

Groupement des Acousticiens de Langue Française



BIBLIOTHÈQUE
Institut de Phonétique
Aix-en-Provence

EXPIU A

8^{ème} JOURNÉES D'ÉTUDES du GROUPE de la
" COMMUNICATION PARLÉE "

Avec la participation
du CNRS et du GALF



Volume II

Compte-rendu des rapporteurs

Date d'édition : DECEMBRE 1979

Journées d'Études organisées par
l'INSTITUT DE PHONÉTIQUE
d'Aix-En-Provence

Aix - En - Provence
25 - 27 Mai 1977.

Groupement des Acousticiens de Langue Française



BIBLIOTHÈQUE
Institut de Phonétique
AIX-EN-PROVENCE



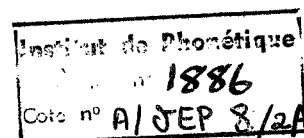
8^{èmes} JOURNEES D'ETUDES du GROUPE de la
" COMMUNICATION PARLEE "

*Avec la participation
du CNRS et du GALF*



Volume II

Compte _rendu des rapporteurs



Date d'édition : DECEMBRE 1978

Journées d'Etudes organisées par

L'INSTITUT DE PHONETIQUE
d' Aix- En - Provence

Aix - En - Provence

25 - 27 Mai 1977.

Ce deuxième volume complète les Actes des 8èmes Journées d'Etudes sur la Parole organisées sous le patronage du Groupement des Acousticiens de Langue Française avec la participation du C.N.R.S. par l'Institut de Phonétique de l'Université de Provence les 25,26,27 Mai 1977 à Aix en Provence.

Ce volume contient les exposés de synthèse sur les thèmes de ces journées :

- I. La fréquence fondamentale : études physio-acoustiques et instrumentales.
Rapporteurs : Louis-Jean BOE , Michel CARTIER
- II. La mélodie : études perceptives et synthèse de la parole.
Rapporteur : Jacques GENIN
- III. L'intonation : aspects linguistiques et reconnaissance de la parole .
Rapporteurs : Albert DI CRISTO, Jean Sylvain LIENARD,
Philippe MARTIN

Professeur Mario ROSSI

Par suite d'un certain nombre de retards dans la réception des manuscrits, ce deuxième volume a été édité par le Groupe de la Communication Parlée du G.A.L.F. en Décembre 1978.

TABLE DES MATIERES

THEME 1

LA FREQUENCE FONDAMENTALE : ETUDES PHYSIO-ACOUSTIQUES ET INSTRUMENTALES

- Variations de la périodicité
- Régulation de la vibration des cordes vocales

Louis-Jean BOE 3

- Détection de F_0 et du voisement

Michel CARTIER 19

THEME 2

LA MELODIE : ETUDES PERCEPTIVES ET SYNTHÈSE DE LA PAROLE

- Rôle des traits prosodiques dans la transmission et la synthèse de la parole

Jacques GENIN 33

THEME 3

L'INTONATION : ASPECTS LINGUISTIQUES ET RECONNAISSANCE DE LA PAROLE

- Méthodes et modèles d'analyse dans les recherches sur l'intonation

Albert DI CRISTO 41

- Méthodes et modèles d'étude des rapports intonation-syntaxe

Philippe MARTIN 61

- Prosodie et reconnaissance automatique de la parole

Jean-Sylvain LIENARD 75

THEME 1

La fréquence fondamentale : études physio-acoustiques et instrumentales.

Sous-thèmes :

- Variations de la périodicité
- Régulation de la vibration des C.V.
- La détection des Fo et du voisement.

Président : R. CARRE
Rapporteurs : L.J. BOE, M. CARTIER

Sous-thème :

- Variations de la périodicité
- Régulation de la vibration des C.V.

Rapporteur : *Louis-Jean BOE (*)*
Institut de Phonétique de Grenoble

(*) Mes remerciements vont à Christian ABRY avec qui cet exposé a été largement discuté.

VARIATIONS DE LA PERIODICITE REGULATION DE LA VIBRATION DES CORDES VOCALES

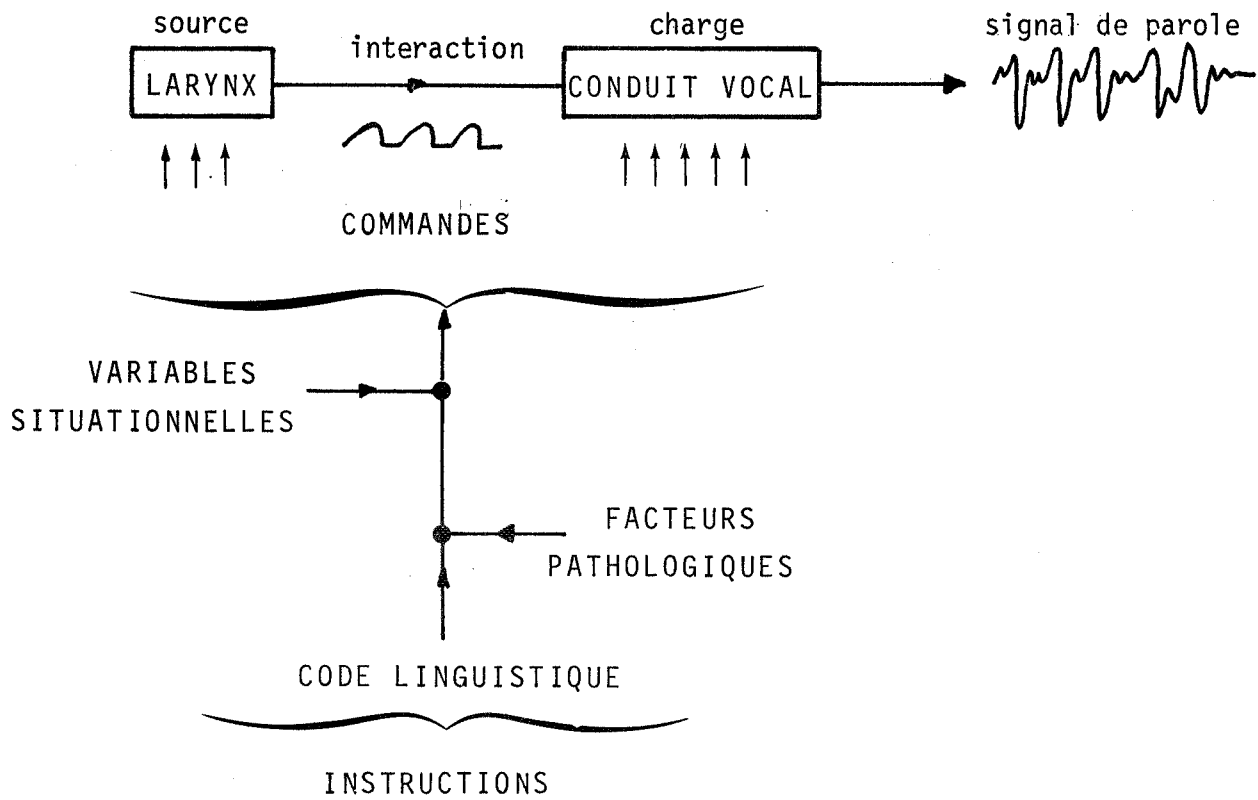
Sous ce thème ont été regroupées les communications qui se rapportent à des sujets différents et qu'il sera difficile de présenter sur tout leur contenu d'autant plus que certaines d'entre elles sont à cheval sur deux thèmes.

Je vais estimer que les communications ont déjà été lues par les participants et je vais me pencher sur des points qui me semblent importants ou à discuter.

Pour essayer de classer les communications de la première partie de ce thème dans un ordre relativement cohérent, je vais me servir d'un organigramme qui ne se veut pas un modèle de production du signal de la parole mais un simple schéma de présentation. On peut considérer, dans le cas des sons voisés, réalisations pour lesquelles les cordes vocales vibrent, que le larynx est une source qui excite une charge, le conduit vocal. Le larynx produit alors un signal que - par extension de notion - on qualifie de périodique, en arguant du fait qu'il ne varie que modérément d'un cycle à un autre. Dans le cas le plus stable on observe des variations de 2% à 7% d'une période à l'autre. C'est de ces variations de fréquence que nous allons parler en les situant pour l'essentiel dans un domaine physiologique et/ou acoustique laissant aux participants du thème III le soin d'en discuter à un niveau plus linguistique.

En opérant une classification on peut dire que les variations de F_0 , la fréquence de vibration des cordes vocales, sont dues :

- à des variations des commandes du larynx pour l'essentiel :
 - . la pression subglottique P_s
 - . la disposition des cordes vocales
 - . la tension des cordes vocales
- à des effets de couplage entre source et charge. Cette interaction peut être de nature acoustique lorsque l'impédance d'entrée du conduit vocal n'est plus négligeable devant l'impédance de la source. Ils peuvent être aussi de nature physiologique dans la mesure où les mouvements des articulateurs (langue, mâchoire) peuvent entraîner, par interaction, des variations de la tension des cordes vocales.



Organigramme permettant de classer les communications du thème I.

Ces commandes du larynx sont modifiées à la suite d'instructions émises par le système nerveux central opérant dans le cadre de la communication afin de satisfaire au code linguistique.

Ces instructions peuvent être influencées par deux types de paramètres :

- des variables situationnelles :
variables liées aux conditions de la communication, par exemple s'il s'agit d'un exposé, d'un dialogue, de lecture, etc. Interviennent aussi l'état d'émotivité du locuteur, sa fatigue, etc. et enfin
- des facteurs pathologiques
qui peuvent plus ou moins affecter la transmission.

Dans l'ordre, nous allons aborder :

1. La commande et ses manifestations physiologiques :
 - . MAEDA
 - . GUERIN & BOË
2. La forme de l'onde glottique :
 - . GUERIN & BOË
 - . LANDERCY
3. Les variations de F_0 qui ne sont pas dues au couplage et pas forcément aux commandes :
 - . LANDERCY
4. Les variations de F_0 dues au couplage :
les caractéristiques intrinsèques
 - . DI CRISTO & CHAFCOULOFF
5. Les communications qui proposent une stratégie pour faire la part respective des évolutions de F_0 :

. Nina THORSEN

. MAEDA

6. L'influence des variables situationnelles :

. FONAGY & al.

7. L'influence sur F_0 de facteurs pathologiques:

. CHEVRIE-MULLER & DECANTE

. GUBRYNOWICZ & al.

1 - LES PROBLEMES DE LA COMMANDE ET DES EFFETS PHYSIOLOGIQUES.

Pour simplifier le problème nous dirons qu'il existe deux types de commande :

a) La pression subglottique générée par les poumons.

Cette pression est relativement accessible et on en a une idée assez précise grâce à de multiples procédés de mesure :

- on peut la mesurer chez un sujet qui respire avec une canule de trachéotomie,
- avec un trocard piqué dans la trachée sous le cricoïde,
- à l'aide d'un ballonnetoesophagien,
- à l'aide d'une sonde introduite par le nez.

On constate que durant l'expiration phonatoire la pression subglottique évolue de quelques cm d'H₂O autour de 10 cm d'H₂O.

D'autre part les variations de F_0 dues aux évolutions seules de la pression subglottique ont été relativement bien chiffrées :

5 Hz/cm d'H₂O

1/2 octave / 7,5 cm d'H₂O selon les auteurs.

F_0 au cours d'une phrase énonciative variant en moyenne d'une octave, on voit que les variations de P_s ne sont pas seules à intervenir.

b) Le deuxième type de commande est la tension des cordes vocales

Mais il faut bien voir qu'il existe deux types de tension :

- une tension passive : sous l'effet du basculement du cricoïde par rapport au thyroïde, les cordes vocales s'allongent, s'amincissent et se tendent. Ce basculement du cricoïde est provoqué par la contraction du muscle cricothyroïdien.

- une tension active : les cordes vocales sont constituées d'une muqueuse, d'un ligament et de deux faisceaux de muscles. Sous l'effet de la contraction du muscle thyroaryténoïdien les cordes vocales ont tendance à se tendre, à s'épaissir et à raccourcir.

Les études E.M.G. montrent que tension passive et tension active se produisent en même temps et que leur conjugaison diminue la longueur des cordes vocales en augmentant la valeur de la fréquence fondamentale.

Sur des larynx excisés il est possible, après avoir effectué des préparations, de mesurer la tension des cordes vocales et son influence sur F_0 mais uniquement pour la tension passive.

Donc en pratique, on ne dispose que des valeurs relatives à la longueur des cordes vocales, à leur épaisseur, à l'activité E.M.G. des muscles qui commandent la tension, mais non des données relatives à cette tension elle-même.

Une communication, celle de MAEDA présente des relevés qui sont liés à ce problème. D'habitude la mesure de la longueur des cordes vocales se fait sur un film laryngoscopique obtenu à l'aide d'un miroir disposé le plus loin possible dans la cavité buccale ou à l'aide d'une fibre optique introduite dans le pharynx par la cavité nasale. MAEDA utilise des films radiographiques sur lesquels il mesure la longueur des ventricules (les

ventricules sont les espaces entre les vraies cordes vocales et les fausses) et il constate qu'au cours d'une phrase énonciative la longueur des cordes vocales décroît régulièrement avec une variation totale de 3,8 mm.

Cette méthode permettant d'estimer la longueur des cordes vocales semble plus facile à mettre en oeuvre que les autres. Mais elle présente une difficulté que l'auteur a lui-même soulignée : la partie postérieure des ventricules est souvent masquée par les aryténoïdes. Je voudrais qu'il nous dise s'il considère que c'est une méthode assez précise pour estimer la longueur des cordes vocales.

Dans ce problème des commandes du larynx se situe aussi la communication de GUERIN & BOË.

En utilisant un modèle à deux masses (modèle ISHIZAKA & FLANAGAN), les auteurs, dans un premier temps, ont déterminé les relations qui apparaissent entre les masses et leurs raideurs lorsque l'on fait varier tous ces paramètres indépendamment : le critère retenu était la forme de l'onde glottique.

Après un traitement statistique, les auteurs ont retenu deux paramètres de commande :

P_s la pression subglottique ,

Q qui correspond à une tension entraînant une diminution de l'épaisseur des cordes vocales.

Ensuite, ils ont étudié la part respective de ces deux commandes pour des variations de F_0 se situant dans un intervalle d'une octave correspondant à la voix de poitrine chez un locuteur masculin.

On constate que les variations de F_0 avec le modèle sont plus imputables au paramètre Q qu'au paramètre P_s .

Ce résultat peut donc être utilisé pour la synthèse paramétrique avec une source de ce type. Il reste évidemment à déterminer les occurrences respectives des interventions de P_s et de la tension des cordes vocales dans la parole continue et cela par rapport à une analyse linguistique.

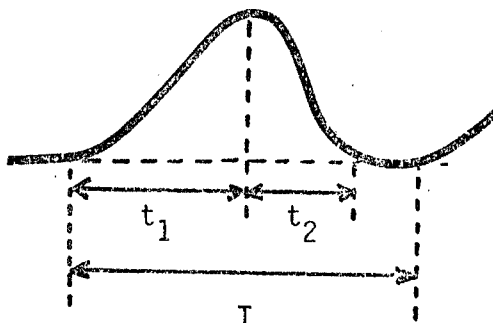
2 - FORME DE L'ONDE GLOTTIQUE

Le larynx fonctionne comme un générateur de débit plutôt que comme une source de pression dans la mesure où on estime généralement que l'impédance d'entrée de la charge, c.a.d. le conduit vocal est faible par rapport à l'impédance interne de la source. En fait il semble bien que ce ne soit pas toujours le cas. Mais si la connaissance de l'impédance d'entrée du conduit vocal a progressé assez bien, en particulier avec les travaux de MRAYATI & al. (1976) (parus dans ACUSTICA), l'impédance de la glotte n'est pas facile à estimer.

