

Variations sur un peigne

Jean-Sylvain Liénard

LIMSI-CNRS, bp 133, 91403 Orsay Cedex
Tél. : ++33 (0)1 69 85 81 13 / ++33 (0)1 45 87 09 61
jean-sylvain.lienard@limsi.fr

ABSTRACT

Philippe Martin designed the Spectral Comb method for the measurement of the speech fundamental frequency in the late 70's, for the purpose of his prosodic researches. We present some theoretical considerations showing that any F_0 measuring system has to deal with specific types of errors, and we explain how those errors have been contained in Martin's Spectral Comb. We also present a new development of the same technique, using several variant families - the missing teeth combs and the negative teeth combs - in order to specifically cancel the main types of F_0 errors.

1. INTRODUCTION

La mesure de la fréquence fondamentale de la parole est un problème très ancien de la phonétique acoustique. Jusqu'aux années 70 on utilisait une mesure directe sur le signal, qui avait le mérite de fonctionner en temps réel mais qui produisait des erreurs. Lorsque Philippe Martin, déjà passionné par l'analyse des contours mélodiques, a présenté sa méthode du peigne spectral, ce fut non seulement un progrès scientifique car la mesure était plus précise et plus robuste qu'auparavant, mais aussi un exploit technique car elle tournait en temps réel sur les microordinateurs des années 80.

Dans cet article nous revenons sur les fondements de la méthode. Celle-ci revient à prélever dans le spectre harmonique du signal une série de valeurs discrètes multiples d'une même valeur F_c , que l'on fait varier dans tout l'intervalle à explorer jusqu'à trouver la meilleure mise en correspondance, attribuée à F_0 . Nous montrons que cette idée, appliquée sans précautions, ne conduit pas à un système utilisable. Philippe Martin a trouvé d'emblée la solution qui permet d'en faire un véritable instrument de mesure, fiable et rapide. Nous présentons également des travaux menés dans le même esprit, qui nous ont conduits à développer des familles de peignes très particuliers dans le but de traiter des mélanges de signaux de parole.

2. LE PEIGNE DE MARTIN DANS SON CONTEXTE

La mesure de la fréquence fondamentale de la voix est un problème difficile, pour de multiples raisons détaillées dans l'ouvrage de référence de Wolfgang Hess [Hes83].

Jusqu'à la généralisation des méthodes numériques dans les années 70 on cherchait surtout à isoler le fondamental des harmoniques pour en mesurer la fréquence, avec deux problèmes majeurs: d'une part F_0 peut varier selon les locuteurs et les instants dans un intervalle très vaste (3 à 4 octaves), et d'autre part ce fondamental peut être d'intensité très faible ou franchement absent. Cette stratégie ne fonctionnait que si l'on connaissait d'avance l'ordre de grandeur de F_0 , et si sa variation restait limitée.

On savait pourtant, par l'observation des spectrogrammes en bande étroite et par des considérations auditives ou mathématiques [Sch68], que l'intervalle entre harmoniques, par définition égal à F_0 , pouvait constituer une meilleure base d'investigation. On savait aussi que l'autocorrélation était plus proche de la définition de la périodicité que la recherche du fondamental. Mais ces deux approches ne pouvaient pas donner naissance à de nouveaux systèmes d'estimation de F_0 avant que les technologies numériques ne soient accessibles pratiquement aux chercheurs en phonétique ou en traitement du signal.

Philippe Martin, alors ingénieur dans un laboratoire de phonétique, avait déjà mis au point un système de détection de F_0 , avec les technologies électroniques de l'époque [Mar73]. Il en connaissait les défauts, mais il appréciait aussi l'avantage du temps réel, qui permet au phonéticien de voir immédiatement le résultat de la mesure et de travailler en interaction avec son instrument. Il était aussi attiré par les nouvelles techniques numériques et s'était initié à la programmation en langage assembleur, au plus près du système.

