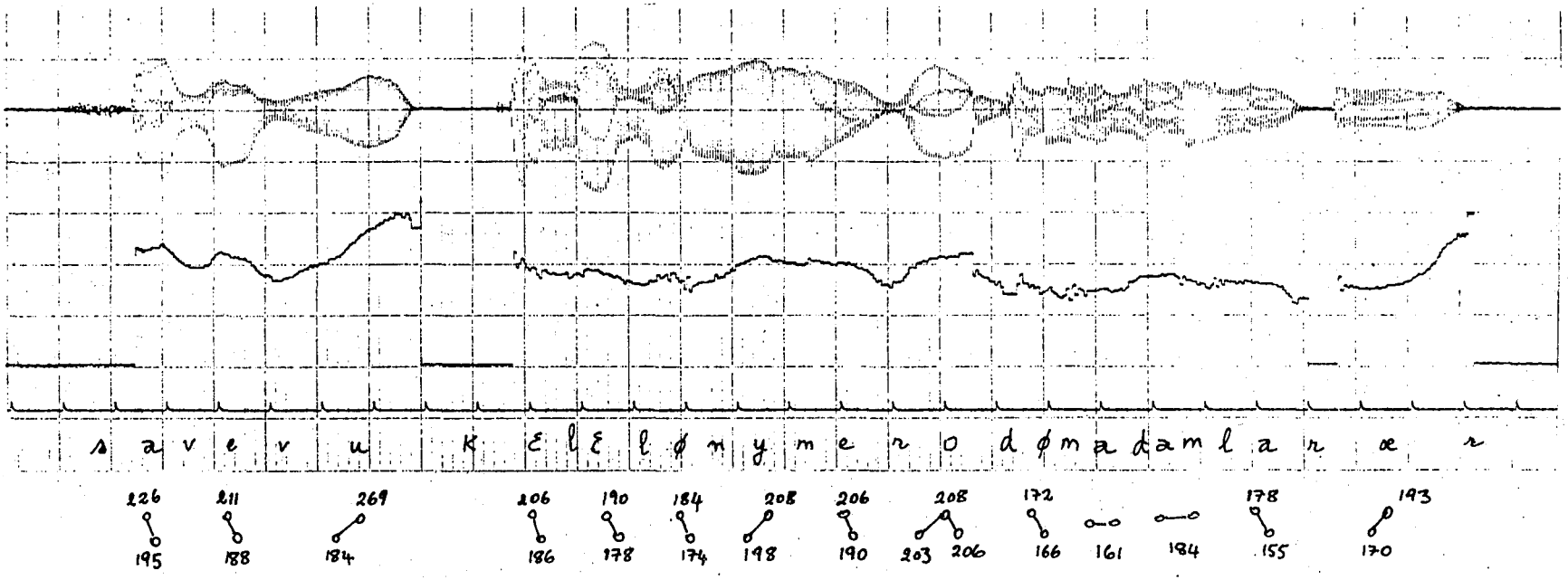
a - Est-ce que vous partez demain ?