Pour en revenir à l'onde glottique, il est possible de l'estimer par des procédés que je classerai en deux grandes catégories, selon qu'ils opèrent:

- à partir du signal de la parole avec tous les procédés de filtrage inverse. Mais ceci suppose la connaissance exacte de la fonction de transfert du conduit vocal pour le son sur lequel on opère le filtrage inverse,
- à partir de la mesure de l'aire de la glotte et de la valeur de la pression subglottique. Mais il faut faire quand même certaines approximations.

De nombreux travaux ont été menés dans ce domaine et on estime que l'onde de débit a la forme suivante :



La forme de l'onde glottique peut-être caractérisée par deux facteurs : le quotient d'ouverture $Q.O = t_1 + t_2 / T$, c'est à dire le temps d'ouverture des cordes vocales ramené à la période et le quotient de dissymétrie $Q.D = t_1/t_2$ rapport du temps d'écartement des cordes vocales au temps de rapprochement.

La simulation du modèle à deux masses (ISHIZAKA & FLANAGAN) donne un signal qui correspond bien à la forme relevée. LANDERCY dans sa communication a opéré une mesure de l'onde glottique : à partir de la 1^o méthode. Cette onde de débit, il l'a reconstituée par une méthode numérique d'analyse par synthèse après avoir dans un premier temps calculé la fonction de transfert du conduit vocal. Mais il obtient un signal pour lequel le temps de montée est plus court que le temps de descente c'est-à-dire inversé par rapport à la forme d'onde. Alors j'aimerais qu'il nous dise ce qu'il en pense.

3 - VARIATIONS QUI NE SONT PAS DUES AUX COUPLAGES ET PEUT-ETRE PAS AUX COMMANDES MAIS PLUTOT A L'EXECUTION DE LA COMMANDE

On sait qu'il existe des variations à court terme de la forme de l'onde glottique et de sa période d'un cycle à l'autre. De telles fluctuations ont été relevées, par exemple, par LIEBERMAN pour l'anglais et par RISBERG pour le suédois et le japonais. LANDERCY a fait des mesures pour des voyelles soutenues prononcées isolément. Il relève sur la période des fluctuations de 2 à 7 % et pour la forme de l'onde glottique des évolutions de 4 à 7 % pour t_1 et de 2 à 13 % pour t_2 .
Question : à quel phénomène peut-il relier ces observations ?

4 - VARIATIONS DUES AU COUPLAGE

DI CRISTO & CHAFCOULOFF ont mesuré les caractéristiques intrinsèques des consonnes sonores constrictives et occlusives, celles des voyelles et les influences des consonnes sur les voyelles.

En effet on sait que:

- toutes choses égales par ailleurs, les voyelles fermées ont une F_0 plus élevée que les voyelles ouvertes,
- les consonnes sonores ont une F_0 plus basse que les voyelles, ce phénomène étant nettement marqué pour les occlusives et les constrictives,
- au voisinage des consonnes la fréquence fondamentale des voyelles est influencée : après les consonnes sourdes elle est plus aiguë qu'après les sonores.

Ces variations sont au maximum de l'ordre du ton (12 %); on les appelle les caractéristiques intrinsèques dans la mesure où l'on considère qu'elles sont inhérentes au processus de production et qu'elles ne participent pas directement au code linguistique, système extrinsèque au processus de phonation.

Les caractéristiques intrinsèques ont été bien étudiées pour le français (en particulier à Grenoble et au CNET) et introduites pour la synthèse par IBM, le CNET, l'ENSERG.

Les travaux de DI CRISTO & CHAFCOULOFF reprennent certaines analyses et les complètent.

Je ne vais pas insister puisqu'ils font partie du thème II mais je voudrais leur poser une question. Elle est relative à l'origine de ces caractéristiques intrinsèques.

Pour les consonnes on sait bien expliquer les origines de ces variations de F_0 . Elles sont dues à l'effet de couplage entre la source et le larynx. Lors d'une occlusion ou d'une constriction, la pression supraglottique augmente et devient supérieure à la pression atmosphérique. La pression intraglottique diminue, ce qui est entraîné donc un abaissement de F_0 .

Cette évolution contamine aussi les voyelles adjacentes.

Pour les voyelles, l'explication est plus difficile :

- couplage acoustique ou
- couplage mécanique ?

L'impédance d'entrée étant liée à la valeur du 1^o formant, on a pu avancer l'hypothèse que c'était son influence qui entraînait en jeu. Les travaux de GUERIN réalisés avec une simulation à deux masses laissent à penser que ce n'est pas le cas.

Dans les voyelles ouvertes la langue est plus basse que dans les voyelles fermées. Cette position de la langue agirait sur la tension des cordes vocales ?

Questions :

. Pour quel type de couplage penchent les auteurs ?

. L'évolution de la F_0 pendant la tenue des consonnes intervocaliques est fonction du contexte prosodique dans lequel s'insère la consonne suivant que le schéma est montant ou descendant.

Les auteurs ont-ils une hypothèse à formuler sur le plan de la réalisation syllabique ?

5. STRATEGIE CONCERNANT L'INTERPRETATION DES EVOLUTIONS DE LA FREQUENCE FONDAMENTALE POUR UN ENONCE.

Nina THORSEN propose une procédure permettant de décomposer les variations de F_0 pour une phrase en différents constituants :

- l'intonation de l'énoncé,
- les faits d'accent,
- les caractéristiques intrinsèques.

Il s'agit donc d'une procédure traditionnelle. Nous pensons que l'auteur aurait intérêt à utiliser des corrections chiffrées en pourcentage, ou en tons, pour qu'elles puissent être utilisables quelle que soit la hauteur de la voix des locuteurs (voix masculines ou féminines en particulier).

Le travail de MAEDA est plus ambitieux; tout d'abord l'auteur propose une description des variations de F_0 pour les phrases énonciatives (en anglais-américain) schématisées par deux types d'évolutions superposées :

- le premier affecte les éléments de la phrase : montée, descente ou plateau (hat pattern),
- le second s'étend à toute la phrase : la ligne de déclinaison continue (base line).

Mais l'auteur tient à démontrer que cette analyse correspond effectivement à une réalité physiologique et en cherche les corrélats. Comme influence prépondérante concernant cette évolution globalement décroissante de F_0 est écartée la pression subglottique. Puis sont étudiées, en particulier, les variations de position verticale du thyroïde et de l'os hyoïde, la longueur des cordes vocales (étudiée, comme nous l'avons vu, par radiographie). Les corrélations montrent que décroissance de F_0 et écourtement de la longueur des cordes vocales sont bien corrélées.

L'auteur avance alors une hypothèse : cette diminution régulièrement décroissante de la longueur serait due essentiellement au basculement du cricoïde par rapport au thyroïde, mouvement provoqué par une augmentation de la tension trachéale.

Sans vouloir discuter cette dernière hypothèse, je voudrais revenir sur les étapes de la démonstration concernant l'existence même de la "base line". Pour l'anglais-américain (comme pour le français) il est possible de schématiser les évolutions de F_0 d'une phrase énonciative composée d'un syntagme nominal sujet SN et d'un syntagme verbal (SV $\rightarrow$ V + SN) par une montée réalisée sur SN et une descente sur SV.



Il est certain que si SN sujet est très court par rapport à SV, comme pour la phrase présentée par MAEDA (SN : Ken SV : raises the light yellow sheep) l'évolution de F_0 peut être alors réduite à une descente.

Je voudrais demander à MAEDA si dans son travail il a utilisé des phrases comportant des syntagmes nominaux plus importants et s'il a trouvé alors une approximation linéaire tout aussi satisfaisante ? Sur ce point j'aimerais avoir l'avis de COLLIER.

6 - INFLUENCE DES CONDITIONS DE COMMUNICATION SUR F_0 .

Pour le travail de FONAGY & al. nous ne traiterons ici que ce qui relève du thème I, le reste sera certainement abordé par les rapporteurs du thème III.

Les auteurs ont étudié les influences des conditions de communication sur la fréquence fondamentale. Ils ont enregistré 4 locuteurs dans trois situations différentes : exposé de type magistral, lecture d'un texte dans un studio d'enregistrement, quasi-monologue à la suite d'une question d'un interviewer. Les résultats ont été établis pour des enregistrements d'au moins 90 secondes chacun; ils portent sur la fréquence laryngienne moyenne, son écart type (le registre étant défini à $\pm 2\sigma$), l'hypothèse de distribution normale étant vérifiée.

Les résultats montrent de très nettes tendances : si on passe de l'exposé à la lecture puis à l'interview, la fréquence laryngienne moyenne décroît (de 1 ton à 1,5 ton) ainsi que le registre (de 2 à 3 tons).

Je demande aux auteurs si les variations ainsi observées ne seraient pas directement liées à des différences de niveau d'intensité sonore : on parle généralement plus fort pour un exposé que pour répondre à un seul interlocuteur ?

7 - ETUDE DE F_0 EN PSYCHIATRIE, NEUROLOGIE ET PATHOLOGIE.

Mme CHEVRIE-MULLER & DECANTE et GUBRYNOWICZ & al. nous proposent des procédures systématiques pour étudier, sous forme numérique, les évolutions de F_0 , dans le cadre de la psychiatrie, neurologie et pathologie du larynx.

Les premiers ont mis au point et utilisé une statistique très bien adaptée à ce genre de problèmes: ils mesurent la fréquence laryngienne moyenne, son écart type, la durée voisée, la durée des pauses et leur histogramme. Afin de caractériser précisément les fluctuations d'une impulsion à l'autre du signal glottique ils chiffrent le coefficient de VON NEUMAN qui est la valeur moyenne des carrés des différences d'une mesure à l'autre:

$$\Delta^2 = \frac{1}{n-1} \sum_{j=1}^n (x_{j+1} - x_j)^2$$

Ces mesures se doublent d'une appréciation subjective de la "quantité de mélodie" (variations de F_0 pour la phrase) et du débit, chacun des deux paramètres étant noté sur une échelle de 1 à 9.

Les résultats montrent une bonne corrélation entre la variance et l'appréciation subjective de la richesse mélodique, la durée des pauses et l'impression de débit. D'autre part il est possible d'interpréter le rapport Δ^2 / σ^2 en fonction du degré d'atteinte de la maladie de Parkinson.

GUBRYNOWICZ & al. présentent un travail qui s'intègre dans un système d'analyse acoustique de la voix pathologique en vue de l'établissement d'un diagnostic sur le comportement phonatoire du larynx. F_0 est étudié statistiquement après numérisation : calcul de la moyenne, écart type, moment, aplatissement, visualisation de la densité de probabilité. Il s'agit d'un outil bien adapté pour la phoniatry.

En m'adressant à Mme CHEVRIE-MULLER je voudrais lui demander si elle pense augmenter sa vitesse d'échantillonnage, 50 ms me semblant un peu long, ainsi que la durée du corpus ?

Il me semble, et là ma réflexion appelle une réponse de tous les auteurs de ce sous-thème, que si l'analyse statistique globale est une étape indispensable, l'analyse segmentale détaillée pourrait certainement constituer une deuxième partie riche d'interprétation. En effet de nombreux travaux physiologiques permettent de relier les variations du fondamental, par exemple dans le cas de l'accent ou de l'intonation, à telle ou telle action musculaire. L'étude de la localisation des pauses me semble aussi intéressante.

Communications rapportées:

- Programme pour le traitement automatique des données obtenues par extraction du fondamental de la parole (application à la pathologie de la prosodie en psychiatrie et en neurologie).
C. CHEVRIE-MULLER & P. DECANTE.
- Incidence des variables situationnelles sur quelques paramètres de l'intonation.
I. FONAGY & V. LUCCI, L-J. BOË.
- Evaluation de l'état pathologique des cordes vocales d'après l'analyse des variations du fondamental.
R. GUBRYNOWICZ, W. MIKIEL & P. ZARNECKI.
- Variations à court terme de la fonction de source vocale.
A. LANDECY.
- La régulation de la vibration des cordes vocales : simulation à l'aide d'un modèle à deux masses.
B. GUERIN & L-J. BOË.
- Sur les corrélatifs physiologiques de la fréquence du fondamental de la parole.
S. MAEDA.
- Les faits microprosodiques du français : voyelles, consonnes, coarticulation.
A. DI CRISTO & M. CHAFCOULOFF.
- On the interpretation of raw fundamental frequency tracings.
N. THORSEN.

Sous-thème :

- Détection de F_0 et du voisement

Rapporteur : *Michel CARTIER*
C.N.E.T. - Lannion

MESURE DU FONDAMENTAL :

PASSE, PRESENT,... FUTUR ?

INTRODUCTION

Le problème de la mesure du fondamental est l'un des "serpents de mer" du traitement de la parole. Il le restera sans doute longtemps puisqu'il s'agit d'attribuer une fréquence à un signal qui n'est pas strictement périodique. Cependant, périodiquement (*), le hasard ou les progrès de l'informatique conduisent à une nouvelle idée ; parallèlement, l'évolution de la technologie permet une réalisation impensable quelques années plus tôt ; enfin, l'une des raisons du récent regain d'intérêt se trouve sans doute dans l'analyse-synthèse par prédiction linéaire, qui a confirmé que la qualité d'un vocodeur est très fortement liée à son "détecteur de pitch". Il n'est donc pas étonnant que la littérature sur le sujet se soit encore fortement enrichie ces dernières années, en particulier lors des sixièmes, septièmes et huitièmes Journées d'Etudes sur la Parole.

RAPPELS ET PRESENTATION DES COMMUNICATIONS

Les idées de base sont très anciennes. Cependant, en fonction de l'évolution des techniques, les différents procédés ont connu des fortunes diverses et se sont renouvelés au fur et à mesure. De nombreuses synthèses, plus ou moins condensées, ont été faites (par exemple : Mc Kinney, 1968 ; Léon et Martin, 1969 ; Pirogov, 1974 ; Rabiner et al., 1976). Notre propos n'est bien sûr pas de faire un exposé complet sur la question : le chapitre de Pirogov, quoique incomplet sur certains points, comporte plus de quatre-vingt-dix pages... Il s'agit simplement de rappeler l'évolution de la question et de situer les neuf communications présentées.

Parmi diverses classifications possibles, adoptons-en une qui soit commode

(*) pourquoi pas ?!

en fonction du menu de ce jour : nous distinguerons les méthodes spectrales, les méthodes de décalage temporel et les méthodes de détection d'événements.

L'analyse spectrale est difficile à mettre en oeuvre et, malgré son intérêt théorique, a connu peu de développements. Le procédé le plus connu dans cette catégorie est le "cepstre" (Noll, 1968). Signalons également le "spectre de produit" (Schroeder, 1968). Le système "Ciphon" (Potage et Bourgenot, 1975) utilise pour la mesure du fondamental une analyse spectrale des variations d'amplitude du signal filtré dans une bande située au milieu du spectre.

Les méthodes de décalage sont associées à l'idée de ressemblance entre périodes voisines. Si la fonction de ressemblance est une fonction de corrélation, l'équivalent spectral apparaît immédiatement, à moins que la corrélation soit précédée d'un traitement non-linéaire, ce qui est généralement le cas (Gill, 1961 ; Sondhi, 1968 ; Frei, 1973 ; Galand, 1974). Un procédé de philosophie très voisine est celui de la fonction de différence (Miller, 1956 ; Meo et Righini, 1971 ; Ross et al. 1974) où l'on calcule simplement la valeur moyenne de la différence entre les valeurs du signal et du signal décalé. L'article de Meo et Righini présente le problème en termes de reconnaissance de formes, celui de Ross associe à la fonction de différence un algorithme de lissage et de correction (AMDF : average magnitude difference function).

Nous nous intéresserons davantage à la troisième catégorie à laquelle se rattachent pour l'essentiel toutes les communications qui nous intéressent. Il s'agit des méthodes fondées sur la détection d'événements $i(t)$. Ces événements peuvent être des passages par zéro, ou des maxima ou minima locaux. Ils suivent toujours un prétraitement du signal brut et sont suivis d'un dispositif de mesure(s) et d'une logique de décision et, ou, de lissage. Cette dernière partie du traitement n'est parfois qu'évoquée, voire invoquée.

