

DETECTEUR DIGITAL DE TONES MULTIFREQUENCES

Guy Roulier
Claude Galand
Robert Vermot-Gauchy

Helene Cerf

Dpt. Traitement de Signal
Laboratoire IBM
06610 La Gaude, France

IBM France
Centre Scientifique
36 Av. Poincaré
75116 Paris, France

Nous avons mis au point et implémenté un détecteur numérique de signaux multifréquences, qui utilise un algorithme de compression fondé sur la technique de la prédiction linéaire (LPC) et se déroule concurremment. Ainsi la puissance de calcul spécifique nécessaire au récepteur est réduite à environ 0.5 MIPS avec une bonne probabilité de détection ainsi qu'un bon taux de réjection des tones simulés par la voix. Les tones sont détectés sur une base de 20ms et validés sur une base de 2 détections parmi n possibles. Des fenêtres de 20ms décalées sont de plus utilisées pour améliorer les résultats et diminuer la durée théorique minimum d'un tone à 30ms.

1.0 INTRODUCTION.

L'usage des téléphones à touches devient très répandu. Ces appareils permettent à l'utilisateur de transmettre des données (12 ou 16 symboles) à un récepteur qui peut ensuite effectuer des actions suivant la séquence des symboles reçus. Les symboles sont codés à l'aide d'une combinaison de deux sinusoides pures appartenant à deux groupes de fréquences mutuellement exclusives : (697, 770, 852, 941 Hz) and (1209, 1336, 1477, 1633 Hz).

Les détecteurs numériques commercialisés utilisent soit des bancs de filtres soit un algorithme fondé sur la Transformée de Fourier Rapide. L'approche par Prédiction Linéaire utilise le fait que les coefficients directs A_i sont des descripteurs précis du contenu spectral d'un signal stationnaire. Les signaux multifréquences faisant partie de ces signaux stationnaires, il est possible d'évaluer la transformée de Fourier pour certaines fréquences caractéristiques avec très peu de calculs. Si une paire de pics énergétiques est présente et que ces pics répondent à certains critères, alors un signal est dit détecté. Plusieurs détections et l'étude de la stationnarité du signal (à l'aide des coefficients d'autocorrélation) peuvent conduire à la validation de la détection.

Dans une première partie nous décrivons les algorithmes fondés sur la Transformée de Fourier et sur la Prédiction Linéaire. Dans un deuxième temps nous décrivons une mise en oeuvre en temps réel associée au classique VEPC (Voice Excited Predictive Coder)/1/ et nous explicitons certains points d'intérêts. Enfin nous donnons des résultats de performances obtenus lors de l'analyse d'une bande test suivant la recommandation du CCITT pour les mesures de détecteurs multifréquences/2/.

2.0 DESCRIPTION DES ALGORITHMES.

2.1 ALGORITHME FONDE SUR LA TFR.

Soit $x(n)$, $[n=0, \dots, N-1]$ le signal d'entrée correspondant à un bloc de N échantillons. Généralement N est égal à 160 pour les vocodeurs (représentant ainsi 20 ms) mais peut varier de 128 à 256 si l'on s'intéresse à la TFR.

An efficient digital multifrequency receiver was designed and implemented which uses data computed by an LPC based speech compression algorithm, and runs concurrently. This way DTMF specific processing is reduced to roughly 0.5 MIPS with a good detection probability as well as a good rejection of voice simulated digits. Multifrequencies are detected on 20ms basis and validated on a 2 detections out of n possible basis. Overlapping 20ms windows are also used to improve results and decrease the theoretical minimum tone duration to less than 30ms.

Pour $N=128$ et un échantillonnage à 8 KHz, la résolution en fréquence est d'environ 30 Hz (4000/128), ce qui est suffisant pour différencier deux fréquences appartenant au même sous-groupe.

La TFR est un algorithme permettant de calculer rapidement la Transformée de Fourier Discrète d'un signal $x(n)$ qui s'exprime comme suit :

$$(1) \quad X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-2j \cdot \pi \cdot kn/N} \quad k=0, \dots, N-1$$

La détection est effectuée en comparant les valeurs de $X(k)$ entre elles et si nécessaire avec des quantités fixes. La validation de la détection dépend de la politique choisie : n détections successives parmi m fenêtres ou n détections parmi m fenêtres.