C'est ainsi que son travail l'amène à mettre au point, dans la fin des années 70, un système révolutionnaire à plusieurs égards. Ce système fonctionne sur des segments de signal de durée 32 ms prélevés toutes les 10 ms. Voici comment il en décrit le principe, le traitement étant appliqué à chaque prélèvement (ou trame) de signal [Mar81]: "Le principe de la méthode du peigne réside dans la recherche de valeurs du spectre situées à des fréquences harmoniques et dont la somme soit maximale pour un intervalle fréquentiel donné. L'intercorrélation peigne-spectre revient au calcul de la somme des composantes spectrales correspondant à une structure harmonique donnée, dont la fondamentale varie dans une gamme prédéterminée (par exemple 70 à 1000 Hz)."

Dans ce même article Philippe Martin mentionne plusieurs points techniques, dont le plus important est d'assurer une décroissance des dents du peigne. Un autre point se trouve dans le prétraitement du spectre, visant à éliminer toute énergie autre que celle des harmoniques. Le résultat est parmi les meilleurs de l'époque: le système peut traiter la voix de divers locuteurs, dans des conditions difficiles (bruit, voix téléphonique, fondamental absent). Cette version rédigée en Fortran tourne sur un miniordinateur en 200 fois le temps réel.

L'année suivante, il présente une évolution du système, notamment en ce qui concerne la détection du voisement, et des résultats d'évaluation plus complets [Mar82], incluant une comparaison avec la méthode du cepstre, très en vogue à l'époque. Le peigne donne de meilleurs résultats sur tous les plans: moins d'erreurs sur divers signaux de parole, capacité à traiter des signaux réduits à une fréquence pure ou enregistrés via le canal téléphonique, résistance au bruit, rapidité.

Au fil des années qui suivent il perfectionne sa méthode et en fait le coeur d'un logiciel d'analyse de la parole dénommé WinPitch, distribué par la société Pitch Instruments [Mar96]. S'adressant à des phonéticiens, des chanteurs, des pédagogues de la voix, le fonctionnement en temps réel est indispensable, et c'est un tour de force qu'il accomplit, du point de vue de la programmation, pour faire tourner ce logiciel sur les micro-ordinateurs grand public, avec une interface graphique sophistiquée.

Depuis Philippe Martin a introduit dans les versions successives de WinPitch de nombreuses améliorations de son peigne de base, comme la brosse [Mar00], qui consiste à aligner plusieurs spectres successifs sur le premier d'entre eux, de façon à rendre plus robuste la détection locale de F_0 . Ce principe a été développé récemment pour déterminer une trajectoire locale optimale reposant sur l'alignement de cinq spectres successifs [Mar08]. Très récemment il a mis au point une approche multi-algorithmes [Mar12], qui vise à intégrer à WinPitch plusieurs algorithmes (en complément du Peigne et de la Brosse), chacun pouvant être meilleur que les autres dans des conditions particulières ou sur certains segments de signal entachés de diverses distorsions telles que bruit, écho, compression, sons simultanés.

3. FONDEMENTS DU PEIGNE SPECTRAL

Pour bien comprendre à la fois l'originalité et les développements du peigne de Martin, nous allons dans cette section tenter de montrer les difficultés liées à l'utilisation d'un peigne spectral, qui ressemblent à celles rencontrées par tout détecteur de F_0 , et les solutions qui permettent d'en tirer le meilleur parti.

3.1 La fonction de pitch (FP)

La plupart des détecteurs de pitch font aujourd'hui usage, souvent de manière implicite, d'une "fonction de pitch" (désignée FP dans la suite), prenant une certaine valeur à toute fréquence comprise dans l'intervalle de pitch exploré.

De manière équivalente, la FP peut être exprimée en termes de durée de la période. Cette fonction porte des noms divers selon la manière de la calculer et selon les auteurs: fonction d'intercorrélation spectre-peigne chez Martin [Mar81], histogramme des périodes ou produit spectral chez Schroeder [Sch68], ou encore fonction d'autocorrélation dans les méthodes du même nom. La valeur F_0 cherchée est la valeur pour laquelle la FP est maximale (ou minimale dans certaines méthodes comme l'AMDF). Le problème que révèle la FP est que plusieurs maxima parasites peuvent accompagner, voire dépasser en amplitude, le maximum recherché correspondant à F_0 . Voyons ceci sur des exemples schématiques.

