Université de Montréal
Département de linguistique

XIIe Journées d'études sur la parole

organisées à l'Université de Montréal
avec la collaboration du GALF
les 25, 26 et 27 mai 1981

TABLE DES MATIERES

SEANCES AFFICHEES

EFFETS DU COUPLAGE SOURCE-CONDUIT VOCAL SUR LES CARACTERISTIQUES DE L'ONDE DE DEBIT GLOTTIQUE

A. AL ANSARI, B. GUERIN

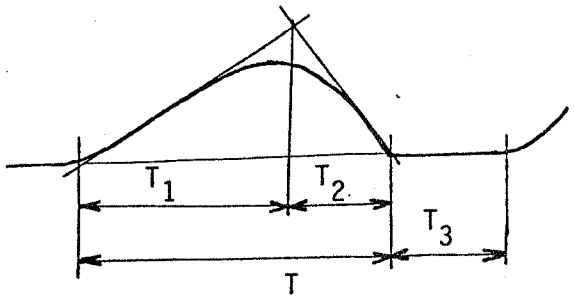
Laboratoire de la Communication Parlée - ERA 366 - E.N.S.E.R.G. 23, rue des Martyrs 38031 GRENOBLE CEDEX FRANCE

Nous avons étudié quelques effets du couplage source-conduit vocal sur la fréquence fondamentale d'oscillation du modèle de la source vocale à deux masses simulé numériquement et sur les caractéristiques de forme de l'onde de débit. Les études ont été poursuivies pour 10 voyelles orales françaises : (u, o, ɔ, a, ɛ, i, y, ϕ,æ)

Le conduit vocal est représanté, pour chacune des voyelles, par son impédance d'entrée, impédance d'entrée dont les caractéristiques ont été redéfini avec précision.

L'évolution de la fréquence d'oscillation intrinsèque du modèle de la source vocale avec les différentes configurations du conduit vocal est identique à celles déjà mesurée par GUERIN et al : F_0 croit légérement avec F_1. cette augmentation est plus faible que dans les simulations précédentes : on peut attribuer cet écart à la définition nouvelle et plus précise de l'impédance d'entrée. Elle confirme donc les conclusions déjà citées quant à l'origine de la fréquence intrinsèque des voyelles naturelles : le couplage acoustique ne peut l'expliquer.

La forme de l'onde de débit glottique peut-être caractérisé, principalement, par deux valeurs : le quotient d'ouverture et le quotient de dissymétrie (cf figure).



$$\text{quotient d'ouverture} = \frac{T1 + T2}{T}$$

$$\text{quotient de dissymétrie} = \frac{T1}{T2}$$

L'onde de débit est approché par un triangle et on évalue : QO = T1 + T2/T et QD = T1/T2 pour les 10 voyelles citées pour 9 valeurs des paramètres de commande du modèle. On constate que pour toutes les voyelles considérées, le quotient d'ouverture QO est plus faible que celui obtenu lorsque la source vocale n'a pas de charge. Par contre le quotient de dissymétrie est généralement plus élevé quant il y a couplage. On peut également mesurer la pente α du spectre de l'onde de débit. On montre qu'elle est bien correlée avec le rapport Q.O/QD : plus QD est grand, plus les composantes hautes fréquences seront élevées et plus la pente du spectre est faible en valeur absolue.

L'évolution de QO, QD et α en fonction des paramètres de commande de la source vocale a également été étudié.

Bibliographie

ISHIZAKA K., FLANAGAN J. 1972 - BSTJ 51 1233-1268
GUERIN B, BOE L.J. 1980 Phonetica 37 n° 3 159-169
SWAJ W.G. 1979 J.A.S.A 66(2) 358-362

SOURCE-TRACT COUPLING EFFECTS ON THE GLOTTAL SIGNAL CHARACTERISTICS

A. AL ANSARI, B. GUERIN

Laboratoire de la Communication Parlée - ERA 366 - E.N.S.E.R.G. 23, rue des Martyrs 38031 GRENOBLE CEDEX - FRANCE

The glottal signal is calculated from the two-mass model of the vocal cords loaded by the input tract impedance. A study of various coupling effects is desmonstrated such as the modification of fundamental frequency and that of the volume velocity form characteristics. Our investigation is limited to the following 10 French vowels : (u, o, ɔ, ɛ, e, i, y, ø, œ).

For each vowel, we determined the input vocal tract impedance with a high precision.

The intrinsic frequency evolution of the model with different configurations of the vocal tract is an identical one with that measured recently by GUERIN et al : F_0 increases a little with F1, but this increase is slightly small compared with the previous results. This difference is due to a more accurate definition of input vocal tract impedance, hence this coupling cannot explain the intrinsic frequency behaviour for natural vowels.

The caracteristics of the glottal volume velocity may be mainly discribed by two factors : the opening ratio and the dissymmetry ratio (or open quotient and speed quotient) as indicated in figure 1.

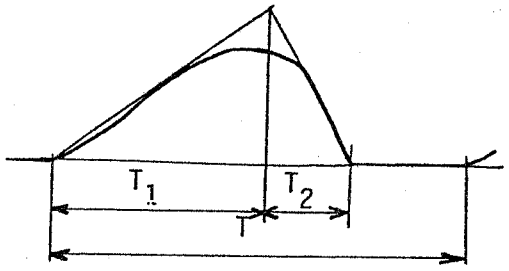


Fig. 1 Definition of opening ratio $Q.O = \frac{T1 + T2}{T}$ and dissymmetry ratio $Q.D. = \frac{T1}{T2}$

For the later one, the signal shape is approximated by a triangle and both ratio can be easily found for the mentioned vowels while 9 command parameter values are used. For all the considered French vowels, we can ascertain that the loaded vocal source has a smaller opening ratio Q.O. than that of the unloaded one. On the other hand, the dissymmetry ratio is remarkably increased by the acoutic coupling. The spectrum slope α which was directly measured using FFT is a function of Q.O/Q.D.. Due to the increase of Q.D., the high frequency component increases and consequently the absolute value of the spectrum slope. The different measurements have also given an opportunity to study the evolution of all the caracteristics : Q.O., Q.D. and α as a function of the command of the vocal cords parameters.

Bibliography
ISHIZAKA K., FLANAGAN J. 1972 - BSTJ 51 1233-1268
GUERIN B., BOE L.J. 1980 Phonetica 37 n°3 159-169
EWAN W.G. 1979 J.A.S.A. 66(2) 358-362

Etude longitudinale d'un cas d'anarthrie "pure": analyses préliminaires

Danièle Archambault*, Michèle Bergeron**, Laurent Santerre*, Denise Lafont***

* Département de linguistique et philologie, Université de Montréal
** Ecole d'orthophonie et d'audiologie, Université de Montréal
*** Centre du langage, Hôtel-Dieu de Montréal

Il s'agit de montrer, premièrement, comment les déviations phonétiques et/ou phonémiques présentes dans le discours anarthrique s'accompagnent aussi de perturbations au niveau du débit syllabique et de la durée syllabique moyenne et, deuxièmement, que ces deux types de comportement semblent interdépendants tout au long de leur évolution.

L'analyse des segments a été faite sur une base acoustique et auditive et les différentes réalisations ont été regroupées selon trois types de déviations:

1. déviations phonétiques (ex.: /s/→ [s̠])
2. déviations phonémiques (ex.: /s/→ [ʃ])
3. déviations phonétiques / phonémiques (ex.: /s/→ [s ʃ])

On a alors procédé à une étude quantitative et qualitative de ces changements, d'abord à l'intérieur de chacun des corpus, puis de façon longitudinale; cette dernière étude montre une amélioration considérable au cours de l'évolution de la maladie.

Avant d'effectuer les analyses de débit syllabique et de durée syllabique moyenne, nous avons décomposé le temps d'élocution des messages de la façon suivante:

1. temps de pauses
2. temps d'interjections
3. temps de conduites d'approches et de révision
4. temps de parole réelle

On a ensuite réalisé une étude comparative des corpus du point de vue de la répartition du temps d'élocution (tableau I), de la durée syllabique moyenne et du débit syllabique (i.e. nombre moyen de syllabes par seconde) (tableau II). Les résultats témoignent d'une évolution marquante entre le premier et le dernier corpus, soit huit mois plus tard.

Tableau I

	Corpus I	Corpus II
temps de pauses	28%	22.5%
temps d'interjections	4.49%	5.3%
temps de conduites d'approches	11%	1.9%
temps de parole réelle	57%	70.2%
durée totale	135.4 sec	126.7 sec

Tableau II

	Corpus I	Corpus II
débit syllabique	1.7 syll/sec	3.3 syll/sec
durée syllabique moyenne	40 cs/syll	26 cs/syll

Matériel disponible: bandes sonores
tableaux d'analyses

Longitudinal study of a case of "pure" anarthria: preliminary analyses

Danièle Archambault*, Michèle Bergeron**, Laurent Santerre*, Denise Lafont***

* Department of Linguistics and Philology, Université de Montréal
** School of Speech and Hearing, Université de Montréal
*** Language Centre, Hôtel-Dieu Hospital, Montreal

We want to show, first, how the phonetic and/or phonemic deviancies present in anarthric speech are accompanied by disturbances in syllabic rate of utterance and mean syllabic length and secondly, that these two types of behavior seem to be interdependant throughout the resolution of the anarthria.

Segmental analysis was made on an acoustic and auditory basis and the different productions were classified into three types of deviancy:

1. phonetic deviancies (e.g. /s/→[s̪])
2. phonemic deviancies (e.g. /s/→[ʃ])
3. phonetico-phonemic deviancies (e.g. /s/→[s͡ʃ])

We then undertook a quantitative and qualitative study of these changes, first within each sample, then longitudinally. The longitudinal study shows a considerable improvement during the course of the disorder.

Before carrying out the analyse of syllabic rate of utterance and mean syllabic length, we broke down the message enunciation time as follows:

1. pause time
2. interjection time
3. time occupied by preparatory manoeuvres and revisions
4. time occupied by true speech

We then made a comparative study of the samples in terms of the distribution of enunciation time (Table I), mean syllable length and syllabic rate (i.e. mean number of syllables per second) (Table II). The results show a marked change from the first sample to the second, taken eight months later.

Table I

	Sample I	Sample II
pause time	28%	22.5%
interjection time	4.49%	5.3%
preparatory manoeuvres	11%	1.9%
true speech time	57%	70.2%
total duration	135.4 sec	126.7 sec

Table II

	Sample I	Sample II
syllabic rate of utterance	1.7 syll/sec	3.3 syll/sec
mean syllabic length	40 cs/syll	26 cs/syll

Materials available: tape recordings
tables of analysis

SYNTHÈSE DE PAROLE, INDICES ACOUSTIQUES ET HANDICAP AUDITIF.

L'étude de la pathologie de la perception de la parole est fondamentale pour les handicapés auditifs, mais aussi pour la recherche théorique et technologique en communication parlée.

1/ APPORTS A LA DEFINITION ET A LA HIERARCHIE DES INDICES ACOUSTIQUES, pondérant : les analyses spectrographiques ou oscillographiques en fonction du seuil d'audibilité (OWENS, BENEDICT, 1972) ; les résultats des tests de perception (STEVENS, BLUMSTEIN, 1979) en fonction des modifications pathologiques des seuils différentiels (temps, fréquence, intensité) ou de l'allongement des temps de latence ou de la compression du champ dynamique. Ex : le pôle de fréquence à 8 KHz des continues graves /f-v/ ne peut avoir autant d'importance que les pôles plus graves en raison du seuil d'audibilité moyen de la population de plus de 50 ans (CHAFCOULOF, DI CRISTO, SEIMANDI, 1975).

2/ PERCEPTION DE LA PAROLE NATURELLE PAR LES HANDICAPES AUDITIFS :
- mise en évidence des confusions d'identification phonémique par des tests basés sur le principe des paires minimales, d'après ROSSI, 1975. Ces erreurs sont fonction du contexte vocalique, de l'âge du patient et du type de surdité (BERAHA, 1976, 1977, 1978).
- analyse spectrographique des mots des tests précédents filtrés par des appareils de correction auditive (BERAHA, 1976, 1977).

3/ PERCEPTION DES INDICES ACOUSTIQUES DES CONSONNES CONTINUES NON-VOISEES.
A partir des travaux précédents, élaboration de mots synthétiques V-C-V sur synthétiseur à formant. Le but est d'étudier l'influence sur l'identification des consonnes continues non-voisées des paramètres :
a- fréquence et largeur de bande des pôles de bruit ou des bursts des consonnes.
b- loci et tempo des transitions des formants vocaliques.
Les deux voyelles (durée = 200 ms) sont identiques (a-C-a) et contiennent 3 formants (F1-F2-F3). 3 contextes vocaliques sont étudiés : aigu - compact - grave. La fréquence des pôles de bruit de la consonne varie de 1000 Hz à 9000 Hz par pas de 500 Hz. Les mots, enregistrés sur magnétophone NAGRA, en ordre aléatoire, sont présentés par haut-parleur en champ libre au sujet dont la tâche consiste à répéter les mots. Il est en effet difficile d'utiliser des tests de choix multiples avec des patients souvent âgés et en cours de consultation prothétique.

Nous ne donnerons ici que les résultats pour les transitions horizontales qui sont les plus significatives pour les applications de cette expérimentation (BERAHA, 1978, à paraître) : les "monstres" sont souvent mieux identifiés que les copies du naturel. En contexte vocalique aigu, un pôle de bruit centré sur 1000 Hz est parfaitement identifié comme /s/ par les déficients auditifs (dans la mesure où le /θ/ n'existe pas en français) ; un pôle de bruit à 2000 Hz est identifié /f/ ; un pôle de bruit à 4000 Hz est identifié /ʃ/.

4/ CONCLUSIONS.
Ces résultats, très partiels, montrent cependant que le domaine de la recherche en pathologie de la perception de la parole est vaste, largement inexploré (tout au moins pour le français) et pourrait aboutir à des applications immédiates très utiles pour les handicapés auditifs si les chercheurs étaient plus nombreux à s'y intéresser. Les chercheurs ou étudiants désireux de travailler dans ce domaine trouveraient l'aide immédiate et efficace des audioprothésistes pour l'élaboration et la mise en oeuvre des protocoles expérimentaux, la sélection et le recrutement des sujets handicapés auditifs bénévoles, le contact avec les industriels français et étrangers de l'audiologie prothétique.

MATERIEL ACOUSTIQUE PRESENTE : bande magnétique, vitesse 38 cm/s.
MATERIEL VISUEL : photocopies de sonagrammes.

JP BERAHA, Laboratoire d'Audiologie - TOULON, France,
représentant l'ASSOCIATION DES AUDIOPROTHESISTES FRANCAIS.

SPEECH ANALYSIS, ACOUSTICAL FEATURES AND DEAFNESS.

The study of speech perception pathology is of primary interest to deaf persons and to theorical and technological research in the field of speech communication.

1/ DEFINITION AND HIERARCHY OF ACOUSTICAL FEATURES,
weighting : spectrographic and oscillographic analyses depending on the audibility threshold (OWENS, BENEDICT, 1972); results of perception tests (STEVENS, BLUMSTEIN, 1979) depending on the pathological modifications of differential thresholds (time, frequency, intensity) or on the latency time lengthening or on the dynamic compression. E.g : the 8 KHz friction noise of grave continuants /f-v/ may not have the same weight as lower noises owing to mean hearing threshold of persons over 50 years of age (CHAFCOULOF,DI CRISTO, SEIMANDI, 1975).

2/ NATURAL SPEECH PERCEPTION BY DEAF PERSONS :
- study of phonemic confusions by means of diagnostic rhyme tests (ROSSI, 1975). Confusions are explained by vowel context, age of subjects, type of deafness (BERAHA, 1976,1977, 1978).
- spectrographic analysis of the words of the above mentioned tests filtered by hearing aids (BERAHA, 1976, 1977).

3/ PERCEPTION OF ACOUSTICAL FEATURES BY NON-VOICED CONTINUANTS.
From the previous experimentations, synthetic V-C-V words have been worked out on a formant synthetiser, in order to study the impact on identification of non-voiced continuants by the following parameters :
a- frequency and bandwith of friction noise or bursts of consonants.
b- loci and tempo of vowel formant transitions.
The two vowels (duration : 200 ms) are identical (a-C-a) and are made of 3 formants (F1-F2-F3). 3 vowel contexts are [redacted] under study : acute - compact - grave Consonant noise friction frequency is modified from 1000 Hz to 9000 Hz by 500 Hz steps. Words, tape recorded on NAGRA, on a random order, are presented by a loudspeaker in a free field to the subject who repeats the words. The use of multiple choice of tests is impossible with subjects, often elderly, and under prosthetic examination.

Results for horizontal transitions will only be given here : they are the most significant for application of this experimentation (BERAHA, 1978) : "monsters" are often best identified than natural copies. In an acute vowel context, a friction noise centered on 1000 Hz is well identified as /s/ by deaf persons (as far as /θ/ does not exist in french) ; a friction noise of 2 000 Hz is identified /f/ ; a friction noise of 4000 Hz is identified /ʃ/.

4/ CONCLUSIONS.
Those results show, partly, that research in the field of speech perception pathology is large, far from beeing fully investigated (at least for french) and could give way to immediate applications to deaf persons if there were more scientists to work at it. Scientists and students, interested in this field, would receive the immediate and efficient help from hearing aid acousticians to work out experimental methodology, and selection of voluntary hard-of-hearing subjects, contacts whith hearing aids industry.

ACOUSTICAL MATERIAL : tape, 38 cm/s
VISUAL MATERIAL : sonagrams.

JP BERAHA, Laboratoire d'Audiologie - TOULON, France.
on behalf of ASSOCIATION DES AUDIOPROTHESISTES FRANCAIS.

ETUDE STATISTIQUE DE LA DYNAMIQUE ARTICULATOIRE DE PHONEMES VOCALIQUES

Renée Béland, Université de Montréal.

Des recherches en modélisation articulatoire montrent qu'il est possible de reproduire la plupart des positions de la langue par une combinaison de deux ou trois configurations principales. Si ces modèles statiques et/ou dynamiques parviennent à générer l'ensemble des phonèmes du français ou de l'anglais, la majorité d'entre eux manquent de précision dans la définition des mouvements de la lame et de la pointe de la langue.

Nous avons refait l'analyse en composantes principales proposée par S.Maeda (1) à laquelle nous avons ajouté l'aspect dynamique articulatoire. La qualité des films cinéradiologiques que nous avons dépouillés assure une plus grande précision dans la partie antérieure de la cavité buccale. Le corpus comprend 50 rencontres de type C_0VC où V est toujours l'une des trois voyelles [i], [a], [u].

Pour les 50 rencontres C_0VC, nous avons tracé à l'aide d'un projecteur les contours linguaux de chacune des images de la durée de la voyelle. En moyenne chaque voyelle compte 11 images. L'ensemble des quelque 530 tracés a été soumis à une grille qui divise le conduit vocal en 15 sections réparties dans une demi-circonférence. Nous avons opéré des divisions trois fois plus fines dans la partie antérieure que dans la partie postérieure du canal buccal. Les 15 coordonnées forment autant de rayons dont nous étudions les variations de longueur dans le temps. La mesure d'un rayon est définie par la distance séparant l'origine fixe du point d'intersection avec le contour lingual.

Le programme statistique "STATIS" que nous utilisons permet d'effectuer une analyse factorielle en fonction d'un ou plusieurs paramètres. Nous obtenons ainsi les contours principaux relatifs à chacune des voyelles [i], [a], [u] pour chacune des deux locutrices et l'évolution de ces contours dans le temps.

---

(1) Maeda Shinji, Un modèle articulatoire de la langue avec des composantes linéaires dans 10 ièmes journées d'Etude sur la Parole, Grenoble 30 mai- 1er juin 1979, pp. 152-162

12èmes JOURNEES D'ETUDES SUR LA PAROLE

MONTREAL 25, 26, 27 mai 1981

Définition Emg des voyelles du français

Anne Brunelle
(Université de Montréal)

Nous voulons dans cette étude décrire les propriétés mécaniques des lèvres supérieure et inférieure pendant la production des voyelles du français. L'activité électromyographique des muscles orbicularis oris inferioris, orbicularis oris superioris, buccinator et mentalis a été observée au moyen d'électrodes de surface, et ce pour sept locuteurs. Une analyse statistique a été faite au moyen d'un ordinateur, afin de définir certaines constantes électromyographiques pour les voyelles du français. Les résultats obtenus nous permettent de proposer une définition de la labialité en trois traits, soient [±rond], [±écarté] et [±projeté], chacun de ces traits correspondant à une activité musculaire spécifique.