Bien que cette détection ne demande pas le calcul de toutes les valeurs complexes $X(k)$, mais seulement pour certaines fréquences, le calcul explicite de la relation (1) représente déjà une haute complexité. En effet il y a $N \cdot \log(2)N$ multiplications complexes, plus un calcul de module pour chaque fréquence.

2.2 ALGORITHME FONDE SUR LA PREDICTION LINEAIRE.

L'analyse par prédiction linéaire de la séquence $x(n)$, $[n=0, \dots, N-1]$, donne un ensemble de coefficients $a(i)$, $[i=0, \dots, p]$, représentatif du filtre prédictif $P(Z)$ de la séquence $x(n)$:

$$(2) \quad P(Z) = \frac{1}{A(Z)}$$

$$(3) \quad A(Z) = \sum_{i=0}^p a(i) \cdot Z^{-i}$$

On peut montrer /3/ que les coefficients $a(i)$, $[i=0, \dots, p]$ permettent de calculer une bonne estimation de la densité spectrale de la séquence $x(n)$, $[n=0, \dots, N-1]$, dans le cas où le signal $x(n)$ peut être modélisé par un processus autorégressif (AR). La parole ayant cette propriété (sur des durées de 10 à 30 ms) ainsi que les signaux multifréquences, l'estimation sera possible.



$$(4) \text{TFD}(f(k)) = \sum_{n=0}^p a(n) \cdot e^{-2j \cdot \pi \cdot n \cdot f(k) / f_s} \quad k=0, \dots, p$$

TFD(f(k)) représente la Transformée de Fourier Discrète du signal x(n) pour un jeu donné de fréquences f(k) et la fréquence d'échantillonnage (f_s = 8000Hz). Les valeurs sont particulièrement précises autour des pôles du signal original x(n). L'estimation s'améliore aussi lorsque l'ordre p du prédicteur augmente.

L'ordre du prédicteur est fixé à au moins deux fois le nombre de pôles du signal x(n). Pour la parole en téléphonie cet ordre est souvent égal à 8 ou 10. Dans notre application nous avons choisi 8.

Les TFD des deux séquences x(n) et a(i) donnent une estimation précise des résonances. Cependant la séquence a(i), [i=0, ..., p] est beaucoup plus courte que la séquence x(n), [n=0, ..., N-1]. Ainsi la relation (1) peut être évaluée de façon beaucoup plus efficace.

En fait il suffit d'évaluer TFD(f(k)) dans la relation (4) pour le jeu de fréquences normalisées par le CCITT pour les signaux multifréquences ainsi que pour une ou deux fréquences de référence.

GROUPE 1	$\begin{cases} f(1) = 697 \text{ Hz} \\ f(2) = 770 \text{ Hz} \\ f(3) = 852 \text{ Hz} \\ f(4) = 941 \text{ Hz} \end{cases}$	GROUPE 2	$\begin{cases} f(5) = 1209 \text{ Hz} \\ f(6) = 1336 \text{ Hz} \\ f(7) = 1477 \text{ Hz} \\ f(8) = 1633 \text{ Hz} \end{cases}$
----------	--	----------	--

Fréquence de référence f(0) = 500Hz

Les algorithmes de détection et de validation utilisés pour la méthode fondée sur la FFT, restent valables dans cette méthode.

3.0 MISE EN OEUVRE.

La figure 1 représente l'implémentation d'un détecteur de signaux multifréquences et présuppose l'existence d'un codeur à prédiction linéaire. Ce codeur peut être soit le VEPC décrit succinctement plus loin ou n'importe quel codeur LPC (Linear Predictive Coder) ou APC (Adaptive Predictive Coder).

