Un système de détection du voisement et du F0 (fréquence fondamentale)

Robert Espesser

...

...

UN SYSTÈME DE DÉTECTION DU VOISEMENT ET DE Fo

Résumé

On présente un système de détection du voisement et de Fo, construit sur deux méthodes présentées antérieurement (9, 16). Le voisement est d'abord détecté par classification automatique avec distance adaptative, puis le Fo est calculé sur les seules séquences voisées, par intercorrélation du spectre de puissance et d'une fonction peigne.

A SYSTEM FOR VOICED-UNVOICED DECISION AND Fo MEASUREMENT

Abstract

A system for voiced-unvoiced (V/NV) decision and Fo measurement is described. It is based on two methods previously presented (9, 16). First, V/NV detection is carried out by a clustering procedure, using adaptative distances; then Fo is computed on the voiced part of the signal; this method is based on the cross correlation between the power spectrum and a comb function.

This system is implemented on a SEMS T 1600 mini computer.

...

...

UN SYSTEME DE DETECTION DU VOISEMENT ET DE Fo

Robert ESPESSER
Institut de Phonétique d'Aix, L.A. 261

INTRODUCTION

On décrit le système de détection du voisement et de Fo utilisé à l'Institut depuis 1981, construit sur deux méthodes antérieurement exposées (9, 16); c'est un système à deux passes : d'abord, détection du voisement par classification automatique, puis, sur les seules séquences voisées, détection du Fo par la méthode du "peigne".

L'ensemble est implanté sur un mini-calculateur SEMS T 1600.

Les détections du voisement et du Fo utilisent un signal de parole filtré passe-bande à 80-5000 Hz, échantillonné à 10 kHz avec une résolution de 12 bits.

1. - DETECTION DU VOISEMENT

Les techniques de reconnaissance de forme ont été utilisées ces dernières années pour la détection du voisement (1, 3, 4, 14, 20, 22, 24, 25). A la différence de la plupart de ces méthodes, celle décrite ici ne nécessite pas d'apprentissage et/ou ne fait pas référence à des données préalablement obtenues par celui-ci, évitant ainsi les difficultés dues à des conditions d'enregistrement ou à un locuteur différant trop de ceux présents lors de l'apprentissage (1, 22).

1.1. - Calcul des paramètres

Toutes les 10 ms, on calcule :

- l'énergie RMS sur une fenêtre de 10 ms, élevée à la puissance 0.05,
- le taux de passage par zéro (TPZ) sur une fenêtre de 10 ms, élevé à la puissance 0.35,
- l'erreur de prédiction linéaire normalisée (EPN) (15), élevée à la puissance 0.0125; (méthode d'autocorrélation sur une fenêtre de 20 ms, avec 12 poles, sans préemphase).

Ces puissances ont été déterminées en utilisant des phrases et des locuteurs différents de ceux employés pour le test d'évaluation de la méthode.

Les triplets ainsi obtenus sont considérés comme des vecteurs d'un espace à trois dimensions.

Pour des raisons d'efficacité de calcul et de simplicité, nous n'utilisons que ces trois paramètres :

- comme mentionné dans (22), augmenter le nombre de paramètres n'accroît pas toujours la discrimination entre classes;
- parmi les paramètres usuels pour la détection voisé - non voisé, ces trois paramètres sont ceux qui sont le moins corrélés entre eux (1); cette situation est préférable en classification (26). En particulier, le coefficient d'autocorrélation à retard unitaire n'a pas été retenu, car il est équivalent, en première approximation, au TPZ (voir 9).

1.2. - Classification

1.2.1. - Algorithme

Il s'agit de classer en deux groupes, voisé d'une part, non voisé et silence d'autre part, l'ensemble des vecteurs précédemment obtenus.

La procédure utilisée relève de la "méthode des nuées dynamiques"; nous employons la technique dite des "distances adaptatives"; il s'agit de la méthode des "centres mobiles" (FORGY, E., décrite dans 12) combinée avec l'emploi de la distance de Mahalanobis pour l'affectation d'un vecteur à une classe (pour une présentation complète de ces techniques, voir 5, 6, 7).

Dans la suite, E désigne l'ensemble des vecteurs à classer; d(x,y) la distance entre deux vecteurs x et y de E, selon la métrique alors définie sur E.