This study is an attempt to describe the mechanical properties of upper and lower lip movement during vowel production in French. Emg activity from orbicularis oris superioris, orbicularis oris inferioris, buccinator and mentalis muscles was observed, and this for seven subjects. A computer was employed to produce a statistical analysis to determine emg constants. Results suggested a definition of labiality in three features: [±round], [±spread], [±projected], each of these features being the product of a specific muscular activity.

ETUDE DE QUELQUES PARAMETRES ACOUSTICO-PHONETIQUES EN VUE DE L'IDENTIFICATION DU LOCUTEUR.

CAELEN GENEVIEVE, LABORATOIRE C.E.R.F.I.A., 118 route de Narbonne, 31062 TOULOUSE CEDEX.

Un texte formant une unité sémantique de quatre phrases et d'une cinquantaine de mots, a été enregistré cinq fois par des locuteurs masculins, sur une période d'un an. Un soin attentif a été porté aux conditions d'enregistrement, concernant tant les locuteurs (absence de rhumes), que le choix des lieux, ou les techniques (distance source/micro, réglages identiques...).

Les spectrogrammes issus d'un modèle d'oreille (Jean CAELEN, Laboratoire C.E.R.F.I.A., 1974) ont été segmentés manuellement en vue de l'analyse acoustico-phonétique. Cette étude prend en considération des indices, tels que, concernant les occlusives sourdes, les phénomènes de sonorité à l'implosion et l'explosion, l'énergie différentielle à l'explosion, et concernant les voyelles, les phénomènes de sourdité initiale après un contexte consonantique sourd, ou durée d'établissement de F_o. Cette étude prend également en considération la répartition de l'énergie dans les graves et les aigus du spectre des liquides, des consonnes nasales et des voyelles, les phénomènes relatifs à la stabilité spectrale , à la stabilité énergétique, ou à la nasalité.

Un traitement statistique de ces données nous permet de proposer des tableaux comparatifs inter et intra locuteurs de pourcentages d'occurrence de ces divers phénomènes, de leurs durées, ou des rapports de durées entre les différents avatars intraphonémiques et les phonèmes concernés, des différentes mesures de ces paramètres, et de présenter, à partir des différences constatées, quelques résultats en matière d'identification des locuteurs.

A STUDY OF A FEW ACOUSTICO-PHONETIC PARAMETERS FOR SPEAKER RECOGNITION

CAELEN GENEVIEVE, LABORATOIRE C.E.R.F.I.A., 118 Route de Narbonne, 31062 TOULOUSE CEDEX.

A text making up a semantic unit of four sentences and about fifty words, was recorded five times by male speakers, over a period of one year. Attention was payed to recording conditions, concerning as well the speakers (free from colds), the choice of the room, or the techniques (distance source/micro, identical adjustements...).

The spectrograms issued from a model of ear (Jean CAELEN, Laboratoire C.E.R. F.I.A., 1974) have been manually segmented in order to the acoustico-phonetic analysis. This study consider cues such as, about"surd stops", the phenomena of sonority during implosion and burst, in sonorant context, the differential energy in the burst, and, about vowels, the phenomena of initial surdity after an unsonorant consonantic context. This study consider too the repartition of energy in the low and high frequencies of the spectrum of the liquids, the nasal consonants and vowels, the phenomena relative to the spectral stability and energetic stability, or to nasality.

A statistical treatment of these data enable us to purpose comparative inter and intra-speakers tables of occurrence percentages of these various phenomena, of their duration, or timing relations between the various intraphonemic characteristics and the concerned phonems, of different measures of these parameters, and to present, from differences observed among talkers, a few results for speaker recognition.

12EMES JOURNEES D'ETUDE SUR LA PAROLE

MONTREAL, 25-27 MAI 1981

ANALYSE, SYNTHESE ET TRANSMISSION DE LA PAROLE : CANAUX OU PREDICTION LINEAIRE ?

M. CARTIER CNET - Lannion A 22301 LANNION FRANCE

Trente ans après l'invention du vocodeur par Dudley, l'introduction de la prédiction linéaire dans le traitement de la parole a sans nul doute représenté une étape importante dans l'évolution des techniques d'analyse, de synthèse et de transmission de la parole (ASTP). De là à considérer les vocodeurs à canaux comme périmés, il n'y a qu'un pas. Avant de le franchir essayons de poser le problème de façon aussi précise et complète que possible.

Il convient de s'interroger sur l'éventualité d'une "filière" homogène d'ASTP, c'est-à-dire d'un ensemble de composants compatibles entre eux, qui seraient répartis dans un réseau de communication homme-machine pour transmettre la parole, l'analyser, la stocker, la restituer, la fabriquer.

En tout cas le problème de la comparaison canaux/P.L. se décompose en trois questions relativement indépendantes liées respectivement à l'analyse (→ RAP), à la synthèse (→ SAP) et à la transmission.

Pour l'analyse, on avance parfois comme avantage de principe pour les coefficients Parcor de traduire le rapport de sections adjacentes du conduit vocal ; parallèlement, un banc de filtres pourrait présenter l'avantage, tout aussi théorique, de réaliser une analyse fréquentielle au même titre que la cochlée. Pratiquement on dispose de données contradictoires sur le plan des performances. Que se passe-t-il en présence d'un bruit ambiant ?

Des synthétiseurs à prédiction linéaire existent sous forme de circuits intégrés. Les difficultés rencontrées lors de la manipulation de données en synthèse par diphones, si elles ne constituent pas un obstacle infranchissable, amènent cependant à s'interroger. L'utilisation de synthétiseurs à canaux ne se heurte qu'à des problèmes technologiques (celui de la qualité est sur le point d'être résolu) ; elle est liée à l'existence de vocodeurs à canaux intégrés.

Pour la transmission, le débat est ouvert. A 4800 bit/s la prédiction linéaire peut donner une parole d'excellente qualité. Inversement des résultats tout aussi intéressants ont été obtenus avec des vocodeurs à canaux ; ceux-ci résistent mieux au bruit et aux erreurs de transmission ; la technologie des dispositifs à transfert de charge et des capacités commutées permet leur intégration.

On propose de débattre de ces questions à partir de quelques données chiffrées et d'un jeu "Qui, quand, combien ?" où l'on devra écouter des phrases et indiquer pour chacune d'elles l'auteur et l'origine, le locuteur, le système de codage et son débit.

12EME JOURNEES D'ETUDE SUR LA PAROLE

MONTREAL, 25-27 MAI 1981

ANALYSIS, SYNTHESIS AND TRANSMISSION OF SPEECH : CHANNEL VOCODERS OR LINEAR PREDICTION ?

M. CARTIER CNET - Lannion A 22301 LANNION FRANCE

Thirty years after Dudley's invention of the vocoder, for use in analysis, synthesis and transmission of speech, the development of linear prediction threatened to render it obsolete. With more than ten years of experience with linear prediction now behind us, we shall consider the question of whether channel vocoders still have a role to play.

We pose the problem in the context of an eventual integrated system, capable of analyzing, synthesizing, storing and transmitting speech. However, the question of which we should use, channel vocoders or linear prediction, still decomposes into three relatively independent parts, corresponding to analysis, synthesis and transmission.

For analysis, proponents of prediction coefficients point out that they can be directly translated as a representation of the vocal tract by a series of adjacent cylindrical sections. However, a filter bank can claim kinship with cochlea, which performs the same sort of frequential analysis. Experience with speech recognition systems has not so far shown clearcut performance advantages for either technique. As far as we know, results about ambiant noise sensitivity are lacking.

There are already commercialy available integrated circuits which synthesize speech by linear prediction. Difficulties encountered on manipulating speech samples for the purpose of synthesis by diphones, while perhaps not insuperable, nevertheless give one pause. There is a problem of voice quality with existing filter bank synthesizers, but its solution seems imminent. Of course the utility of such synthesizers will depend on the possibility of realizing them as integrated circuits.

For transmission, the question remains open. At 4800 bits/sec, linear prediction can give speech of excellent quality. On the other hand, good results have also been obtained with channel vocoders, which are more resistant to noise and transmission errors. They can also be put into integrated circuits using CCD or switched capacitors.

We propose to discuss these questions in terms of some quantitative data and a game where contestants will listen to sentences and try to guess who wrote, who spoke them, how they were coded and at what bit rate.

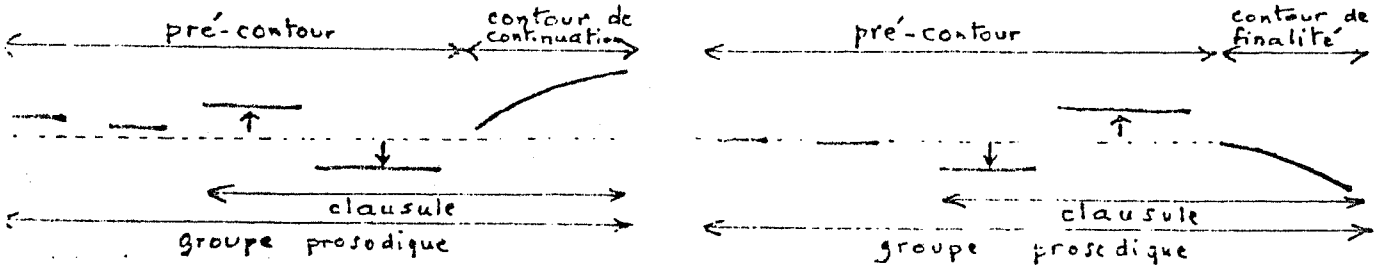
Fernand CARTON (Nancy)
12e J.E.P. Montréal MAI 1981

TYPOLOGIE DES CLAUSULES RYTHMICO-MÉLODIQUES DANS LES VARIÉTÉS DE FRANÇAIS

1. BUT DE LA RECHERCHE: trouver des indices prosodiques caractérisant les diverses variétés géographiques et socio-culturelles de français. Objectiver les qualificatifs "traînant","chantant",etc. appliqués aux "accents".

2. DEFINITION : clausules = arrangement de longues/brèves, de montées/descentes (éventuellement d'ictus d'intensité) affectant 2 ou 3 syllabes à la fin de groupes prosodiques polysyllabiques. De complexe de configurations peut se stéréotyper, quelle qu'en soit l'origine (phonologique, morphologique, lexicale). La clausule fonctionne alors comme marque et comme surmarque (utilisable par des imitateurs du parler en question), et caractérise un sociolecte (catégorie socio-professionnelle ou culturelle) et/ou un géolecte (français national, régional, dialectal).

3. DIFFERENCE ENTRE CLAUSULE ET CADENCE (prétonique + tonique). La prosodie dialectale française utilise souvent un effet de "double bascule" (contraste portant sur trois syllabes).



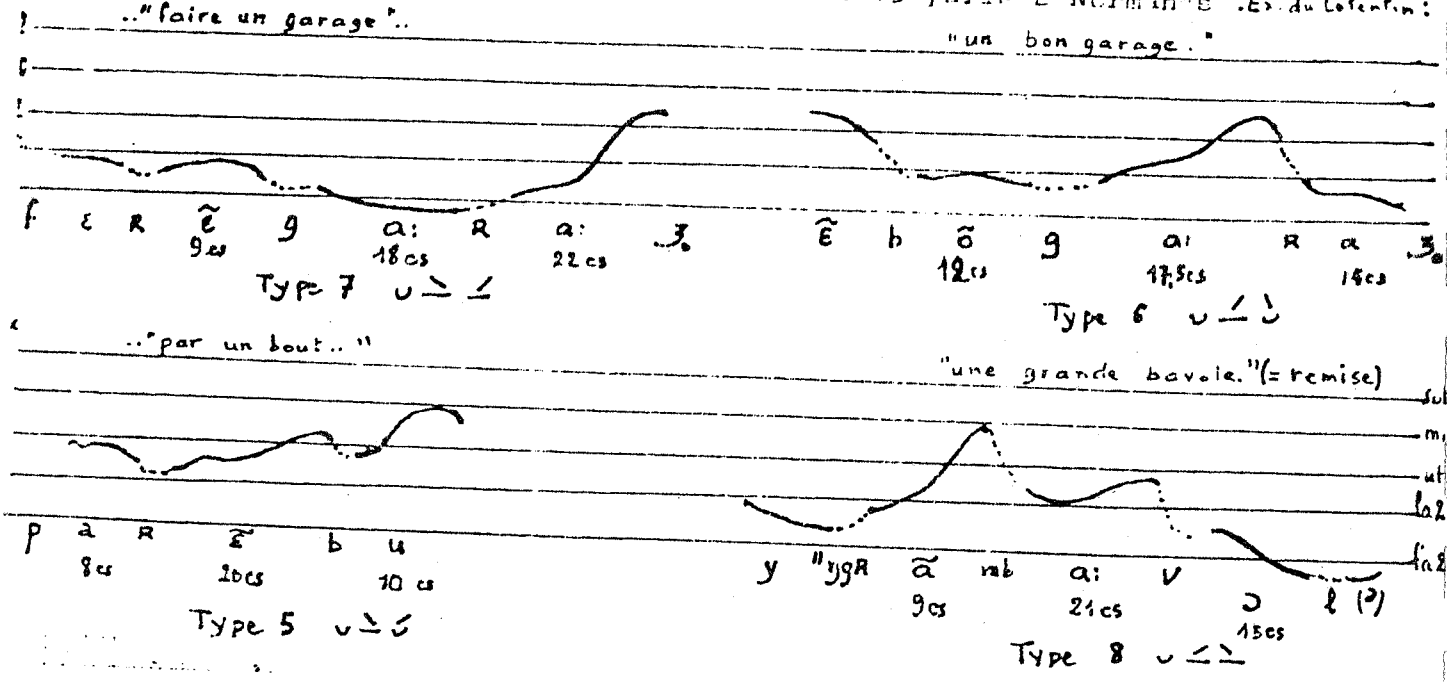
4. NOTATION. S'inspire de la métrique de la prose gréco-latine (et des neumes du Haut Moyen-âge) pour constituer des patrons rythmico-mélodiques (patterns); exemple: "Atmosphère!.. Est-ce que j'ai une gueule d'atmosphère?" (réplique d'Arletty dans Hôtel du Nord, film de Carné, 1939 : a) ∪ ⁄ ⁄ b) ∪ ⁄ ⁄

5. PROBLEMES DE TRANSPOSITION à partir de l'analyse acoustique. Calculs de durée moyenne, transposition perceptive de la mélodie.

6. TYPOLOGIE. Théoriquement, pour 3 syllabes, on pourrait avoir 4^3 =64 possibilités. Des contraintes réduisent évidemment ce nombre. Les plus récurrents dans 3 régions de la France septentrionale :

1. ∪ ⁄ ⁄
2. ∪ ⁄ ⁄
3. ∪ ⁄ ⁄
4. ∪ ⁄ ⁄
5. ∪ ⁄ ⁄
6. ∪ — ⁄
7. ∪ ⁄ ⁄
8. ∪ ⁄ ⁄
9. — ⁄ ⁄
10. — ⁄ ⁄
11. — ⁄ ⁄
12. — ⁄ ⁄
13. — ⁄ ⁄

7.APPLICATION. 7 publications s'inspirent de [illegible] dans l'Est(Lorraine), dans le Nord, dans l'Ouest (3 parlers Normands). Ex. du Cotentin :



Séries minimales et évaluation du codage acoustique-phonétique des système de communication parlée.

Gérard CHOLLET*, Alain ASTIER**, Mario ROSSI*.

* Institut de Phonétique, C.N.R.S. LA 261, 29 ave R. Schuman 13621 AIX EN PRVCE
** R.N.U. Renault, DSI, B. P. 103 - 92109 BOULOGNE BILLANCOURT

Un locuteur, auditeur d'une langue est capable de distinguer des unités parlées qui ne diffèrent que par un seul phonème (ex : kapo/kabo/kato/kado/kano/kaʒo/kavo/kaʃo/kajo/kalo/kaʁo/kao/). De telles unités parlées forment une série minimale. Nous utilisons des séries minimales pour évaluer l'intelligibilité des systèmes de synthèse et de transmission, et les performances des systèmes de reconnaissance automatique de la parole. En synthèse et transmission, cette technique n'est pas nouvelle (VOIERS, 1967; ROSSI et PECKELS, 1973; CARTIER et ROSSI, 1973). Elle permet d'établir un diagnostic sur les éléments éventuellement défaillants d'un système de codage. Le même principe peut être utilisé pour les systèmes de reconnaissance (CHOLLET et al, 1981).

Une bande de test comprenant sur une piste : 9 séries minimales répétées 6 fois par un locuteur; et sur l'autre piste un bruit blanc. Cette bande est disponible sur simple demande aux auteurs. Le phonème variable est soit une voyelle (pour 2 séries) soit une consonne (pour les 7 autres). Les unités parlées sont de 1 à 2 syllabes. Les voyelles sont testées en position interconsonantique et finale. Les consonnes sont testées à l'initiale, en position intervocalique, et finale.

La méthode consiste à établir une matrice de confusion entre éléments de la série (les confusions inter-séries sont négligeables). Des matrices ont été obtenues pour des systèmes de reconnaissance de mots isolés et de parole continue, et pour un échantillon d'auditeurs humains. L'analyse des confusions est effectuée en termes de traits phonétiques. D'une façon générale la nasalité pose le plus de problème pour la reconnaissance automatique des voyelles, alors que l'être humain a plus de difficultés avec l'opposition grave/aigu. Les systèmes testés confondent les consonnes de façon diverses. Une utilisation systématique des techniques de comparaison dynamique a pour conséquence la confusion de l'opposition interrompu/non interrompu. D'autres systèmes ont des difficultés avec les traits de voisement et de nasalité, oppositions très résistantes chez l'humain.

La méthode d'évaluation proposée est limitée au niveau acoustique-phonétique. Nous pensons cependant qu'elle permet de contrôler l'évolution des performances, de comparer les systèmes et de choisir les algorithmes les plus performants.

Références :

- CARTIER M., ROSSI M. "Le test de diagnostic par paires minimales mises en oeuvre et résultats" Symp. Int. de la parole, LIEGE (1973).
- CHOLLET G., ASTIER A., ROSSI M. "Evaluating the performance of speech recognizers at the Acoustic-Phonetic level" ICASSP, Atlanta (1981).
- ROSSI M., PECKELS J.P., "Le test de diagnostic par paires minimales" Revue d'Acoustique 27, 245-262 (1973).
- VOIERS W.D. "Performance of speech processing devices, III : diagnostic evaluation of speech intelligibility" Final report AFCRL-0101 (1967).

Minimal series and the evaluation of speech communication systems at the acoustic-phonetic level

Gérard CHOLLET*, Alain ASTIER**, Mario ROSSI*.

* Institut de Phonétique, C.N.R.S. LA 261, 29 ave. R. Schuman 13621 AIX EN PRVCE
** R.N.U. Renault, DSI, B. P. 103 - 92109 BOULOGNE BILLANCOURT

A fluent speaker of a language can distinguish utterances that differ by one phoneme only (ex : the series of french words /kapo/kabo/kato/kado/kano/kaʒo/kavo/kaʃo/kajo/kalo/kaʁo/kao/). Such a set of utterances is called a minimal series. We use minimal series to evaluate the intelligibility of speech synthesis and transmission systems, and the performance of automatic speech recognizers (ASR). This technique is being used routinely to test vocoders (VOIERS 1967; ROSSI & PECKELS, 1973; CARTIER & ROSSI, 1973). It leads to a diagnostic evalution pointing out defficient algorithms or parts. A similar technique is being used to evaluate ASR systems (CHOLLET et al, 1981).

A tape is being distributed. Nine minimal series repeated 6 times at random by one speaker are recorded on the left channel. White noise is available on the right channel. This tape is available upon request. Two series concern oppositions between vowels; the others are opposing consonants. The utterances are 1 or 2 syllables long. Vowels are tested in interconsonantal and final positions. Consonants are tested in inital, intervocalic, and final positions.