1. *Extraction du fondamental*. C'est peut-être le procédé le plus ancien (ex : Grützmacher, 1937) ; il a constamment évolué et s'est beaucoup perfectionné. De nombreuses variantes sont possibles selon le nombre de filtres utilisés, leur adaptativité éventuelle, les procédés de détection de crêtes ou de passages par zéro et les algorithmes de choix.

La communication de B. Teston et M. Rossi dépasse le cadre de la détection de F_0 ; on décrit un appareillage où l'estimation de période est faite sur l'in-

tervalle séparant les passages par zéro du signal filtré. Dans cette version, le filtre de mise en forme est fixe. Ce procédé fonctionne de façon satisfaisante si l'on peut commuter manuellement le filtre en fonction de l'utilisation. On trouvera des descriptions voisines dans des articles de Carré et al. (1963), et Boë et Rakotofiringa (1971).

La commutation des filtres peut être remplacée par un filtre unique variable. C'est l'option décrite dans la communication de J. Caelen et P. Cazenave. Le filtre adaptatif, du second ordre, est accordé en fonction des valeurs estimées précédentes, du voisement et d'une logique de décision. Le programme correspondant fonctionne en temps réel sur IRIS 80 ; il a été utilisé conjointement avec l'appareil de Teston cité précédemment, dans l'étude des corrélations entre mélodie, intensité et durée (8èmes JEP).

Dans la même ligne générale (filtrage et passages par zéro) Ph. Martin préfère une troisième option qui évite commutations et adaptations : l'appareil décrit comporte un banc de filtres dont la différence des temps de propagation de groupe respectifs est compensée par des lignes à retard. Plusieurs autres particularités intéressantes sont indiquées : l'emploi d'une détection de seuil symétrique, d'une logique de décision programmée, et la visualisation de courbes mélodiques qui permet d'utiliser l'appareil pour la recherche phonétique et l'enseignement des facteurs prosodiques. Les principes qui ont guidé le choix des filtres sont exposés dans Léon et Martin (1968) ; ils permettent une grande précision de la mesure.

2. *"Régénération" du fondamental.* Le principe de base consiste à fabriquer une ou plusieurs pointes par période grâce à un prétraitement suivi d'une détection de pics plus ou moins élaborée.

La détection de pointes est l'un des principes les plus anciens, les plus étudiés... et les plus éprouvés. Les premiers travaux remontent loin (ex : Gruenz et Schott, 1949). Des perfectionnements ont été apportés par la mise en cascade de réseaux de détection et de différentiation (Dolansky, 1955). Plusieurs détecteurs indépendants peuvent être utilisés en parallèle, avec un algorithme de choix parmi les diverses valeurs estimées à partir de chaque détecteur (Gold et Rabiner, 1969). Le procédé décrit par J.F. Zurcher, implanté en temps réel, est sans doute l'un des plus élaboré. L'essentiel a été présenté aux 6èmes JEP (J.F. Zurcher et al., 1975). La communication décrit en détail la détection de

crêtes : après filtrage et régulation d'amplitude, la détection se fait par un circuit bouclé qui optimise le seuil de détection et permet de favoriser celle-ci lorsque la pente du signal est forte. On évite ainsi la confusion entre F_0 et F_1 dans le cas où le premier harmonique de F_0 coïncide avec F_1 . Les résultats indiqués sont excellents pour des conditions correctes de prise de son quelle que soit la voix, masculine ou féminine. Les causes principales des dégradations examinées sont un bruit important pour les locuteurs féminins et un filtrage très sévère à 300 Hz pour les locuteurs masculins. La communication de I. El Mallawany et J.F. Zurcher reprend le principe précédent. Elle décrit la simulation numérique de celui-ci, étudie des simplifications et des perfectionnements compatibles avec des calculs en virgule fixe. Les filtres d'entrée ont été modifiés ; quelques erreurs ont été observées au début de certains segments voisés, qui impliquent, comme dans la version analogique (Zurcher, 1975), une détection de crêtes sur deux voies positive et négative. Le pourcentage d'erreurs observées est très faible. L'introduction des calculs en virgule fixe et d'un algorithme parallèle pour accélérer le traitement n'a pas diminué les performances.

Au lieu de travailler sur le signal $s(t)$, la détection de maxima peut opérer sur la puissance à très court terme (Maïssis, 1973). C'est l'option décrite dans la communication de D. Dours, R. Facca et G. Pérennou. N'ayant pas eu de problème pour la mesure de F_0 proprement dite, les auteurs se sont préoccupés de la fermeture glottique. Celle-ci peut d'ailleurs s'obtenir, dans des perspectives d'application différentes, directement sur le locuteur par un glottographe qui mesure des variations d'impédance électrique de chaque côté du cartilage thyroïde (voir la communication de S. El Samkari et F. Hamelin). L'algorithme de Dours fonctionne avec un automate de segmentation. La fonction dont on détecte les pics est une valeur absolue moyenne, à la différence du procédé de Maïssis qui utilise une valeur efficace. La détection fonctionne en général de façon satisfaisante. Des corrections sont effectuées en temps différé en fonction de la position des zones irrégulières dans le segment voisé correspondant. Cet algorithme est utilisé dans le projet de reconnaissance A.R.I.A.

Une autre façon de régénérer le fondamental est de procéder par filtrage inverse du signal limité aux basses fréquences. Ce traitement peut être suivi d'une détection de crêtes ou d'une corrélation (Markel, 1972). Le filtrage inverse est lourd à mettre en oeuvre. Une solution simplifiée, utilisant le signal d'erreur d'un prédicteur d'ordre deux et la stabilité de ce dernier, a été exposée aux 6èmes JEP (Le Roux, 1975). Une autre méthode, également assez simple est présentée

dans la communication de R. Hess : après un premier filtre passe-bas fixe, on élimine le premier formant par un filtre en peigne variable déterminé à partir d'un histogramme de passages par zéro. La méthode, qui concilie simplicité et efficacité, permet de localiser la fermeture glottique. L'algorithme fonctionne en 20 fois le temps réel sur PDP 11/45 et peut être implanté en temps réel après une adaptation des étapes de correction.

3. *Signal analytique.* Notre panorama souffrirait d'un grand vide sans le signal analytique. Ici, l'événement détecté est le retour sur lui-même, après un intervalle de temps égal à la période du fondamental, d'un vecteur à $2N$ composantes : N composantes spectrales du signal et leurs transformées de Hilbert (Rader, 1964). Si les N filtres sont assez étroits, le procédé se ramène à une autocorrélation (Sondhi, 1964). La trajectoire du signal analytique à deux dimensions a également été étudiée, et appliquée à la détection de périodicité (Lobanov, 1971).

4. *Modèle d'audition.* Les procédés précédents se réfèrent à un traitement du signal purement physique. La communication de L. Bastet, J.M. Dolmazon et B. Guérin apporte un éclairage différent : leur méthode est fondée sur un modèle fonctionnel du système auditif. Le but du travail, plus que l'obtention de tracés de F_0 , est d'améliorer la connaissance du fonctionnement du système auditif périphérique en testant la validité de modèles. Le modèle présenté comporte vingt voies de détection, dont chacune est constituée de deux filtres résonants voisins et d'un détecteur qui se déclenche lorsque les sorties des deux filtres sont suffisamment déphasées, ce qui ne se produit que dans une bande étroite de fréquences. La conclusion essentielle du travail est que les déphasages en sortie de la membrane basilaire permettent d'évaluer F_0 . Les performances du détecteur seront améliorées par l'augmentation du nombre de cellules et le perfectionnement de la logique de décision. On retrouve là, toutes proportions gardées, une philosophie d'organisation qui rappelle celle de Ph. Martin : batterie de filtres, détecteurs à seuil, microprocesseur...

DISCUSSION

Malgré l'unité du thème "mesure du fondamental", les centres d'intérêt étaient trop divergents chez les auteurs pour permettre une discussion constructive.

Les commentaires ont été axés sur les points suivants :

1. *Le cadre d'application.* En effet, le choix d'une méthode est le résultat de compromis entre le temps de réponse, la complexité, la compatibilité avec les équipements connectés, les possibilités d'ajustage manuel et de corrections après coup, la qualité de la prise de son.

2. *La "cuisine".* La qualité des résultats obtenus est plus fonction de l'énergie consacrée à la question et de l'utilisation de raffinements empiriques que des propriétés intrinsèques de la philosophie qui a guidé le choix de tel ou tel procédé.

3. *Lissage et corrections.* Plus le retard accepté est grand, plus les corrections sont aisées. En fait, tout dépend du but recherché : précision de la mesure, fiabilité du résultat, importance des erreurs.

4. *Problèmes de principes et résultats.* Il serait intéressant de comparer les résultats pour des signaux d'entrée d'origines diverses. Une telle étude a été faite aux Etats-Unis, dont il n'est pas facile de déduire des pourcentages de fonctionnement correct malgré les nombreuses précisions publiées (Rabiner et al., 1976). Une seconde étude a d'ailleurs montré que les relations entre résultats objectifs et subjectifs ne sont pas évidentes (Mac Gonet al et al., 1977)...

CONCLUSION

Il n'y a pas eu ces dernières années de progrès révolutionnaires, mais il existe maintenant des procédés éprouvés qui, moyennant quelques contraintes d'emplois, fonctionnent effectivement pour les recherches en phonétique, en rééducation, en reconnaissance, en synthèse, et pour la transmission de la parole. Une évaluation comparative serait difficile à mettre en oeuvre et surtout à définir, tant sont différents, en fonction des applications envisagées, les signaux à traiter et le sort des valeurs obtenues.

A moins qu'un modèle d'audition ne remette en cause la notion de fondamental, toute nouvelle méthode apparaîtra comme une cousine (à la mode de Bretagne ?) d'un procédé vieux de plus de dix ans. Envisageons donc modestement l'avenir : d'abord, enrichir notre connaissance des erreurs de mesure et de leurs conséquences

en fonction des autres paramètres acoustiques et du contexte phonétique, le meilleur critère d'appréciation paraissant être l'écoute d'un signal synthétisé à partir de la courbe estimée en comparaison avec une courbe de référence ; ensuite, en fonction des critères qui pourraient se dégager de cette connaissance, améliorer et optimiser les algorithmes connus vis-à-vis des détériorations acoustiques et électriques que rencontrent inévitablement les signaux de parole "idéaux".

BIBLIOGRAPHIE

- BOE L.J. et RAKOTOFIRINGA H., 1971 : Exigences, réalisation et limites d'un appareillage destiné à l'étude de l'intensité et de la hauteur d'un signal acoustique - Revue d'Acoustique, 4ème année, n° 14, pp. 104 - 114.
- CARRE R. et al., 1963 : Etude et réalisation d'un détecteur de mélodie pour l'analyse de la parole - l'Onde Electrique, n° 434, pp. 556 - 562.
- DOLANSKY L.O., 1955 : An instantaneous pitch-period indicator - J.A.S.A., vol. 27, n° 1, pp. 67-72.
- FREI A.H. et al., 1973 : Adaptive predictive speech coding based on pitch-controlled interruption-reiteration techniques - Int. Conf. on Communications, 11 - 12 - 13 juin, Seattle, Washington.
- GALAND C. et al., 1976 : Détection de la mélodie par autocorrélation non linéaire - 7èmes J.E.P., G.A.L.F., Nancy 19 - 21 mai, pp. 333 - 345.
- GILL J.S., 1961 : Automatic extraction of the excitation function of speech - 3ème C.I.A.
- GOLD B. et RABINER L.R., 1969 : Parallel processing techniques for estimating pitch periods of speech in time domain - J.A.S.A., Vol. 46, pp. 442 - 448.
- GRUENZ O.O. et SCHOTT L.O., 1949 : Extraction and portrayal of pitch of speech sounds - J.A.S.A., vol. 41, n° 5, pp. 487-495.
- GRUTZMACHER M. et LOTTERMOSER W., 1937 : Uber ein Verfahren zur trägheitsfreien Dufzeichnung von Melodiekurven - Akustische Zeitschrift, 5es Heft, Band 2, pp. 242-248.
- LEON P.R. et MARTIN Ph., 1969 : Prolégomènes à l'étude des structures intonatives, Ed. Didier, Montréal, Paris, Bruxelles.
- LE ROUX J., 1975 : Une méthode synchrone d'extraction en temps réel du fondamental - 6èmes J.E.P., G.A.L.F., Toulouse 28-30 mai, pp. 2 - 12.

- LOBANOV B., 1971 : Automatic discrimination of noisy and speech sounds by the phase-plane method - Soviet Physics-Acoustics, vol. 16, n° 3, pp. 353-356.
- MAISSIS A., 1973 : Une méthode d'extraction du fondamental - L'Onde Electrique, vol. 53, n° 3, pp. 110 - 112.
- Mc GONEGAL et al., 1977 : A subjective evaluation of pitch detection methods using LPC synthesized speech - IEEE-ASSP, vol. 25, n° 3, pp. 221-229.
- Mc KINNEY N.P., 1968 : Laryngeal frequency analysis for linguistics - Research communic. Sci. lab. Univ. Michigan, Report 14, Contract 1224 (22) NR 049/122.1968.
- MARKEL J.D., 1972 : The SIFT algorithm for fundamental frequency estimation - I.E.E.E. AU-20, n° 5, pp. 367-377.
- MEO A.R. et RIGHINI G., 1971 : A new technique for analyzing speech by computer - Acustica, vol. 25, n° 5, pp. 261-268.
- MILLER R.L. et WEIBEL E.S., 1956 : Measurement of the fundamental period of speech using a delay line, 51st ASA Meeting, DC 10.
- NOLL A.M., 1967 : Cepstrum pitch determination - J.A.S.A., vol. 41, pp. 293 - 309.
- PIROGOV A.A., 1974 : Vokodernaja telefonija, Ed. Svjaz', Moscou, pp. 309 - 418.
- POTAGE J. et BOURGENOT J.S., 1975 : Transmission de la parole à 1200 eb/s ; le vocodeur "Ciphon" - Colloque F.A.S.E., 29-9 au 3.10.75, Paris, Ed. C.N.E.T., pp. 339-349.
- RABINER L.R. et al., 1976 : A comparative performance study of several pitch detection algorithms - I.E.E.E. ASSP , vol. 24, n° 5, pp. 399 - 418
- RADER C.M., 1964 : Vector pitch detection - J.A.S.A., vol. 36, n° 10, p. 1963.

ROSS M.J. et al., 1974 : Average Magnitude Difference Function pitch extractor - I.E.E.E. ASSP, vol. 22, n° 5, pp. 353 - 362.

SCHROEDER M.R., 1968 : Period histogram and product-spectrum : new methods for fundamental frequency measurement - J.A.S.A. , Vol. 43, n° 4, pp. 829-834.

SONDHI M.M., 1964 : Equivalence of "vector" and autocorrelation pitch detectors - J.A.S.A., vol. 36, n° 10, p. 1964.

SONDHI M.M., 1968 : New methods for pitch extraction - I.E.E.E. AU-16, n° 2, pp. 262-266.

ZURCHER J.F. et al., 1975 : Détection et mesure du fondamental - 6èmes J.E.P., G.A.L.F., Toulouse 28-30 mai, pp. 11-21.

THEME 2

La mélodie : études perceptives et synthèse de la parole.

Sous-thèmes :

- La perception de la hauteur et de ses variations (micromélodie)
- Corrélations entre l'intensité, la durée et la fréquence fondamentale
- Rôle des traits prosodiques dans la transmission et la synthèse de la parole.

Président : M. ROSSI
Rapporteurs : J. GENIN, M. WAJSKOP (*)

(*) M. WAJSKOP était absent pour raisons de santé

Sous-thème :

- Rôle des traits prosodiques dans la transmission et la synthèse de la parole.