Le déroulement de l'algorithme est le suivant :

a) Initialisation

Deux éléments de E, g_1 et g_2 sont choisis au hasard, sous une contrainte que nous verrons plus tard; ils agiront comme centres des deux classes. Au départ, la distance d est la distance euclidienne.

b) Partitionnement itératif

Chaque vecteur de E est affecté au centre le plus proche au sens de la distance d en cours d'utilisation; les deux classes obtenues P_1 et P_2 sont donc définies par :

$$P_k = \{ x \in E \ / \ d(x,g_k) \leqslant d(x,g_l), \quad l \neq k, \quad k \in \{1,2\}, \quad l \in \{1,2\} \}$$

le critère à minimiser, mesure de la "qualité" de la participation, est défini par :

$$W = \sum_{k=1}^{2} \sum_{x \in P_k} d(x,g_k) \quad \text{où } g_k \text{ centre de la classe k}$$

ce critère représente la somme des inerties de chaque classe par rapport à son centre g_k au sens de la métrique d.

Les centres de gravité g_1 de P_1 et g_2 de P_2 sont alors calculés, ils agiront comme nouveaux centres de classe.

Les matrices de covariance Q_1 de P_1 et Q_2 de P_2 sont calculées, et la métrique désormais utilisée est la distance de Mahalanobis, définie par :

$$d(x,g) = (x - g)^t \, M \, (x - g) \qquad x,g \in E$$

où $M = |Q|^{1/3} . Q^{-1}$ avec

$$Q = Q_1 + Q_2 \qquad \text{et} \qquad Q_k = \sum_{x \in P_k} (x - g_k) \, (x - g_k)^t$$

On reprend alors la phase d'affectation en b), et ce jusqu'à obtenir un état stable (partition stable).

c) Contrainte finale

Si les deux centres de gravité finals vérifient aussi la contrainte men-

tionnée en a), le classement est retenu.

La procédure totale (a, b, c) est répétée de 5 à 10 fois, et la meilleure partition au sens du critère W (c'est-à-dire ayant le W le plus faible) est retenue.

On voit donc qu'au cours de la phase b) la distance est adaptée à chaque itération; la présence d'une inversion de matrice justifie l'utilisation d'un petit nombre de paramètres pour limiter le temps de calcul.

La matrice Q a toujours été inversible, et il n'a donc pas été nécessaire d'utiliser la procédure décrite dans (5) pour Q singulière.

Grâce à l'emploi de la distance de Mahalanobis, le classement est invariant pour toute transformation linéaire régulière opérée sur E.

1.2.2. - Contrainte "acoustique"

Les centres initiaux de les centres de gravité finals doivent vérifier la contrainte suivante :

le centre de la classe voisé a :

- une énergie RMS $>$ énergie RMS moyenne,
- un TPZ $<$ TPZ moyen,
- une EPN $<$ EPN moyenne.

Le centre de la classe non voisé - silence doit vérifier les conditions opposées.

Les centre de gravité finals doivent de plus satisfaire :

$$|RMS(g_v) - RMS(g_{nv})| > 0.05\ \sigma_{RMS}$$
$$|TPZ(g_v) - TPZ(g_{nv})| > 0.05\ \sigma_{TPZ}$$

où g_v est le centre de gravité de la classe voisé,
et g_{nv} le centre de gravité de la classe non voisé - silence.

Moyennes et écarts-types sont calculés sur la totalité de E.

Cette contrainte a été introduite pour guider l'algorithme et éviter ainsi certains classements "acoustiquement" aberrants rencontrés seulement pour des

durées d'analyse brèves (inférieures à 1 s.) : en l'absence de contrainte, nous avons obtenu seulement 7 classements de ce type sur 60, ces classements ne la vérifiant pas ou rarement, à la différence des classements "corrects" (moins de 20 % d'erreurs) qui vérifient toujours cette contrainte, quelle que soit la durée d'analyse. Compte tenu de cette remarque, il n'est pas surprenant que la contrainte ait aussi pour effet d'accélérer la convergence de l'algorithme.

Elle sert également à identifier les classes.

1.3. - Résultats, discussion

Le corpus de test consistait en deux phrases d'environ 3.3 s. chacune :

"la pipe de Jean s'est cassée en tombant de ta gabardine"
"la fille de Charles Sablon a voulu un petit chien en guise de cadeau"

Ces deux phrases présentent l'ensemble des consonnes bruissantes du français; chacune d'elles a été prononcée par 5 femmes et 5 hommes, soit au total 20 phrases. Les zones voisées / non voisées ont été segmentées manuellement par observation visuelle et parfois auditive, à l'aide d'un éditeur de signal (8), pour constituer une référence.