The experimental procedure consists of recording confusions between items of a series (confusions across series are very rare). Matrices of confusions will be posted for isolated word and continuous speech recognizers, as well as human listeners . Distinctive feature theory can be used for a simple but linguistically significant analysis of the confusions. In general terms, nasality is difficult to recognize automatically in vowel. Consonants are confused in many different ways. Word recognizers based on dynamic pattern matching have a tendancy to miss the feature "interrupted". Other systems detect improperly voicing or nasality, two strong features in human perception.

Our methodology is limited to acoustic-phonetic evaluation. We believe it is most valuable during the development phase of a recognizer because it allows a meaning ful choice of algorithms. We think it should be part of a standard evaluation procedure.

References :

- CARTIER M. ,ROSSI M. "Le test de diagnostic par paires minimales mises en oeuvre et résultats" Symp. Int. de la parole, Liège (1973).
- CHOLLET G., ASTIER A., ROSSI M."Evaluating the performance of speech recognizers at the Acoustic-Phonetic level" ICASSP, Atlanta (1981).
- ROSSI M., PECKELS J.P., "Le test de diagnostic par paires minimales" Revue d'Acoustique 27, 245-262 (1973).
- VOIERS W.D. "Performance of speech processing devices, III : diagnostic evaluation of speech intelligibility" Final report AFCRL-0101 (1967).

12ièmes Journées d'Etudes sur la Parole. MONTREAL 25-27 Mai 1981

COURVILLE L.*, FLOUTIER D.**, et MARCHAL A.*
*Université de MONTREAL et **Université de MONTPELLIER

"UTILISATION D'UN MICROPROCESSEUR DANS UN SYSTEME D'ELECTROPALATOGRAPHIE"

L'électropalatographie permet l'étude des aspects dynamiques de l'activité linguale pendant la production de la parole. La grande quantité d'informations obtenues (6400/sec) rend nécessaire en vue d'une interprétation phonétique le recours à un traitement informatisé. Nous avons développé à l'Université de Montréal un système EPG à 64 électrodes qui est assisté par un microprocesseur.

L'acquisition des données, leur traitement, leur stockage et l'impression des données EPG sont contrôlés par un AIM 65.

L'utilisateur n'a pas à connaître la programmation pour pouvoir l'utiliser. Il a simplement à répondre à quelques questions simples concernant les options disponibles. Ainsi il peut demander l'acquisition et l'impression des valeurs de courant sur chaque électrode : "les comptes" (Mode lux : information alphanumérique) ou l'état de chaque contact (touché; non-touché (Mode Norma : information binaire). Cette tâche demande un traitement plus élaboré puisque chaque valeur doit être comparée à un seuil. Le seuil pour chaque électrode est rentré par le clavier. Cette option permet l'utilisation de palais construits avec des normes différentes et ajoute encore à la versatilité de ce système. Cet appareil portable d'électropalatographie se révèle très performant. Il est actuellement utilisé pour la description articulatoire des consonnes.

12ièmes Journées d'Etudes sur la Parole. MONTREAL 25-27 Mai 1981

COURVILLE L.*, FLOUTIER D.**, et MARCHAL A.*
*Université de MONTREAL et **Université de MONTPELLIER .

"USE OF A MICROCOMPUTER IN AN EPG SYSTEM".

Electropalatography enables the investigation of the dynamic aspects of lingual activity during speech. The large amount of data (6400 informations/sec) and the need for a phonetic interpretation require a computer assisted processing.

We have developped at the Université of Montréal a system with 64 electrodes assisted by a microcomputer. Acquisition, treatment, stokage and printing of EPG data are performed and controlled by an AIM 65.

User does not need to know anything about programming. He has only to answer a few easy questions about available options. It may ask for an acquisition and printing of the current value on each electrode "the scores" (Lux mode : alphanumeric information) or the state of each contact : touched, non-touched (Norma mode : binary information). This program asks for a more elaborate treatment since each value has to be compared to a reference level before a decision can be taken. The decision level is entered for each electrode through the keyboard. This facility permits the use of non-standard palates and adds to the whole versatility of the system. This EPG portable unit proves to be very performant. It is at this time mainly used for the phonetic description of consonants.

VERIFICATION DU LOCUTEUR PAR COMPARAISON DYNAMIQUE DE MOTS ISOLES

R. DE FORAS - J.S. LIENARD

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
(L.I.M.S.I. - C.N.R.S.)

Il s'agit, par une méthode inspirée de la reconnaissance de mots isolés, de vérifier l'identité d'un locuteur coopératif, à partir d'une ou plusieurs émissions d'un mot convenu à l'avance.

Le signal est analysé au moyen d'un banc de 32 filtres, dont 12 seulement sont utilisés, couvrant la bande de 300 à 3400 Hz. Le spectre est mesuré 100 fois par seconde, et les amplitudes sont estimées selon une loi quasi-logarithmique. Une normalisation en amplitude est effectuée ensuite de façon à obtenir la même valeur moyenne quelle que soit la voie d'analyse considérée, ainsi que divers lissages en temps et en fréquence. Les processus d'apprentissage et de reconnaissance ne comportent ni compression ni normalisation temporelle, et le cheminement dans la matrice de comparaison est laissé "libre" : les progressions horizontales et verticales sont autorisées. Mais les écarts à la diagonale sont cumulés et fournissent, en fin de parcours, une note qui traduit les différences de rythme existant entre les deux séquences comparées. Cette note de rythme est utilisée, avec la note de ressemblance spectrale, pour classer la séquence provenant d'un locuteur inconnu par rapport aux séquences mémorisées, provenant des locuteurs autorisés.

Une partie du travail a été effectuée sur un corpus, constitué à partir d'une trentaine de locuteurs ayant prononcé chacun au moins deux fois un mot donné (le mot "informatique"), au cours de sessions d'enregistrement distinctes. Les performances obtenues sont fonction de seuils, qui traduisent un certain compromis entre la sécurité (pas d'erreurs mais des rejets) et la commodité (peu de rejets mais quelques erreurs). Voici un résultat typique, sur 60 comparaisons : si l'on règle les seuils de manière à obtenir un taux d'imposture nul, on doit tolérer 11 rejets de locuteurs autorisés, et on obtient 49 vérifications correctes.

Le même programme est expérimenté en temps réel, sur des données renouvelables. Le problème de la constitution d'un dictionnaire de mots-références à partir d'apprentissages multiples est à l'étude. C'est pour des raisons matérielles que nous n'avons pas utilisé l'évolution de F_0, qui ne pourra qu'améliorer les résultats dans un système ultérieur. Cette étude doit aboutir prochainement à des applications industrielles.

REFERENCE

R. DE FORAS, 1980, Vérification du locuteur par une méthode de comparaison dynamique - Thèse de 3e cycle, Université de Paris XI.

SPEAKER VERIFICATION USING DYNAMIC COMPARISON OF ISOLATED WORDS

R. DE FORAS - J.S. LIENARD

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
(L.I.M.S.I. - C.N.R.S.)

This paper deals with the identification of a given cooperative speaker, using a method inspired by isolated word recognition, one or several enunciations of a word agreed upon in advance are employed as the bases of the identification.

The signal is analysed by a bank of 32 filters, only twelve of which are used. These represent the band from 300 to 3400 cps. The spectrum is measured 100 times per second, and the amplitudes are estimated through a quasi-logarithmic law. Next, amplitude normalisation is carried out so as to obtain the same mean value for all possible types of analyses, as well as for the various time and frequency curve smoothings. The training pass and recognition processes use neither compression nor time normalisation, and the direction in which the system progresses through the comparison matrix in left "free" - horizontal and vertical steps are permitted. But diagonal distances are accumulated and, at the end of the total distance covered, a score is made up, representing the differences in rhythm which exist between the two signals. This "rythm" score is used, along with the spectral ressemblance score, to classify an utterance coming from an unknown speaker as compared to the memorized utterances from authorized speakers.

Part of the work was done on a corpus made up of approximately thirty speakers, each of which pronounced the word "informatique" ("computer") at least twice during separate recording sessions. The performances obtained are a function of thresholds which represent a certain compromise between security (no errors but several rejections) and convenience (few rejections but several errors). Here is a typical result based on 60 comparisons : if we tune the thresholds in a way that allows no impostors, we must be ready to tolerate 11 rejections of authorized speakers - we therefore obtain 49 correct identifications in this case.

The same program is being experimented on line, with the possibility of using new utterances. The problem of setting up a dictionary of reference words based on multiple training passes is under study. Due to material difficulties, we have not yet been able to incorporate the use of F_0 evolution. This parameter will doubtlessly ameliorate our results in a future system. This project will be applied to industrial uses.

REFERENCE

R. DE FORAS, 1980 - Verification du locuteur par une méthode de comparaison dynamique - Ph. D. Thesis, University of Paris XI.

J.M. DOLMAZON, P. ESCUDIER -21-

Laboratoire de la Communication Parlée - ERA 366 - E.N.S.E.R.G. 23, rue des Martyrs 38031 GRENOBLE CEDEX - FRANCE

ETUDE DES PHENOMENES DE PROPAGATION DANS LES FIBRES NERVEUSES CONNECTEES AUX CELLULES CILIEES

Dès lors que nous considérons l'information acoustique comme portée par une grandeur électrique (des potentiels par exemple), il devient naturel de traiter la propagation en terme de ligne électrique. Dans un précédent travail DOLMAZON (1978,1980) modélise l'intéraction cellules ciliées internes-cellules ciliées externes et montre l'importance des valeurs relatives des vitesses de propagation sur la membrane basilaire d'une part et dans les fibres d'autre part. Chacune des fibres est traitée comme un conducteur à constantes électriques réparties : R, G et C sont respectivement la résistance, la conductance et la capacité linéique de la dendrite.

La détermination des caractéristiques de la ligne se fait à partir des valeurs numériques obtenues sur le calmar géant (POTTOLA 1973). Considérant la dendrite comme un conducteur cylindrique nous connaissons la loi de variation des constantes en fonction du diamètre de la dendrite. La connaissance des diamètres respectifs chez le calmar et chez l'homme permet la détermination numérique des constantes pour l'homme à partir des mesures effectuées sur le calmar.

L'étude de la réponse impulsionnelle conduit aux résultats suivants: Une fibre radiale de longueur 0,2 mm issue d'une cellule ciliée interne introduit un retard de 12 µs ce qui correspond à une vitesse moyenne de propagation de 16,6 m/s. Par contre une dendrite de 1 mm de long constituant une fibre spirale introduit un retard de 0,3 ms ce qui conduit à une vitesse de propagation de 3 m/s. Les ordres de grandeurs précédants sont raisonnables et correspondent à ceux observés dans la littérature. De plus, négligeable dans une fibre radiale, le retard dû à la propagation devient déterminant dans une fibre spirale.

L'étude du régime harmonique montre, que dans ce modèle, chaque fibre se comporte comme un filtre passe-bas dont la fréquence de coupure est d'autant plus faible que la fibre est longue. En basse fréquence les deux types de fibre transmettent les valeurs instantanées alors qu'en haute fréquence elles ne transmettent que les valeurs moyennes. La limite fréquentielle entre ces deux types de fonctionnement déjà proposée par DOLMAZON (1980), se situe vers 2 KHz.

Ces phénomènes doivent jouer un rôle important dans le processus de codage du message véhiculé vers les centres supérieurs.

REFERENCES

.J.M. DOLMAZON, L. BASTET (1979)
Hair cell receptor modeling
J. Acoust. Soc. Am., 65, Supp N°1, S14

. J.M. DOLMAZON (1980)
Modèle de fonctionnement du système auditif
State Thesis, I.N.P.G. Grenoble (FRANCE)

. E.W. POTTALA, T.R. COLBURN, D.R. HUMPHREY (1973)
A dendritic compartment model neuron
IEEE Trans., BME 20, N°2, 132-139

J.M. DOLMAZON, P. ESCUDIER

Laboratoire de la Communication Parlée - ERA 366 - E.N.S.E.R.G 23 rue des Martyrs 38031 GRENOBLE CEDEX - FRANCE

A STUDY OF PROPAGATION PHENOMENA IN NERVE FIBERS CONNECTED TO HAIR CELLS

Since we consider that the acoustical message is conveyed by an electrical support (such as a potential), an electrical line appears to be a suitable way of introducing propagation mechanisms in the fiber. In recent publications DOLMAZON (1978, 1980) proposes a hair cell receptor model assuming an interaction between inner and outer hair cells ans shows the importance of the relative values of velocities on the basilar membrane and on the fibers. We consider one fiber as a conductor with electrical characteristics uniformly distributed. One length unit is represented by a quadripole with a resistance R, a conductance G and a capacitance C.

The characteristics of the line are deducted form data obtained on Giant Quid (POTTOLA 1973). If we describe the dendrit of the fiber as a cylindric conductor we can infer the relation between the constants of the line and the diameter of the dendrit. The comparison of the values of the diameter for man and for Giant squid allow us to determine the constants for man.

The study of impulse response lead us to the following results: A 0,2 mm long radial fiber (connected to one inner hair cell) shows a 12 µs delay which corresponds to a velocity of 16,6 m/s. A 1 mm long dendrit (one of these connected to the spiral fibers) shows a 0,3 ms delay which corresponds to a velocity of 3 m/s.

The estimated mean propagation velocities on fibers are in rather good agreement with data found in specialized litterature. Delay, almost negligible for a radial fiber, becomes an important parameter for a spiral fiber.

So, in this model, each fiber plays the role of a low pass filter, with a cut-off frequency which becomes lower as the fiber is longer. With low-frequency signals, radial and spiral fibers convey instantaneous values while only mean values are transmitted for high frequencies. The frequency bondary between these two kinds of behavior already proposed by DOLMAZON (1980) is closed 2 KHz.

These phenomena must play an important part in the mechanisms of coding of the auditive message.

RLFERENCES :

. J.M. DOLMAZON, L. BASTET (1979)
Hair cell receptor modeling
J. Acoust. Soc. Am., 65, Supp N°1, S14

. J.M. DOLMAZON (1980)
Modèle de fonctionnement du système auditif
State thesis, I.N.P.G., GRENOBLE (FRANCE)

. E.W. POTTALA, T.R. COLBURN, D.R. HUMPHREY (1973)
A dendritic compartment model neuron
IEEE Trans., BME 20, N°2, 132-139

Détection du voisement par une technique de classification automatique

(Robert ESPESSER, Institut de phonétique, Université de Provence,LA 261 CNRS, 13621 AIX EN PROVENCE, FRANCE)

Plusieurs auteurs ont montré l'interêt des techniques de reconnaissance/classification de formes pour la détection du voisement (B.S. ATAL & L.R. RABINER, I.E.E.E. Trans. ASSP-24, 201-212 (1976)) (L.J. SIEGEL, I.E.E.E. Trans. ASSP-27, 83-89 (1979)). Cependant, les methodes employées nécessitaient un apprentissage préalable, ou la connaissance d'informations a-priori. Nous présentons une méthode de détection du voisement évitant ces problèmes en utilisant une technique de classification automatique.

L'énergie RMS, le taux de passage par zéro (TPZ) et l'erreur de prédiction normalisée (erreur LPC) sont calculés toutes les 10 ms sur le segment de parole, échantillonné à 10 kHz et filtré passe-bande à 80-5000 Hz (le coefficient d'autocorrélation à un échantillon de décalage n'a pas été retenu, car il est en première approximation équivalent au TPZ; une explication en est proposée). Chacun de ces triplets est considéré comme un vecteur d'un espace à 3 dimensions. Cet ensemble de vecteurs est alors classé en deux groupes par une procédure de réallocation itérative relevant de la "méthode des nuées dynamiques" (DIDAY & coll. I.N.R.I.A. (1979)); nous employons la méthode des "centres mobiles" (E.W. FORGY, Biometrics 21, 768-769 (1965)) combinée avec l'emploi de la distance de Mahalanobis pour l'affectation d'un vecteur à l'une des deux classes (distances adaptatives, DIDAY op. cité). Cette métrique, qui nécessite le calcul de la matrice de variance-covariance de chaque classe, est donc redéfinie à chaque itération de l'algorithme; à l'initialisation de l'algorithme, on utilise la distance euclidienne. On impose aux centres initiaux des classes, choisis aléatoirement, de satisfaire la contrainte "acoustique" suivante: le centre voisé a une énergie supérieure à la moyenne, un TPZ inférieur à la moyenne, une erreur LPC inférieure à la moyenne (moyennes relatives à chacune des 3 dimensions, et calculées sur l'ensemble des vecteurs); le centre non-voisé doit satisfaire les conditions opposées. Cette contrainte doit être aussi satisfaite par les centres de gravité des classes finales. Elle a pour effet d'éviter certains classements "acoustiquement" aberrants, et accélère d'autre part la convergence de l'algorithme. Plusieurs classements sont effectués sur un même segment de parole, et le meilleur au sens d'un critère défini est retenu.

Le corpus de test était constitué de deux phrases d'environ 3.3 s chacune, prononcées par 5 femmes et 5 hommes; ces phrases comportent toutes les consonnes bruissantes du français. Pour chacun de ces 20 segments de parole, l'analyse a été faite sur des durées de 250, 500, 1000, 1500, 2000 ms et sur le segment entier. Avec un rapport voisé/silence (S/B) de 32 dB, l'erreur moyenne est de 3.6 %, et de 5.5 % pour un rapport S/B de 20 dB (obtenu par adjonction de bruit blanc à l'enregistrement original), et ce pour des durées supérieures à 1 s environ. La réduction de la durée analysée augmente peu l'erreur moyenne (+1 % environ) et influe surtout sur sa dispersion.

Voiced/unvoiced decision by a clustering procedure

(Robert ESPESSER, Institut de phonétique, LA 261 C.N.R.S. ,Université de Provence, 13621 AIX EN PROVENCE, FRANCE)

Several authors have demonstrated the interest of pattern recognition/ classification approach for the V/UV decision (B.S. ATAL & L.R. RABINER, I.E.E.E. Trans. ASSP-24, 201-212 (1976)) (L.J. SIEGEL, I.E.E.E. Trans. ASSP-27 , 83-89 (1979)). Nevertheless, the methods which were used need either a training set of data or a-priori-information. A method avoiding these problems is presented, using a clustering procedure.

Energy RMS, zero-crossing rate (TPZ) and LPC normalized error are computed every 10 ms over the speech signal, which is band-pass filtered (80-5000 Hz) before sampling at 10 kHz (the normalized autocorrelation coefficient at unit-sample delay has not been retained, because it is, in first approximation, equivalent to the TPZ; an explanation is proposed). Each of these triplets is considered as a vector in a 3-dimensionnal space. This set of vectors is then classified in two groups by an iteration reallocation procedure clustering , according to the "dynamic clusters method" (DIDAY & coll. , I.N.R.I.A. (1979)); the method of FORGY (Biometrics 21, 768-769 (1965)) is employed, and Mahalanobis distance is used for allocating a vector to a class (adaptative distances, DIDAY, op. cit.). This distance, which requires the calculus of the variance-covariance matrix of each class, is redefined at each iteration of the algorithm. At the initialization of the procedure, Euclidean distance is used. The following "acoustical" constraint is imposed on the randomly choosen initial centers of the classes: the voiced-center has an energy greater than the mean energy; a TPZ smaller than the mean TPZ; and an LPC error smaller than the mean LPC error (these averages are computed over the set of vectors, for each dimension respectively); the unvoiced/silence center must satisfied the opposite conditions. This constraint must be satisfied also by the centroids of the final classes. This procedure avoids very wrong classifications on the acoustical plan, and speeds up the convergence of the algorithm. Several partitions are computed for a speech segment and the best one according to some defined criteria is retained.

The test material consisted of two sentences, both of them being about 3.3 sec. long; these sentences contain all unvoiced/voiced plosives and fricatives in French. Five females and five males were used as speakers. For each of these 20 sentences, analysis was performed for a 250, 500, 1000, 1500, 2000 ms duration and for the duration of the whole segment. For a voiced-to-silence ratio (S/N) of 32 dB, mean error rate is 3.6 %; it is 5.5 % for a S/N of 20 dB (ratio obtained by adding white noise on the original record). These error rates are obtained for segments longer than 1 sec.; for brief durations, the mean error rate is slighty increased(+1 %), but this shortening has a stronger effect on its dispersion.