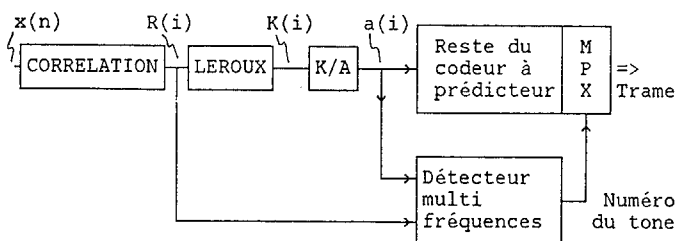


Figure 1. Implémentation de la fonction Détection de Signaux Multifréquences dans un codeur à prédicteur linéaire.

Pour chaque bloc de 160 échantillons du signal x(n), [n=0, ..., N-1], l'analyse par prédiction linéaire nous fournit un jeu de 9 coefficients d'autocorrélation R(i), [i=0, ..., 8], qui sont ensuite utilisés par l'algorithme de Leroux-Gueguen pour calculer les coefficients PARCOR K(i), [i=1, ..., 8], qui sont eux-même convertis afin d'obtenir les coefficients a(i), [i=0, ..., 8] dans leur forme directe.

A ce point nous avons suffisamment d'informations sur le signal d'entrée pour commencer les calculs spécifiques au détecteur.

1. La TFD est calculée pour les 9 fréquences en utilisant l'équation (4). Les minima de chaque groupe sont successivement comparés à une moyenne pondérée des neuf modules de la TFD puis au deuxième minimum de chaque groupe afin de vérifier l'étroitesse des raies du spectre qui doivent caractériser un tone.
2. Afin de prévenir des fausses détections dans les fenêtres de voix où le contenu spectral pourrait être identique à celui d'une fenêtre de tone, nous calculons la différence quadratique des coefficients d'autocorrélation d'une fenêtre de 10ms à la suivante. La différence est comparée à un seuil. Les résultats de la fenêtre courante et des fenêtres précédentes nous permettent éventuellement de démarrer le processus de validation et de préparer ou autoriser des calculs supplémentaires que nous appelons calculs complémentaires et dont nous parlerons davantage plus loin.
3. Connaissant la valeur du tone détecté (paire de fréquences), la stabilité des coefficients a(i) (pôle à l'intérieur ou à l'extérieur du cercle unité), la "stabilité" de l'autocorrélation (passage d'un silence à du signal, de signal à silence ou continuité) et l'historique des fenêtres précédentes, le tone est validé ou non et l'historique est mis à jour.
4. Si les calculs complémentaires sont autorisés, certaines fonctions sont de nouveau réalisées mais sur une fenêtre de 20ms contenant les 80 derniers échantillons de la précédente fenêtre normale et les 80 premiers de la fenêtre courante.
5. Si un tone a été validé, son numéro est passé à un programme qui dans l'environnement du VEPC construira une nouvelle trame.
6. Le tone reste validé aussi longtemps qu'aucune transition n'est détectée lors de l'étude des coefficients d'autocorrélation.

Nous allons maintenant détailler certains points.

4.0 CALCUL COMPLEMENTAIRES.

4.1 SPECIFICATIONS DU CCITT.

La durée du tone a été spécifiée par le CCITT :

- Le tone doit durer au moins 40ms.
- Le tone doit être suivi d'un silence d'au moins 40ms.
- La durée d'un groupe (tone + silence) ne doit pas être plus courte que 85ms.

Comme on peut voir sur la figure 2, le tone recouvre en général plusieurs fenêtres de 20ms, le recouvrement étant tout à fait aléatoire. Des simulations intensives nous ont prouvé que 10 à 12ms devraient suffire dans la majorité des cas pour détecter un tone. Dans ce cas, nous devrions donc toujours avoir au moins deux "bonnes" fenêtres pour effectuer la validation. Cependant le manque de précision dans le calcul des coefficients a(i) impose d'obtenir deux détections afin de valider un tone.