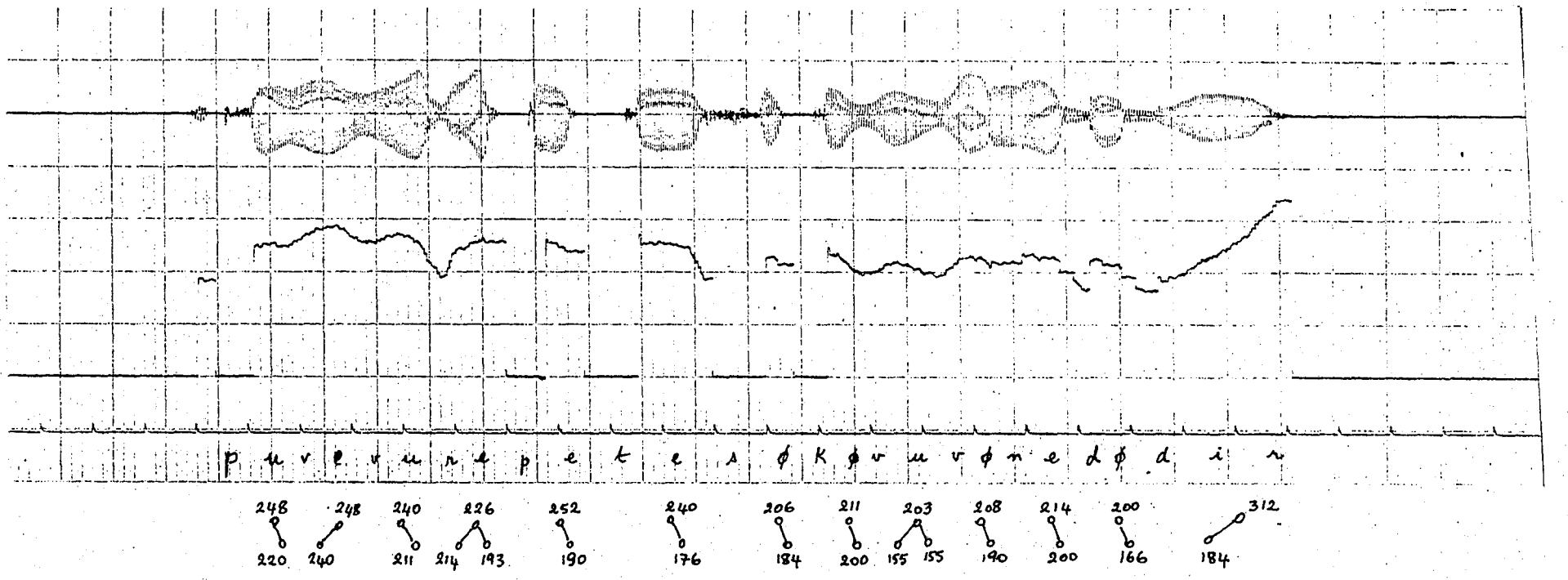
a - Vous pouvez répéter ?



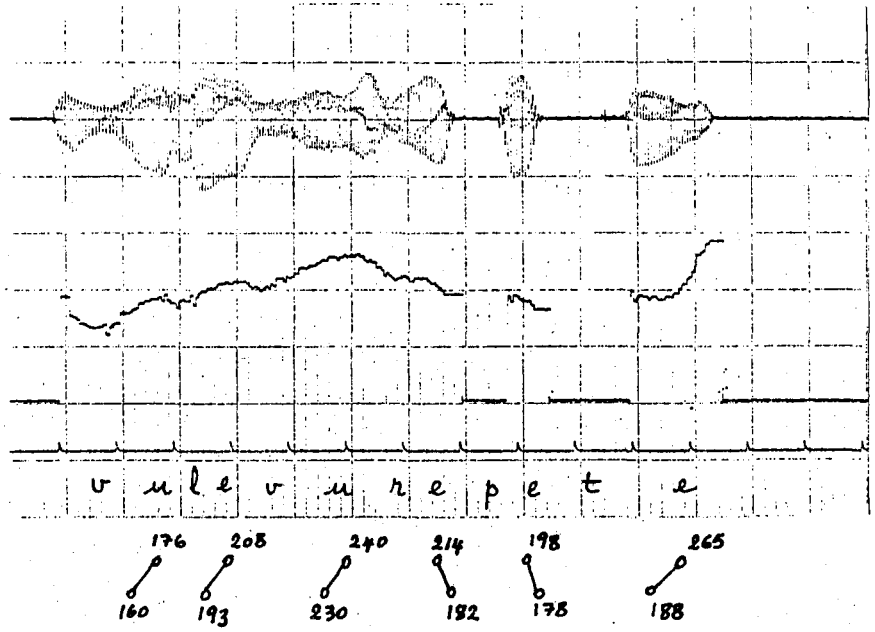
a - C'est ce que vous vouliez savoir ?