EXPERIENCES EN RECONNAISSANCE DE PAROLE CONTINUE

M.T. GIORGETTI - M. LAMOTTE - M.J. VIGNERON
Laboratoire d'Electricité et d'Automatique
Université de NANCY I
C.O. 140 - 54037 NANCY Cédex (FRANCE)

Le but de ces travaux est la reconnaissance, en temps différé, de courtes phrases obéissant à une syntaxe simplifiée (vingt et une règles syntaxiques) et utilisant un vocabulaire réduit (cinquante mots).

Dans le système réalisé au Laboratoire d'Electricité et d'Automatique, trois niveaux de reconnaissance ont été considérés : acoustique, lexical et syntaxique. Le niveau sémantique n'a pas été abordé.

Le traitement acoustico-phonétique conduit à une représentation phonémique de la phrase à deux dimensions : sous forme d'une matrice à colonnes de dimensions variables regroupant horizontalement les segments reconnus et verticalement ceux qui leur sont le plus semblables. Ces segments sont numérotés.

On repère, dans la chaîne, des "séquences caractéristiques de segments" qui recouvrent une ou plusieurs syllabes de certains mots du vocabulaire. Lorsque la meilleure des séquences caractéristiques est détectée, elle constitue un point d'ancrage. De là, le processus de recherche lexicale permet d'identifier le mot correspondant. Puis, par une méthode "ascendante-descendante", on recherche, de proche en proche, les voisins de droite et de gauche, en consultant les données propres au langage utilisé (règles de réécriture, ...).

A chaque noeud de l'arborescence, il existe un certain nombre n de mots candidats. Pour la mise en oeuvre du processus, on utilise un système de "blocs". Chaque bloc contient toutes les informations relatives aux résultats antérieurs et celles nécessaires à l'étape suivante :

- liste des mots candidats éventuellement classés ;
- un pointeur pour le tableau des solutions ;
- des indicateurs repérant la solution en cours par son numéro et son adresse ;
- des indicateurs de branchement à des sous-programmes traitant les désinences verbales, les pluriels, le cas des mots courts (articles, etc...).

A chaque nouvelle solution envisagée, on crée un bloc. Si le mot correspondant est retenu, le bloc suivant est engendré et le processus se poursuit. Sinon, on efface le bloc précédent et on revient en arrière. On génère ainsi un empilement de blocs, analogue à une pile LIFO, et à tout instant les solutions partielles sont évaluées pour développer les plus prometteuses.

EXPERIMENTS IN CONTINUOUS SPEECH RECOGNITION

M.T. GIORGETTI - M. LAMOTTE - M.J. VIGNERON
Laboratoire d'Electricité et d'Automatique
Université de NANCY I
C.O. 140 - 54037 NANCY Cédex (FRANCE)

The aim of this work is to find a time-delayed procedure for recognition of short sentences. These sentences are restricted to those which obey a simplified syntax and use quite a reduced lexicon (21 syntactic rules and 50 words).

In the present procedure, recognition has been carried out at three levels : acoustical, lexical and syntactic. The semantic aspect has not been considered so far.

The acoustical-phonetic analyse leads to a two dimensional phonemic network which represents each sentence. In the first line of this network are gathered the recognized segments and each column contains the more similar segments. Segments are numbered.

Then, one try first to find typical sequences of segments which can be easily recognized in some syllables of certain words. The best detected sequence will be used as an anchor point in the following. Thus the corresponding word is identified and using a top-down procedure, left side and right side words are recognized through the characteristic rules of the language (rewrighting rules, ...). Each crossing-point of a tree-like pattern corresponds to a number n of candidate words.

Every stage of the procedure results in an "information block" containing all information obtained so far during the whole set of previous stages along with all the necessary information to proceed :

- list of candidate words (classified or no) ;
- a pointor to explore the solution table ;
- indicators to signal the in process solution through its numero and its address ;
- branching flags to subroutines dealing with verbal desinences, plurals, short words (as articles, ...).

In a try and failure procedure, a block pile is generated like a LIFO pile. Each potential solution is evaluated and the more promising ones are developped.

TITRE DE LA COMMUNICATION PRESENTEE

Consultation d'un annuaire téléphonique au moyen du poste de l'abonné.

Nom de l'auteur — Centre de Recherche
HUTIN Michel — CNET-LANNION A

INTRODUCTION

Les études en renseignement téléphonique se poursuivent au CNET suivant deux directions en vue de permettre la consultation directe par l'abonné : l'une en utilisant un terminal spécifique, c'est l'approche visuelle, l'autre en utilisant le poste téléphonique de l'abonné, c'est une approche vocale (voix enregistrée et voix synthétique).

C'est cette seconde approche qui fait l'objet de cette présentation.

ANALYSE DU PROBLEME

Vu la pauvreté des possibilités du poste téléphonique en entrée d'information, il est probable que le service rendu ne puisse répondre à la totalité des demandes. De plus, si on souhaite fournir des éléments de concordance entre la demande et la réponse, c'est-à-dire ne pas se contenter d'énumérer un ou plusieurs numéros, il est nécessaire de disposer d'une synthèse de la parole par diphonèmes pour formuler le (ou les) nom(s) de l'abonné trouvé et ses attributs. La chaine de traduction orthographique plus synthèse vocale apportera également une dégradation par rapport à une voix naturelle du fait des particularités de prononciation régionale, du timbre et de l'élocution d'une voix synthétique.

RESULTATS ESCOMPTES

Nous connaissons a priori quelques unes des difficultés qui risquent de se présenter à l'usager d'un tel service. Notre objectif est d'en cerner les causes plus précisément pour essayer d'y remédier et aussi d'évaluer le service rendu : nombre et type de demandes satisfaites pa rapport à l'ensemble des demandes formulées, enfin de cerner la population intéressée par un tel service.

C'est pourquoi, nous allons faire une évaluation mi-1981 sur cette maquette.

REFERENCES

DAII/TAI/98 : Le futur réseau du renseignement
NT/DAS/2 : Système de renseignement téléphonique S4. Description du projet.
NT/LAA/TSS/41 : Spécification de définition de la maquette S4 sur MINARET
NT/DAS/ETA/53 : Résultats d'un test sur l'intelligibilité des lignes d'annuaire en parole de synthèse.

LAHLOU Moncef, Institut de Phonétique
Aix-En-Provence.

Caractéristiques intrinsèques des voyelles : Application à une langue arabe : le marocain.

Les langues à substrat arabe, dont le marocain, sont caractérisées par leur richesse consonantique : 28 consonnes, et par un système vocalique réduit aux trois voyelles de base : /a/,/u/,/i/. Mais paradoxalement, ce sont les voyelles qui posent le plus de problèmes dans l'analyse, vu leur timbre instable.

Nous avons étudié les caractéristiques intrinsèques de ces trois voy. dans un corpus relativement étendu de mots en CVCVCV, avec toutes les combinaisons possibles des voy.: $3^3=27$, et dans quatre contextes consonantiques différents en faisant varier le facteur mode phonatoire (voisé/non voisé), et le mode articulatoire (constrictives/occlusives).

Résultats:

Fo : Calculée sur la moyenne de chaque voy.dans les syllabes d'attaque. /i/et /u/ ont une Fo supérieure à celle de /a/, mais l'écart VH/VB atteit à peine 4%. On peut penser que ce faible écart est dû à l'absence d'autres voy. dans la langue, qui fait que /a/ se réalise souvent [æ] ou [ɛ], que /i/ tend vers [e], et que /u/ se prononce parfois [o] voire [ɔ]. Les trois voy. se réalisent en fait très rarement comme voy. cardinales, leur timbre dépendant beaucoup de leur entourage consonantique.

Intensité:/a/ a en moyenne une intensité supérieure de 4 dB à celle de /u/ ou de /i/ (différence significative à l'analyse de variance, $p<0.01$). L'intensité de /u/ est à son tour supérieure à celle de /i/, mais cette différence n'est pas significative.

Durée: /a/ se caractérise en moyenne par une durée supérieure à celle de /u/ et de /i/. Cette différence peut atteindre 20% en position accentuée (/a/=120 ms, /u/=101 ms, /i/=98 ms). Ce pourcentage correspond à celui défini par Rossi(1) pour le seuil différentiel de durée.

Conclusion: Ces résultats devraient nous permettre, après les travaux complémentaires actuellement en cours, de proposer, sur le modèle de A. Di Cristo, D. Hirst, et Y. Nishinuma(2), une méthode d'évaluation des caractéristiques intrinsèques des voyelles marocaines à partir de leurs valeurs formantiques.

---

(1) : cf M.Rossi, Le seuil différentiel de durée
in Papers in linguistics and phonetics to the memory of P. Delattre
ed. A.Valdman, Mouton, 1972, pp 435-450

(2) cf A. Di Cristo, D. Hirst, Y. Nishinuma,
L'estimation de la Fo intrinséque des voyelles : étude comparative.
Travaux de l'institut de phonétique d'Aix-En-Provence
Volume 6, 1979, pp 147-176.

LAHLOU Moncef
Institut de phonétique,
Aix-En-Provence.

Intrinsic characteristics of vowels in Moroccan Arabic.

Arabic languages, such as Moroccan, ara characterized by a rich consonant system : 28 consonants, and a reduced vowel system with th three basic vowels : /a/,/u/, /i/; Paradoxically, it is the vowels which present most problems in the analysis, due to their unstable quality.
We studied the intrinsic characteristics of these three vowels in a relatively large corpus of words with the pattern CVCVCV; includingeach of the 27 (3^3) combinations of vowels, in four consonantal contexts (voiced vs unvoiced, stops vs fricatives).

Results :

Fo: Taking the mean value for each vowel in the initial syllable gives a higher Fo for /i/ and /u/ than for /a/, but the difference is barely 4%, probably owing to the absence of other vowels in the language. This has, as a result that /a/ is often pronounced [æ] or even [ɛ], /i/ tends towards [e], and /u/ tends to be [o] or [ɔ], depending on the consonantal contexts; the pronunciation as cardinal vowels is, in fact, quite rare.

Intensity : The mean intensity for /a/ is 4 dB higher than that of /u/ or /i/(the analysis of variance shows a significant difference, $P < 0.01$). The mean intensity of /u/ is also higher than that of /i/, but the difference is not significant in this case.

Duration : The mean duration of /a/ is greater than that of /u/ and /i/. This difference reaches 20 % in stressed position (/a/= 120 ms, /u/= 101 ms, /i/= 98ms). This difference is comparable to that found by Rossi for the differential threshold for duration(1). The difference between /u/ and /i/ is not significant.

Conclusion :

These results should allow us, after finishing complementary research we are now doing, to propose, following the model of A. Di Cristo, D.J. Hirst, and Y. Nishinuma, a method of evaluating intrinsic characteristics of Moroccan vowels once their formant values precisely determined.(2)

---

(1) cf Rossi,
le seuil differentiel de durée.
in Papers in linguistics and phonetics to the memory of P. Delattre
ed by A. Valdman, Mouton, 1972, pp 435-450

(2) cf A. Di Cristo, D.J. Hirst, and Y. Nishinuma,
L'estimation de la Fo intrinsèque des voyelles : étude comparative.
Travaux de l'institut de phonétique d'Aix-En-Provence volume, 6 , 1979, pp 147-176.

Analyse de la nasalité vocalique

A. Landercy - G. Hattiez.
Université de Mons

Les caractéristiques acoustiques de la nasalité vocalique sont difficilement mises en évidence car elles dépendent à la fois du locuteur, de la voyelle orale de référence et du degré de couplage entre les conduits oral et nasal. Aussi, avons-nous tenté une étude perceptive instrumentale, à savoir l'analyse contrastive des spectres de sonie des voyelles orales et nasales.

L'étude a porté sur l'analyse de 960 spectres de sonie (10 sujets masculins francophones prononçant 48 logatomes du type CVCV dans lesquels étaient combinés les 4 couples de voyelles orales et nasales françaises (ɔɑœɛ/ɔ̃ɑ̃œ̃ɛ̃) avec les 6 consonnes occlusives orales et nasales (bdg mnɲ).

Nous avons tout d'abord calculé, pour l'ensemble des réalisations de chaque voyelle, la sonie moyenne par bande. Ensuite, afin d'essayer d'éliminer les différences dues au locuteur ou à l'entourage consonantique, nous avons comparé chaque opposition voyelle orale - voyelle nasale et traduit les différences par bark en 9 classes d'intervalle égal à 0,5 sone.

L'examen de ces résultats montre que :
- pour les voyelles /ɔ/ et /ɑ/, postérieures, sombres, dont F_1 et F_2, ainsi que F_3 et F_4, sont suffisamment proches, la nasalisation produit une sorte de fusion entre ces formants vocaliques de sorte que les spectres de sonie de /ɔ̃/ et /ɑ̃/ se caractérisent par deux zones formantiques, l'une précisément située entre les zones des deux premiers formants vocaliques et l'autre étalée entre les zones des troisième et quatrième rendant ainsi les spectres de /ɔ̃/ et /ɑ̃/ nettement plus diffus que ceux de /ɔ/ et /ɑ/;

- pour les voyelles /œ/ et /ɛ/, antérieures, plus claires, dont F_1 et F_2 sont plus écartés, la nasalisation ne permet pas la fusion de ces deux formants mais les rapproche considérablement l'un de l'autre, rendant ainsi les spectres de /œ̃/ et /ɛ̃/ beaucoup plus compacts que ceux de /œ/ et /ɛ/.

L'étude des spectres de sonie permet de mettre en évidence les caractéristiques acoustiques les plus pertinentes pour l'audition. Elle permet notamment d'éliminer l'influence des antirésonances qui sont masquées par les résonances précédentes. De manière générale, cette étude ne fait pas apparaître des constantes de la nasalité (formant nasal fixe aux environs de 50 Hz ou de 1000 Hz) mais des différenciations au sein de chaque voyelle. En résumé, nous constatons que, d'une manière générale, les voyelles nasales présentent deux zones de résonance importantes situées
- toutes deux entre F_1 et F_2 des orales correspondantes, pour les voyelles /œ̃/ et /ɛ̃/;
- l'une entre F_1 et F_2 et l'autre entre F_3 et F_4 des orales correspondantes pour /ɔ̃/ et /ɑ̃/.

Bibliographie sommaire

FANT, G., Acoustic Theory of Speech Production, The Hague, Mouton, 2ème éd., 1970 (1ère éd., 1960).

LANDERCY, A., RENARD, R., Eléments de Phonétique, Didier, Bruxelles, 1977.

MRAYATI, M., CARRE, R., Acoustic Aspects of French Nasal Vowels, 89th Meeting of ASA, Austin Texas, 1975.

Analysis of nasal vowels

The acoustical cues of nasal vowels have already been described and it has been demonstrated that these cues depend on the speaker, on the oral vowel of reference and on the degree of nasal coupling. We have tried an instrumental perceptive study, based on the contrastive analysis of loudness spectrums of the oral and nasal French vowels.

The study is based on the analysis of 960 loudness spectrums (10 male French speakers pronouncing 48 CVCV utterances in which were combined the four pairs of oral and nasal vowels (ɔ ɑ œ ɛ ɔ̃ ɑ̃ œ̃ ɛ̃) with the 6 oral and nasal occlusive consonants (b d g m n ɲ).

For all the realisations of each vowel, we calculated first the average loudness in each band. After that, with a view to eliminating the individual performances and also the influence of consonantal environment, we compared each nasal vowel with the corresponding oral vowel for the same speaker and the same environment and we distributed the differences for each bark in 9 classes of 0,5 sone.

The results show :

a) for the back vowels /ɔ/ and /ɑ/, with F_1 near F_2, and F_3 near F_4, the nasalisation results in a kind of fusion of the vocal formants F_1F_2 and F_3F_4. Thus the loudness spectrum for /ɔ̃/ and /ɑ̃/ are caracterized by two formant bands : the first is situated between the oral F_1 and F_2 and the second between the oral F_3 and F_4;

b) for the front vowels /œ/ and /ɛ/, whose F_1 and F_2 stand further apart the nasalisation doesn't result in the actual fusion : it only brings the formants closer to each other.

The study of the loudness spectrum shows the most pertinent acoustic cues for auditory perception. It makes it possible to eliminate the influence of zeros in the spectrum, which are masked by the preceding poles. Our study doesn't bring out any nasalisation constants (nasal formant at 500 Hz or 1000 Hz) but only spectrum differences in each vowel. Briefly, we can see that nasal vowels present two important frequency resonance bands situated

a) between oral F_1 and F_2 for /œ̃/ and /ɛ̃/;

b) one between oral F_1 and F_2 and the other between oral F_3 and F_4 for /ɔ̃/ and /ɑ̃/.

EVOLUTIONS RECENTES DU SYSTEME DE COMPREHENSION DE LA PAROLE CONTINUE ESOPE

J. MARIANI - LIMSI/CNRS - BP 30 - 91406 ORSAY CEDEX

Le système ESOPEØ, modèle probatoire réalisé en 1977 (1), utilise une stratégie Prédiction-Vérification (ou descendante), du niveau pragmatique au niveau phonétique, avec conservation des meilleures solutions intermédiaires (best-few), sans retour arrière, et analyse de la gauche vers la droite. Les niveaux pragmatique , syntaxique , lexical et phonétique prédisent des hypothèses en nombre limité. Les phonèmes prédits sont alors comparés aux 4 meilleurs candidats du treillis phonétique reconnu en tenant compte de 4 éventualités : présence du phonème, confusion (ressemblance donnée par une matrice de confusion), ajout, élision (possibilité décrite par une matrice d'élision) . Pour chaque hypothèse, seule l'éventualité donnant la meilleure note est retenue. Les hypothèses sont rangées en 3 classes de solutions intermédiaires, avec élimination de celles dont la note diffère trop de la solution intermédiaire ayant la meilleure note .

Le système ESOPE 1 (2), utilise la même stratégie de façon plus systématique. A chaque instant, sont conservées les meilleures solutions intermédiaires (limitées en nombre $\leqslant$N), classées suivant leur note . On prédit alors les phonèmes correspondants aux 3 éventualités (bonne segmentation, doublement, élision), c'est à dire,pour chaque solution intermédiaire,les deux phonèmes suivants. Ces hypothèses sont également limitées en nombre ($\leqslant$M), et sont faites en suivant l'ordre de classement des solutions intermédiaires. A chaque hypothèse est affectée la note de reconnaissance du phonème prédit par comparaison avec les variétés phonétiques reconnues pour le segment considéré. On classe ensuite les hypothèses pour ne conserver que les N meilleures solutions intermédiaires.

ESOPE 1-1 effectue la prédiction jusqu'au niveau acoustique en utilisant un dictionnaire de diphonèmes. Pour chaque phonème prédit, le système examine 3 éventualités (ajout, progression normale, élision), suivant un algorithme de comparaison dynamique. Les hypothèses ainsi effectuées sont classées et les meilleures solutions intermédiaires sont retenues. Parallèlement, nous étudions l'apprentissage automatique d'un modèle centiseconde pour diminuer la taille du dictionnaire de diphonèmes et faciliter l'adaptation au locuteur .

Dans ESOPE 2, la stratégie est à la fois montante et descendante, suivant une méthode de Prédiction-Vérification-Induction. Le système effectue une prédiction à chaque niveau, mais peut également induire des phonèmes, puis des mots. Ce qui lui permet un apprentissage limité soit de combinaisons syntaxiquement imprévues, soit d'éléments lexicaux syntaxiquement prévus. Ce système est en cours de réalisation.

Enfin, dans nos essais d'orthographication d'une chaîne phonétique exacte (3), avec une syntaxe de type langue naturelle, et un vocabulaire de 170 000 formes, nous avons utilisé une stratégie montante, avec conservation des meilleures solutions en parallèle, la décision du choix en cas d'ambiguïtés ($<$5 %), étant reportée à une analyse sémantique ultérieure.

Il nous semble donc que lorsque le langage est contraint(syntaxe très rigide, lexique limité), et que la reconnaissance phonétique est médiocre, une strtégie descendante soit à prescrire. Par contre, plus le langage est étendu, plus la reconnaissance phonétique est bonne, plus une stratégie montante est nécessaire et nous l'expérimentons dans la réalisation d'ESOPE 2, en tenant compte des récents résultats recueillis en Psycholinguistique.