3.2 Le peigne uniforme infini

Considérons (figure 1) un spectre harmonique $|S|$ formé de bosses successives de même amplitude, placées aux fréquences pF_0 multiples de F_0 . Considérons également un peigne spectral infini C , formé de valeurs discrètes et unitaires (les dents du peigne), aux fréquences qF_c multiples de la fréquence du peigne F_c . Les indices p et q sont des entiers positifs.

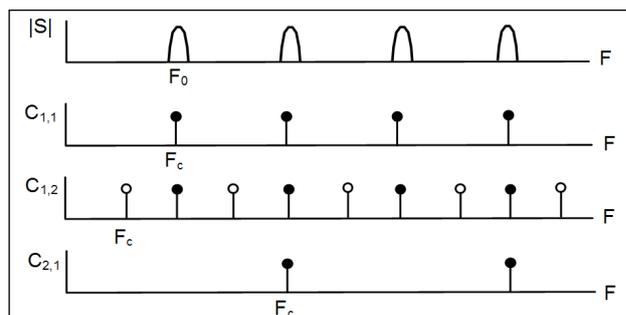


Figure 1: Représentation schématique d'un spectre harmonique uniforme de fréquence F_0 et d'un peigne uniforme infini de fréquence F_c , pour trois valeurs de F_c : $F_c=F_0$, $F_c=F_0/2$ et $F_c=2F_0$.

Lorsque F_c est calé sur F_0 (donc $p=q=1$, cas $C_{1,1}$) chaque dent du peigne est alignée avec une bosse spectrale. La somme des valeurs captées par le peigne est donc maximale.

Lorsque F_c est calé sur la sous-octave $F_0/2$ ($p=1$, $q=2$, cas $C_{1,2}$), seules les dents paires du peigne sont alignées sur des harmoniques. Mais comme chaque bosse spectrale est mise en correspondance avec une dent, la valeur totale captée par le peigne reste identique au cas précédent.

Lorsque F_c est calé sur l'octave $2F_0$ ($p=2$, $q=1$, cas $C_{2,1}$) les choses sont différentes: seules les bosses paires sont en correspondance avec des dents du peigne et la valeur totale captée diminue de moitié.

On voit là le problème posé par le peigne infini: le maximum correspondant à F_0 n'est pas unique. Détecter le maximum de la FP a toutes les chances de produire des erreurs sous-harmoniques à $F_0/2$, $F_0/3$ etc.

Ce point est illustré par la fonction de pitch de la figure 2, obtenue à partir d'un son harmonique réel (train d'impulsions de $F_0=250$ Hz, fenêtre de Hanning 50 ms) et

d'un peigne infini. Le pic principal, pour lequel $F_c=F_0$, est étiqueté par $(p=1, q=1)$, en abrégé $(1,1)$. On voit apparaître des pics parasites sous-harmoniques $(1,2)$, $(1,3)$... $(1,q)$, des pics parasites harmoniques $(2,1)$... $(p,1)$, et des pics parasites fractionnaires $(2,3)$, $(4,3)$, $(3,2)$... d'amplitude plus faible.

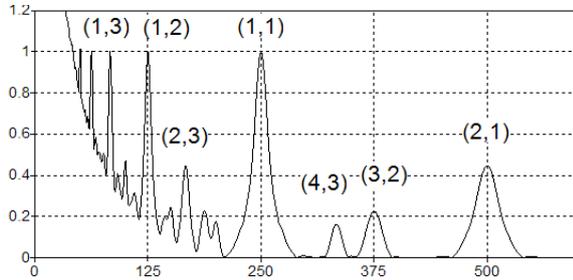


Figure 2: FP d'un son harmonique à $F_0=250$ Hz analysé par un peigne uniforme infini. Les pics sont indicés par (p,q) selon $F_c=(p/q)F_0$.

La remontée de la ligne de base de la fonction vers les fréquences basses correspond au cas où les dents sont tellement serrées qu'il s'en trouve plusieurs dans la largeur d'une bosse spectrale: le peigne ne se comporte plus comme un sélecteur de bosses mais comme un système intégrateur.