Afin d'étudier la méthode sur des durées différentes, l'analyse a été menée, pour chacun des 20 phrases, sur des durées de 250, 500, 1000, 1500, 2000 ms. et sur la phrase entière.

Dans la suite, on note voisé par V, non voisé - silence par NVS.

L'erreur globale est calculée selon deux méthodes :

- erreur E1 : nombre de séquences mal classées / nombre total de séquences. E1 favorise la détection des séquences V.

- erreur E2 : les taux d'erreur de détection, V d'une part, NVS d'autre part, sont pondérés par la quantité d'information relative à chaque classe, respectivement (voir 10) :

Si TV est le taux moyen de V, TNVS le taux moyen de NVS (avec TNVS = 1 - TV), alors :

$QV = - TV \log_2 TV$ quantité d'information de la classe V

$QNVS = - TNVS \log_2 TNVS$ quantité d'information de la classe NVS

Avec TEV : taux d'erreur sur le V (nombre de séquences V mal classées / nombre total de séquences V)

TENVS : taux d'erreur sur le NVS (nombre de séq. NVS mal classées / nombre total de séquences NVS).

On a :

$$E2 = (QV \cdot TEV + QNVS \cdot TENVS) / (QV + QNVS)$$

E2 favorise la détection des séquences NVS, porteuses de plus d'information que les séquences V. Remarquons que pour des durées supérieures à 1 s., le TV de notre corpus est de 70 %, valeur proche de celui mentionné dans (10).

Ces erreurs ont été calculées pour 2 rapports voisé/silence (S/B) :

Fig. 1 : S/B = 32 dB, enregistrement original en chambre sourde.

E1 en % courbe o
E2 en % courbe Δ

Fig. 2 : S/B = 20 dB, obtenu par adjonction de bruit blanc à l'enregistrement original

E1 en % courbe o
E2 en % courbe Δ

(Voir également en annexe).

Au-delà de 1 s., les résultats placent favorablement notre méthode par comparaison à celles évaluées (il s'agit de l'erreur E1) dans (21) dont les taux varient entre 4 % et 12 %.

Pour les durées inférieures à 0.5 s., c'est surtout la dispersion du taux d'erreur qui augmente (voir 9); cette durée paraît être une limite inférieure pour obtenir des résultats très précis.

Plusieurs points restent à développer :

- optimisation des puissances utilisées pour le calcul des paramètres;
- étude du comportement de la méthode pour des taux de voisement proches de 100 %; elle est évidemment inopérante sur des séquences totalement voisées; dans ce cas il paraît possible de calculer un critère mesurant la facilité