-(1)- MARIANI J.,LIENARD J.S. , ESOPE Ø : un programme de compréhension automatique de la parole procédant par prédiction-vérification aux niveaux phonétique,lexicale et syntaxique. Congrès AFCET-IRIA Reconnaissance des Formes. 1978
-(2)- MARIANI J.,LIENARD J.S. , Eléments linguistiques et cognitifs dans un système de communication vocale homme-machine. Séminaire "Syntaxe et Sémantique en reconnaissance de la parole" GALF-AFCET 1980
-(3)- ANDREEVSKY A. & Col. ,Les dictionnaires en forme complète...10émes JEP

RECENT DEVELOPMENTS IN THE ESOPE CONTINUOUS SPEECH UNDERSTANDING SYSTEM

J. MARIANI - LIMSI/CNRS - BP 30 - 91406 ORSAY CEDEX

The ESOPE Ø system, a trial model designed in 1977 (1), uses a top-down strategy from the pragmatic level down to the phonetic one, using a best few exploration strategy with no-backtracking and left-to-right analysis. The pragmatic, syntactic, lexical and phonetic levels predict a limited number of hypotheses. The predicted phonemes are the compared to the four-best candidates recognised from the phoneme lattice, taking four possibilities into account : presence of the phoneme, confusion (ressemblance given by a confusion matrix), addition, elision (possibility described by an elision matrix). For each hypothesis, only the possibility with the best score is kept. The hypotheses are arranged in 3 intermediary solution classes, with elimination of those having a score which is too different from that of the intermediary solution having the best score.

The ESOPE 1 (2) system uses the same strategy in a more systematic manner. At each instance, the best intermediary solutions (limited to be $\leqslant N$), are ordered by their respective scores. We then predict the phonemes corresponding to the 3 possibilities (good segmentation, counting one segment as two, elision), that is, for each intermediary solution, the two following phonemes. These hypotheses are also limited in number ($\leqslant M$) and are generated according to the classification order of the intermediary solutions. For each hypothesis, a score for the recognition of predicted phonemes is established by comparison with the phonetic varieties that have been recognised for the given segment. We then order the hypotheses so as to only keep the N best ones.

ESOPE 1-1 carries out the prediction down to the acoustic level using a diphone dictionary. For each predicted phoneme, the system examines 3 possibilities (addition, normal progression, elision)on the basis of a dynamic comparison algorithm. The hypotheses thus formed are ordered and the best intermediary solutions are retained. On a parallel to this, we are studying automatic training passes for a centisecond model so that the size of the diphone dictionary may be reduced and so that the adaptation to the speaker may be easier.

In ESOPE 2, the strategy is bottom-up and Top-down, according to a Prediction-Verification-Induction method. The system makes a prediction at each level, but may also introduce phonemes and, further up , words. This permits itto have a limited amount of training in either the number of combinations of words syntactically unpredicted, or the number of lexical elements syntactically hypothesized. This system is presently being designed.

Finally, in our trials of the writing out of an exact phonetic string (3), with a syntax of a natural language nature, and a vocabulary of 170 000 forms, we have used a bottom-up strategy with parallel, best-few treatment, in the case of ambiguity ($<5\%$) leaving the decision to a later semantic analysis.

It therefore seems to us that when the language is constrained (very rigid syntax, limited vocabulary), and that phonetic recognition is mediocre, a top-down strategy is necessary. On the contrary, the more the language is generalised, and the better the phonetic recognition, the more a bottom-up strategy is called for. Nevertheless, a mixture of the two types of strategies is necessary, and we are trying this possibility out in ESOPE 2, taking into account recent results found in psycholiguistics.

-(1)- MARIANI J.,LIENARD J.S. , ESOPE Ø : un programme de compréhension automatique de la parole procédant par prédiction-vérification aux niveaux phonétique,lexicale et syntaxique. Congrès AFCET-IRIA Reconnaissance des Formes. 1978
-(2)- MARIANI J.,LIENARD J.S. , Eléments linguistiques et cognitifs dans un système de communication vocale homme-machine. Séminaire "Syntaxe et Sémantique en reconnaissance de la parole" GALF-AFCET 1980
-(3)- ANDREEVSKY A. & Col. ,Les dictionnaires en forme complète...10émes JEP

12ièmes Journées d'Etudes sur la Parole. MONTREAL 25-27 Mai 1981

MARCHAL A.*, FLOUTIER D.**, et COURVILLE L.*
*Université de MONTREAL et **Université de MONTPELLIER

"EXAMEN ELECTROPALATOGRAPHIQUE DE L'ENCHAINEMENT DES OCCLUSIVES EN FRANCAIS"

Nous présentons quelques résultats préliminaires obtenus à l'aide de l'électropalatographe de Montréal.

Ce système assisté par un microprocesseur est présenté plus en détail dans la communication conjointe (COURVILLE, FLOUTIER, MARCHAL, XIIè J.E.P. I98I).

Nous avons examiné la question de la force d'articulation des occlusives du français et parvenons aux conclusions suivantes:

I- Toutes choses par ailleurs égales, les consonnes les plus fortes sont les sourdes, suivies des sonores; les plus faibles sont les nasales.

2- L'accent joue un rôle important quant à la force d'articulation. Il a plus d'importance que le mode et détermine le nombre de contacts de la langue au palais.

3- Dans le cas des assimilations de voisement ou d'assourdissement, il était généralement admis que le changement pour les occlusives françaises se limitait à la dimension de [voix] et que le trait [$\pm$ Fortis] demeurait inchangé. Nous n'avons pas trouvé de telle distinction dans nos données et les occlusives assourdies présentent les mêmes patrons de contact que les sourdes.

4- Dans le cas de l'enchaînement de 2 occlusives, le degré d'élévation de la langue au palais est plus grand lorsqu'une consonne antérieure précède une consonne postérieure. Dans la majorité des cas, il y a une double occlusion.

5- Lorsqu'une consonne postérieure est suivie par une consonne antérieure, nous observons dans tous les cas le relâchement de l'occlusion de la première consonne avant que ne s'établisse le barrage de la consonne suivante.

6- Pour les occlusives sonores, le relâchement de la première occlusion se produit toujours avant la deuxième occlusion, quelque soit la direction de l'enchaînement articulatoire.

12ièmes Journées d'Etudes sur la Parole. MONTREAL 25-27 Mai 1981

MARCHAL A.*, FLOUTIER D.**, et COURVILLE L.*
*Université de MONTREAL et **Université de MONTPELLIER

"ELECTROPALATOGRAPHIC INVESTIGATION OF COARTICULATION IN FRENCH STOPS".

In this paper, we present and discuss electropalatographic data obtained by the micro-computer assisted system developped at the University of MONTREAL. We examined the coarticulation in french stops and the force of articulation.

1- All things being equal, the strongest consonants are the voiceless, then the voiced and the weakest are the nasals.

2- The accent plays a significant role. It has a greater influence on the force of articulation than the phonemic status. It determines the number of tongue-palate contacts.

3- In cases of voicing or devoicing assimilations, it was commonly believed that the change for the french stops concerned only the voicing dimension and that the fortis-lenis feature remained unchanged. We found no such a distinction in our data where devoiced stops show the same contact pattern than voiceless.

4- When two voiceless stops follow each other, the degree of height of the tongue is larger when a front consonant comes first. In most cases, there is a double occlusion.

5- When a back stops is following by a front one, we observe, in all cases, the release of the closure of the first consonant prior to the second closure.

6- For voiced stops, the release of the first closure occurs always before the second consonant independantly of the direction of coarticulation.

SESSION AFFICHEE

H.MELONI, J.GUIZOL G.I.A. FACULTE DE LUMINY MARSEILLE

Nous illustrons, au moyen d'exemples, les phases de notre système de reconnaissance automatique de parole continue.

1 - ACQUISITION DE LA PAROLE

Les phrases sont enregistrées, dans un environnement peu bruité, sur un magnétophone UHER 4200 à l'aide d'un microphone amplifiant d'environ 3 dB les fréquences supérieures à 2000 Hz. Le signal, filtré entre 80 et 4800 Hz, digitalisé à 10 KHz et codé sur 12 bits est stokcé dans la mémoire périphérique (disque magnétique) d'un mini-ordinateur SOLAR 16-65.

2 - TRAITEMENT DU SIGNAL - PARAMETRES

Certains paramètres sont obtenus directement à partir du signal (énergie,densité des passages par zéro, rapport R1/R0), mais la plupart sont issus de spectres calculés toutes les 10 mS à travers une fenêtre de 25 mS, à l'aide d'une transformée rapide de Fourier sur les 14 premiers coefficients de LPC (méthode d'autocorrélation). Les principaux paramètres mesurent les caractéristiques suivantes : répartition spectrale de l'énergie (basse fréquence, moyenne, maximale dans une fenetre de 800 Hz), maxima du spectre (fréquence, amplitude, largeur de bande à -3 dB), émergence de F1, instabilité spectrale (spectres situés à 20 mS et 40 mS), fréquence fondamentale, erreur normalisée.

3 - SEGMENTATION ET MACRO-CLASSES PSEUDO-PHONETIQUES

La segmentation est opérée à partir des fonctions d'instabilité spectrale auxquelles s'ajoutent, pour les zones de variations "lentes", les effets amplificateurs des modifications des valeurs de F1 et de la densité des passages par zéro.

Les macro-classes caractérisent le plus fréquemment un phonème et ses transitions vers ceux qui lui sont immédiatement contigus. Cependant ces unités limitent dans certains cas des zones correspondant à une portion de phonème, une transition entre deux phonèmes, une séquence de plusieurs phonèmes acoustiquement proches. Les étiquettes retenues (VO, OC, FR et CC, CO, SI, VC) désignent respectivement les voyelles, les consonnes occlusives, les consonnes constrictives, les consonnes vocaliques, les silences et des segments de nature indéterminée.

4 - IDENTIFICATION DES TRAJECTOIRES FORMANTIQUES

A partir des zones de maximum de stabilité des segments "VO" les trajectoires possibles des formants dans les cadres voisées sont évaluées à l'aide des pics du spectre. Lorsque plusieurs solutions subsistent, elles sont toutes conservées.

5 - MACRO-TRAITS PSEUDO-PHONETIQUES

Pour chacune des macro-classes un ensemble de traits spécifiques lui est associé. Ils sont quelquefois assortis d'un degré traduisant une hiérarchie (trait ouvert/fermé), mais le plus souvent ils prennent l'une des trois valeurs suivantes : positif, négatif, indéterminé.

6 - NOYAUX ACCENTUES ET SHEMA PROSODIQUE GENERAL

La durée normalisée (dans la phrase et en fonction de la valeur de F1) des segments "VO" ainsi que les variations de F0 permettent de particulariser certains d'entre eux en leur associant l'un des labels suivants : FNTMA, FNTMI, FT, AL, MB. Ils désignent des unités qui marquent des frontières prosodiques (non terminales majeure ou mineure, terminale) et des éléments mis en relief par un allongement ou un maximum de F0.

Les valeurs de F0 à l'attaque et à la finale d'une phrase permettent de lui affecter le caractère interrogatif, affirmatif ou indéterminé.

7 - REGLES D'INTERPRETATION DES EVENEMENTS

Ces règles permettent d'associer à une chaine de phonèmes l'ensemble des séquences d'évènements pseudo-phonétiques susceptibles de la décrire. Les productions de ces règles sont liées dans la plupart des cas au contexte phonémique immédiat de la chaine à dériver.

8 - LEXIQUE

La description lexicale des mots comporte des informations syntactico-sémantiques (genre, nombre, traits, nature et particularités des compléments, etc...) et les représentations phonémique et graphique ainsi que leurs variantes contextuelles.

9 - ANALYSE SYNTAXIQUE

La grammaire d'un sous ensemble du franÇais comportant les structures fondamentales qui permettent la description et l'interrogation d'une banque de données est décrite dans un systeme qui autorise également la programmation de stratégies de conduite de l'analyse en fonction du contexte (gauche-droite, droite-gauche, à partir de points d'ancrages).

Les stratégies sont déterminées par des informations "descendantes" proposées par les modules de traitement de la sémantique et de la pragmatique ainsi que par des informations "remontantes" induites par la détermination de la structure prosodique et de la situation des noyaux accentués.

10 - SEMANTIQUE ET PRAGMATIQUE

Ces modules interagissent avec l'analyse syntaxique pour d'une part valider ou infirmer les structures reconnues et, d'autre part pour proposer des structures particulières déterminées par le contexte (interrogation, portion de phrase, traits de certains éléments attendus, etc...).

Leur dépendance étroite du contexte ne modifient cependant pas les options générales retenues pour conduire l'analyse d'une phrase.

12èmes JOURNEES D'ETUDE SUR LA PAROLE

MONTREAL 25.26.27 mai 1981

STABILITE ET INSTABILITE DES DEVIATIONS PHONETIQUES ET/OU PHONEMIQUES DES APHASIQUES. INSUFFISANCE D'UN MODELE STATIQUE D'ANALYSE.

Jean-Luc Nespoulous * ** & André Roch Lecours ** ***

* Département de Linguistique et Philologie. Université de Montréal.
** Centre de Rééducation du Langage et de Recherche Neuropsychologique. Hôtel-Dieu de Montréal.
*** Centre de Recherches en Sciences Neurologiques.

.RESUME

La quasi-totalité des études antérieures sur les déviations phonétiques et phonémiques des aphasiques reposent sur l'utilisation d'un modèle descriptif dont le concept opératoire fondamental est le trait distinctif. Ainsi se trouve systématiquement évalué, en un point précis de la chaîne parlée, le nombre de traits différenciant l'élément-cible de l'élément réalisé. Cette évaluation se fait, bien entendu, par référence à une description abstraite et idéalisée du système phonologique(au niveau de la Langue)et a souvent tendance à minimiser-- sans toutefois le négliger complètement-- le rôle éventuel que peut jouer le contexte élocutoire(au niveau de la Parole) dans l'engendrement des erreurs.

La présente étude a également recours à un modèle classique de ce type. Elle en souligne l'intérêt-- particulièrement dans la description des transformations produites par les aphasiques de Broca-- mais aussi les limites. Ces dernières sont surtout évidentes lorsqu'un tel modèle statique est utilisé dans la description des perturbations phonémiques présentées par les aphasiques de conduction, patients chez lesquels la production de phonèmes erronés semble être, en grande partie, la résultante de phénomènes de contamination contextuelle(/otoRite/ → / ototite/). Dans ce dernier cas, une analyse strictement paradigmatique-- statique-- peut être considéré comme insuffisante dans la mesure où elle ne permet pas de saisir ce qui semble constituer le déficit majeur de cette catégorie de patients : une difficulté dans l'organisation dynamique du discours.

. ABSTRACT

Most previous analyses of phonetic and phonemic deviations in aphasic speech rely upon a descriptive model based on distinctive features. Thus, are systematically estimated -- for each and every deviation in the patients' speech-- the number of features differentiating the phonemes which are uttered from the target-phonemes.

The present study relies as well on such a classical model. It emphasizes its usefulness as far as the description of deviant phenomena observed in Broca's aphasia is concerned, but it clearly states that the validity of such a model is limited when one comes to account for deviations observed in conduction aphasia, where the main deficit seems to result from a difficulty in the dynamic ordering of phonemes (/otoRite/→ /ototite/).

ETUDE PERCEPTIVE DE LA VOIX.

Expérience préliminaire : analyse de proximité de voix masculines

Dominique PASCAL CNET Lannion A
Département Codage et Modèles de Communications
Route de Trégastel 22301 LANNION Cedex.

Quels processus utilisons nous pour valider nos décisions perceptives concernant les voix humaines, quels attributs distinctifs nous permettent de différencier et évaluer les emissions vocales de differents locuteurs? Le problème des fondements perceptifs de cette opération, et de leurs corrélats acoustiques, reste trés largement irrésolu. Il a surtout été abordé dans le cadre d'études sur la reconnaissance du locuteur. Une approche directe est envisagée ici: l'analyse des indices de similarité entre "qualitésde voix", ou timbres (selon l'acception courante du terme) de locuteurs distincts.

Selon la théorie descriptive de Laver (79), la notion de qualité de voix traduit un déterminant non-segmental des sons de parole, constitué d'une composante organique, reflétant l'anatomie et la physiologie du locuteur, et d'une composante phonétique, résultant des adaptations musculaires habituelles de celui-ci : ces comportements acquis limitent notre potentiel de performances vocales.

Une étude récente (Singh et Murry (78)) démontre que la dimension dominante pour la discrimination perceptive des voix est une catégorisation selon la notion tacite féminin- masculin. De plus, à l'intérieur de chaque catégorie, les paramètres discriminants diffèrent. Il importe donc d'envisager séparément pour les voix masculines et féminines la détermination du nombre et de la nature de ces paramètres.

Un enregistrement de mots plurisyllabiques prononcés en isolation par dix locuteurs masculins a été réalisé. Dix auditeurs avaient pour tâche d'évaluer directement le degré de proximité des paires de voix qui leur étaient présentées, sur une échelle à sept notes. Ils devaient par la suite effectuer une notation psychophysique (sept notes) de chaque voix en des termes descriptifs usuels. Ces derniers sont donnés par la littérature des études sur la voix (hauteur, rugosité, effort, nasalité...). Les jugements de similarité ont été soumis à une analyse multidimensionnelle via l'analyse des correspondances et/ou INDSCAL et les dimensions obtenues interprétées à l'aide des jugements psychophysiques et de mesures acoustiques relatives à la fréquence fondamentale Fo et à la tessiture formantique des parties vocaliques. Les résultats seront discutés lors de la session affichée.

Les facteurs discriminants devraient être différemment amplifiés par différents contextes (d'une simple voyelle à une phrase). Leur émergence relative semble être fonction du processus perceptif mobilisé par la tâche à laquelle nous sommes confrontés : identification ou évaluation de la voix émettrice. Les deux expériences seront envisagées par la suite, ainsi que l'influence de divers systèmes de traitement de parole sur l'évolution de l'émergence de ces paramètres.

LAVER (79) The description of voice quality in general phonetic theory. Work in Progress n° 12 Edinburgh University p. 30-52.
SINGH andMURRY (78) Multidimensional classification of normal voice qualities. JASA 64 (1) p. 81-87.

PERCEPTUAL STUDY OF VOICE QUALITY

Preliminary study : analysis of proximities among male voices.

Dominique PASCAL CNET Lannion A
Département Codage et Modèles de Communications
Route de Trégastel 22301 LANNION Cedex.

Which strategies are used to make perceptual decisions concerning human voice, which distinctive features allow us to distinguish and rate the vocal production of different speakers ? The question of the perceptual bases for this process as well as their acoustical correlates is still not solved. It has been principally approached through the speaker recognition domain. A study of voice quality similarity is directly considered here.

According to Laver (79) voice quality is conceived as a non-segmental index of speech sounds derived from to separate factors in vocal performance : an organic feature expressing the talker's anatomy and physiology, and a phonetic feature resulting from his usual muscular adjustments. These learned settings reduce our potentialities of vocal production.

A recent study (Singh and Murry (78)) showed that dominant dimension for perceptual discrimination among normal voices was the male-female categorization. Moreover this discrimination within the male-female categories utilized distinct parameters. So it is essential to separately determine the number and nature of these parameters.

The voice of ten male speakers were recorded ; they were instructed to clearly and naturally pronounce plurisyllabic words in isolation. The task of ten listeners was to rate the similarity of voice pairs on a seven-point-equal-appearing interval scale. The listeners were then instruted to make psychophysical ratings according to usual descriptive terms given in the literature (pitch, roughness, effort, nasality...). The similarity judgments were submitted to multidimensional analysis through the analysis of correspondences and/or INDSCAL, and the resulting dimensions were interpreted by means of acoustical measurements derived from fundamental frequency and formantic structure of vocalic parts of speech signal, as well as psychophysical judgments. The results will be discussed during the poster session.

LAVER (79) The description of voice quality in general phonetic theory. Work in Progress n° 12. Edinburgh University p. 30-52.
SINGH and MURRY (78) Multidimensional classification of normal voice qualities. JASA 64 (1) p. 81-87.