Rapporteur : *Jacques GENIN*
C.N.E.T.-Lannion

ROLE DES TRAITS PROSODIQUES DANS LA TRANSMISSION ET LA SYNTHÈSE DE LA PAROLE

Cette session pêche par manque d'homogénéité. Les communications 1, 3, 5 et 7 n'ont pas de rapport entre elles et ne peuvent donc pas faire l'objet d'une discussion comparative. Les textes correspondants ayant été publiés par ailleurs il est inutile de les rappeler ici.

Les communications 2, 4 et 6 constituent par contre le noyau de la session.

Prosodie automatique pour la synthèse par diphonèmes (C. CHOPPY, J.P. LIENARD).

- Analyse de la structure intonative de la phrase en français (D. LARREUR, F. EMERARD).

- Utilisation d'un modèle prosodique pour la synthèse par concaténation de mots en français et en anglais (P. MARTIN).

On y ajoutera la communication suivante, présentée mais non publiée dans les actes des Journées d'Etudes :

- Génération automatique de la mélodie et de la micro-mélodie. Contribution à un système de synthèse par règle (X. RODET, L. POIROT).

Les différents laboratoires qui travaillent en synthèse de la parole détiennent actuellement des résultats de qualité suffisante pour nombre d'applications. Cependant, le problème de la synthèse correcte d'une phrase quelconque reste imparfaitement résolu. Une part importante en revient à la commande de l'évolution de la prosodie.

Les communications présentées dans cette session donnent en trois approches différentes l'essentiel des travaux effectués sur cette question pour la langue française.

ANALYSE DE LA PROSODIE

Sur le plan de l'analyse on opposera la recherche d'une prosodie moyenne (CHOPPY, LIENARD) à l'étude d'une voix particulière (LARREUR, EMERARD). La première solution est la plus intéressante pour la recherche fondamentale, la seule qui puisse donner des résultats valables aux yeux des phonéticiens. Un écueil, cependant, réside en ce que cette prosodie moyenne puisse ne pas exister, ne pas correspondre à un phénomène réel. Les auteurs ne méconnaissent pas ce problème et le signalent par exemple pour les phrases interrogatives ; sont-ils sûrs de ne pas l'avoir négligé ailleurs ?

Par contre la prise en compte d'un seul locuteur est avancé comme la meilleure garantie d'obtenir une voix naturelle, concernant sa personnalité au mieux de ce que permettent la technologie et les méthodes employées. Ce caractère est à rechercher tant dans le sens d'évolution des paramètres que dans leurs valeurs quantitatives.

MODELE PROSODIQUE

Au-delà du choix et de l'analyse des données retenus, il convient de bâtir un modèle de production prosodique. On s'aperçoit ici que la syntaxe gêne suffisamment les auteurs pour ne pas croire à sa totale inutilité...

CHOPPY et LIENARD prennent le parti de l'ignorer totalement, faisant en fait une analyse syntaxique sommaire dans l'avouer par la recherche de mots grammaticaux classés dans un dictionnaire. Ils arrivent par ce biais à bâtir un contour prosodique pour des phrases simples bien découpées. En effet, dans l'exemple présenté dans le texte "elle", "a", "une" et "très" permettent de localiser les composants majeurs de la phrase.

Que donnera ce modèle avec un sujet agrémenté d'épithètes et un verbe sans auxiliaire complété d'un adverbe, s'il faut croire avec LARREUR, EMERARD et MARTIN que la position du verbe est fondamentale pour obtenir un bon résultat ?

Par exemple "la jeune fille blonde achètera demain une robe". Il n'est pas question de condamner la méthode. Elle est sans doute indispensable pour faire la synthèse d'un texte orthographique elle n'est suffisante que pour des énoncés simples.

LARREUR, EMERARD et MARTIN confient à d'autres l'analyse syntaxique mais prennent le parti d'en utiliser les résultats. On aboutit ainsi à une étude de la prosodie pour les structures syntaxiques les plus complexes de la langue française (et analyse pour MARTIN). Certains points de désaccord apparaissent entre ces deux communications. Cependant le choix d'un locuteur unique d'un côté, l'ignorance où l'on se trouve de la nature du "corpus" de l'autre interdisent de conclure. Un comportement prosodique différent ne dénoterait une erreur que si toutes les analyses avaient été faites sur des moyennes...

SYNTHESE DE LA PROSODIE

Enfin, il faut se souvenir que la prosodie est un problème assez complexe pour que l'application à la synthèse des résultats de l'analyse soit assez difficile. Pour cela nous attacherons une valeur particulière aux travaux qui ont réellement été vérifiés en synthèse automatique de la parole, ceux de CHOPPY - LIENARD implantés sur l'Icophone et repris par RODET au C.E.A., ceux de LARREUR-EMERARD implantés sur le synthétiseur à canaux au C.N.E.T.

Ceux de MARTIN appliqués à une synthèse par mots paraissent plus un outil d'analyse qu'un système de synthèse.

THEME 3

L'intonation : aspects linguistiques et reconnaissance de la parole.

Sous-thèmes :

- Intonation et syntaxe
- Rôle de l'intonation dans la segmentation de l'énoncé et la reconnaissance de la parole
- Intonation et communication parlée.

Président : G. FAURE

Rapporteurs : A. DI CRISTO, J.S. LIENARD, Ph. MARTIN

Sous-thème :

- Méthodes et modèles d'analyse dans les recherches sur l'intonation.

Rapporteur : A. DI CRISTO,
Institut de Phonétique Aix-en-Provence

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

Le nombre élevé des communications présentées dans le cadre du troisième thème de ce Colloque, ainsi que leur regroupement au sein des trois sous-thèmes suivants :

- intonation et syntaxe,
- rôle de l'intonation dans la segmentation de l'énoncé et la reconnaissance de la parole,
- intonation et communication parlée,

peuvent, à première vue, donner l'impression que nous sommes en présence de recherches très diversifiées. Un examen attentif des communications révélera, cependant, que la diversité n'est qu'apparente. Les recherches qui traitent des relations entre l'intonation et la syntaxe ne concernent à vrai dire, que l'un des multiples aspects de cette problématique : l'étude de l'actualisation des unités syntaxiques de la phrase par les éléments prosodiques. De même, il est manifeste que le vaste domaine des fonctions assumées par l'intonation dans la communication parlée est à peine effleuré par les auteurs. En vérité, la lecture des différentes communications fait apparaître que les recherches ont été généralement orientées vers un même but : la définition des indices prosodiques (acoustiques et perceptuels) qui permettraient de localiser des frontières relatives aux niveaux syllabique, lexical, sémantique et syntaxique et qui pourraient, de ce fait, être intégrés dans un processus d'analyse syntaxico-sémantique de la phrase.

Dans cette perspective et en dépit de la diversité des méthodes mises en oeuvre, il semble qu'un consensus entre les auteurs soit établi au moins sur deux points fondamentaux :

- l'hypothèse d'une relation forte (ou semi-forte) entre la structuration syntaxique et la structuration prosodique,
- l'utilité des marqueurs prosodiques pour la reconnaissance automatique de la parole.

Nous laissons à nos collègues, Ph. MARTIN et J.S. LIENARD, le soin de présenter, respectivement, les modèles théoriques et les exemples d'application proposés par les différents auteurs.

Nous nous limiterons à une présentation synthétique des méthodes d'analyse employées par ces derniers. Cet exposé, au cours duquel nous serons amené, parfois,

à évoquer des communications qui ne figurent pas dans le thème 3, portera sur les points suivants :

- le matériau analysé,
- les modèles et les méthodes d'analyses présentés par les auteurs.

I. LE MATERIAU ANALYSE

On ne manque pas d'être frappé, en premier lieu, par le fait que presque toutes les communications se fondent sur l'analyse de corpus lus (phrases ou textes plus importants). On est en droit de se demander, à ce sujet, si les résultats obtenus sont vraiment représentatifs de l'objet qu'on se proposait d'étudier : les structures intonatives d'une langue donnée ou s'ils ne contribuent pas plutôt à établir l'inventaire des stéréotypes prosodiques de la lecture.

La question mérite d'être posée si nous considérons, par exemple, les conclusions présentées par FONAGY, LUCCI et BOE (1), au terme de leur analyse comparative de l'incidence des variables situationnelles sur la prosodie. Ces auteurs montrent, en effet, que les trois types de message qu'ils ont examinés se caractérisent par des divergences intonatives nettement marquées.

De même, MARTIN (2) signale, à juste titre, que la lecture entraîne inévitablement la mise en oeuvre de redondances prosodiques. La prise en considération de ces redondances peut être bénéfique à la synthèse, dans la mesure où elle est susceptible d'accroître son efficacité ; mais elle risque d'être néfaste à la reconnaissance automatique, car elle peut conduire à fonder celle-ci sur la détection d'indices qui ne seront pas forcément présents dans la parole.

A la lumière de ces réflexions, il nous semble que l'étude de SAINT-BONNET et BOE (3), qui traite du problème intéressant et si peu étudié en français des relations entre le débit d'élocution et la durée des pauses et des groupes rythmiques, mais dont le corpus est constitué par des lectures d'un texte littéraire, demeure, malgré sa qualité indiscutable, très limitée dans ses applications.

Le choix du corpus lu présente, à notre avis, un autre inconvénient majeur.

(1) cf. Volume 1 des Actes, pp. 14-17.

(2) cf. Volume 1 des Actes, pp. 248-54.

(3) cf. Volume 1 des Actes, pp. 337-43.

Il conduit trop souvent à forger des exemples qui n'ont qu'un rapport très éloigné avec les énoncés attestés dans la conversation usuelle.

Il serait facile d'illustrer cette remarque en donnant ici la liste des phrases analysées par les différents auteurs.

Nous remarquerons que cette tendance est dangereuse, car elle limite les investigations à l'analyse de métalangages ou de langages restreints, dont la manipulation est sans doute commode, mais qui ne permettent pas de dégager de façon exhaustive les lois qui régissent le code prosodématique (1).

Certes, l'analyse prosodique de langages restreints est en partie justifiée, lorsqu'on envisage une application spécifique comme celle du dialogue entre l'homme et l'ordinateur. Elle n'a plus aucune raison d'être si la finalité de la recherche n'est pas asservie à ce type d'application.

Le choix d'un corpus de phrases lues soulève une dernière question : est-il légitime d'entreprendre des recherches sur l'intonation en faisant abstraction de la situation et du contexte ?

Des recherches que nous avons effectuées à l'Institut de Phonétique d'AIX font apparaître que l'intonation d'une même phrase peut être réalisée différemment, selon que cette phrase est produite dans un cadre situationnel et dans un contexte donnés ou isolément. Ces recherches révèlent également qu'une intonation particulière n'est pas interprétée de la même façon en contexte et hors contexte.

Nous citerons, dans la même perspective, le travail de HAZAEL-MASSIEUX(2) qui montre, notamment, que, à de rares exceptions près, les phrases isolées sont, d'une part, difficiles à produire avec naturel et, d'autre part, relativement mal décodées.

Enfin, nous n'abandonnerons pas la discussion sur les critères de validité du corpus sans évoquer brièvement le problème des locuteurs.

(1) cf. les remarques de M. GROSS, à ce sujet : Remarques sur le rôle de la prosodie en reconnaissance, Actes des 6èmes Journées d'Etude sur la Parole, II, pp. 61-62.

(2) cf. Volume 1 des Actes, pp. 289-95.

Dans les communications présentées à l'occasion de ce Colloque, il est rarement fait état du nombre de locuteurs retenus pour la recherche et, encore moins, de la représentativité de ces derniers.

On remarque, cependant, que leur nombre est souvent très restreint. C'est ainsi, par exemple, que LARREUR et EMERARD (1) présentent une analyse de la structure intonative de la phrase française en se fondant sur un corpus de 400 phrases lues par un seul sujet.

Tous ceux qui ont déjà effectué des recherches dans le domaine concerné n'ignorent pas que les sources de variabilité qui conditionnent la réalisation des paramètres prosodiques diffèrent, dans leurs effets, d'un sujet à l'autre. C'est pourquoi nous pensons que les études prosodiques devraient porter sur un nombre important de locuteurs. Cette condition est indispensable si l'on veut éviter des confusions entre deux types de variations paramétriques : celles qui dépendent des caractéristiques individuelles ou socio-culturelles et celles qui sont engendrées par les instructions linguistiques.

Nous sommes surpris, à vrai dire, qu'un problème aussi fondamental que celui que nous venons d'évoquer ne soit pas abordé dans ce Colloque.

II. MODELES ET METHODES D'ANALYSE

La seconde partie de notre exposé est consacrée aux modèles et aux méthodes d'analyse proposés par les différents auteurs. Nous y traiterons successivement : des approches théoriques, des niveaux d'analyse et du choix des unités prosodiques retenues pour l'interprétation fonctionnelle.

1. LES APPROCHES THEORIQUES

Les communications présentées dans le thème 3 peuvent être regroupées en deux catégories, selon que la démarche adoptée par l'auteur est du type inductif ou déductif.

Le modèle d'analyse proposé par MARTIN (2) appartient au second type. Il

(1) cf. Volume 1 des Actes, pp. 228-36.

(2) cf. Volume 1 des Actes, pp. 305-11.

se fonde essentiellement sur une application des principes de présupposition réciproque et unilatérale aux unités sémantiques, syntaxiques et prosodiques de l'énoncé (1). Ces principes sont formalisés par un ensemble de règles qui opèrent sur des contours caractéristiques. Ces derniers sont définis par une matrice de traits binaires, établie sur la base de contrastes mélodiques de direction (+ montant / - montant), de durée (+ long / - long) et d'amplitude (+ ample / - ample) (+ restreint / - restreint).

MARTIN définit ensuite des règles qui permettent de dériver l'agencement syntagmatique des contours à partir de la structure prosodique parenthésisée.

L'algorithme de reconnaissance défini par MARTIN est d'un grand intérêt, car il permet, pour la première fois à notre connaissance, une intégration des mécanismes de neutralisation qui affectent les contours dans leur fonction identificatrice de la structure prosodique.

MARTIN ne cherche pas, du moins dans sa communication, à tester la valeur opérationnelle de son modèle. Cette expérience est tentée partiellement par THERRIEN-ROY, JACQUEMIN et BOE (2), qui entreprennent de vérifier la conformité acoustique des contours engendrés par la théorie déductive de MARTIN. Leurs résultats ne confirment que partiellement les hypothèses de ce dernier : certains des contours postulés par MARTIN ne sont pas attestés par l'analyse acoustique. LARREUR et EMERARD aboutissent, par ailleurs, à une conclusion similaire (3).

Il convient de remarquer cependant qu'aucun de ces auteurs ne tient compte des faits de neutralisation signalés par MARTIN. C'est pourquoi leurs critiques ne nous semblent pas totalement justifiées.

Bien que le modèle présenté par HIRST soit fondamentalement différent, dans sa démarche, de celui de MARTIN, il procède néanmoins d'une théorie entièrement axiomatisée.

HIRST (4) définit d'une part, un ensemble de traits intonatifs abstraits

(1) pour plus de détails, cf. MARTIN (1975).

(2) cf. Volume 1 des Actes, pp. 371-376.

(3) cf. Volume 1 des Actes, pp. 227-237.

(4) cf. Volume 1 des Actes, pp. 297-303.

(déterminés par la structure syntaxique) et, d'autre part, une matrice de traits prosodiques concrets dont l'application se veut universelle. Il montre ensuite, à l'aide d'exemples empruntés à l'anglais, que les traits prosodiques ne peuvent être affectés directement à partir de l'indicateur syntagmatique, mais doivent être déduits d'un autre niveau de représentation. Ce modèle présente de très sérieux avantages pour la synthèse et la reconnaissance, car il permet une dérivation directe des indices prosodiques à partir des traits intonatifs déterminés par la structure syntaxique.

ROSSI et DI CRISTO (1) posent les bases théoriques d'un modèle interprétatif qui s'apparente au modèle perceptuel d'analyse par synthèse défini par STEVENS, 1972.