3.3 Le peigne limité

C'est le nombre de dents, supposé infini, qui cause l'ambiguïté exposée ci-dessus. Une solution évidente est de limiter ce nombre. Reprenons la figure 1 et imaginons que le peigne ne comporte que 3 dents. Quand $F_c=F_0$ le peigne capte les 3 premières bosses du spectre. La valeur de la FP est donc moins forte qu'avec le peigne infini, puisqu'on laisse de côté les harmoniques de rang supérieur à 3. Mais lorsque $F_c=F_0/2$ seule une bosse spectrale est captée, et le pic correspondant de la fonction de pitch se trouve 3 fois plus faible que le pic principal. Donc les erreurs sous-harmoniques sont beaucoup moins probables qu'avec le peigne infini.

Pourtant cette solution n'est pas la panacée: lorsque $F_c=2F_0$ le peigne à 3 dents capte les bosses 2, 4 et 6 du spectre. La fonction de pitch prend en ce point la même valeur que celle du pic principal, et ce sont alors les erreurs harmoniques qui deviennent les plus probables...

3.4 Le peigne à dents décroissantes, ou Peigne de Martin

On a compris d'après ce qui précède que le peigne infini produit des erreurs sous-harmoniques, et le peigne limité des erreurs harmoniques. La solution proposée par Philippe Martin se trouve à mi-chemin: elle consiste à faire décroître l'amplitude des dents en fonction de leur rang. La fonction de décroissance règle le compromis entre les deux types d'erreurs et tient compte du fait que le spectre de la parole voisée est lui-même décroissant en moyenne au delà de 800 Hz. Philippe Martin a choisi une décroissance exponentielle avec un nombre de dents compris entre 5 et 10.

La figure 3 montre la fonction de pitch obtenue avec le même son que précédemment et un peigne à 10 dents d'amplitude décroissant en fonction inverse de la racine carrée de leur rang. Les deux pics parasites les plus dangereux sont à l'octave $(2,1)$ et la sous-octave $(1,2)$ du pic principal. Le recul de ces pics par rapport au pic principal n'est pas très grand (respectivement 6 et 3 dB environ). En soi le fait que l'émergence du pic principal soit faible n'empêche nullement le système de bien fonctionner si le signal ne contient qu'une voix et si la mise en œuvre garantit que le maximum est unique et correspond à F_0 .

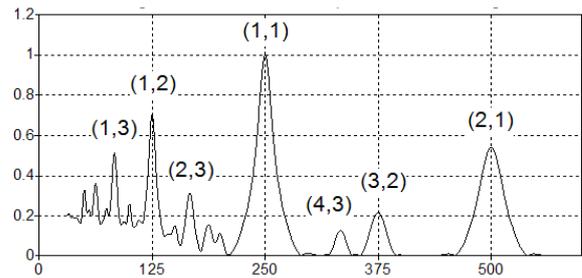


Figure 3: FP d'un son harmonique à $F_0=250$ Hz analysé par un peigne à 10 dents d'amplitude décroissant selon l'inverse de la racine carrée de leur rang.

3.5 La mise en œuvre du peigne

La manière dont le spectre est préparé avant son analyse par le peigne a pour objet de garantir l'unicité du maximum, d'obtenir une mesure précise et de limiter le temps de calcul. Ici comme partout ailleurs le diable est dans les détails. Dans ses premiers articles Philippe Martin utilise des amplitudes en décibels, un seuillage pour éliminer les composantes trop faibles, une interpolation parabolique dans l'entourage des maxima locaux. De plus, la largeur de la fenêtre temporelle gouverne la résolution fréquentielle, la complexité du calcul doit être adaptée à la puissance du processeur disponible pour atteindre le temps réel, les problèmes de précision de calcul imposent de choisir un algorithme plutôt qu'un autre pour calculer la fonction souhaitée, etc. C'est un travail d'artisan, qui ne peut être effectué que par quelqu'un qui a une profonde connaissance de tous les aspects du processus, et qui sait exactement à quel résultat il veut aboutir.