LES PARAMETRES SUPRASEGMENTAUX PEUVENT A EUX SEULS DISTINGUER LES TEMPS EN FRANCAIS QUEBECOIS

Dans le langage populaire, on remarque que les trois formes temporelles suivantes du verbe être, employé comme auxiliaire ou au sens de "aller", peuvent être réduites à la même suite segmentale [teta]; ainsi [tetapo] signifie: 1. T'es à Pau - 2. T'étais à Pau - 3. T'as été à Pau. L'analyse acoustique montre que la production de cinq montréalais varie en durée, en intonation et en intensité selon qu'ils se proposent de faire entendre un sens ou un autre. Dans le but de savoir si la variation de l'intonation ou de la durée prise isolément pouvait suffire à distinguer les trois "temps", j'ai synthétisé, au moyen des programmes de Klatt à M.I.T., trois énoncés /tetapo/ où seule varie l'intonation ou la durée. Des tests auprès de deux groupes d'auditeurs montréalais (n = 24 et n = 103) ont montré qu'ils distinguent les trois phrases par l'intonation ($p \simeq .95$) et par la durée des syllabes ($p \simeq 1$).

(Recherche subventionnée par le Conseil de recherches en sciences humaines du Canada)

Laurent Santerre, Université de Montréal
J.-L. Chandon, Université du Québec à Montréal

SUPRASEGMENTAL PARAMETERS DISTINGUISH TENSES IN TENSES IN MONTREAL FRENCH

In rapid connected speech, we observe that the forms of 3 tenses may be reduced to the same segmental string, for example [teta]; [tetapo]: 1. T'es à Pau (you're in Pau) - Pau is a town of France - 2. T'étais à Pau (you were in Pau). 3. T'as été à Pau (you've been in Pau or you went to Pau). Sonagraphic analysis of the production of five Montrealers asked to say [tetapo] with the intention of meaning the different tenses reveals that both intonation and duration, and, to a certain extent, intensity, as well as the vocalic spectrums, vary with the tense on the three syllables. In order to determine whether the variation of the intonation and of the duration alone is sufficient to distinguish the three tenses, three sentenses /tetapo/ were synthesized by means of the Klatt programs at M.I.T. Tests were done with two groups of Montreal-French listeners (n = 24 and n = 103), and it was found that they were able to distinguish the three tenses on the basis of the variation of intonation ($P \simeq .95$) and of duration on the syllables, $P \simeq 1$.

(Work supported by CRSHC)

Laurent Santerre, Univ. de Montréal

J.-L. Chandon, Univ. du Québec à Montréal

12èmes JOURNEES D'ETUDE SUR LA PAROLE

MONTREAL 25, 26, 27 mai 1981

Corrélations acoustiques et articulatoires entre différents variphones du phonème /R/

Claude Tousignant
(Université du Québec à Chicoutimi)

Le français montréalais comporte une large variété de /R/. Ceux-ci diffèrent de par leurs caractéristiques articulatoires et acoustiques. On ne compte pas moins d'une dizaine de /R/ montréalais, dont certains sont articulés avec la langue placée vers l'avant de la bouche ou encore dans la zone du palais dur, du palais mou ou du voile du palais.

De plus, les différences articulatoires se situent au niveau du mode du phonème. Certains sont occlusifs, pluri-occlusifs ou tout simplement constrictifs. Ces différences articulatoires et bien d'autres impliquent évidemment une large variété de réalisations acoustiques. Ainsi, certains /R/, de par leur position postérieure de la langue, posséderont une résonance antérieure alors qu'on notera le contraire dans le cas des phonèmes plus antérieurs.

On note également que la pression d'air provenant des poumons, la largeur du contact de la langue avec la paroi supérieure, la force articulatoire, etc., sont autant de paramètres qui modifieront le résultat acoustique de ce phonème.

Quelles sont les relations existant entre l'existence de tels paramètres et la réalisation acoustique de différents /R/ montréalais? Quelles constantes acoustiques retrouve-t-on entre deux /R/ dont certains paramètres articulatoires sont les mêmes? Quelles constantes articulatoires note-t-on entre deux /R/ dont le résultat acoustique est à peu près le même? C'est à ce genre de questions que nous tenterons de répondre.

12èmes JOURNEES D'ETUDE SUR LA PAROLE

MONTREAL 25, 26, 27 mai 1981

Montreal French has a large variety of /R/, which differ by their articulatory and acoustic characteristics. We can find about ten different types of /R/ in Montreal, some articulated with a front tongued position, others with a back tongued position, under the hard palate, soft palate or velum. We also observe that the articulatory differences can depend upon the articulatory mode. Some phonemes are occlusive, pluri-occlusives or fricative. These articulatory differences create obviously a large variety of acoustic realisations. Due to their posterior position, some /R/ have an anterior resonance and others have a posterior resonance because of an anterior lingual position. We also note that the air pressure coming from lungs, the tongue's contact width, the articulatory strenght, etc., are parameters that can modify the acoustic result.

What relations exist between the existence of such parameters and the acoustic relaization of different Montreal /R/? What acoustic constancies can we find between two /R/ whose articulatory parameters are the same? What articulatory constancies can we find between two /R/ whose acoustic result is almost the same? These are the problems we will try to solute.

ESSAI D'OPTIMISATION D'UN VOCODEUR A CANAUX

ZURCHER FREDERIC - CNET LANNION FRANCE.

Au CNET-LANNION, les premières études et réalisations de vocodeurs à canaux remontent à 1965. Depuis cette époque plusieurs versions se sont succédées ; mais les améliorations étaient plus d'ordre technologique que fonctionnel.

Aussi avons-nous entrepris un essai d'optimisation de différentes fonctions du vocodeur à canaux, pour le mettre en concurrence avec les techniques de prédiction linéaire, qui ont été l'objet de nombreuses études récentes.
Les travaux réalisés sont les suivants :

- un nouvel échelonnement des filtres d'analyse et de synthèse a été étudié. Par rapport à l'échelonnemnt précédent, ses caractéristiques se rapprochent davantage de celles des bandes critiques de l'oreille et apportent une meilleure résolution de l'analyse dans le bas du spectre de fréquences. Des tests d'écoute ont montré une très nette amélioration de l'agrément d'écoute, surtout pour les voix masculines.

- la bande passante globale a été élargie de 1 000 Hz vers le haut par adjonction de deux canaux supplémentaires.

- un algorithme de correction et de lissage de la fréquence fondamentale a été mis au point et implanté en temps réel. Pour des conditions de fonctionnement satisfaisantes du détecteur de fondamental, il ne subsiste pas de défaut nettement perceptible.

- une interpolation fine du signal d'excitation des filtres de synthèse ainsi qu'une addition de bruit à ce signal (dans les zones de voisement), dosé canal par canal, ont permis d'obtenir une parole synthétique plus naturelle.

- un algorithme de compression des données d'analyse, basé sur la reconstitution d'un échantillon d'analyse à partir des deux échantillons qui l'encadrent, a permis de passer d'un débit de 3 600 bit/s à un débit de 2 400 bit/s, sans perte notable de qualité.

Compte tenu de ces améliorations, il semble établi que la technique de vocodeurs à canaux permet l'obtention de liaisons téléphoniques de qualité acceptable avec un débit de 2 400 bit/s.

ON AN IMPROVED CHANNEL VOCODER

ZURCHER Frédéric CNET LANNION FRANCE

The first research on channel vocoders and its production at the CNET in Lannion goes back to 1965. Since then, several versions have been produced, but the improvements have been in the technological domain rather than in the performance domain.

We have undertaken, therefore, the improvement of various components of the channel vocoder, in order to upgrade its transmission quality. We hope that such improved channel vocoders can compete, in terms of the quality, with LPC vocoders which have been the subject of much research recently.

The improvements introduced are the following :

- A new distribution for analysis and synthesis filters was studied. This new distribution system is close to that of the critical bands of the ear and provides a better analysis accuracy at low frequencies.

- A listening test showed a clear preference for the new system of distribution in comparison with a channel vocoder with our conventional distribution system. The preference was even more marked in the case of male speakers.

- The transmission band was increased by 1000 Hz through the addition of two extra channels.

- A correction and smoothing algorithm for the fundamental frequency (Fo), which is working in real time, was perfected. For satisfactory functioning conditions of the fundamental detector, no clear audible defects were perceived.

- In the synthesis process, Fo values specified every 20 ms were smoothed by interpolation ; a noise is also injected into each channel to improve the naturalness of the synthesized speech.

- Channel data are analyzed every 20 ms, and then a compression algorithm reduces the sampling rate by one half in such way that the missing analysis data can be reconstructed from the information received every 40 ms. This compression algorithm reduces the transmission rate from 3600 bit/s to 2400 bit/s without notable loss in the quality.

Taking into account the improvements achieved, telephon links of an acceptable standard with a bit rate of 2400 bit/s are possible.

12èmes JOURNEES D'ETUDE SUR LA PAROLE

MONTREAL 25.26.27 mai 1981

DECODAGE LEXICAL ET COMPOSANTE PHONOLOGIQUE DANS ARIAL II

GUY PERENNOU
MARTINE DE CALMES
MINH BUI VAN

Laboratoire CERFIA
Université Paul Sabatier - Toulouse

.**RESUME**

Nous exposons les idées principales concernant l'organisation du système de reconnaissance de la parole A.R.I.A.L. II (Analyse et Reconnaissance des Informations Acoustiques et Linguistiques) développé au Laboratoire C.E.R.F.I.A.

ARIAL II est un système de reconnaissance de parole continue dictée, sans adaptation notable au locuteur, pour l'entrée de données textuelles ou numériques. C'est un système expérimental qui doit permettre l'essai de divers modèles phonétiques, syntaxiques et sémantiques. A cette étape de notre projet la compréhension n'est pas le but poursuivi. Cependant, un niveau linguistique est utilisé pour contribuer à la reconnaissance :

a) par la prédiction de sous-lexiques
b) par le rejet de suites de mots hypothétiques incorrectes.

Dans cet article nous décrivons l'organisation de l'unité de décodage lexical utilisant les contraintes syntaxiques et sémantiques. Nous donnons ensuite un exemple de modèle phonologique incluant la syllabe et son utilisation dans le décodage lexical.

. ABSTRACT

We intend to give the main ideas of the organization of a speech recognition system : ARIAL II (Analysis - Recognition - Informations - Acoustical - Linguistic), developped in the CERFIA laboratory.

ARIAL II is a recognition system for dictated connected speech, without adaptation to the speaker. We aim at the entry of textual and numerical data. Furthermore, ARIAL II will allow the experimentation on phonetic and linguistic models. The comprehension is not the purpose. However, a linguistic level is needed in order to contribute to the recognition :

- first by prediction of sublexicon
- secondly by rejection of incorrect hypothetic strings.

In this paper we first describe the organization of the lexical decoding unit using the syntactic and semantic constraints, give afterwards an example of phonological model including the syllabic unit and its use in the lexical decoding.

INTRODUCTION

Les systèmes ARIAL (Analyse et Reconnaissance des Informations Acoustiques et Linguistiques) sont développés au laboratoire CERFIA ; ces recherches sont soutenues par l'ADI. (*)

ARIAL II est un système de reconnaissance de la parole continue dictée sans adaptation du locuteur. Son organisation est adaptée à l'entrée vocale de données textuelles et numériques sur ordinateur. L'introduction de pauses pendant la dictée permet de fractionner la parole continue en petits groupes de mots, ce qui facilite la reconnaissance.

Dans l'étape actuelle, l'objectif n'est pas la compréhension de la parole. Cependant un niveau linguistique - syntaxe et sémantique - est utilisé pour contribuer au décodage lexical :

- en filtrant les hypothèses : celles dont on peut prouver simplement l'incorrection sont abandonnées,
- en prédisant les suites possibles (un sur-ensemble) attachées à chaque hypothèse : cela revient à sélectionner un sous-lexique.

Un des objectifs de ARIAL II est de permettre l'expérimentation de modèles phonologiques, lexicaux et linguistiques ainsi que des stratégies de reconnaissance.

C'est pourquoi le système est conçu de façon très modulaire, chaque composante ayant une autonomie importante pour le contrôle comme pour l'exécution des tâches qui lui sont soumises. Cela permettra en outre une implémentation multiprocesseur très naturelle.

Dans cette communication, nous nous proposons de présenter les idées directrices d'ARIAL II et de préciser l'organisation du niveau lexical et phonologique. On trouvera dans PERENNOU (1978, 1980), PERENNOU, FATHOLAHZADEH , DE CALMES (1980) un exposé plus systématique des idées directrices.

I - PRESENTATION GENERALE

ARIAL II est un système de reconnaissance d'entrées vocales prévu pour s'insérer dans un système plus grand comportant (cf fig. 1) :

- un module de gestion du dialogue,
- le système de reconnaissance ARIAL II,
- un module d'analyse et d'interprétation linguistiques.

L'objectif de ce dernier est de poursuivre l'analyse et l'interprétation des sorties d'ARIAL II pour :

- écarter les hypothèses comportant des incohérences non détectées par ARIAL II,
- affiner l'analyse syntaxique et sémantique des sorties non détectées comme incorrectes afin de permettre, s'il y a lieu :
  - de les retranscrire de manière orthographique,
  - de restituer aux phrases leur signification dans le cadre de l'application en cours. (ex : "cinq" → 5 € N)

(*) Aides à la recherche ADI numéros 44-79 et 80-444

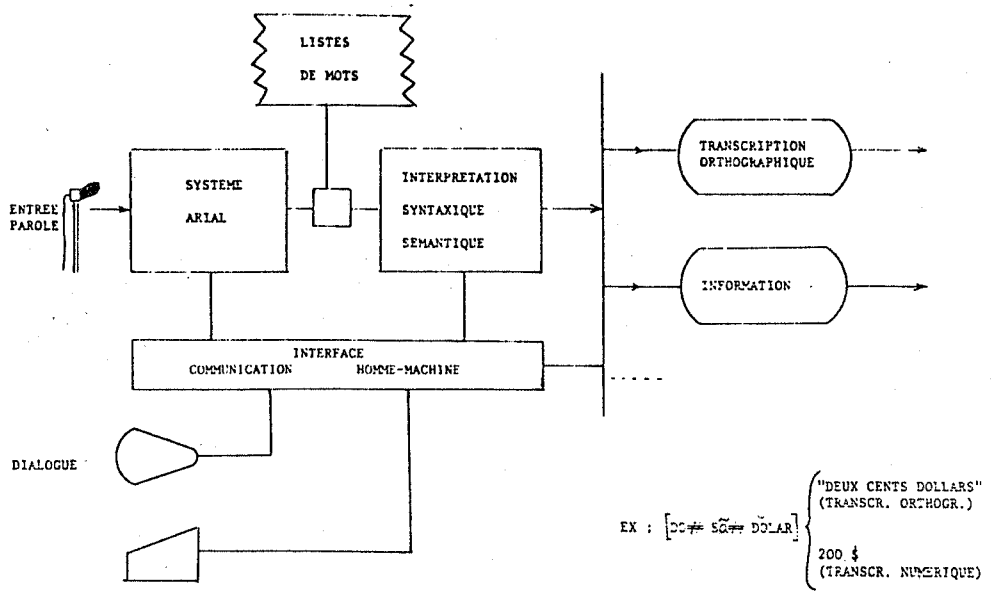


Fig. 1 *Présentation générale du système et de son environnement.*

L'une des fonctions du module de dialogue est d'assurer l'entrée de termes ou de phrases inconnus de ARIAL II.

De plus, lorsque des difficultés apparaissent à la reconnaissance et à l'interprétation, il peut être préférable de procéder à l'entrée directe par clavier.

A noter également les tâches de recompilation des réseaux lexical et phonologique lors de l'introduction ou la suppression de mots ou syllabes. Des mises à jour doivent être également possibles aux niveaux linguistiques -voir organisation ARIAL II ci-après.

Organisation d'ARIAL II

Le système est organisé de manière modulaire pour permettre une adaptation très souple aux nouvelles applications et l'amélioration indépendante de chaque partie. Il se décompose en trois modules (cf fig. 2) :

- Analyse acoustique et phonétique,
- Reconnaissance lexicale,
- Contrôle syntacticosémantique.

Ces fonctions se partagent deux à deux les ressources communes suivantes :

- les informations phonétiques,
- les listes d'hypothèses,
- les listes de mots partiellement reconnues,
- les listes de mots acceptées.

Notons que le contrôle est réparti sur le système. En effet pour gérer l'accès aux ressources partagées et ordonnancer le séquencement des actions, nous introduisons une hiérarchie de contrôle :

- contrôle sur les services les plus extérieurs (traitement du signal, reconnaissance lexicale...) par processus nommé arbitre ou contrôleur,

- contrôle sur les services imbriqués (calcul du score,...) par le service qui les contient.

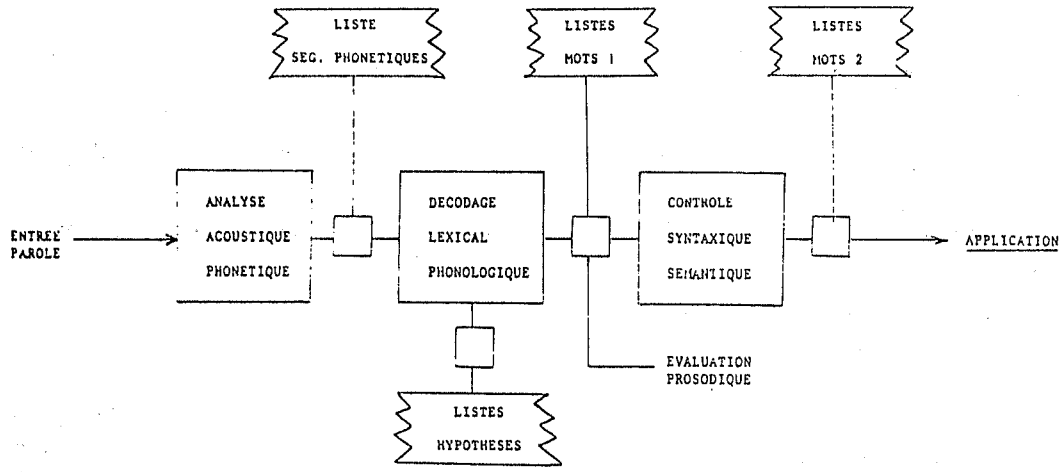


Fig. 2 *Organisation générale des systèmes ARIAL.*

II - ORGANISATION GENERALE DES NIVEAUX LEXICAL ET PHONOLOGIQUE

On trouvera dans D. PL. KLATT (1979), M. ROSSI (1980) et G. PERENNOU (1979) diverses approches sur le lexique et la composante phonologique en traitement automatique de la parole. Indiquons également LENNIG et MERMELSTEIN (1979) qui présentent un système analogue à ARIAL II, mais avec une imbrication importante des niveaux linguistique et phonologique. Notons d'abord que lexique et phonologie sont interdépendants car le schéma d'encodage (et/ou de décodage) des termes, qui rend compte de ses diverses réalisations phonétiques en contexte, dépend en effet du choix des unités et des règles phonologiques. (cf. PERENNOU, TEP (1979) à ce sujet).

Dans un système de reconnaissance de la parole, compte tenu des exigences d'optimisation, il faut adapter les structures de données du lexique et des composantes phonologiques en vue du décodage.