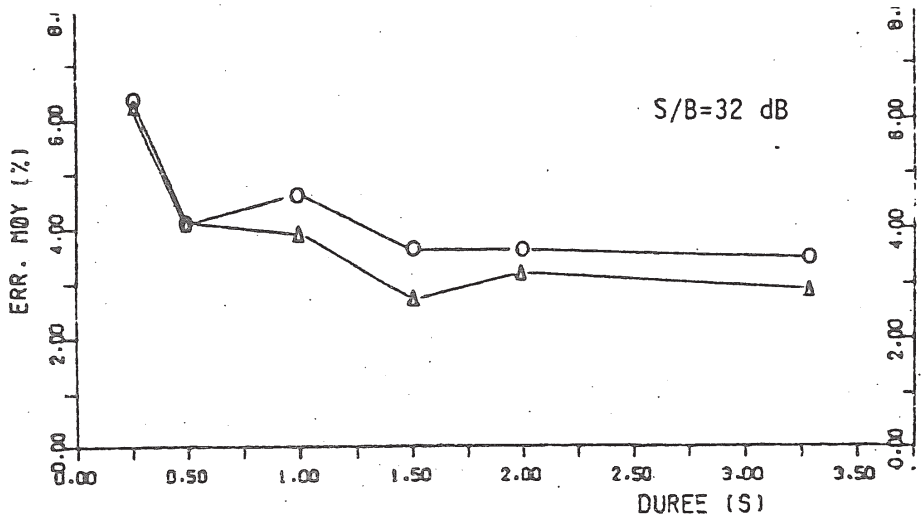


fig 1

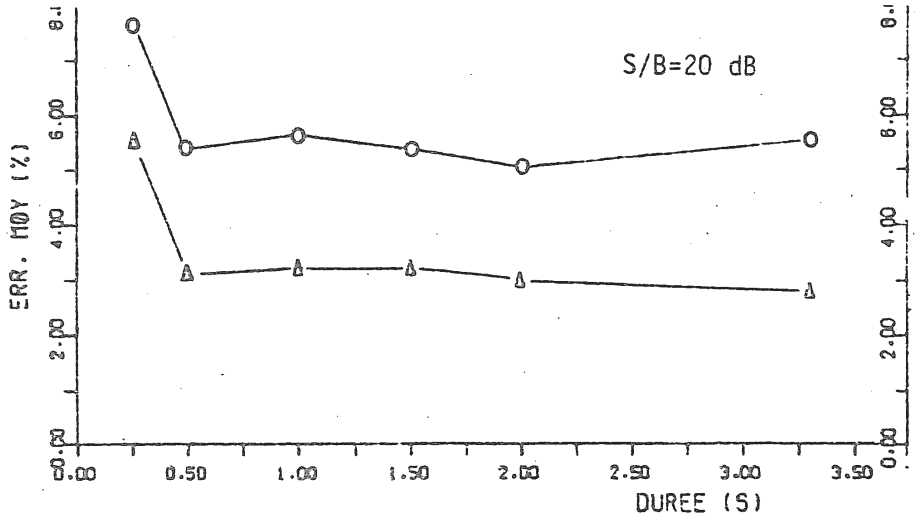


fig 2

Erreur moyenne de classement V/NV

erreur E1: O

erreur E2: Δ

avec laquelle la classification sous contrainte s'est effectuée : une classe NVS assez importante (selon l'algorithme) et un petit nombre de classements satisfaisant la contrainte acoustique (voir 1.2.2.) sont contradictoires, et indiquent un classement "acoustiquement" aberrant ou impossible (cf. remarque faite au 1.2.2.); les premiers essais ont donné des résultats prometteurs.
- test de la méthode sur un corpus phonétiquement équilibré (2, 13).

2. - DETECTION DU Fo

Il s'agit d'une nouvelle approche du principe mis en oeuvre dans les techniques "somme harmonique" ou "produit harmonique" (17, 18, 23) : dans ce dernier, on calcule pour chaque fréquence f la fonction s(f) définie par :

$$s(f) = \sum_{k=1}^{K} \log |S(kf)|^2 \qquad \text{avec } |S(f)|^2 \text{ spectre de puissance}$$

K : constante, ≈ 5

s(f) présentera un maximum lorsque les kf correspondront aux harmoniques du Fo, donc pour f = fo.

Ces méthodes ont tendance à détecter des harmoniques du fo :

Soit une fonction parfaitement périodique de fondamental p; s(f) aura un maximum pour f = p (contribution de p, 2p, 3p, 4p, 5p) et un maximum pour f = 2p (contribution de 2p, 4p, 6p, 8p, 10p), etc.

2.1. - Méthode du "peigne"

Pour chaque fréquence f, on intercorrèle une fonction du spectre de puissance et une fonction peigne :

$$C_f(\tau) = \int_{fi}^{fs} T(|S(\nu)|^2)\ P_f(\nu - \tau)\, d\nu \qquad (1)$$

fi : limite inférieure de fréquence
fs : limite supérieure de fréquence
$|S(\nu)|^2$: spectre de puissance
T : fonction appliquée à $|S(\nu)|^2$

$P_f(v)$: fonction peigne
= 1 pour $v = kf$, avec $k \in \{1,2,\ldots n\}$
= 0 sinon.

Dans notre implémentation, on ne calcule que $C_f(0)$ (intercorrelation à retard nul) noté désormais C(f).

Alors, pour T = I fonction identité, on retrouve la somme harmonique,
pour T = log on retrouve le produit harmonique.

C(f) est calculé pour $v \in [fi,fs]$; par principe on n'a plus de problème d'harmoniques, mais un problème de sous-harmoniques : en reprenant l'exemple d'une fonction parfaitement périodique à p Hz on a :

$$C(p/2) \geqslant C(p) \geqslant C(2p)$$

(C(p/2) ayant plus de termes que C(p), etc...)