Le nombre de possibilités est donc réduit et nous avons besoin d'un moyen pour augmenter artificiellement le nombre de fenêtres de signal. Deux solutions s'offrent à nous :

1. Utiliser des fenêtres de 10ms.
2. Utiliser des fenêtres de 20ms recouvrant les fenêtres classiques.

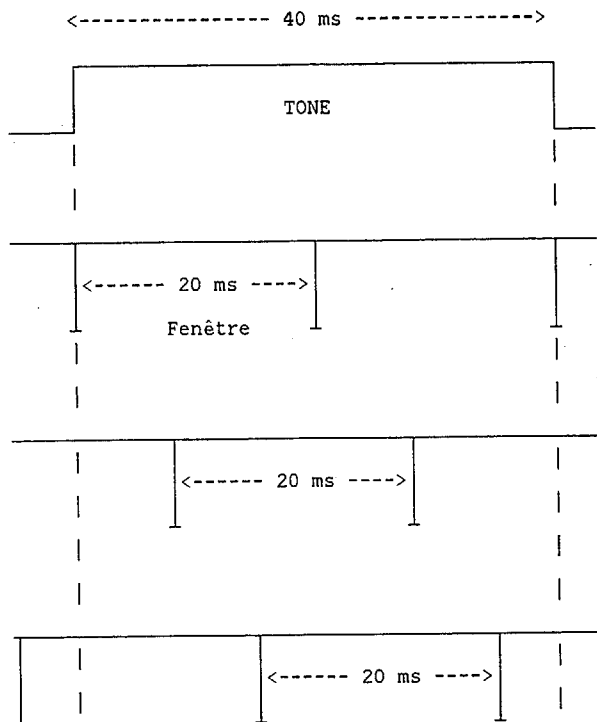


Figure 2. Exemples d'apparition d'un tone par rapport aux fenêtres de traitement.

- La stabilité des coefficients d'autocorrélation est calculée sur des fenêtres de 10ms.
- Les calculs de TFD ont lieu sur des fenêtres de 20ms.
- Le calcul complémentaire utilise des fenêtres de 20ms décalées par rapport aux fenêtres régulières.

4.5 DECISION POUR LES CALCULS COMPLEMENTAIRE:

Ce que nous appelons "stabilité de l'autocorrélation" consiste à calculer la différence quadratique entre deux jeux successifs de coefficients d'autocorrélation. Si cette différence est inférieure à un seuil donné, l'autocorrélation est dite stable, autrement elle est dite instable.

$$(5) \text{ Stab}[0-79](t) = \sum_{n=1}^7 (\text{Rn}[0-79](t) - \text{Rn}[0-79](t-1))^2$$

$$(6) \text{ Stab}[80-159](t) = \sum_{n=1}^7 (\text{Rn}[80-159](t) - \text{Rn}[80-159](t-1))^2$$

Rn[0-79] est le nième coefficient de l'autocorrélation calculé avec les 80 premiers échantillons du bloc d'entrée.

Rn[80-159] est le nième coefficient d'autocorrélation calculé avec les 80 derniers échantillons du bloc d'entrée.

Pour chaque fenêtre de 20ms nous calculons 2 fois la stabilité de l'autocorrélation, ce qui donne 2 indicateurs (un pour chaque tranche de 10ms). Les calculs des indicateurs ont lieu séparément ce qui permet des comparaisons entre le premier indicateur et le deuxième de la fenêtre précédente. Différentes actions sont entreprises suivant les valeurs des indicateurs :

1. Stab[80-159](t-1) ou Stab[0-79](t) indique une instabilité (Ils sont au-dessus du seuil). Nous pourrions être au début ou à la fin d'un tone donc l'historique et la variable contenant le numéro du tone sont réinitialisés sauf la première position. Si Stab[80-159](t-1) indique une instabilité et Stab[0-79](t) indique une stabilité, le tone a sans doute commencé lors de la fenêtre précédente. Si un tone est détecté l'historique est mis à jour autrement il est réinitialisé (uniquement la première position).
2. Stab[80-159](t-1) ou Stab[0-79](t) indique une stabilité. Nous sommes probablement dans un tone et pouvons préparer le bloc de données pour les calculs complémentaires, c'est à dire stocker les 80 derniers échantillons du bloc d'entrée.
3. Stab[0-79](t) et Stab[80-159](t-1) indiquent une stabilité. Le signal est stationnaire et les calculs complémentaires peuvent être autorisés. Le bloc de données est complété par les 80 premiers échantillons du bloc d'entrée.