La figure 3 donne le schéma d'organisation du sous-système de Décodage Lexical et Phonologique - ou DLP - de ARIAL II. En voici les différentes parties :

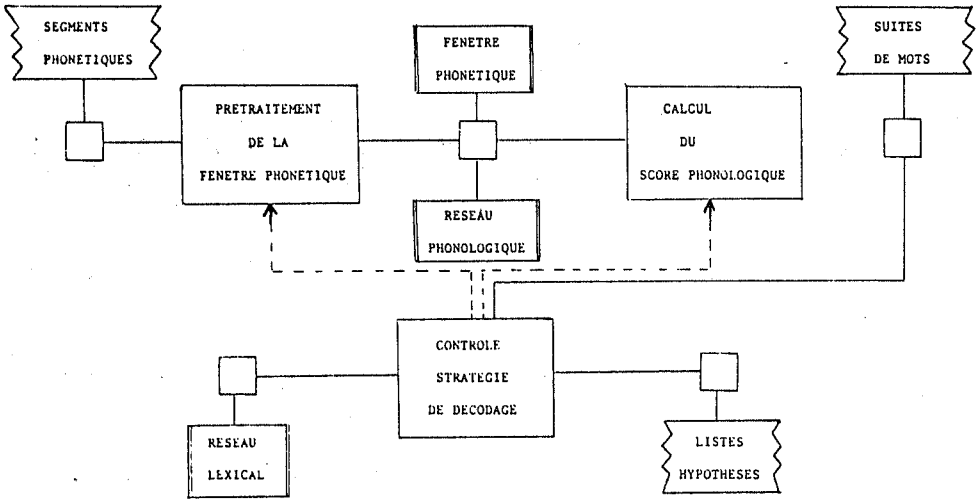


Fig. 3 *Organisation du Décodage Lexical et Phonologique (Sous-système DLP).*

1) Le module de Contrôle de la Stratégie de Décodage - ou CSD - qui en fonction de la stratégie adoptée assure :

- la distribution des tâches (notamment d'évaluation de score, de compatibilité linguistique...)
- la mise à jour des listes d'hypothèses de décodage (ajout, suppression, modification),
- la liaison avec les niveaux linguistiques,
- l'accès au réseau lexical.

2) Le module de Calcul du Score Phonologique - ou CSP - qui détermine la compatibilité des unités phonologiques (communiqués par CSD) avec les données phonétiques d'entrées ; pour cela il doit accéder aux descriptions des unités phonologiques et mettre en oeuvre une stratégie donnée. La compatibilité étant obtenue, le nouveau score de l'hypothèse concernée est calculée.

3) Le module de Prétraitement de la Fenêtre Phonétique - ou PFP - qui assure :

- l'acquisition d'une nouvelle suite de segments phonétiques lorsque la précédente est traitée - une fenêtre couvre environ deux syllabes ; deux syllabes successives se chevauchent,
- l'élaboration ascendante du décodage phonologique ; cette élaboration s'arrête dès qu'apparaissent des ambiguïtés.

4) Le Réseau Lexical qui organise les connaissances lexicales en un arbre factorisant les parties gauches communes (cf Fig. 4). Chaque noeud contient :

- la référence à une unité phonologique principale (partie croissante et coeur syllabique antérieur ou à une fin de mot (voir § IV à ce sujet),
- les acceptions syntaxiques et sémantiques : pour utiliser le noeud, une hypothèse devra avoir au moins une caractéristique linguistique compatible.

5) Le Réseau Phonologique contient les descriptions phonologiques nécessaires à la détermination des compatibilités. Il traduit en fait les connaissances phonologiques. Nous décrirons plus loin ce réseau.

III - STRATEGIE GENERALE DE DECODAGE

Une hypothèse H correspond à une tentative d'interprétation des entrées phonétiques par une suite cohérentes de mots. Les diverses hypothèses envisageables sont évaluées en parallèle et progressivement. A chaque nouvelle fenêtre phonétique leur score est réévalué. Certaines hypothèses sont abandonnées :

- si elles conduisent à des impasses du point de vue linguistique,
- si elles voient leur score devenir trop faible.

De plus le nombre total des hypothèses est limité. D'une manière générale nous avons donc une stratégie de recherche en faisceau.

Une hypothèse est représentée par un 5 - uplet H = (N, SC, HSYSE, DT, PTST) avec

N : un noeud du réseau lexical, le dernier utilisé par H,

SC : le score obtenu lors du cheminement jusqu'à N,

DT : date du dernier segment phonétique d'entrée utilisé par H,

PTST : le pointeur vers le dernier élément de la suite de mots auquel se chaînera le mot recherché.

Les hypothèses se rangent dans deux types de listes. Si on laisse de côté le problème d'optimisation de la gestion de places libres et de vitesse de recherche nous avons deux listes :

$\mathcal{H}_A$= hypothèses anciennes avec DT antérieur au coeur syllabique en cours d'analyse,

$\mathcal{H}_N$= hypothèses nouvelles avec DT postérieur au coeur syllabique.

Le contrôle de stratégie de décodage (CSD) active la procédure SUIV (H) pour toute hypothèse H prise dans $\mathcal{H}_A$. Dans ses grandes lignes, cette procédure fonctionne comme suit :

Procédure SUIV

1er cas VALIDATION - NOEUD (N) si le noeud N possède des noeuds successeurs.

Pour tout successeur N de N' tel que HSYSE' $\neq \emptyset$, la procédure active le module "calcul du score" (CSP) et le module "acquisition date" pour déterminer le score SC' et la date DT' - HSYSE', ensemble des catégories du mot recherché, résulte de l'intersection de HSYSE et de l'ensemble des possibilités syntactico-sémantiques de N'. SC' et DT' représentent le score et la date après validation de l'unité phonologique PH de N à partir de la date DT et du score SC déjà réalisé-. Si le nouveau score est suffisant (SC'$\geq$ seuilΘMIN), l'hypothèse H' = (N', SC', HSYSE', DT', PTST') est créée dans $\mathcal{H}_A$ ou $\mathcal{H}_N$ selon que DT' est antérieure ou postérieure au noyau en cours d'analyse.

L'hypothèse H est ensuite supprimée.

2ème cas VALIDATION - MOT (Q, R) si le noeud N associé à H est terminal (autrement dit toutes les unités phonologiques d'un mot M d'adresse ad (N) ont été validées).

Un message Q = (ad (N), HSYSE, PTST) propose M pour suivant d'une liste (éventuellement vide) de mots. PTST définit l'adresse du mot auquel M devra se chaîner. M satisfait aux contraintes prédictives (cf 1er cas) ; HSYSE contient l'ensemble des catégories syntactico-sémantiques que l'on peut associer à M, chacune pouvant donner lieu à une réponse si les contraintes a posteriori sont satisfaites. En cas de succès, deux types de message réponse sont possibles :

a) R = (accepte, HSYSE', PTST') où HSYSE' traduit les contraintes pour la recherche du mot suivant, PTST' l'adresse à laquelle il devra se chaîner. Cette réponse provoque la création d'une hypothèse H' = ((0, SCD, HSYSE', DT, PTST')) dans $\mathcal{H}_A$. L'hypothèse initialise une nouvelle recherche de mot à partir de la racine 0 du réseau lexical avec un score initial SCD, les possibilités syntactico-sémantiques sont données par HSYSE'.

b) R = (accepte, ##, PTST') où ## déclenche une vérification de pause.

Lorsque toutes les transitions ont été examinées, une réponse R = (transition terminée) provoque la suppression de l'hypothèse H. A noter que ce dernier message peut être le seul, auquel cas l'hypothèse est abandonnée.

Contrôle de la stratégie de décodage (CDS)

Ce contrôle concerne :

A - l'initialisation du processus de décodage,

B - le traitement des fenêtres avec parole.

L'arrêt de la procédure A (resp. B) déclenche l'activation de la procédure B (resp. A) tant qu'il reste de la parole à examiner.

CSD fonctionne pour l'essentiel comme suit

Initialisation

Une hypothèse H_o = (0, SCD, $HSYSE_o$, DT_o, $PTST_o$) est créée et $\mathcal{H}_A$ et $\mathcal{H}_N$ sont initialisés par :

$\mathcal{H}_A = (H_o), \mathcal{H}_N = (\emptyset)$.

La procédure de recherche de la première fenêtre phonétique contenant de la parole est activée.

Traitement des fenêtres avec parole

Itérations tant qu' : - il existe des fenêtres à traiter
- il existe des hypothèses valides

Chaque itération comporte :

- un prétraitement de la fenêtre phonétique,
- le traitement des hypothèses de $\mathcal{H}_A$,
- des tâches de terminaison d'itération.

. Prétraitement de la fenêtre phonétique

Le décodage phonétique n'est pas l'objet de l'article, indiquons cependant que le prétraitement détermine :

- le noyau syllabique, s'il existe clairement ;
- l'emplacement vraisemblable de voyelles dans les continuums vocaliques tels que: [ynlimite] ("une limite") ; les oppositions grave-aigu et ± stable sont prises en compte dans ce cas ;
- la localisation des consonnes fortes (occlusives, fricatives) ;
- l'absence de noyau vocalique et la possibilité de pause (##).

. Traitement des hypothèses de $\mathcal{H}_A$

Pour chaque hypothèse de $\mathcal{H}_A$, la procédure SUIV détermine les hypothèses filles à retenir (en liaison avec le niveau linguistique s'il y a lieu).

. Traitement d'itération

- Une syllabe ayant été traitée on fait appel à de nouveaux segments phonétiques et on abandonne ceux qui dans toute hypothèse sont antérieurs à DT.
- Les hypothèses de $\mathcal{H}_N$ deviennent les hypothèses anciennes ($\mathcal{H}_A \leftarrow \mathcal{H}_N, \mathcal{H}_N \leftarrow (\emptyset)$)

Calcul du Score (CSP)

Le type de calcul de score qui peut figurer ici, n'est pas figé. Il n'est d'ailleurs pas déterminant quant à la structure de notre système. Donnons cependant quelques indications sur les aspects liés à la structure d'ARIAL II :

a) Le décodage fournit une unité phonologique (actuellement apparentée à la syllabe)

b) La description des différentes variantes de cette unité phonologique est connue du module d'élaboration du score

c) Dans la fenêtre en cours d'étude, se trouve une suite de segments phonétiques où le prétraitement a repéré les éléments principaux

d) L'hypothèse à l'origine de la demande de vérification a fourni la date DT de la dernière unité phonologique validée ainsi que le score SC déjà obtenu

e) Une compatibilité est d'abord établie entre l'unité demandée et la suite segments phonétiques postérieure à DT. Cette compatibilité peut être déterminée de différentes manières :

- méthode type programmation dynamique sur segments phonétiques,
- recalage sur les segments les plus importants.

f) Le score SC' élaboré est une combinaison linéaire faisant intervenir SC et la compatibilité obtenue.

IV - UN EXEMPLE DE REPRESENTATION PHONOLOGIQUE

IV.1 - Présentation générale

Voici un exemple de représentation phonologique utilisée dans ARIAL II montrant les relations entre réseau lexical, composante phonologique et détermination de score. Les phonèmes sont répartis comme suit :

Q ⟶ p|t|k
Ø ⟶ b|d|g
S ⟶ f|s|ʃ
Z ⟶ v|z|ʒ
Co ⟶ Q|Ø|S|Z
N ⟶ m|n
L ⟶ l|r
SV ⟶ j|ɥ|w
V ⟶ toutes les voyelles
U ⟶ i|y|u

Remarque : le phonème (ɲ) est traité comme (nj) .

Nous utiliserons le symbole ∧ pour désigner le phonème vide et poserons :

L' ⟶ L | ∧
SV' ⟶ SV| ∧

Nous devons définir des unités réutilisables autant que possible dans divers contextes. C'est pourquoi nous nous appuyons d'abord sur la syllabe stable définie comme suit :

$SYLL_{ST} \longrightarrow DSY + NOY \mid NOY$
$DSY \longrightarrow DSY_0 \mid DSY_1$
$DSY_0 \longrightarrow C_0 + L'$
$DSY_1 \longrightarrow N \mid L$
$NOY \longrightarrow SV' + V$

Règles d'interdiction :

- SYLL $\longleftarrow$ $C_0 + L + j + V$
- NOY $\longleftarrow$ w + {o,ɔ,... } (listes extensibles sur des
y + {o,ɔ,ø,œ,... } applications limitées)

Le codage d'un mot s'effectuera à partir de ses parties stables et éléments restant qui sont de trois types :

- éléments internes fermant les syllabes,
- éléments fermant le mot,
- éléments initiaux.

Les "e" succeptibles d'être élidés naturellement sont considérés comme ne donnant pas de noyau de partie stable. Avec ces conventions voici quelques décompositions de mots :

"dimanche" ⟶ (|di) (|mɑ̃) (ʃə |60)
"divisé" ⟶ (|di) (|vi) (ze |) (|01)
"mardi" ⟶ (|ma) (r |di) (|00)
"mars" ⟶ (|ma) (rs |20)
"stabilité" ⟶ (s |ta) (|bi) (|li) (|té) (|01)
"hibou" ⟶ (h |i) (|bu) (|01)
"samedi" ⟶ (|sa) (mə |di) (|00)
"aérosol" ⟶ (- |ae) (|ro) (|so) (l |01)

On observera que chaque parenthèse comporte deux parties représentant :

1) le contexte gauche à prendre en charge (cette partie peut être vide),

2) la partie syllabique stable ou la condition de terminaison.

Cette dernière se présente sous forme ij, i signifiant une modalité phonétique (Ex : encha^{s}nement, liaison, nasalité...) et j une variation d'origine syntaxique (Ex : invariable, variation en nombre, en genre...).

Pour résoudre au mieux quelques cas de variations phonétiques, il est nécessaire d'établir un certain nombre de règles.
Voici (de manière non exhaustive) deux cas :

a) Ouverture de la voyelle en syllabe ouverte :

En reconnaissance de la parole, l'opposition e/ɛ (ou ø/œ, ou ɔ/o) n'est pas toujours facile à apprécier. On peut donc choisir un symbole qui représente l'ensemble des deux phonèmes - soit E par exemple - Ainsi on aura :

"vertu" ⟶ (|vE) (-r |ty) (|01)
"vérité" ⟶ (|vE) (- |ri) (|tE) (- |01)

On peut ainsi mettre en commun (|vE) dans les deux mots (Voir organisation du réseau lexical). Dans l'unité suivante, grâce à l'indication "-" nous pouvons selon que la syllabe est fermée ou ouverte, vérifier s'il y a lieu que E est [ɛ]

ou [e] .

b) Nasalisation - Dénasalisation

Considérons les variations du genre bõ/bɔn(a).Il est intéressant, la distinction entre ɔ/õ n'étant pas toujours aisée, de rendre commune la racine "bon".

"bon" ⟶ (|bɔ) (~n |12)

12 signifiant liaison et variation en genre et nombre. Le symbole ~ doit alors se traduire ainsi : si le [n] se réalise (liaison ou féminin) vérifier que c'est [bɔ] qui a lieu (ou [bɔn]), sinon vérifier au contraire que c'est bõ.

IV. 2 - Traduction du système phonologique dans le réseau lexical d'ARIAL II et dans le système d'évaluation de score

Réseau lexical

La figure 4 montre sous forme de graphe comment s'organise un tel réseau, avec mise en facteur des parties gauches communes.

On remarquera aux noeuds les deux constituants donnés précédemment dans une parenthèse. Les autres informations représentent les catégories syntactico-sémantiques compatibles avec le noeud considéré (les codes sont explicites - ex : MAN pour [mã])

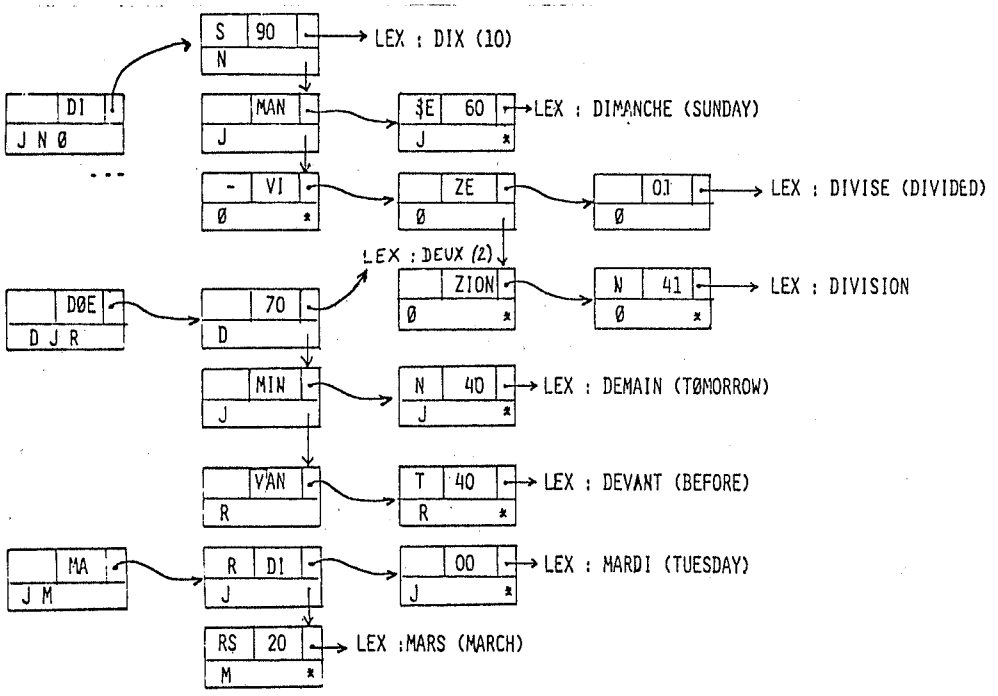


Fig. 4 *Un réseau Lexical (ci-dessus)*

Structure d'un noeud (ci-contre)

FERMETURE SYLLABIQUE	SYLLABE OU TYPE FIN-LIAISON	POINTEUR VERS LE NOEUD FILS AINE
CATEGORIES SYNTACTICO-SEMANTIQUES		POINTEUR VERS LE NOEUD FRERE (=* SI NOEUD BENJAMIN)

Réseau phonologique

La figure 5 montre sous forme de graphe ET/OU comment est traduite la structure syllabique pour l'évaluation du score. Chaque noeud terminal est lui-même décomposé sur le même principe afin d'aboutir à une description phonétique. Cette partie de notre projet sort du cadre de l'exposé actuel qui se limite au niveau phonologique abstrait traduisant déjà le type de démarche de la reconnaissance.

L'opérateur noté $\overset{\curvearrowright}{ET}$ est en fait la concaténation mais les procédures de traitement des graphes ET/OU restent valables. On voit qu'ils permettent de rendre compte de la structure syntagmatique de la syllabe en même temps que des choix paradigmatiques.

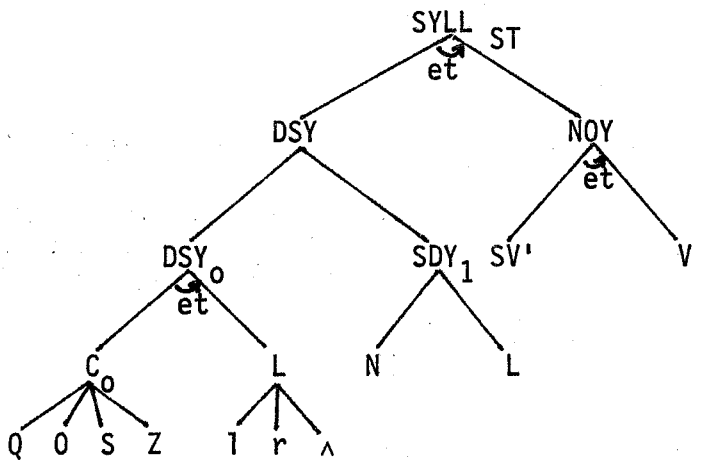


Fig. 5 *Graphe ET/OU représentant la structure d'une syllabe stable* $SYLL_{ST}$

CONCLUSION

Notre système ARIAL II doit se substituer à ARIAL I actuellement disponible en version conversationnelle sur IRIS 80. Il pallie un certain nombre de limitations provenant de l'absence d'une articulation phonologique.

Actuellement le compilateur de réseaux est opérationnel et le module de décodage lexical en cours d'essai. Ils sont écrits dans un premier temps en PASCAL sur IRIS 80 ; l'implémentation multimicroprocesseur sera étudiée dans un deuxième temps.

REFERENCES

CAELEN (J.), 1974, "Un modèle de cochlée et son application au signal" ; Thèse de docteur-ingénieur, Université Paul Sabatier, Toulouse. pp. 1-150.

CAELEN (J.), 1979, "Un modèle d'oreille. Analyse de la parole continue. Reconnaissance phonémique" ; Thèse d'état, Université Paul Sabatier, Toulouse. pp. 1-200.

CAELEN (J.), PERENNOU (G.), 1978, "Indices et traits acoustiques dans un système de reconnaissance de la parole continue : quelques résultats" ; 9ème J.E.P. (G.A.L.F.), Lannion. 7 pages.

CAELEN (J.), PERENNOU (G.), 1979, "Etage phonétique d'un système de reconnaissance et de compréhension de la parole continue" ; 9ème Congrès des Sciences

Phonétiques, Copenhague. 12 pages.