Pour éviter cela on procède à :

- un "nettoyage" des $T(|S|^2)$; ce qui a pour effet de tendre vers la relation $C(p/2) = C(p) > C(2p) > \ldots$ etc.

  Les $|S|^2$ sont calculés par FFT à partir du signal de parole mentionné précédemment, filtré numériquement passe-bas à 1.8 kHZ, déséchantillonné à 4 kHz, et multiplié par une fenêtre de Kaiser-Bessel de 128 points (soit 32 ms.).

  Chaque spectre est "nettoyé" par :
  - un seuillage : les composantes inférieures de plus de 70 dB au maximum de $|S|^2$ sont annulées; (il s'agit d'amplitude au carré);
  - les pics du spectre sont interpolés paraboliquement (pour obtenir une "résolution" de 1 Hz), et on ne garde qu'une bande de $\pm$ 16 Hz autour de chacun d'eux.

- une pondération de la fonction peigne par une fonction décroissante, pénalisant ainsi les C(f) de f faible.

Soient SI(f) les spectres nettoyés; on a pris pour T la fonction identité; sous forme discrète et après simplification, (1) s'écrit :

$$C(f) = \sum_{k=1}^{nf} SI(kf)/k^{1/8}$$

avec $60 \leqslant f \leqslant 1000$, de Hz en Hz

nf : nombre maximum d'harmoniques de f dans la bande $[60,2000]$

$1/k^{1/8}$: pondération de la fonction peigne.

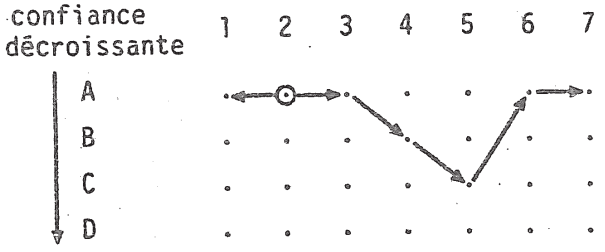
Théoriquement, le Fo correspond au C(f) maximal.

2.2. - Suivi de Fo

Aux corrections classiques par lissage (ex. : median smoothing, 19) ou interpolation (voir suivi de Fo dans la méthode SIFT, 15), il a été préféré un choix parmi plusieurs candidats possibles pour chaque fenêtre.

On garde comme candidats les 4 plus grands C(f), classés par amplitude décroissante, ordre correspondant à un degré de confiance décroissante; les erreurs sont en général des erreurs d'harmoniques, et la "bonne" valeur est pratiquement toujours parmi les 4 plus grands C(f) .

Sur une fenêtre de 7 mesures consécutives (à l'intérieur d'une même séquence voisée), soit 70 ms., on détermine la médiane des 7 premiers candidats; elle est prise comme point d'ancrage, à partir duquel on progresse en retenant le candidat le plus proche en valeur absolue; celui-ci est à son tour pris comme point d'ancrage, etc.; on a donc un cheminement dans un treillis de 4 x 7 candidats; exemple :



⊙ : médiane (point (A,2)) des 7 premiers candidats ((A,1) à (A,7))

ex. : (C,5) est le plus proche en valeur absolue de (B,4).

Les fenêtres de 70 ms sont indépendantes, ceci pour éviter des propagations d'erreur; 70 ms parait être une durée sur laquelle on peut s'attendre à une certaine "continuité" du Fo.

2.3. - Résultats et discussion

2.3.1. - Taux d'erreur

Des résultats plus détaillés sont donnés dans 16; sur une phrase de 3.1 sec dite par 3 hommes et 3 femmes (enregistrement de type "studio") le taux d'erreur harmonique est de 1.7 %; il est d'environ 12 % sur de la parole filtrée passe-haut (téléphone, télévision).

En l'absence de données quantitatives sur le suivi de Fo, mentionnons seulement qu'il s'est révélé efficace (en particulier pour les erreurs d'harmoniques) et "discret", évitant des corrections ou lissages excessifs; le suivi est inopérant (et donne des résultats fantaisistes) si :

- la valeur d'ancrage est fausse, ce qui implique que 4 premiers candidats sur 7 soient faux, au minimum

  et/ou

- la valeur correcte du Fo n'est pas parmi les 4 plus grands C(f)

Ceci correspond dans les deux cas à une très mauvaise détection initiale : étant donné l'utilisation faite de ce système (recherche), ceci est "sain" : au-dessus d'un certain taux d'erreur initial, tout suivi est dangereux ou illusoire.