3.6 Universalité de la fonction de pitch

Comme cela a été évoqué plus haut (§3.1) une forme de fonction de pitch est présente dans pratiquement tous les systèmes contemporains de détermination de F_0 , même si cela n'apparaît pas explicitement. Comme la fréquence fondamentale d'un signal est habituellement unique, physiquement et auditivement (on peut toutefois construire artificiellement des signaux à périodicité multiple), il paraît tout naturel de choisir un extrémum unique, souvent le premier d'une série, les autres étant considérés comme parasites.

En réalité, ces autres extrémums ne tombent pas du ciel. Ils sont produits par un processus d'interférence entre deux objets mathématiques périodiques: le signal analysé d'une part, avec sa fréquence fondamentale fixe, et le signal - ou la structure de calcul récurrente mise en œuvre pour la mesure - dont on fait varier la fréquence ou la cadence pour déterminer sa ressemblance avec le signal. Cette ressemblance peut être à caractère cumulatif, et on cherche alors un maximum de la fonction de pitch, ou à caractère compensatoire, et on cherche alors un minimum. Mais quelles que soient les variations entre méthodes, on retrouve toujours les mêmes propriétés interférentielles qui se traduisent par les pics de la FP, repérables par deux indices tels que les entiers p et q définis ci-dessus. Le point important ici est que les pics parasites n'ont pas une origine aléatoire; ils sont la conséquence de la comparaison de deux processus périodiques. Au lieu de subir cette contrainte, on peut peut-être essayer d'en tirer parti pour améliorer la mesure. C'est ce que nous allons exposer maintenant.

4. NOUVEAUX PROBLEMES, NOUVEAUX PEIGNES

4.1. De nouveaux problèmes ?

Aujourd'hui les meilleurs détecteurs de pitch donnent des taux d'erreur (erreurs dites "grossières", c'est-à-dire s'écartant de plus de 20% de la fréquence locale du signal) inférieurs à 1% sur des bases de données de référence (voir p.ex. [Cam08] ou [Sig09]), pour autant que les comparaisons aient un sens car chacun peut donner satisfaction dans une application et s'avérer insuffisant dans une autre. Cet ordre de grandeur paraît minime mais masque de graves insuffisances. En particulier en ce qui concerne le voisement et la résistance au bruit.

Outre sa signification du point de vue de la phonation (vibration des cordes vocales), le terme de voisement a deux acceptions bien distinctes. Du point de vue phonologique c'est un trait distinctif, c'est-à-dire une qualité abstraite, binaire, caractérisant deux grandes classes de phonèmes. Du point de vue acoustique c'est un degré de périodicité approximative, apprécié sur un segment de quelques dizaines de millisecondes, dans un intervalle convenu, par exemple 60 à 600 Hz. Les limites temporelles d'un segment voisé ne sont pas parfaitement définies: commence-t-il à la première période détectée, à la seconde, ou après un intervalle de temps défini ? Les périodes, en matière de voix parlée ou chantée, ne sont jamais exactement identiques à elles-mêmes; quel degré de ressemblance doivent-elles avoir pour conférer à l'instant la qualité "voisé" ou "non-voisé" ? Le problème est encore plus flou en fin de segment voisé, où souvent le voisement s'éteint lentement pour devenir une simple oscillation quasi-sinusoïdale et se confondre avec la réverbération ambiante.

Comme la notion de voisement acoustique est floue, on a du mal à évaluer la performance d'un système de détection du voisement - et tout système de détection du

pitch est aussi et avant tout un système de détection du voisement. C'est pour cette raison que l'on s'en tient en général à la performance en matière de F_0 , comptée sur les instants indiscutablement bien voisés, au cœur des syllabes. Mais ceci conduit à un paradoxe: un système élaboré, détectant correctement le voisement aux limites des segments voisés, a toutes les chances de faire des erreurs de pitch en ces zones où, précisément, le voisement est imparfait et la fréquence fondamentale incertaine. Il sera pénalisé sous l'aspect du taux d'erreurs sur F_0 par rapport à d'autres qui ne considèrent comme voisés que les noyaux syllabiques.