Le calcul complémentaire peut avoir lieu à peu près n'importe où durant l'analyse car il est dissocié du reste de l'algorithme de compression, mais de préférence avant la construction des trames envoyées sur la ligne afin de réduire le délai entre validation et transmission.

Les calculs sont pratiquement identiques à ceux effectués sur une fenêtre normale de 20ms, les mêmes sous-programmes étant utilisés comme on peut le voir sur l'organigramme de la figure 3. Les coefficients d'autocorrélation sont calculés puis les PARCOR ainsi que les coefficients a(i) nécessaires au calcul des 9 points de la Transformée de Fourier Discrète. Le sous programme de détection est aussi appelé. Seul le calcul sur la "stabilité" des coefficients d'autocorrélation est court-circuité ainsi que la partie permettant de décider ou non d'effectuer des calculs complémentaires.

4.2 UTILISATION DE FENETRES DE 10MS.

Des simulations ont prouvé la faisabilité de ce traitement qui permet d'obtenir au moins 4 fenêtres et donc 4 décisions. Cependant, il amène une dégradation du taux de validation, essentiellement dû à la trop grande imprécision dans le calcul des PARCOR (Algorithme de LEROUX-GUEGUEN).

Cette dégradation pourrait être en partie éliminée en combinant un traitement sur 20ms.

4.3 FENETRES DE 20MS RECOUVRANT LES FENETRES USUELLES.

Une autre possibilité d'augmenter le nombre d'échantillons est d'en prendre dans deux fenêtres contiguës. Les 80 derniers échantillons de la fenêtre N-1 sont sauvegardés dans un bloc avec les 80 premiers échantillons de la fenêtre N. Ce bloc peut alors être traité comme un bloc de données normal.

Cette sauvegarde n'aurait lieu que lorsque une fréquence au moins à déjà été détectée et que les coefficients d'autocorrélation indiquent la stabilité du signal. Ceci ne nécessite que peu de modifications, n'autorisant ce double calcul que lorsque la probabilité de détection est grande.

Cependant cette méthode conduit à un nombre minimum de fenêtres plus faible (seulement au moins 3) et le nouveau signal est évidemment corrélié avec les deux précédents.

4.4 ALGORITHME CHOISI.

De nombreuses simulations sur 10 et 20ms ont confirmé notre hypothèse que les coefficients d'autocorrélation étaient très "stable" durant la période d'apparition d'un tone. En conséquence nous avons décidé de capitaliser sur ce point et d'utiliser une solution hybride des deux précédentes.

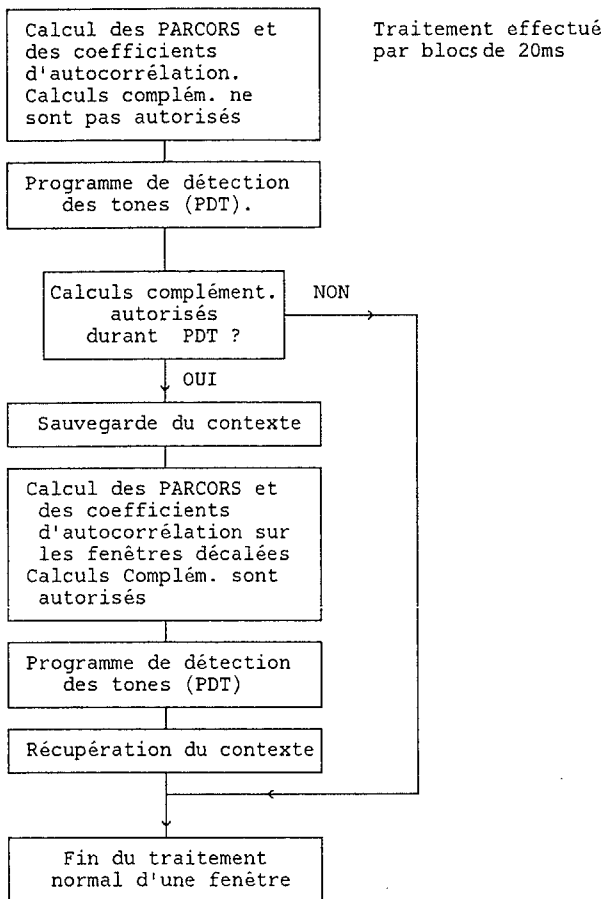


Figure 3. Organigramme du traitement sur 20ms.