2.3.2. - Dynamique du système

De même que les méthodes classiques travaillant sur une fenêtre (SIFT, cepstre, autocorrélation, etc.), la méthode du "peigne" offre une certaine inertie à une variation rapide du Fo. De plus, le suivi de Fo a également une tendance "conservatrice"; pour évaluer ceci, nous avons synthétisé (LPC) une voyelle neutre /ə/ avec un Fo variant de 100 Hz en 150 ms (selon une courbe spline, voir 11). La fig. 3 montre que le système suit très bien : la courbe réelle et la courbe détectée se superposent presque totalement. A cet égard, l'emploi d'une fenêtre de 32 ms seulement est un élément favorable (contre 40 ms pour SIFT, 51 ms pour le cepstre, dans l'implémentation "étalon" de RABINER et al., voir 21). Le système ne suit pas des variations plus rapides, qui existent cependant dans la parole naturelle (jusqu'à 145 Hz en 70 ms).

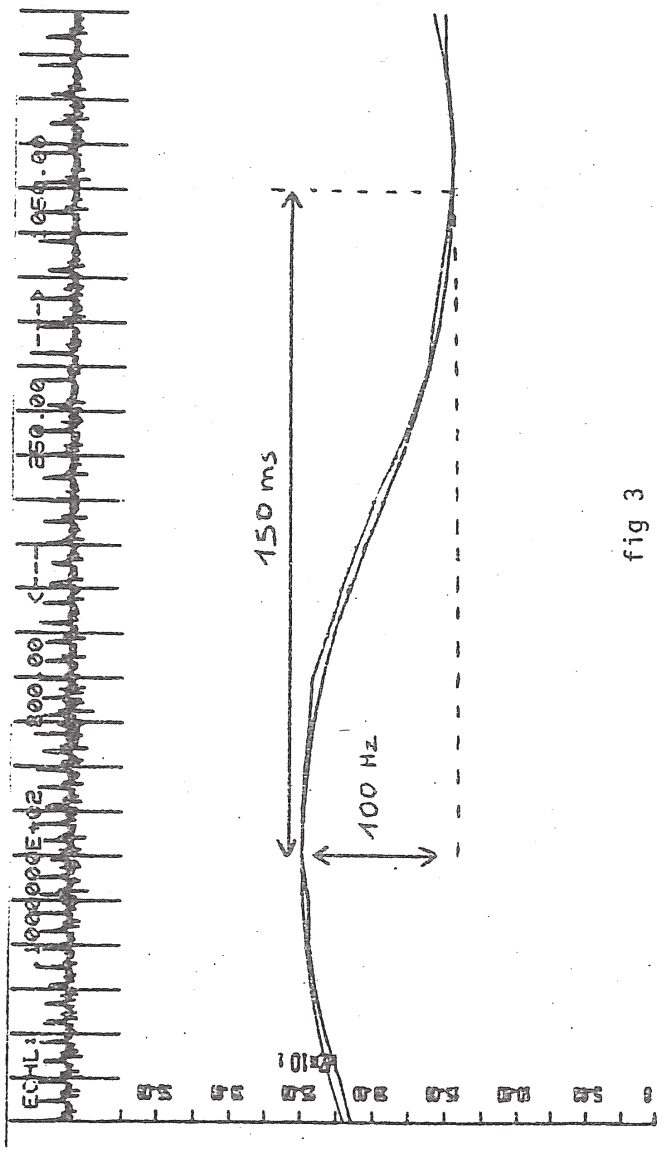


fig 3

test de dynamique de la détection du F_o

2.3.3. - Développement

Il serait utile de

- Faire des essais systématiques pour déterminer une fonction T (voir 2.1.) optimale, en liaison avec la fonction pondération. La fonction T utilisée (identité, donc en travaillant avec le carré du module du spectre) fait chuter peut-être trop vite l'amplitude des harmoniques, risquant ainsi de perdre des coïncidences pour le calcul des C(f); prendre à l'inverse une fonction log. risque de trop amplifier les composantes inter-harmoniques.
- Tester la méthode sur un corpus de voix chantée, vu le très bon résultat (mais trop partiel) mentionné dans 16.