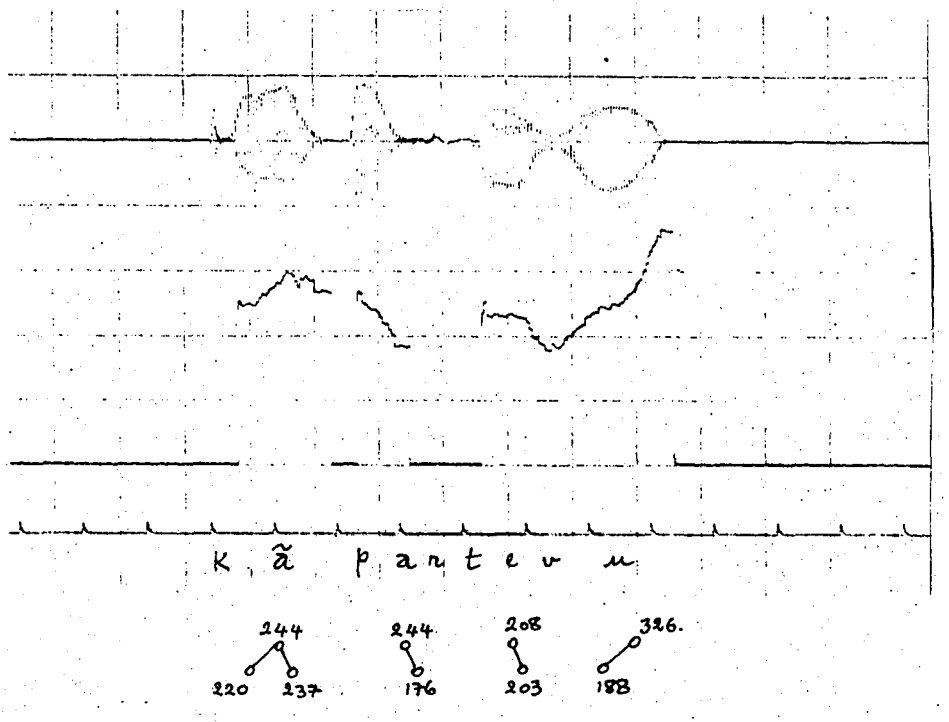
b - Savez-vous quel est le numéro de Madame LARREUR ?



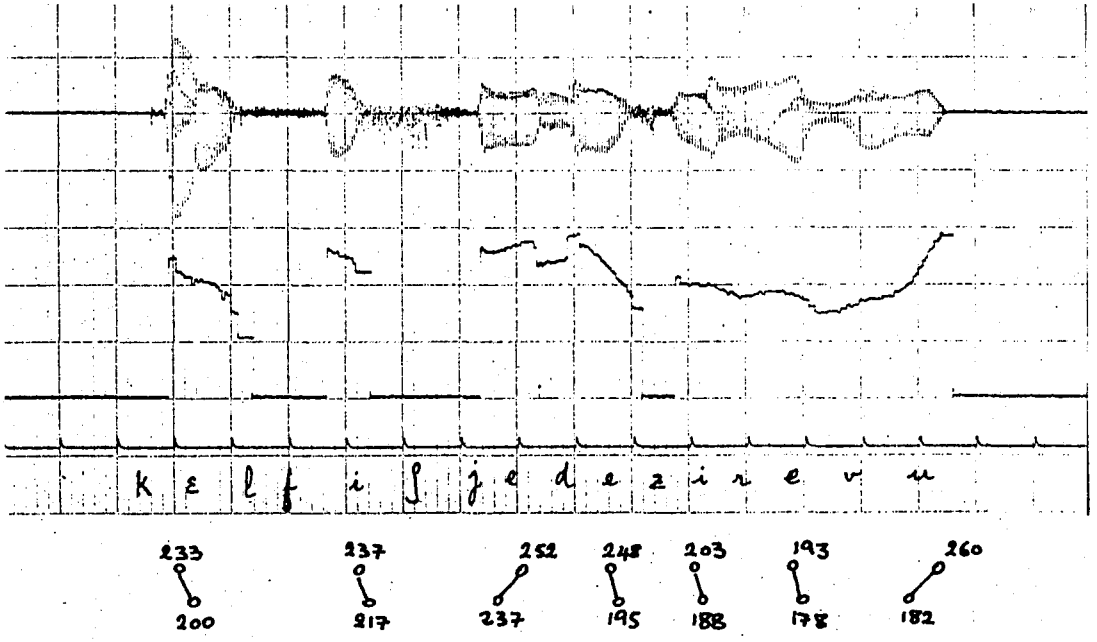
b – Pouvez-vous répéter ce que vous venez de dire ?