Durée du tone	Probabilité (erreur)
41ms	4.10E-3
45ms	9.10E-4
47.5ms	2.10E-4
50ms	4.10E-5
60ms	4.10E-5
70ms	< 1.10E-5

Grâce au moniteur utilisé et au programme de génération de sinusoides utilisés nous avons pu générer plus de 50.000 tones différents (déphasage et amplitudes).

Comme prévu les performances sont proches des spécifications du CCITT qui demande un taux d'erreur de $10E-4$ pour la durée la plus courte.

5.3 BANDE MAGNETIQUE MITEL.

Une batterie de tests utilisant une bande magnétique MITEL /5/ a donné les résultats suivants:

- Capacité de détecter des tones aussi court que 33ms.
- 20 dB de dynamique globale pour les amplitudes.
- Bonne immunité au bruit (-24 dBV) mais le nombre de fausses détections augmente avec l'augmentation du niveau du bruit (-18 et -12 dBV).
- Très bonne immunité à une atténuation différenciée pour la fréquence haute (17 dB) et moyenne pour la fréquence basse (5 dB).
- Excellente immunité au signal de parole (13 détections alors que 30 sont considérées comme acceptables).

5.0 RESULTATS DES TESTS.

5.1 LE V.E.P.C.

Le VEPC /1/ est un algorithme de compression de la voix opérant à un débit de 7.2 Kbit/s et utilisant un prédicteur à court terme pour extraire un signal résiduel d'excitation qui est lui même codé en utilisant une technique de codage en sous bandes et de filtre miroir en quadrature pour la bande de base. La bande haute de ce signal résiduel est codée par de l'information sur son énergie.

Lors de la décompression, la bande de base du signal résiduel est recomposée, la bande haute est régénérée par distorsion non linéaire et réajustement à l'aide des valeurs d'énergie transmises. Les deux signaux sont combinés et utilisés comme signal d'excitation du filtre prédictif dérivé lors de l'analyse.

Comme le VEPC utilise un prédicteur (d'ordre 8), les coefficients d'autocorrélation et ceux du prédicteur sont déjà disponibles, ce qui permet d'économiser un grand nombre de calculs.

Le programme a été implémenté sur un système de développement à base de PC utilisant un processeur de signaux IBM. Ce processeur est une machine 16 bits utilisant un multiplieur 12 par 12, résultat sur 20 bits en permanence et ayant une puissance de 10 MIPS /4/.

5.2 SIMULATION.

Les tones ont tout d'abord été simulés par logiciel afin d'en maîtriser les différents paramètres.

Les résultats qui suivent donnent la probabilité qu'un tone généré soit mal ou pas reconnu par l'algorithme pour des durées différentes :

6.0 CONCLUSIONS.

Un algorithme permettant de détecter un signal multifréquences à l'entrée d'un système de compression du signal vocal a été présenté. Il utilise un schéma très rapide de calcul de la Transformée de Fourier Discrète ce qui permet de diminuer considérablement la charge de travail.

Pour tenir compte de la longueur relative des fenêtres de travail et du signal multifréquences, la possibilité d'effectuer des calculs complémentaires sur des fenêtres décalées par rapport à celles de travail a été développée.

Les performances obtenues dépendent de la durée du signal multifréquence. L'immunité à la fausse détection est excellente.

7.0 REFERENCES.

/1/ C. Galand & al. 'Voice Excited Predictive Coder, Implementation on a 10 MIPS signal processor'. IBM Journal of Research and Development, March 1985.

/2/ International Standard CCITT Vol. VI recommendation Q23.

/3/ Digital Processing of Speech Signals, L. Rabiner & R. Schafer, Prentice-Hall Editor.

/4/ Personal Instrument (PI) - A PC based Signal Processing System, G. Schichman, IBM Journal of Research, March 1985.

/5/ MITEL CM7291, Tone Receiver Test Tape.