3. - CONCLUSION

Ces deux méthodes, mises en oeuvre conjointement, constituent un système qui s'est révélé très satisfaisant quant à la qualité des résultats, moins agréable quant au temps de calcul (environ 200 fois le temps réel : comparable au cepstre). Ce logiciel est interfacé à un éditeur de signal (8) (voir fig. 4, un exemple de résultat tel qu'il apparaît à l'utilisateur : en haut, oscillogramme de la phrase - en italien - "le chat est en train de boire", en bas V/NV et Fo détectés) et un vocodeur LPC, l'ensemble constituant un outil essentiel pour l'étude du Fo et sa modélisation.

ANNEXE

		S/B = 32 dB		S/B = 20 dB	
durée (ms)	TV (%)	E1 (%)	E2 (%)	E1 (%)	E2 (%)
250	61	6.4	6.2	7.6	5.59
500	70	4.1	4.08	5.4	3.12
1000	72	4.65	3.85	5.65	3.30
1500	71	3.6	2.65	5.4	3.29
2000	72	3.6	3.15	5.1	3.1
Totalité (6721 décisions)	77	3.38	2.77	5.43	2.84

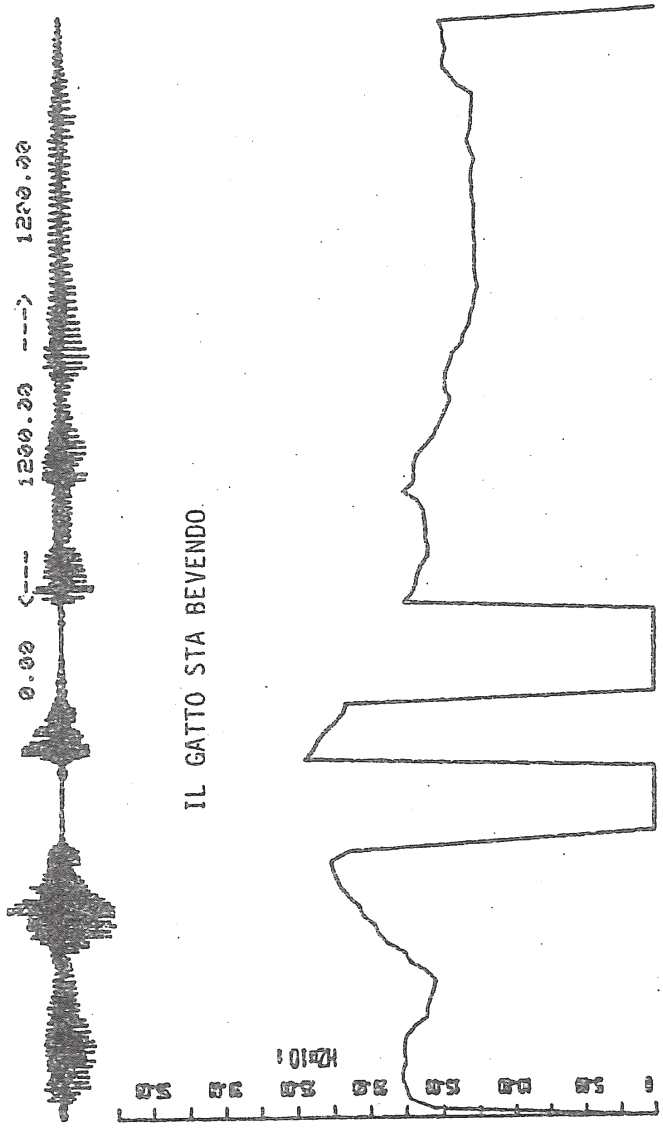


fig 4

exemple de détection du V/NV et du F_o

REFERENCES

(1) B.S. ATAL & L.R. RABINER, "A pattern recognition approach of voiced-unvoiced-silence classification with applications to speech recognition", IEEE Trans. ASSP-24, pp. 201-212, 1976.

(2) P. COMBESCURE, "20 listes de dix phrases phonétiquement équilibrées", Revue d'Acoustique, 56, pp. 34-38, 1981.

(3) B.V. COX & L. TIMOTHY, "Non-parametric rank-order statistics applied to robust voiced-unvoiced-silence classification", IEEE Trans. ASSP-28, pp. 550-561, 1980.