b - Voulez-vous répéter ?



c - Quand partez-vous ?



c - Quel fichier désirez-vous ?

EQUIVALENCE PITCH/FREQUENCE

PÉRIODE								FRÉQUENCE
1	15625	65	240	129	121	193	80	
2	7812	66	236	130	120	194	80	
3	5208	67	233	131	119	195	80	
4	3906	68	229	132	118	196	79	
5	3125	69	226	133	117	197	79	
6	2604	70	223	134	116	198	78	
7	2232	71	220	135	115	199	78	
8	1953	72	217	136	114	200	78	
9	1736	73	214	137	114	201	77	
10	1562	74	211	138	113	202	77	
11	1420	75	208	139	112	203	76	
12	1302	76	205	140	111	204	76	
13	1201	77	202	141	110	205	76	
14	1116	78	200	142	110	206	75	
15	1041	79	197	143	109	207	75	
16	976	80	195	144	108	208	75	
17	919	81	192	145	107	209	74	
18	868	82	190	146	107	210	74	
19	822	83	188	147	106	211	74	
20	781	84	186	148	105	212	73	
21	744	85	183	149	104	213	73	
22	710	86	181	150	104	214	73	
23	679	87	179	151	103	215	72	
24	651	88	177	152	102	216	72	
25	625	89	175	153	102	217	72	
26	600	90	173	154	101	218	71	
27	578	91	171	155	100	219	71	
28	558	92	169	156	100	220	71	
29	538	93	168	157	99	221	70	
30	520	94	166	158	98	222	70	
31	504	95	164	159	98	223	70	
32	488	96	162	160	97	224	69	
33	473	97	161	161	97	225	69	
34	459	98	159	162	96	226	69	
35	446	99	157	163	95	227	68	
36	434	100	156	164	95	228	68	
37	422	101	154	165	94	229	68	
38	411	102	153	166	94	230	67	
39	400	103	151	167	93	231	67	
40	390	104	150	168	93	232	67	
41	381	105	148	169	92	233	67	
42	372	106	147	170	91	234	66	
43	363	107	146	171	91	235	66	
44	355	108	144	172	90	236	66	
45	347	109	143	173	90	237	65	
46	339	110	142	174	89	238	65	
47	332	111	140	175	89	239	65	
48	325	112	139	176	88	240	65	
49	318	113	138	177	88	241	64	
50	312	114	137	178	87	242	64	
51	306	115	135	179	87	243	64	
52	300	116	134	180	86	244	64	
53	294	117	133	181	86	245	63	
54	289	118	132	182	85	246	63	
55	284	119	131	183	85	247	63	
56	279	120	130	184	84	248	63	
57	274	121	129	185	84	249	62	
58	269	122	128	186	84	250	62	
59	264	123	127	187	83	251	62	
60	260	124	126	188	83	252	62	
61	256	125	125	189	82	253	61	
62	252	126	124	190	82	254	61	
63	248	127	123	191	81	255	61	
64	244	128	122	192	81	256	61	

VOCODER

Fréquence (Hz) = $\frac{1}{\text{pitch} \times 64\,\mu s}$