Leur méthode consiste à opérer une sorte de "filtrage inverse" du signal, afin de détecter les traits qui seraient liés de façon univoque aux instructions linguistiques. Le modèle interprétatif comprend 4 ensembles de règles :

- R1 : règles d'effacement microprosodique,
- R2 : règles de conversion perceptuelle,
- R3 : règles intonosyntaxiques,
- R4 : règles intonotactiques.

Les auteurs présentent un exemple d'application de leur modèle à la segmentation en groupes intonatifs et à la reconnaissance des constituants syntaxiques.

Le modèle proposé par HATON, MARI, MESSENET, PIERREL et SANCHEZ (2) est intéressant à plus d'un titre. D'une part, les auteurs s'efforcent d'intégrer la prosodie à un système général de reconnaissance déjà constitué (programme MYRTILLE). D'autre part, ils projettent d'utiliser les indices prosodiques pour formuler des hypothèses au cours de la segmentation et au cours de l'analyse syntaxico-sémantique de la phrase.

Dans cette double perspective, les marqueurs prosodiques servent à localiser le point de départ du processus de reconnaissance et à limiter le nombre rhédictoire de solutions que l'analyseur attribue aux phrases.

(1) cf. Volume 1 des Actes, pp. 323-329.

(2) cf. Volume 1 des Actes, pp. 331-36.

Le programme est particulièrement alléchant, mais nous laisse quelque peu sur notre faim. Nous aurions souhaité (peut-être sommes-nous trop exigeant ?) que les auteurs ne se bornent pas à des considérations théoriques et nous présentent des résultats plus tangibles. Mais sans doute ces derniers nous apporteront-ils des précisions complémentaires, lors de la discussion.

La démarche adoptée dans les autres études est généralement du type inductif et analytique. La plupart des travaux reposent, en effet, sur l'analyse de données et sur leur interprétation directe. Il convient cependant de distinguer entre les recherches qui se limitent à des investigations purement acoustiques et celles qui soumettent les données objectives à une interprétation perceptuelle.

2. LES NIVEAUX D'ANALYSE

a) analyse acoustique

Parmi les recherches qui se fondent sur l'analyse des données acoustiques, on note une tendance croissante à une intégration plus large des différents paramètres, tendance qui semblait, jusqu'à présent, être plutôt l'apanage des recherches sur les faits d'accent et sur l'intonation expressive.

Cette tendance bénéfique se manifeste dans l'étude de CAELEN et MAURAND (1) où les auteurs s'attachent à préciser la nature des corrélations qui lient les paramètres de (F_0), de durée et d'intensité, dans l'intonation de la phrase énonciative française.

Il se dégage notamment de leur travail que les paramètres de (F_0) et d'intensité évoluent parallèlement dans les groupes prosodiques terminaux et de façon antagoniste au sein des groupes non finaux. Il semblerait donc, d'après ces résultats, que chaque groupe prosodique soit affecté d'une mélodie et d'une énergie spécifiques.

Cette recherche que nous estimons très fructueuse appelle cependant deux remarques :

- le choix de critères purement distributionnels conduit parfois les auteurs à faire abstraction de la hiérarchie des constituants syntaxiques. On se demande, à ce propos, s'il est légitime de classer indifféremment dans les groupes initiaux

(1) cf. Volume 1 des Actes, pp. 316-22.

un SN et un SP.

- d'autre part, il semble que la recherche de CAELEN et MAURAND ne soit pas fondée uniquement sur le traitement de données acoustiques brutes. Les auteurs affirment que "les données recueillies (concernant les trois paramètres) ont été minutieusement pondérées, afin de neutraliser les phénomènes microprosodiques et les caractéristiques intrinsèques". Pourrions-nous avoir quelques indications sur la façon dont sont apportées ces pondérations ?

Les variations de l'énergie sont également prises en considération par HATON et al. Mais ces derniers n'utilisent ce paramètre que dans leur algorithme récursif de recherche de syllabes. C'est aussi le cas de VIVES, LE CORRE, MERCIER et VAISSIERE, qui utilisent les sommets d'énergie pour localiser les noyaux syllabiques (1). Nous aimerions savoir si ces auteurs ont l'intention d'utiliser aussi les indices fournis par ce paramètre pour la segmentation des unités constitutives de la phrase.

VAISSIERE (2) s'attache, de son côté à l'utilisation des variations de durée pour la reconnaissance des frontières de mots dans les phrases. Sans doute s'agit-il là, comme elle semble du reste l'admettre fort bien elle-même, d'une tentative prématurée. Nous ne maîtrisons pas assez les nombreux facteurs qui conditionnent la durée vocalique (durée intrinsèque, effets de la coarticulation(3), du nombre de syllabes dans le mot et dans le groupe, etc... pour entreprendre de telles recherches.

D'autre part, le seuil de 120 m/sec qu'elle adopte pour la détection des voyelles initiales de mot nous paraît trop élevé. C'est peut-être la raison pour laquelle les syllabes de mots qui comportent une voyelle intrinsèquement longue (Institut) sont mieux détectées que celles qui comprennent une voyelle intrinsèquement brève (phonétique, Provence).

Pour qu'un algorithme similaire à celui présenté par VAISSIERE puisse être appliqué avec succès, nous pensons que deux conditions, au moins, doivent être remplies :

(1) cf. Volume 1 des Actes, pp. 354-59.

(2) cf. Volume 1 des Actes, pp. 345-52.

(3) cf. LEHISTE (1970), CHEN (1970), NOOTEBOOM (1972a), (1972b), (1972c), (1974), (1975), KLATT (1973), (1975), LEHISTE (1975).

- la définition de coefficients de pondération permettant de corriger les données objectives,
- la possibilité d'une comparaison syntagmatique des données corrigées.

Il est à noter, par ailleurs, que très peu de chercheurs utilisent systématiquement ce paramètre dans la procédure de segmentation en unités prosodiques.

Dans de nombreuses autres études, les analyses acoustiques sont fondées exclusivement sur l'interprétation des variations de la fréquence fondamentale (Fo). C'est le cas, par exemple, des communications présentées par DAGENAI et MARCHAL (1), VIVES et al, THERRIEN-ROY et al. Cette procédure appelle un certain nombre de remarques.

- La première est que l'hypothèse de travail, selon laquelle les variations de (Fo) seraient liées de façon univoque aux traits intonatifs, est peut-être trop hâtivement érigée en postulat. Il n'est pas encore prouvé, en effet, que les autres paramètres acoustiques n'interviennent pas, soit à titre d'indices complémentaires, soit de façon compensatoire, dans la manifestation des traits linguistiques.

Les faits de neutralisation dont fait état MARTIN s'appliquent aux contours de (Fo), mais il n'est pas interdit de penser que les autres paramètres puissent, dans ce cas, assurer un relai fonctionnel. ROSSI et DI CRISTO montrent, par exemple, qu'une frontière prosodique peut aussi être réalisée et identifiée en l'absence de tout relief mélodique (2).

- La seconde remarque est la suivante : les variations temporelles de (Fo) résultent d'une interaction complexe entre les effets dus à plusieurs facteurs : modalité de la phrase réalisée, structure constituante, accents, état d'esprit et attitude du locuteur, sexe, caractéristiques individuelles et socio-culturelles, nature des unités segmentales, concaténation de ces unités.

C'est pourquoi, il est légitime de penser que l'interprétation des tracés acoustiques doit être précédée de recherches destinées à évaluer l'importance res-

(1) cf. Volume 1 des Actes, pp. 275-79.

(2) cf. Volume 1 des Actes, pp. 329.

pective des effets relatifs à ces divers facteurs. La négligence de ces précautions peut conduire à une interprétation difficile, voire erronée des tracés (1). Le chercheur peut, par exemple, être incapable de dissocier, dans une variation paramétrique, ce qui appartient au trait intonosyntaxique étudié et ce qui relève de l'expressivité. Il peut, d'autre part, attribuer une valeur linguistique à une variation paramétrique conditionnée soit en totalité, soit en partie, par des contraintes inhérentes à l'émission de la parole.

- La dernière remarque sur ce sujet est que l'analyse monoparamétrique paraît difficile à justifier dans une perspective d'application à la reconnaissance automatique, car on se prive ainsi de nombreuses sources d'informations, de redondances, qui peuvent, dans certains cas, s'avérer déterminantes pour le choix des critères de segmentation.

Les observations que nous venons de présenter ne doivent pas faire croire que nous contestons le bien fondé de l'analyse objective. Il nous semble, cependant, que si on s'y limite, on ne dépasse pas le niveau de la performance, un niveau qui a subi un certain nombre de transformations et de distorsions, en raison des multiples contraintes que nous avons déjà évoquées.

On peut alors se demander dans quelle mesure l'analyse objective permet véritablement de rendre compte de la compétence du locuteur, c'est-à-dire de la relation des instructions linguistiques aux traits qui dérivent de ces instructions.

On se posera enfin la question de savoir si l'absence de relations entre les structures intonatives et les structures syntaxiques que constatent parfois certains chercheurs correspond à un fait réel ou ne tient pas plutôt à l'inadéquation de la méthode d'analyse employée.

b) analyse perceptuelle

Les méthodes d'analyse perceptuelle appliquées aux recherches sur l'intonation sont illustrées, dans ce Colloque, par deux tendances nettement dif-

(1) voir, à ce sujet, la très intéressante communication de THORSEN, dans le Volume 1 des Actes, pp. 175-181.

férenciées. La première consiste à procéder, en utilisant des techniques de synthèse, à la stylisation de certaines variations mélodiques et à la comparaison de ces dernières avec des variations naturelles, afin de juger de leurs équivalences. Cette méthode, qui a été élaborée par les chercheurs de l'Institut de Recherches sur la Perception d'Eindhoven (COLLIER and 't HART, 1975, 't HART and COLLIER, 1975), est réemployée par VAISSIERE (1) et par DELGUTTE (2).

La seconde tendance, qui consiste à appliquer au signal acoustique un ensemble de règles de conversion perceptuelle, a été développée à l'Institut de Phonétique d'Aix (3).

Les deux méthodes aboutissent à des résultats comparables. Elles révèlent, notamment, que la plupart des variations physiques ne sont pas intégrées et que l'oreille est plus sensible à certaines variations qu'à d'autres. Ces contraintes perceptuelles justifient la stylisation des contours qui serviront de base à l'interprétation fonctionnelle et la limitation des nombreuses variantes à quelques stéréotypes fondamentaux (4).

Dans cette perspective, DELGUTTE propose de vérifier la validité perceptuelle des contours-types recensés par VAISSIERE. Il étudie, en outre, l'influence de la modalité sémantique sur la réalisation et la perception de ces contours. Ses résultats confirment en partie les recherches antérieures et permettent d'affiner les systèmes décrits jusqu'à présent (VAISSIERE, 1975, DI CRISTO, 1975). Ils montrent, d'autre part, que le modèle perceptuel "d'analyse par synthèse", (LIEBERMAN, 1967, HALLE and STEVENS, 1964) s'applique aussi au décodage des traits prosodiques.

La technique employée par DELGUTTE consiste à faire varier systématiquement, à la synthèse, le contour de (Fo) d'une phrase dont les autres paramètres sont maintenus constants. Il demande ensuite à des sujets d'évaluer le naturel des stimuli ainsi engendrés.

Selon DELGUTTE "on peut voir le naturel comme étant analogue à la notion de grammaticalité". Nous souhaiterions avoir quelques éclaircissements au sujet de cette formulation.

(1) cf. Volume 1 des Actes, pp. 183-88.

(2) cf. Volume 1 des Actes, pp. 281-87.

(3) Voir les nombreux travaux de ROSSI sur ce sujet.

(4) cf. les communications de COLLIER et 't HART, dans ce volume, pp. 139-45 et 167-71.

D'autre part, le fait de fonder l'interprétation perceptuelle sur les variations d'un seul paramètre prosodique soulève les questions suivantes :

- Il n'est pas tenu compte de ce que nous appellerons "la perception interactive des paramètres" (relations entre F_0 et durée, F_0 et intensité (1)).

- Ne risque-t-on pas de donner au paramètre testé un rôle plus important qu'il n'a en réalité (c'est-à-dire, dans la parole naturelle) ?

- Le fait de faire varier un paramètre indépendamment d'un autre avec lequel il entretient des relations étroitement solidaires (F_0 et intensité (2)) n'a-t-il pas pour effet d'introduire de nouvelles variables subjectives dont on ne serait plus maître ?

3. CHOIX DES UNITES PROSODIQUES

Nous distinguerons, ici encore, deux tendances nettement distinctes dans les communications qui nous sont présentées.

Pour la commodité de l'exposé, nous qualifierons la première tendance de sélective et la seconde de globaliste.

Les auteurs qui se rattachent à la première tendance ne considèrent généralement, pour l'interprétation fonctionnelle des données acoustiques et perceptuelles, que certains points de la chaîne sonore bien localisés. C'est ainsi, par exemple, que MARTIN ne retient, dans son analyse, que les contours qui affectent les dernières syllabes des groupes accentuels.

De même, DAGENAI et MARCHAL envisagent uniquement les inflexions mélodiques (perçues ?) réalisées aux frontières de propositions.

Ces auteurs perpétuent ainsi la tradition bien établie, suivant laquelle l'information prosodique est manifestée par certains points clés qui sont porteurs de la fonction. D'autres études ont montré, cependant, que l'intégration n'était pas exclusivement fondée sur la détection de ces points privilégiés (FAURE, 1971, AUTESSERRE et DI CRISTO, 1972, HADDING-KOCH and STUDDERT-KENNEDY, 1964, STUDDERT-KENNEDY and KOCH, 1971).

(1) cf. ROSSI, 1976.

(2) cf. la communication de GRAILLOT et BOE, volume 1, pp. 194-99.

Nous croyons personnellement que l'encodage et le décodage des structures prosodiques s'effectuent à partir de relais (ou points d'ancrages), qui correspondent à certaines syllabes du groupe prosodique et dont les plus importants semblent être : l'attaque, la prétonique et la fin perçue du contour final (DI CRISTO, 1976).

Les résultats d'expériences perceptuelles effectuées par CARTON et LONCHAMP, dans leur très intéressante étude sont de nature à confirmer cette hypothèse (1).

La tendance que nous avons qualifiée de globaliste consiste à prendre pour base d'analyse des unités suprasegmentales de dimension supérieure. De très nombreux auteurs opèrent sur des unités qu'ils nomment des groupes prosodiques (GP).

VAISSIERE s'attache à recenser, dans ses recherches, un certain nombre de formes prosodiques types (qu'elle appelle contours, mais qui sont en fait des GP) et à décrire les mouvements élémentaires de (Fo) constitutifs de ces formes types. Cette méthode est reprise par DELGUTTE, qui l'emploie pour définir des patrons perceptuels. En fait, ces auteurs appliquent scrupuleusement au français les techniques d'analyse élaborées par les chercheurs d'Eindhoven (trop scrupuleusement, peut-être, dans la mesure où les termes utilisés pour décrire les mouvements élémentaires du français ne sont pas traduits de l'anglais : Ri pour montée (rise), etc...

Cependant, le concept de groupe prosodique recouvre des réalités différentes, selon les auteurs. Il est coextensif au syntagme, chez les uns, au mot phonique, chez d'autres, à la proposition, chez d'autres encore.

D'autre part, les critères qui s'appliquent à la délimitation des GP ne sont pas toujours définis de façon très explicite. Il convient, néanmoins, de signaler la tentative intéressante présentée par VIVES et al (2). Ces derniers utilisent les pauses et les mouvements élémentaires de (Fo) décrits par VAISSIERE pour identifier les frontières et les éléments constitutifs des GP.

(1) cf. Volume 1 des Actes, pp. 133-38.

(2) cf. Volume 1 des Actes, pp. 353-59.

Nous avons remarqué, toutefois, que les GP n'étaient pas toujours démarqués par des pauses et que certains effets de coarticulation prosodique pouvaient modifier de façon appréciable le modelé des GP. Une forte montée mélodique réalisée au terme d'un GP peut, par exemple, entraîner la réalisation d'une importante chute à l'initiale du GP subséquent.