(4) F. DAABOUL & J.P. ADOUL, "Parametric segmentation of speech into voiced-unvoiced-silence intervals", Proc. IEEE ICASSP, pp. 327-331, Hartford, CT, May 1977.

(5) E. DIDAY & coll., Optimisation en classification automatique, INRIA ed. 1979.

(6) E. DIDAY & G. GOVAERT, "Classification avec distances adaptatives", C.R. Acad. Sc. Paris, 278, série A, pp. 993, 1974.

(7) E. DIDAY & J.C. SIMON, "Clustering analysis", in Digital Pattern Recognition, K.S. FU ed., Springer, Berlin, 1976.

(8) R. ESPESSER, "Un éditeur de signal", Tr. Inst. Phon. Aix-en-Provence, vol. 7, pp. 123-133, 1980.

(9) R. ESPESSER, "Détection de voisement par classification automatique", Symposium prosodie Toronto, 29-30 mai 1981, pp. 104-112. Résumé des 12e Journées d'étude sur la Parole, Montréal, 25-27 mai 1981 (séances affichées), pp. 23-24.

(10) P. GARCIN, J.F. SERIGNAT, "Seuils optimaux en détection de voisement, une nouvelle de détermination", Revue d'acoustique, 15, pp. 138-143, 1982.

(11) D.J. HIRST, "Un modèle de production de l'intonation", TIPA, vol. 7, pp. 297-315, 1980.

(12) G.N. LANCE & W.T. WILLIAMS, "A general theory of classificatory sorting strategies, II. Clustering systems", Computer J., 10, pp. 271-277, 1967.

(13) M. LENNIG, "3 listes de 10 phrases françaises phonétiquement équilibrées", Revue d'Acoustique, 56, pp. 39-42, 1981.

(14) W.C. LIN & C.F. CHAN, "An isolated word recognition system based on acoustic-phonetic analysis and statistical pattern recognition", Proc. IEEE ICASSP, pp. 679-682, Hartford, CT, May 1977.

(15) J.D. MARKEL & A.H. CHAN, Linear prediction of speech, Springer, Berlin, 1976.

(16) P. MARTIN, "Extraction de la fréquence fondamentale par intercorrélation avec une fonction peigne", 12e Journées d'étude sur la Parole, Montréal, 25-27 mai 1981, pp. 221-232.

(17) A.M. NOLL, "The cepstrum and some close relatives", Proc. of the NATO Advanced study Institute on Signal Processing, pp. 11-22, Academic Press, 1973.

(18) A.M. NOLL, "Pitch determination of human speech by the harmonic product spectrum, The harmonic sum spectrum and a maximum likelihood estimate", Proc. of the symposium on Computer Processing in Communication, April 8-10 1969, vol. 19, pp. 779-797. Polytechnic Press, Brooklyn, New York.

(19) L.R. RABINER, R.W. SCHAFER, Digital processing of speech signals, Prentice Hall, 1978.

(20) L.R. RABINER & M.R. SAMBUR, "Voiced-unvoiced-silence detection using the Itakura LPC distance measure", Proc. IEEE ICASSP, pp. 323-326, Hartford, CT, May 1977.

(21) L.R. RABINER, M.J. CHENG, A.E. ROSENBERG & C.A. Mc GONEGAL, "A comparative performance study of several pitch detection algorithms", IEEE Trans. ASSP-24, pp. 399-418, 1976.

(22) V.V.S. SARMA & D. VENUGOPAL, "Studies on pattern recognition approach to voiced-unvoiced-silence classification", Proc. IEEE ICASSP, pp. 1-4, Tulsa, OK, Apr. 1978.

(23) M.R. SCHROEDER, "Period Histogram and product spectrum : new methods for fundamental frequency measurements", I.A.S.A., 43, pp. 829-834, 1968.

(24) L.J. SIEGEL & K. STEIGLITZ, "A pattern classification algorithm for the voiced-unvoiced decision", Proc. IEEE ICASSP, pp. 326-329, Philadelphia, PA, Apr. 1976.

(25) L.J. SIEGEL, "A procedure for using pattern classification techniques to obtain a voiced/unvoiced classifier", IEEE Trans. ASSP-27, pp. 83-89, 1979.

(26) P. SNEATH, R. SOKAL, Numerical taxonomy, W.M. Freeman, San Francisco, 1973.