DOURS (D.), FACCA (R.), PERENNOU (G.), 1977, "Automate de segmentation et de détection du fondamental" ; 8ème J.E.P. 7 pages.

FANT (G.), 1970, "Acoustic theory of speech production" ; 2ème édition Mouton, The Hague, Paris.

FANT (G.), 1973, "Speech sounds and features" ; MIT Press, Cambridge (Mass) and London.

FLANAGAN (J.L.), 1972, "Speech analysis synthesis and perception" ; 2ème édition, Springer Verlag, Berlin - Heidelberg, New-York.

KLATT (D.H.), 1979, "Speech perception : a model of access to phonetic analysis and lexical access" ; Journal of Phonetics 1979, 7. pp. 279-312.

LENNIG (M.), MERMELSTEIN(P.), 1979, "Entrainement lexical semi-automatique d'un système de reconnaissance à base syllabique". 12 pages.

MERMELSTEIN (P.), 1975, "Phonetic context controlled strategy for segmentation of phonetic labelling speech" ; I.E.E.E., ASSP 23, n° 1. pp. 79-82.

PERENNOU (G.), 1978, "Reconnaissance des mots isolés dans le cas des grands vocabulaires" , 2ème Congrès AFCET-IRIA. 9 pages.

PERENNOU (G.), 1980a, "ARIAL II : System for speech recognition" ; IAPR, Miami. 4 pages.

PERENNOU (G.), 1980b, "Lexique et traitement automatique de la parole" ; Revue acoustique (G.A.L.F.), vol. 13 n° 53. pp. 106-113.

PERENNOU (G.), FATHOLAHZADEH(A.), DE CALMES (M.), 1980, "Le filtrage syntaxique pour l'entrée vocale de texte" ; Séminaire, Paimpont. 17 pages.

PERENNOU (G.), TEP (G.), 1979, "Grands lexiques et traitements phonologiques : une structure de composante phonologique adaptée au traitement automatique" ; 10ème J.E.P., Grenoble. pp. 344-351.

ROSSI (M.), 1980, "Rôle de la phonologie dans la reconnaissance de la parole" ; Revue acoustique (G.A.L.F.), vol. 13 n° 1. pp. 54-67.

WEINSTEIN (C.S.), et coll., 1975, "A system for acoustic phonetic analysis of continuous speech". ASSP 23 n° 1. pp. 54-67.

Nous nous bornerons ici à renvoyer aux actes des J.E.P. GALF/AFCET et des Congrès AFCET/IRIA de Reconnaissance des Formes et Intelligence Artificielle pour les systèmes de la reconnaissance de la parole en français :

- MYRTILLE 1 et 2 (J.P. HATON, J.M. PIERREL et coll)
- KEAL (G. MERCIER, P. QUINTON, R. VIVES et coll)
- ESOPE 0 (J.S. LIENARD, J. MARIANI)

SYNTHESE PHONEMIQUE PAR REGLES

S. CASTAN - J.Y. LATIL Laboratoire C.E.R.F.I.A.

ABSTRACT

This paper describes a phonemic synthesis by rules for French sentences. The system is implemented on a mini-computer equipped with a channel vocoder output.

The software makes use of phonemic articulatory, transitory and prosodic level rules.

The possibility to modify the parameters in an interactive manner makes this configuration very flexible.

RESUME

Une synthèse phonémique par règles du français est décrite dans cet article. Ce système implémenté sur un mini calculateur utilise comme organe de sortie un vocoder à canaux.

Le logiciel utilise des règles au niveau phonémique, articulatoire, transitoire et prosodique.

La possibilité de modifier d'une manière interactive les paramètres rend ce système très souple.

SYNTHESE PHONEMIQUE PAR REGLES

S. CASTAN - J.Y. LATIL Laboratoire C.E.R.F.I.A.

INTRODUCTION

Dans le cadre des recherches que nous effectuons sur la communication homme-machine, nous avons réalisé un système de synthèse vocale par règles implanté sur un mini ordinateur.

Ce système, plus complet et d'une utilisation plus simple que la première version décrite au 10ème J.E.P. de Grenoble, permet une production vocale de meilleure qualité.

PRINCIPE GENERAL

On propose la réalisation d'une synthèse phonémique du français par règles, en utilisant comme synthétiseur un vocoder à canaux(*).

Cette méthode utilise des règles pour les niveaux phonémique, coarticulatoire, transitoire et prosodique.

Ces règles permettent de régénérer les trajectoires formantiques ou "squelette" du spectre.

A chacun de ces niveaux, est associé un ensemble de paramètres. Ces paramètres peuvent être éventuellement modifiés par une règle d'un autre niveau.

Le spectre final est obtenu à partir du "squelette" par l'utilisation d'une règle formantique.

REGLES PHONEMIQUES

Le spectre digital du phonème est reconstitué par juxtaposition de un ou plusieurs segments temporels de spectre.

Ces segments sont définis à partir des maxima spectraux contenus dans les "tranches de spectre (plan fréquence - amplitude) de début et de fin de chaque segment ; ces points caractéristiques sont reliés par interpolation linéaire. (exemple figure 1).

(*) Vocoder du C.N.E.T.

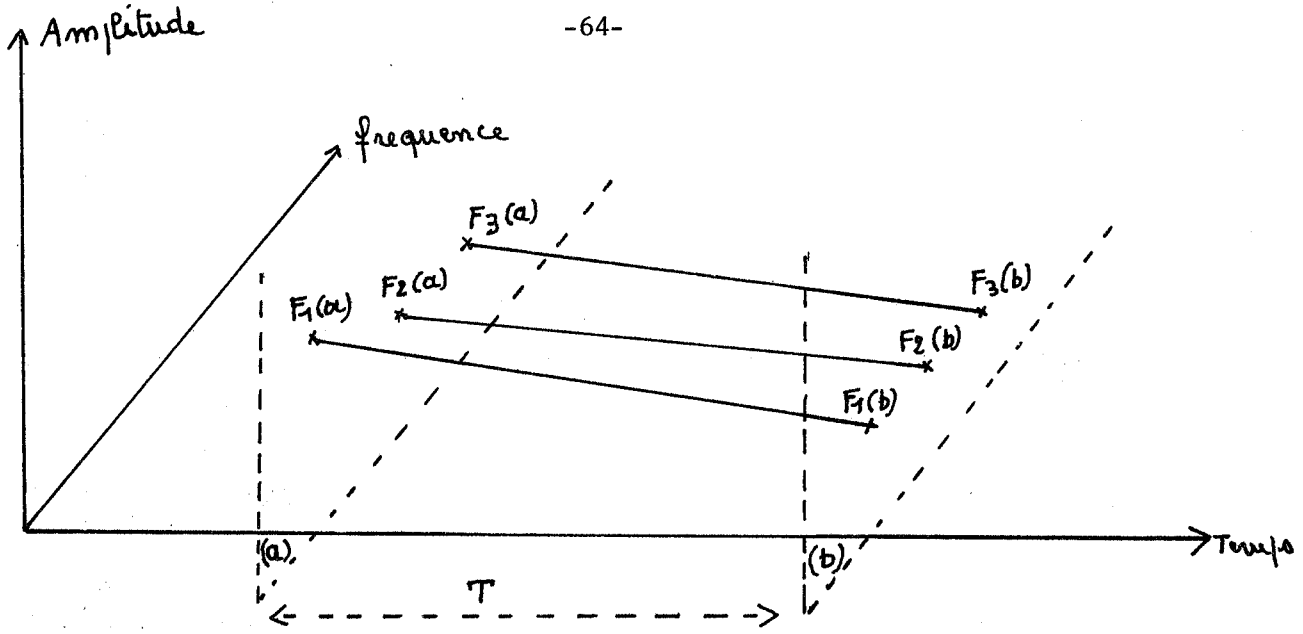


figure 1

Reconstitution des trajectoires formantiques dans un segment temporel de durée T.

(a) plan contenant les maxima spectraux du début du segment $F_1(a)$, $F_2(a)$, $F_3(a)$.

(b) plan contenant les maxima spectraux de fin de segment $F_1(b)$, $F_2(b)$, $F_3(b)$.

Ces points caractéristiques sont obtenus à partir de paramètres associés à chaque phonème.

Ces paramètres sont :

- les fréquences des formants,
- l'amplitude relative de chaque formant,
- une série d'informations permettant de reconstituer jusqu'à dix segments de spectre ; ce qui nous a semblé suffisant pour un phonème.

Elle comporte :

- une information de durée par segment,
- une information de voisement par segment,
- une information d'énergie par segment,
- une information de variation des formants par segment,
- un coefficient permettant d'appliquer une affinité sur la courbe formantique par rapport à l'axe des fréquences.

REGLES DE COARTICULATION

Nous supposons que :

- lors de la production d'une voyelle, la position d'articulation idéale pour cette voyelle est toujours atteinte.

- lors de la production d'une consonne, le mouvement articulatoire est fortement influencé par les deux points d'articulation stables qui l'encadrent, c'est-à-dire la voyelle précédente et suivante. Cette influence s'étend sur tous les paramètres de la consonne.

- variation formantique, amplitude et fréquence pour chacun d'eux, en début et en fin de la production de la consonne : ΔA_n, ΔF_n.

- variation de durée : Δt

- variation d'énergie : ΔE.

A partir des paramètres, (ΔA_n, ΔF_n, Δt, ΔE), associés à chaque combinaison vc ou cv, le système apporte les modifications nécessaires aux paramètres des règles phonémiques de la consonne concernée.

Cela nous permet de générer toutes les combinaisons possibles vcvv.., sauf la combinaison, peu utilisée en français, qui est cc ; pour ce cas particulier, nous avons résolu le problème en insérant entre les deux consonnes adjacentes une voyelle neutre très brève, permettant de définir un point d'articulation entre celles-ci.

REGLES DE TRANSITION

A chaque combinaison de deux phonèmes, est associé un ensemble de paramètres définissant les trajectoires formantiques au cours de la transition, la durée de la transition, ainsi que les altérations de durée, d'intensité et de pitch (microprosodie), apportées par l'adjacence de ces deux phonèmes. (CASTAN S., LATIL J.Y. 5.5.1979). Une interpolation linéaire est alors effectuée entre les points caractéristiques du début et de la fin de la transition.

REGLES DE PROSODIE (EMERARD F. 1977, CAELEN G. 1979)

L'information prosodique est déterminée à partir de courbes prosodiques (de mélodie, de rythme et d'intensité).

Elles sont recadrées, par affinités, sur le spectre à partir de marqueurs insérés dans la chaîne phonémique de commande - ceci au niveau des mots, des groupes de mots et de la phrase.

DETERMINATION DU SPECTRE

L'ensemble de ces règles nous permet de générer le "squelette" du spectre.

On applique un filtrage du 1er ordre sur le "squelette" pour réduire les discontinuités dues à la juxtaposition des segments.

Le spectre est alors reconstitué à partir du "squelette" du spectre et d'une règle formantique.

REGLE FORMANTIQUE

Le vocoder à canaux est ici utilisé comme un synthétiseur formantique parallèle.

Le spectre fréquentiel qui sera utilisé pour générer les commandes des canaux, est reconstitué à partir d'une courbe formantique "type" ; l'allure de cette courbe dérive de la courbe de réponse fréquentielle d'un filtre du 2ème ordre.

En appliquant des anamorphoses successives (en amplitude et en fréquence) sur la courbe formantique "type", on fait coïncider le sommet de celle-ci avec chaque point du "squelette" dans le plan fréquence-amplitude.

En sommant ces différentes courbes ainsi obtenues, pour chaque échantillon de temps, on en déduit la courbe spectrale définitive.

Il a été remarqué que cette méthode donne en partie des résultats similaires à ceux obtenus à partir des règles utilisées par KLATT dans sa simu-

lation de synthétiseur à formants série par un synthétiseur parallèle.

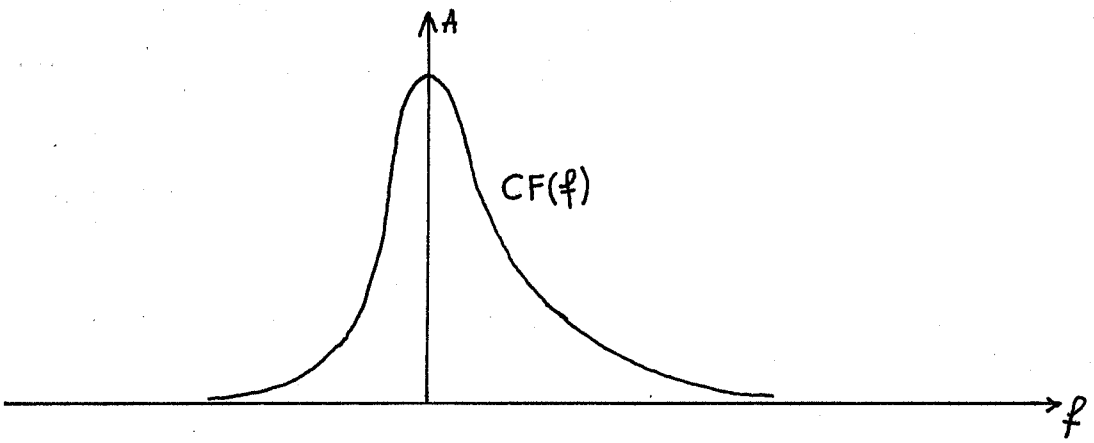
Notamment, dans les faits suivants :

- la variation de la largeur de bande passante des formants est fonction inverse de l'amplitude de ceux-ci.

- le changement de fréquence d'un formant affecte les amplitudes des formants immédiatement supérieurs.

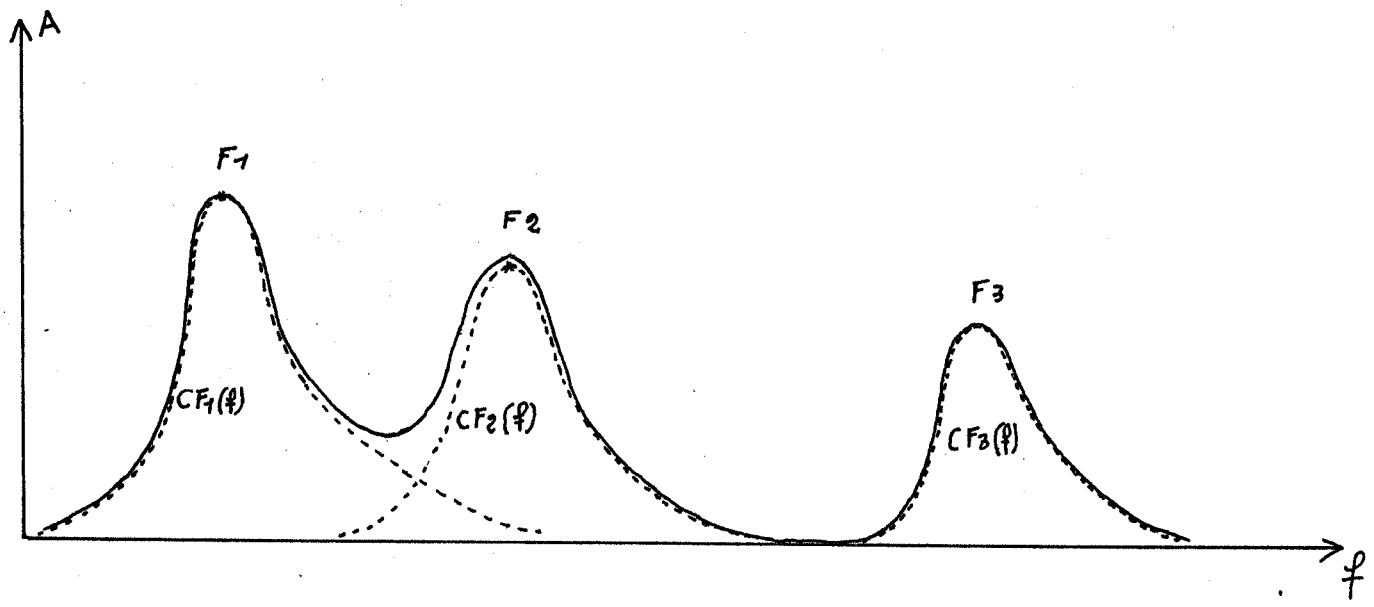
- lorsque deux formants sont proches, leurs amplitudes réciproques sont augmentées.

Soit la courbe formantique $C_F(f)$.



Si à l'instant t, le squelette du spectre comporte F_1, F_2, F_3 le système déduit la courbe spectrale :

$$C(f) = C_{F_1}(f) + C_{F_2}(f) + C_{F_3}(f).$$



MISE EN OEUVRE

Le spectre est généré pour un vocoder virtuel de 40 canaux, couvrant une bande de fréquence de 200 à 4100 Hz, par pas de 100 Hz. Cette définition a été choisie pour permettre une bonne précision du spectre.

Les commandes des quatorze canaux du vocoder se font par prélèvement des informations sur ce spectre dont les fréquences correspondent aux canaux réels.

Pour déterminer les tables de paramètres, nous avons réalisé des outils d'analyse, notamment un banc de filtres dont le spectre résultat a le même format que le spectre synthétique fourni par le système.

CONCLUSION

Le système est implémenté sur un T 1600 télémécanique.

Tous les paramètres peuvent être visualités et modifiés d'une manière interactive par l'utilisateur, constituant ainsi un système d'étude et de synthèse qui fournit une parole synthétique de meilleure qualité que le système précédent.

REFERENCES

CAELEN, J. , MAURANG, G. , 1976, Fréquence intensité et durée : étude comparative des fonctions dans les phrases énonciatives simples et étendues; 7èmes journées du groupe "Communication parlée" NANCY.

CAELEN, J., 1979, "Un modèle d'oreille, analyse de la parole continue, reconnaissance phonémique" ; -Thèse d'Etat, Université TOULOUSE.

CAELEN, J., 1978, Structure prosodique de la phrase énonciative simple et étendue ; -Thèse de 3ème cycle, Université TOULOUSE.

CARRE, R., 1971, Contribution aux études sur l'analyse de la parole, rôle et importance des formants ; -Thèse d'Etat, Université GRENOBLE.

CASTAN, S. , LATIL, J.Y. , 1979 Synthèse par règles du français ; -10èmes journées d'étude sur la parole ; GRENOBLE.

C.N.E.T. , LANNION, 1975, 1976, recherche/acoustique.

COKER, C.H., 1967, Syntheses by rule from articulatory parameters ; -I.E.E.E. conf. on speech communication and processing, CAMBRIDGE, Mass.

DELATTRE, P., 1966, Studies in french and comparative phonetics ; LONDON, The HAGUE, PARIS : Manton and Co.

DELATTRE, P., 1968, From acoustic aves to distinctive features ; -phonetica 18.

DICRISTO, A. , CHAFCAULOFF, M. , 1977, Les faits micro-prosodiques du français : voyelles, consonnes, coarticulation ; -8mes journées du groupe "Communication parlée" AIX EN PROVENCE.

DOURS, D. , FACCA, R. , PERENNOU, G. , 1974 Analyse temporelle du signal de parole comparée à l'analyse fréquentielle du point de vue de la reconnaissance ; -5èmes journées du groupe "Communication parlée" ORSAY.

ENERARD, F., 1977, Synthèse par diphones et traitement de la prosodie ; -Thèse de Docteur de 3ème cycle, GRENOBLE.

KLATT, D., 1980, Software for a cascade/parallel formant synthesizer ; JASA, Vol. 67, n° 3.

LIBERMAN, A.M. , INGEMANN, F. , LISKER, L. , DELATTRE, P. , COOPER, F.S. , 1959 Minimal rules for synthesizing speech ; JASA 31.

LIENARD , J.S., 1972, Analyse synthèse et reconnaissance automatique de la parole ; -Thèse d'Etat, PARIS VI.

LIENARD, J.S., 1977, Les processus de la communication parlée ; -Edition MASSON.

RODET, M.X., 1974, Synthèse de la parole ; -5èmes journées du groupe "Communication parlée ORSAY.

RODET, M.X., 1977, Analyse du signal dans sa représentation amplitude temps, synthèse de la parole par règles ; -Thèse d'Etat, PARIS VI.