Par ailleurs, DELGUTTE a montré que la modalité a pour effet de modifier la forme globale du GP. Il en est de même des accents didactiques.

Nous aimerions savoir si VIVES et al tiennent compte de ces modifications dans leur procédure de segmentation.

Nous souhaiterions aussi que les différents auteurs nous apportent des précisions complémentaires sur la façon dont ils ont procédé pour délimiter les frontières des GP.

Enfin, il serait intéressant que les spécialistes de reconnaissance automatique qui participent à ce Congrès nous disent quelles sont, parmi les nombreuses méthodes de segmentation proposées : identification d'indices ou de faisceaux d'indices ponctuels, contrastes syntagmatiques de contours, définitions de formes prosodiques caractéristiques, celles qui leur paraissent les plus aptes à résoudre les problèmes concrets auxquels ils sont confrontés.

La multiplicité des questions soulevées ici montre à l'évidence que les problèmes de l'analyse intonative sont encore loin d'être tous résolus. Cependant, la diversité et la qualité des communications présentées par les auteurs témoignent de l'effort considérable que connaissent actuellement les travaux sur l'intonation, tant dans le domaine des recherches fondamentales que dans celui de leurs nombreuses applications.

REFERENCES

N.B. Tous les articles qui portent la référence : A-8, GCP se trouvent dans les Actes des 8èmes Journées d'Etude du Groupe de la Communication Parlée (AIX-EN-PROVENCE, 25-27 mai 1977).

AUTESSERRE, D. et DI CRISTO, A. (1972)

Recherches sur l'intonation du français : traits significatifs et non significatifs, Proceedings of the VIIth International Congress of Phonetic Sciences (Montreal), pp. 842-59.

CAELEN, G. et MAURAND, G. (1977)

Corrélations entre mélodie, intensité et durée dans la phrase énonciative, A-8, GCP, pp. 315-22.

CARTON, F. et LONCHAMP, F. (1977)

Validation d'indices perceptifs de l'intonation par analyse multidimensionnelle, A-8, GCP, pp. 133-38.

CHEN, M. (1970)

Vowel length variation as a function of the voicing of the consonant environment, Phonetica, 22, pp. 129-59.

COLLIER, R. (1977)

La perception de l'intonation anglaise par des anglophones et des néerlandophones, A-8, GCP, pp. 139-45.

COLLIER, R. and 't HART, J. (1975)

The role of intonation in speech perception, in : Structure and Process in Speech Perception, COHEN A. and NOOTEBOOM, S.G., eds, pp. 107-23.

DAGENAIS, L. et MARCHAL, A. (1977)

Ligne mélodique et "que" complétif en français de Montréal, A-8, GCP, pp. 275-79.

DELGUTTE, B. (1977)

Expériences perceptives sur l'intonation, A-8, GCP, pp. 281-87.

DI CRISTO, A. (1975)

Recherches sur la structuration prosodique de la phrase française, Actes des 6èmes Journées d'Etude sur la Parole, pp. 96-116.

DI CRISTO, A. (1976)

Indices prosodiques et structure constituante, Cahiers de Linguistique, d'Orientalisme et de Slavistique, 7 : pp. 27-40.

FAURE, G. (1971)

La description phonologique des systèmes prosodiques, Zeitschrift für Phonetik, 24 (5) : pp. 349-59.

FONAGY, I., LUCCI, V. et BOE, L.J. (1977)

Influence des variables situationnelles sur quelques paramètres de l'intonation, A-8, GCP, pp. 14-17.

GRAILLOT, P. et BOE, L.J. (1977)

Relation entre l'évolution de la fréquence laryngienne et l'intensité pour la phrase énonciative du français : premiers résultats, A-8, GCP, pp. 289-95.

GROSS, M. (1976)

Remarques sur le rôle de la prosodie en reconnaissance, Actes des 6èmes Journées d'Etude sur la Parole, II : pp. 55-64.

HADDING-KOCH, K. and STUDDERT-KENNEDY, M. (1964)

An experimental study of some intonation contours, Phonetica, 11 (3-4), pp. 178-85.

HALLE, M. and STEVENS, K.N. (1964)

Speech recognition : a model and a program for research, in : The Structure of Language, FODOR, J.A. and KATS, J.J., eds.

't HART, J. and COLLIER, R. (1975)

Integrating different levels of intonation analysis, Journal of Phonetics, 3 : pp. 235-55.

't HART, J. (1977)

Vers une base psychophonétique de la stylisation intonative, A-8, GCP, pp. 167-73.

HATON, J.P., MARI, J.F., MESSENET, G., PIERREL, C. et SANCHEZ, L. (1977)

Prise en compte d'informations prosodiques dans un modèle de compréhension du discours continu, A-8, GCP, pp. 331-36.

HAZAEI-MASSIEUX, M.C. (1977)

Pour une détermination des fonctions grammaticales de l'intonation : analyse de quelques types d'encodage et de décodage, A-8, GCP, pp. 289-95.

HIRST, D. (1977)

L'intonation générative, A-8, GCP, pp. 297-303.

KLATT, D. (1973)

Vowel duration as a function of the syllabic structure of a word,
J. A. S. A., pp. 312-33.

KLATT, D. et COOPER, W. (1975)

Perception of segment duration in sentence context, in : Structural Process in Speech Perception, COHEN, A. and NOOTEBOOM, S.G., eds, Springer Verlag, pp. 69-86.

LARREUR, D. et EMERARD, F. (1977)

Analyse de la structure intonative de la phrase française, A-8, GCP, pp. 227-36.

LEHISTE, I. (1970)

Suprasegmentals, M.I.T. Press.

MARTIN, Ph. (1975)

Analyse phonologique de la phrase française, Linguistics, 146, pp.35-68.

MARTIN, Ph. (1977)

Utilisation d'un modèle prosodique pour la synthèse par concaténation de mots en français et en anglais, A-8, GCP, pp. 248-54.

MARTIN, Ph. (1977)

Problèmes de neutralisation des marques prosodiques, application à la reconnaissance automatique, A-8, GCP, pp. 305-11.

NOOTEBOOM, S.G. (1972a)

Some timing factors in the production and perception of vowels, Occasional Papers, Language Centre, University of Essex, 13, pp. 49-76.

NOOTEBOOM, S.G. (1972b)

The interaction of some intra-syllable and extra-syllable factors acting on syllable nucleus durations, I.P.O. Annual Progress Report, 7, pp.30-39.

NOOTEBOOM, S.G. (1972c)

Production and Perception of vowel Duration, Doct. Diss., University of Utrecht.

NOOTEBOOM, S.G. (1974)

Some context effects on phonemic categorization of vowel duration,
I.P.O. Annual Progress Report, 9 : pp. 47-54.

NOOTEBOOM, S.G. (1975)

Temporal organisation of connected speech : Prosody, in : Auditory Analysis and Perception of Speech, Academic Press, London, pp. 383-86.

ROSSI, M. (1976)

- Interactions des glissements d'intensité et des glissements de fréquence,
XIVth Congerence on Acoustics, High Tatras, CSSSR.

ROSSI, M. et DI CRISTO, A. (1977)

Propositions pour un modèle d'analyse de l'intonation, A-8, GCP, pp. 323-29.

SAINT-BONNET, M. et BOE, L.J. (1977)

Les pauses et les groupes rythmiques : leur durée et distribution en fonction de la vitesse d'élocution, A-8, GCP, pp. 337-43.

STUDDERT-KENNEDY, M. and KOCH, K. (1971)

Auditory and linguistic processes in the perception of intonation,
Status Report on Speech Research (Haskins Lab.), 27, pp. 153-74.

THERRIEN-ROY, L., JACQUEMIN, D. et BOE, L.J. (1977)

De l'analyse phonologique à l'analyse phonétique des énonciatives en français, A-8, GCP, pp. 371-76.

THORSEN, N. (1977)

On the interpretation of raw fundamental frequency tracings, A-8, GCP, pp. 175-81.

VAISSIERE, J. (1977)

Quelques expériences d'analyse perceptuelle en français, A-8, GCP, pp. 183-89.

VAISSIERE, J. (1977)

Premiers essais de la durée pour la segmentation en mots dans un système de reconnaissance, A-8, GCP, pp. 345-52.

VIVES, R., LE CORRE, C., MERCIER, G. et VAISSIERE, J. (1977)

Utilisation, pour la parole continue, de marqueurs prosodiques extraits de la fréquence fondamentale, A-8, GCP, pp. 353-59.

Sous-thème :

Méthodes et modèles d'étude des rapports
intonation-syntaxe.

Rapporteur : *Ph. MARTIN*

Laboratoire de Phonétique de Toronto (Canada)
Institut de Phonétique Aix-en-Provence

EXPOSE DE SYNTHESE SUR LES COMMUNICATIONS DU THEME 3 :
L'INTONATION : Aspects linguistiques et reconnaissance de la parole

PH. MARTIN

O. En parcourant l'ensemble des travaux réunis dans le troisième thème, on constate qu'il n'y a, en définitive, que fort peu de communications relevant spécifiquement du rapport intonation-syntaxe. On n'y compte guère que celles de D. HIRST, qui s'applique à l'anglais, et de L. DAGENAIS et A. MARCHAL, qui concerne un problème particulier de phonosyntaxe québécoise.

Aussi j'ai choisi de puiser dans l'ensemble des communications de ces journées, et plus particulièrement dans celles traitant de la synthèse et de la reconnaissance automatique tout ce qui se rapprochait implicitement ou explicitement de ce thème.

1. Si une lecture superficielle de ces études donne l'impression que l'étude des rapports entre l'intonation et la syntaxe en est encore, en français, à ses débuts, un examen plus attentif révèle au contraire que, derrière certaines divergences superficielles, dues pour la plus grande part à la diversité des méthodes utilisées, existe une assez grande communauté de vue, qui se manifeste, entre autre, au niveau de la description des données.

2. Pour gagner du temps, plutôt que de passer en revue chacune des communications examinées, je tenterai, en un premier temps, de faire état des traits communs à ces études sous la forme d'une mini-théorie, synthèse des différents traitements du rapport syntaxe intonation. Les divergences constatées seront alors, en un deuxième temps, présentées comme écarts par rapport à cette mini-théorie.

Je situerai cet exposé plus particulièrement dans le plan du code linguistique abstrait, en laissant le soin à A. DI CRISTO d'en évoquer les différentes manifestations possible dans le plan de la substance. Ainsi, lorsque je parlerai de contours, il s'agira en fait d'entités abstraites, qui se manifestent et s'observent principalement par une variation de fréquence fondamentale (ou de hauteur si on adopte le point de vue perceptif), mais aussi éventuellement selon d'autres paramètres prosodiques tels qu'intensité et qualité vocalique par exemple.

3. La première question qui se pose dans l'étude des rapports syntaxe-intonation est : que décrire parmi les faits prosodique? Deux camps ici semblent s'opposer : les partisans du "mot" ou "groupe prosodique", qui décrivent l'évolution de tous les paramètres (M. SAINT BONNET et L.J. BOE, A. DI CRISTO et M. ROSSI, J. VAIS-SIERE),

et ceux du contour -ou du niveau- qui s'en tiennent aux faits prosodiques liés à la seule syllabe accentuée, D. LARREUR et F. EMERARD D. HIRST, Ph. MARTIN). Les premiers répondent à des questions d'ordre phonétique, les seconds à des problèmes plus nettement phonologiques.

M'appuyant sur une tradition déjà longue liant accent et variation mélodique (Bolinger, Rigault, etc...), je dirai que, d'un point de vue théorique, les syllabes accentuées et, partant, les variations prosodiques qui leur sont liées, devraient seules être prises en compte. On peut se demander à ce propos si à la lumière des résultats présentés précédemment aux journées de Toulouse et de ceux discutés au cours de la séance consacrée à la synthèse, les mots prosodiques décrits dans le travail de J. VAISSIERE ne pourraient pas éventuellement être réduits aux seuls contours mélodiques finaux étant donné d'une part que la montée initiale y est toujours la même, et que, d'autre part, la transition entre cette montée et le contour final s'avère perceptuellement peu importante.

Dans la négative, on pourrait donc dire que la notion de mot prosodique peut être provisoirement gardée quitte à la réduire à son seul contour pertinent final dans sa description phonologique.

4. Un concept qui émerge dans plusieurs communications est celui de "structure prosodique", définie comme hiérarchie de mots prosodiques, tout à fait semblable à la structure syntaxique considérée comme classement hiérarchique de mots (B. DELGUTTE, A. DI CRISTO et M. ROSSI, Ph. MARTIN, J. VAISSIERE).

Une structure prosodique (à 2 niveaux dans l'exemple ci-dessous) sera représentée de la manière suivante :

niveau 0 ————— = phrase prosodique

niveau 1 ———

niveau 2. —————

— = "mot" prosodique

Le problème des rapports intonation-syntaxe se trouve ainsi ramené à l'examen des rapports éventuels existant entre cette structure prosodique et la syntaxique de la phrase.

Un énoncé correspondant à l'exemple de phrase prosodique choisi plus haut serait :

La soeur de Maurice a mangé les tartines

Niveau 1() () () () () () () ()
Niveau 2() () () () () () () ()
Niveau 3() () () () () () () ()

Le parenthésage en indique la structure syntaxique, organisée en 3 niveaux.

5. Etablir une correspondance entre les deux structures revient à définir un lien entre le mot prosodique et une certaine entité syntaxique. Diverses notions sont utilisées pour préciser ce lien : groupe de sens, unité de sens, unité de signification, unité accentuable, etc... ainsi qu'en témoignent les études de D. LARREUR et F. EMERARD, M.C. HATON et J.P. HATON, A. DI CRISTO et M. ROSSI.

Pour ma part, je pense que le mot prosodique recouvre l'unité effectivement accentuée. Ainsi l'unité de sens (les bonnes tartines) peut contenir un ou deux mots prosodiques et, partant, ou 2 contours, selon que l'adjectif est accentué ou non.

6. Si on accepte cette solution, il en résulte, pour notre exemple, le parallélisme suivant :

La soeur de Maurice a mangé les tartines

_____ Structure prosodique

Le non parallélisme éventuel entre les 2 structures exprime l'indépendance plus ou moins forte que l'on accorde à la structure prosodique par rapport à la structure syntaxique. Quant à la non accentuabilité d'unités telles que l'article (la), la préposition (de), ou l'auxiliaire (a), elle est due aux liens syntaxiques dits de solidarité qui existent entre ces deux éléments et une autre unité, avec laquelle ils forment une seule groupe accentuable. Le traitement de ce problème par les règles de CH. CHOPPY et J.S. LIENARD en constitue une approximation empirique.

7. Supposons qu'un locuteur donné n'aie pas accentué le premier élément. La structure prosodique supposée parallèle à la structure syntaxique sera alors :

La soeur de Maurice a mangé les tartines

Si le verbe est désaccentué, on aura cette fois :

La soeur de Maurice a mangé les tartines

On retrouve donc, par ce mécanisme, les groupes de sens qui émergent dans la structure syntaxique.

Les deux structures devant être congruentes, on pourrait aussi avoir :

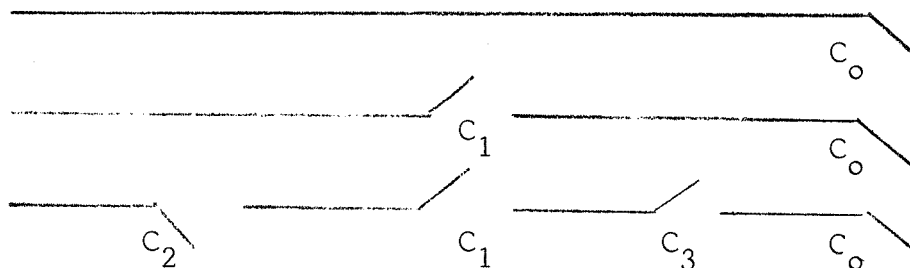
La soeur, de Maurice, a mangé les tartines

Il n'y a, à mon sens, aucune raison logique de rejeter ces possibilités dans les cas où la structure syntaxique n'est pas ambiguë, ou mal indiquée.

Je pense toutefois que dans la synthèse, on a intérêt à rendre les deux structures parallèles, comme le font D. LARREUR et F. EMERARD par exemple. En reconnaissance par contre (VIVES et PH. MARTIN), toutes les configurations possibles doivent être considérées.

8. Le point suivant concerne la recherche des invariants observés dans les courbes prosodiques, ces invariants devant servir à indiquer l'organisation de la structure prosodique.

Si la montée en fréquence au début du mot prosodique peut apparaître comme invariant, pour autant qu'il se réaliserait partout, elle ne saurait en revanche qu'indiquer une structure à un seul niveau, puisque toutes les "marques" des mots prosodiques seraient alors les mêmes. Par contre, le contour final placé sur la syllabe accentuée de l'unité syntaxique présente différentes configurations de pente, durée, amplitude (ou niveau en anglais) propres à assurer l'indication de la structure prosodique. L'identification de ces contours se fait selon le schéma suivant :



Le contour C_0 , descendant et long, est corrélatif de la fin de la phrase prosodique, et indique donc la fin de l'unité de niveau 0. Le contour C_1 montant, opposé par la pente au contour final C_0 , termine toutes les unités de niveau 1, sauf la dernière. Le contour C_2 , descendant, termine les unités de deuxième niveau issues de la division de mots prosodiques terminés par C_1 montant, alors que le contour C_3 , montant et d'amplitude restreinte par rapport aux autres contours montants C_1 , termine les unités de deuxième niveau provenant de la division de mots prosodiques terminés par C_0 descendant. On observe donc à chaque fois un contraste de pente mélodique montant/descendant ou descendant/montant d'un niveau de la structure à l'autre.

La soeur de Maurice a mangé les tartines

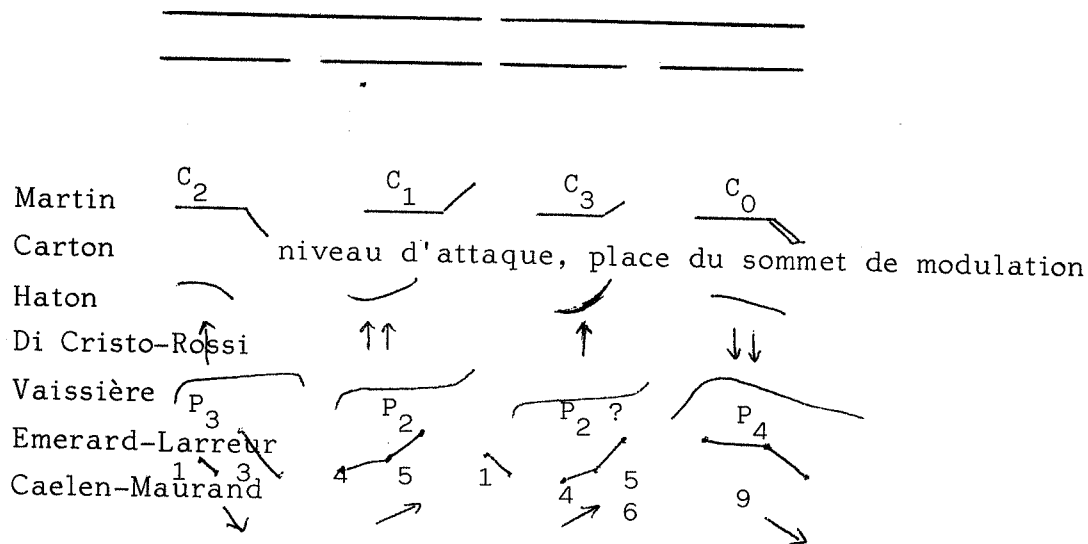


Tableau comparatif des différents systèmes de description des contours ou des mots prosodiques

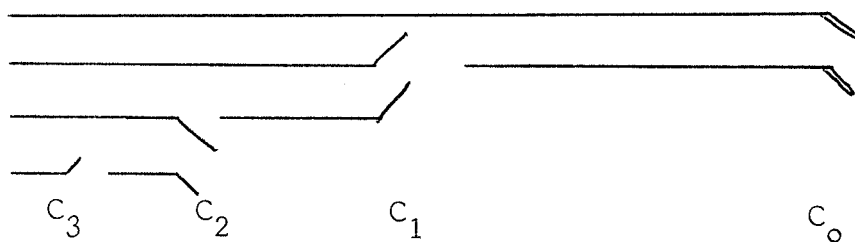
Ces contours invariants sont décrits et désignés différemment selon les auteurs, mais tous (sauf l'équipe de Grenoble) semblent observer ce contraste de pente. De même, le contraste d'amplitude entre les deux contours montant C_1 et C_3 se trouve décrit dans beaucoup de communications, mais non dans toutes. (cf CAELEN et MAURAND).

9. On utilise donc dans cette mini-théorie un ensemble de 4 contours ou mots prosodiques pour indiquer une structure à deux niveaux. Une structure à trois niveaux nécessiterait deux contours supplémentaires (cf. MARTIN), mais se trouve rarement réalisée dans des conditions courantes de production du discours, en dehors de la lecture. Ce nombre de contours, de 4 à 6, correspond au nombre trouvé par t. HART et COLLIER pour le néerlandais en se basant sur des expériences de perception.

10. En cas de désaccentuation ou de non parallélisme entre les deux structures syntaxique et prosodique, la séquence de contour obtenue sera différente (cf plus haut).

Ceci ne sera pas le cas si la structure prosodique est "dominante", c'est-à-dire si elle supplée une défaillance dans l'indication de la structure syntaxique. (type : une montre d'enfant suisse), ou effacement du (que) de la relative en français québécois (L. DAGENAIS et A. MARCHAL). Mais il y a d'autres possibilités de dominance qui peuvent rendre une structure prosodique donnée inacceptable à la synthèse. Ceci a été étudié expérimentalement par B. DELGUTTE dans des cas particuliers pour lesquels les énoncés synthétisés ne sont pas "naturels" quoique toujours compréhensibles.

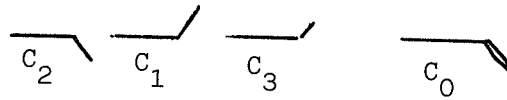
11. Les différences observées par rapport à la mini-théorie sont principalement liées aux objections faites par D. LARREUR et F. EMERARD quant à l'existence d'un contour C_3 indiquent les unités prosodiques de troisième niveau :



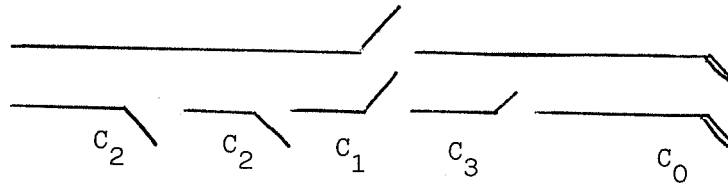
Pour ma part, je pense que l'absence du contour C_3 dans les exemples traités est dû à une désaccentuation de l'élément syntaxique concerné, (a), soit à l'indication d'une structure prosodique non parallèle à la structure syntaxique, (b) comme dans l'exemple ci-après :

Le gros chien blanc mange le fromage

a) ici élément (gros)
désaccentué



b) structure prosodi-
que non parallèle
à la structure
syntaxique.



12. Je parlerai brièvement de la substance prosodique, c'est-à-dire de la manière dont se réalisent les contours, en évoquant la communication d'ABRY et J.L. BOE qui proposent une hypothèse intéressante (déjà utilisée par l'équipe de M. WAJSKOP pour le VOT) s'inscrivant dans le cadre de la description linguistique, mais qui pourrait avoir des applications en reconnaissance automatique. L'existence d'une entité à décrire procède d'une combinatoire de traits, les traits s'explicitant eux-mêmes par des indices, et les indices par des propriétés. Le cas traité est celui du voisement.

13. Enfin, je terminerai en évoquant les problèmes liés aux conditions de production du discours, les contours ou mots prosodiques étant réalisés différemment selon le contexte (au sens large) dans lequel se trouve plongé le locuteur. C'est le domaine de la phonostylistique, où s'inscrivent bon nombre de communications présentées ici (HAZAEL- MASSIEUX, I. FONAGY et al.) ou des nombreux travaux de P. LEON sur le français.

14. Je concluerai en répétant ma remarque liminaire : sous une diversité apparente se retrouve, dans les faits, une grande similitude dans la manière dont le lien intonation-syntaxe apparaît dans ces études.

Il se confirme ainsi que des progrès décisifs ont été accomplis de ce point de vue en l'espace des quelques cinq dernières années.

REFERENCES

Tous les articles qui portent la référence : A-8, GCP sont publiés dans les Actes des 8emes Journées d'Etude du Groupe de la Communication Parlée (Aix-en-Provence, 25-27 mai 1977)

Arbry, C. et Boë, L-J (1977°

Traits, indices et propriétés du voisement : vers une phonologie des unités infra-phonétiques. A-8, GCP, 267-273

Caelen, G. et Maurand, G. (1977)

Corrélations entre mélodie, intensité et durée dans la phrase française énonciative, A-8, GCP, 315-322

Choppy, C. et Lienard, J-S. (1977)

Prosodie automatique pour la synthèse par diphonèmes, A-8, GCP, 211-217

Collier, R. (1977)

La perception de l'intonation anglaise par des anglophones et des néozélandophones, A-8, GCP, 139-146

Dagenais, L. et Marchal, A. (1977)

Ligne mélodique et "que" complétif en français de Montréal, A-8, GCP, 275-279

Delgutte, B. (1977)

Expériences perceptives sur l'intonation, A-8, GCP, 281-287

Di Cristo, A. et Rossi, M. (1977)

Propositions pour un modèle d'analyse de l'intonation, A-8, GCP, 323-329

Fonagy, I. et Sap, J. (1977)

Traits prosodiques distinctifs de certaines attitudes intellectuelles et émotives, A-8, GCP, 237-245

...

- Haton, J.P, Mari, J.F, Messenet, G., Pierrel, J.M et Sanchez, C. 1977
Prise en compte d'informations prosodiques chez l'enfant sourd, A-8, GCP, 364-369
- 't Hart , J. (1977)
Vers une base psychophonétique de la stylisation intonative, A-8, GCP, 167-173
- 't Hart, J. et Collier, R. (1975)
Integrating different levels of intonation analysis, J. of phonetics, (3) 235-255
- Hazaël-Massieux, M-C. (1977)
Pour une détermination des fonctions grammaticales de l'intonation : analyse de quelques tests d'encodage et de décodage, A-8, GCP, 289-295
- Hirst, D. (1977)
L'intonation générative, A-8, GCP, 297-303
- Larreur, D. et Emerard, F. (1977)
Analyse de la structure intonative de la phrase en français, A-8, GCP, 227-236
- Léon, P. (1971)
Essais de phonostylistique, Didier, Paris.
- Martin, Ph (1975)
Analyse phonologique de la phrase française, Linguistics, (146) 35-68
- Martin, Ph. (1977)
Utilisation d'un modèle prosodique pour la synthèse par concaténation de mots en français et en anglais, A-8, GCP, 247-254

Martin, Ph. (1977)

Problèmes de neutralisation des marques prosodiques, application de la reconnaissance automatique, A-8, GCP, 305-311

Therrien-Roy, L., Jaquemin, D. et Boë, L-J. (1977)

De l'analyse phonologique à l'analyse phonétique des énonciatives en français, A-8, GCP, 371-376

Saint-Panet, M. et Boë, L-J (1977)

Les pauses et les groupes rythmiques : leur durée et distribution en fonction de la vitesse d'élocution, A-8, GCP, 337-343

Vaissière, J. (1977)

Quelques expériences d'analyse perceptuelle en français, A-8, GCP, 183-189

Vives, R., Le Corre, C., Mercier, G. et Vaissière, J. (1977)

Utilisation, pour la reconnaissance de la parole continue, de marqueurs prosodiques extraits de la fréquence du fondamental, A-8, GCP, 353-359

Sous-thème :

Prosodie et reconnaissance automatique
de la parole.

Rapporteur : J.S. LIENARD
L.I.M.S.I. Orsay

PROSODIE ET RECONNAISSANCE AUTOMATIQUE DE LA PAROLE

J.S. LIENARD

L.I.M.S.I. (C.N.R.S.) - B.P. 30 - 91406 ORSAY Cedex

L'utilisation de la prosodie dans le domaine de la reconnaissance est relativement récente, et coïncide pratiquement avec le début des systèmes de compréhension automatique de la parole, qui, par définition, s'intéressent à des grandeurs suprasegmentales. Nous tenterons en premier lieu de situer ce problème de compréhension ; nous décrirons les principaux travaux publiés mentionnant l'utilisation d'indices prosodiques, puis nous examinerons les communications présentées sur ce sujet lors de nos Journées d'Etude. Enfin nous essaierons de nous placer sur un plan plus général pour mieux cerner le rôle de la prosodie dans les processus de compréhension de la parole.

I - RECONNAISSANCE ET COMPREHENSION AUTOMATIQUE DE LA PAROLE

Après avoir constaté que l'identification des phonèmes dans le signal de parole était un problème beaucoup plus compliqué qu'ils ne l'imaginaient, les chercheurs se tournèrent, entre 1965 et 1970, d'une part vers la reconnaissance par mots isolés en vue d'applications pratiques (commande vocale), d'autre part vers l'utilisation d'informations de niveau linguistique supérieur (lexique, syntaxe) pour compléter le message vocal reconnu au niveau phonétique. Cette seconde approche prend le nom, quelque peu abusif, de "compréhension automatique de la parole". Elle a été développée récemment aux Etats Unis dans le

cadre d'un vaste projet financé par l'ARPA (1971-1976), dont on peut tirer, très sommairement, les conclusions suivantes :

- Il est nécessaire d'obtenir un bon taux de reconnaissance au niveau acoustique - phonétique ; la perte d'information à ce niveau se paie par des algorithmes très lourds aux niveaux supérieurs, et par des restrictions importantes sur le vocabulaire et la syntaxe utilisables.
- L'utilisation intégrée de plusieurs sources d'information (acoustique, phonétique, phonologique, lexicale, prosodique, etc) pose un problème de stratégie qui relève de l'Intelligence Artificielle. Les méthodes ascendantes (assemblage de phonèmes en mots, de mots en phrases, etc) doivent faire face à une "explosion combinatoire" due à l'indétermination des données phonétiques ; les méthodes descendantes, qui se contentent de vérifier les hypothèses émises aux niveaux supérieurs, sont plus rapides mais très rigides, car le locuteur doit se plier aux contraintes imposées par le système. Pratiquement il faut essayer de combiner les deux approches en limitant les hypothèses à prendre en compte à chaque niveau, au moyen d'informations extraites des autres niveaux.

C'est dans cette optique qu'est envisagée l'utilisation de la prosodie : il s'agit d'extraire du signal des informations prosodiques, et de les utiliser pour limiter les hypothèses lexicales et syntaxiques.

II - TRAVAUX ANTERIEURS

a) Aux Etats Unis

W.A. LEA, M.F. MEDRESS et T.E. SKINNER (équipe de SPERRY-UNIVAC) ont présenté en 1974 une nouvelle stratégie de la compréhension automatique, stratégie "guidée par la prosodie" (1). Il faut distinguer, dans ce travail, ce qui est relatif à la doctrine de ce qui a été effectivement réalisé.

En ce qui concerne la doctrine, les auteurs posent comme hypothèse la prééminence des paramètres prosodiques sur les paramètres phonétiques. Plus précisément ils pensent que la compréhension d'une phrase

parlée, par l'homme ou par la machine, implique l'utilisation précoce de sa structure syntaxique, manifestée dans les paramètres prosodiques, conjointement à l'utilisation de l'information acoustique (segmentale) la plus sûre. Ils proposent donc un modèle de la compréhension essentiellement descendant : les hypothèses formées au niveau syntaxique grâce à la prosodie sont complétées au niveau lexical par les mots sélectionnés d'après leurs principaux caractères segmentaux, et la phrase ainsi construite est comparée plus finement aux données acoustiques, ce qui permet de l'accepter ou de la rejeter.

La réalisation présentée est limitée à la détection des frontières de constituants syntaxiques, à la détection et l'identification grossière des syllabes accentuées, et à la reconnaissance globale de phrases choisies dans un ensemble restreint. Les constituants syntaxiques (groupes prosodiques) sont très simplement délimités par des creux d'amplitude au moins égale à 7% dans l'évolution de F_0 , après prise en compte de l'influence micromélodique de certaines consonnes ; cette procédure a fourni 95% de réussite sur 500 secondes de parole, d'après les auteurs. Les syllabes accentuées sont caractérisées par un maximum d'énergie dans la bande 60-3000 Hz, lié à un mouvement ascendant de F_0 . La première syllabe accentuée d'un groupe prosodique est appelée la "tête". On peut en trouver d'autres qui correspondent alors à une montée de F_0 par rapport à la ligne de déclinaison de F_0 à l'intérieur du groupe. Ces règles auraient permis, d'après les auteurs, de détecter 85% des syllabes accentuées sur un corpus de 400 secondes de parole. La phrase est ainsi divisée en constituants et syllabes accentuées ; chaque syllabe accentuée reçoit une catégorisation, ainsi que certaines consonnes (silences, fricatives, R, etc). La reconnaissance est effectuée par une comparaison globale à un petit nombre de phrases références. Malheureusement aucun résultat chiffré autre que ceux cités plus haut ne vient étayer cette stratégie.

Quatre ans plus tard, en 1977, la même procédure est reprise et complétée par une identification plus fine des voyelles accentuées (2) : les voyelles sont identifiées au centre des zones accentuées, les noyaux prosodiques longs sont considérés comme des diphtongues, et l'identification des voyelles est d'autant plus sûre que l'accentuation est forte. Le système présenté est beaucoup moins ambitieux qu'en 1974.

b) en France

La prosodie intéresse depuis longtemps les chercheurs en langue française, mais c'est seulement en 1975 qu'apparurent des études sur la prosodie dirigées vers son utilisation en reconnaissance automatique. Aux Journées d'Etude tenues à TOULOUSE furent présentés des modèles théoriques (P. MARTIN (3), A. DI CRISTO (4)), et des études préliminaires (J. VAISSIERE (5), L. BUISSON et G. MERCIER (6), DOURS et coll. (7)). Le problème des relations entre prosodie et syntaxe était posé, mais ne pouvait recevoir de réponse expérimentale bien nette. Bien que certains supposent une correspondance assez étroite entre structures prosodiques et structures syntaxiques (p. ex. (4)), d'autres se situent en retrait (voir (6) par exemple). Aux Journées d'Etude de NANCY (1976), J. VAISSIERE proposait une procédure de segmentation en groupes prosodiques reposant essentiellement sur les variations du fondamental (8), et rappelait que les relations entre prosodie et syntaxe n'avaient guère été établies que dans le cas de phrases ambiguës.

III - TRAVAUX PRESENTES AUX JOURNEES D'ETUDE

Les communications concernant notre sujet ont en commun la recherche d'une segmentation de l'énoncé en groupes prosodiques, ce qui apparaissait aussi dans les travaux antérieurs. Chez HATON et coll., la démarche est très voisine de celle de LEA : il s'agit, à partir de l'évolution de F_0 et de l'intensité, de repérer des groupes prosodiques et des noyaux syllabiques. Chez VIVES et coll. on recherche également des profils-types de F_0 pour délimiter les groupes prosodiques à l'intérieur d'une séquence comprise entre deux pauses, et l'on utilise l'intensité pour repérer les noyaux syllabiques ; les résultats positifs obtenus par cette équipe corroborent dans une certaine mesure les résultats annoncés par LEA. VAISSIERE s'intéresse à un paramètre prosodique généralement délaissé, la durée phonémique et syllabique ; bien que portant sur un corpus restreint, ces expérimentations semblent donner de bons indices pour le repérage de mots lexicaux. Enfin ROSSI et DI CRISTO soulignent la nécessité d'une interprétation des paramètres prosodiques bruts en fonction de considérations phonétiques et perceptuelles (ce point est également évoqué par VAISSIERE).

De ces travaux se dégage l'impression que l'on peut utiliser les paramètres prosodiques pour repérer dans le continuum acoustique des entités suprasegmentales telles que syllabes, mots, groupes prosodiques ; et que ces paramètres prosodiques ne sont pas indépendants des paramètres segmentaux. Mais on ne peut qu'être déçu, pour l'instant du moins, par la relative sous-utilisation de la prosodie dans le processus de compréhension. Ceci s'exprime dans "la faible fiabilité des informations prosodiques" (HATON et coll.), la discrétion observée au sujet des règles R3 (intonosyntaxiques) et R4 (intonotactiques) dans l'article de DI CRISTO et ROSSI, les résultats non-significatifs obtenus par VIVES et coll. au niveau syntaxique et, d'une manière générale, l'imprécision régnant sur la notion même de groupe prosodique. Notons également la position de MARTIN, qui étudie les relations entre divers plans d'analyse de l'énoncé (plans sémantique, syntaxique, prosodique et syllabique), et qui estime que les structures prosodiques et syntaxiques ne sont pas nécessairement équivalentes, sauf cas particuliers.

IV - ROLE DE LA PROSODIE DANS LES PROCESSUS DE COMPREHENSION ET DE COMMUNICATION

Nous observerons tout d'abord que les informations prosodiques concernent tous les niveaux de la compréhension, si l'on accepte le modèle représenté sur les figures 1 et 2, et décrit dans (9). Selon ce modèle la compréhension est un processus actif, qui implique plusieurs niveaux de décodage (acoustique, perceptible, phonétique, lexical, syntaxico-sémantique), un double flux d'information (structuration formelle dans le sens ascendant, fonctionnelle dans le sens descendant), une centralisation au niveau X de toutes les chaînes sensorielles de compréhension, et une action immédiate ou différée sur l'environnement.

Au niveau phonétique la prosodie porte sur les éléments segmentaux. Les paramètres prosodiques peuvent être - ou non - pertinents phonétiquement : il n'est pas certain que la micro-mélodie associée à certaines consonnes soit un indice phonétique pertinent ; par contre la durée phonémique est un indice essentiel, par exemple dans la définition du trait de voisement, dans la caractérisation des voyelles na-

sales en français, ou dans l'opposition entre voyelles longues et voyelles brèves en anglais. Sans aller jusqu'au problème des langues à tous, ROSSI (10) a montré que chaque voyelle avait une influence spécifique sur la mélodie, sur l'intensité et sur la durée. Par conséquent l'interaction entre paramètres phonétiques et paramètres prosodiques est très forte à ce niveau. La difficulté, pour la reconnaissance, est que l'utilisation des paramètres prosodiques requiert l'identification préalable des phonèmes sur lesquels ils portent ; si on les utilise en tant qu'indices phonétiques, on ne peut plus leur accorder une valeur prosodique.

Au niveau lexical il existe également une prosodie caractéristique, peu marquée en français (voir cependant le travail de VAISSIERE), mais tout-à-fait essentielle dans les langues accentuées comme l'anglais, au point qu'un mot prononcé avec un accent mal placé en devient incompréhensible.

Au niveau syntaxico-sémantique les choses sont plus floues. S'il est incontestable que la prosodie aide au découpage de la séquence continue, il est difficile d'établir une corrélation étroite entre les groupes prosodiques et la structure de surface de la phrase, sauf dans des cas élémentaires et artificiels. Le problème de fond réside donc dans la nature des groupes prosodiques, qui ne sont pas un simple reflet de la syntaxe mais qui sont liés au sens du message et à la personnalité du locuteur.

Pour approfondir un peu ce dernier point il nous semble utile de souligner la grande différence existant entre la communication parlée et la communication écrite (9). Dans cette dernière l'absence de contexte commun et d'interaction entre les deux individus qui communiquent fait que le message, pour être intelligible, doit être parfaitement structuré : il obéit à une syntaxe et à une orthographe, les mots sont séparés par des blancs, les caractères sont normalisés à l'intérieur d'une même fonte. Rien de tel dans le message parlé, qui est continu, variable, flou, souvent mal bâti syntaxiquement, ou inachevé, tout cela sans perdre son intelligibilité. Cette réalité est généralement masquée en linguistique, si l'on prend pour matériaux des phrases écrites, hors contexte, lues à voix haute. Nous pensons que la fonction essentielle de la prosodie, au niveau supérieur, est de régler les modalités du

dialogue entre deux interlocuteurs ; c'est dans ce cadre de l'interaction entre deux individus, ou entre l'homme et la machine, qu'il faut situer l'étude. On ne peut plus s'étonner, alors, de la mauvaise coïncidence observée entre structures syntaxiques et structures prosodiques.

Pour terminer, nous remarquerons que la reconnaissance - compréhension de la parole, qui tend à devenir l'un des processus privilégiés de la communication homme-machine, rencontre en fait les problèmes qui préoccupent depuis longtemps les spécialistes, psychologues et linguistes, de la communication humaine. On peut espérer que, comme ailleurs, l'ordinateur, outil puissant et stupide, fera progresser la connaissance simplement en obligeant les chercheurs à expliciter et quelquefois remettre en cause ce qui, jusqu'à présent, allait de soi.

V - BIBLIOGRAPHIE

- 1 - W.A. LEA, M.F. MEDRESS, T.E. SKINNER - A prosodically guided speech understanding strategy - IEEE Transactions on ASSP, 23, 1, 1975.
- 2 - M.F. MEDRESS, T.E. SKINNER, T.C. DILLER, W.A. LEA - A system for the recognition of spoken connected word sequences - Proceedings of the IEEE International Conference on ASSP, Hartford, 1977.
- 3 - P. MARTIN - Intonation et reconnaissance automatique de la structure syntaxique - JE 75, Toulouse.
- 4 - A. DI CRISTO - Recherches sur la structuration prosodique de la phrase française - JE 75, Toulouse.
- 5 - J. VAISSIERE - Caractérisation des variations de la fréquence du fondamental dans les phrases françaises - JE 75, Toulouse.
- 6 - L. BUISSON, G. MERCIER - Utilisation de l'information prosodique en segmentation de la parole continue - JE 75, Toulouse.
- 7 - D. DOURS, F. FACCA, Y. LAURENTIE, G. MAURAND, G. PERENNOU - A propos de marqueurs lexico-syntaxiques : quelques exemples commentés de phrases issues de l'analyseur du projet ARIA - JE 75, Toulouse.
- 8 - J. VAISSIERE - Une procédure de segmentation automatique de la parole en mots prosodiques, en français - JE 76, Nancy.

- 9 - J.S. LIENARD - Les processus de la communication parlée ; introduction à l'analyse et à la synthèse de la parole - Masson, Paris, 1977.
- 10 - M. ROSSI - Contribution à la méthodologie de l'analyse linguistique, avec application à la description phonétique et phonologique du parler de Rossano - Thèse d'Etat, Paris, Champion, 1976.

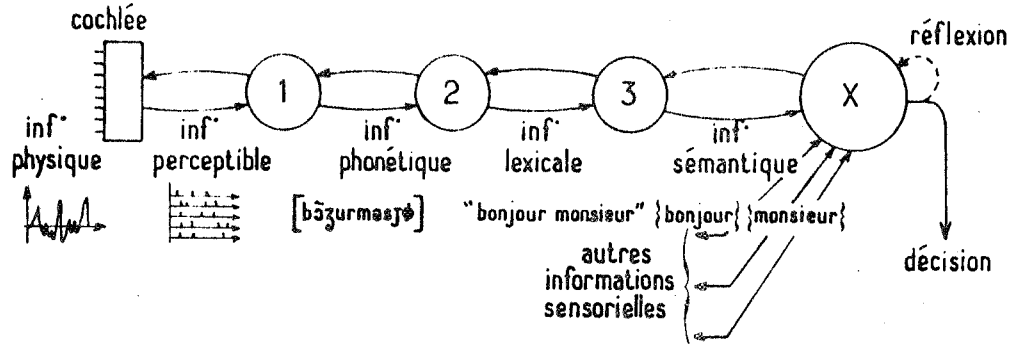


Fig. 1 - La chaîne de compréhension de la parole

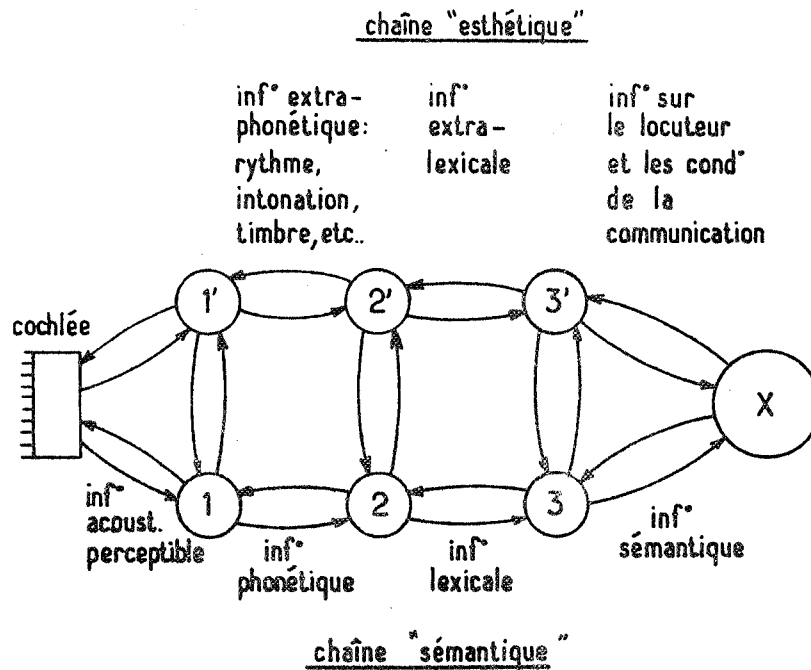


Fig. 2 - Relation entre la prosodie et les divers niveaux de compréhension

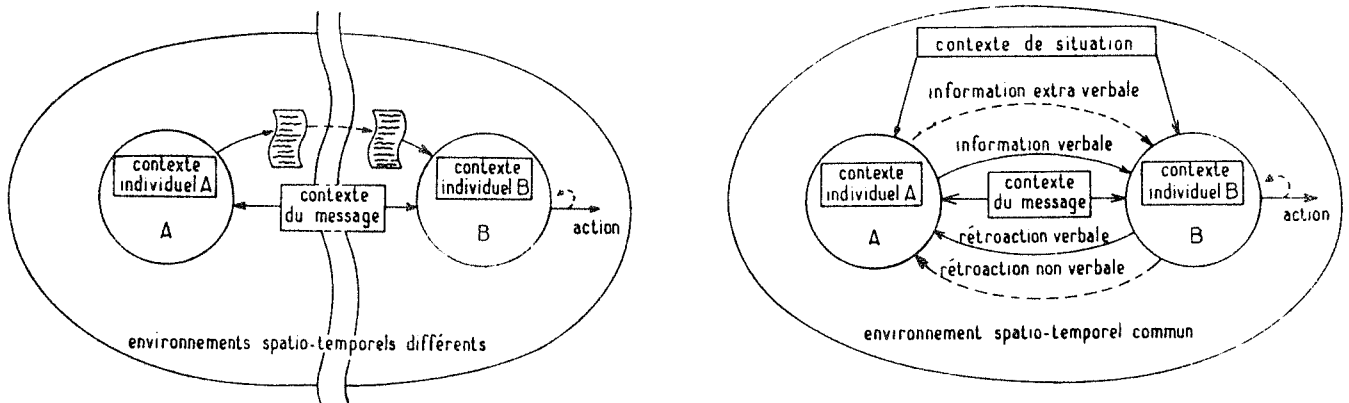


Fig. 3 - Différences entre communication parlée et communication écrite