Par ailleurs, à l'écoute d'un signal de parole, il est très peu usuel de percevoir l'octave, et surtout la sous-octave du fondamental. Or ces erreurs existent, au moins potentiellement, dans tous les systèmes de détection de pitch. On peut donc penser que la perception auditive procède autrement, ou dispose de mécanismes lui permettant de les ignorer ou de les neutraliser.

Un autre problème apparaît dans les situations de bruit, dont chacun sait qu'elles altèrent le fonctionnement de tous les systèmes artificiels, qu'il s'agisse de détermination du pitch et du voisement, ou de reconnaissance automatique de la parole ou du locuteur. On s'arrange habituellement pour contourner le problème en enregistrant en local insonorisé ou en pratiquant une prise de son de proximité. Mais on ne peut oublier que l'audition naturelle, elle, est normalement immergée dans du bruit au point que, bien souvent, le locuteur n'en a même pas conscience. Comment fait-elle pour saisir toutes les nuances de la parole et de la voix malgré les conditions adverses ?

Ce qui vaut pour le bruit vaut aussi pour l'extraordinaire capacité qu'a tout sujet bien entendant à suivre plusieurs conversations à la fois. On sait que l'effet de cocktail party met en jeu un processus attentionnel de haut niveau qui permet de sélectionner à volonté l'une ou l'autre des voix présentes dans le signal. Il n'en reste pas moins que plusieurs flux auditifs sont formés simultanément au bas niveau de la perception auditive. Et qu'un signal totalement imprévu mais hautement significatif peut aussi accéder prioritairement aux niveaux conscients (effet d'émergence soudaine ou *pop-up*). En ce qui concerne la détection de pitch, il faut bien en déduire que le système auditif est capable de former simultanément plusieurs flux vocaux à un niveau préattentionnel, et donc de mettre en œuvre une détection de plusieurs fréquences fondamentales simultanées.

Pour toutes ces raisons nous avons été amenés à nous intéresser à la détection multipitch. Dans cette optique nous avons repris les fondements du peigne spectral en essayant d'améliorer son comportement face à des signaux mélangés.

4.2 Des peignes et des dents

Pour bien comprendre comment fonctionne le peigne spectral il n'est pas inutile de se demander quel est précisément le rôle de chaque dent.

Et d'abord, peut-on imaginer un peigne à une seule dent ? La réponse est oui, bien sûr: appliquer à un spectre quelconque un peigne à une dent d'amplitude unitaire conduit à une fonction de pitch identique au spectre lui-même. Et si l'on applique un coefficient d'amplitude, différent de 1, à cette unique dent, alors le résultat est une fonction de pitch dont toutes les valeurs sont multipliées par ce coefficient, même s'il est négatif.

Le cas du peigne à deux dents est très instructif. La fonction de pitch obtenue représente la somme du spectre avec lui-même, décalé de l'intervalle entre les deux dents. Si les fréquences des deux dents sont dans un rapport du simple au double, soit F_c et $2F_c$, et si le spectre est composé d'harmoniques multiples de F_0 , la fonction de pitch passe par des maxima principaux pour $F_c=F_0$, $F_c=2F_0$ etc, et par des maxima secondaires pour $F_c=F_0/2$, $3F_0/2$ etc. Si la seconde dent est affectée du coefficient -1, alors la fonction de pitch n'est plus le cumul, mais la différence du spectre avec sa version décalée.

D'une manière générale chaque dent ajoutée à un peigne à n dents ajoute à la FP autant de spectres décalés, chacun affecté du coefficient propre à la dent.

Nous allons maintenant utiliser ces observations pour mettre en œuvre des peignes particuliers, destinés à renforcer ou atténuer les pics indésirables de la fonction de pitch. Deux familles de peignes sont présentées ci-dessous, les peignes à dents manquantes et les peignes à dents négatives.

4.3 Peignes à dents manquantes (pdm)

Dans le peigne uniforme infini à dents manquantes de la figure 4 on a supprimé les dents de rang pair. L'effet de ce peigne noté pdm2 (ou peigne impair) est de ne prendre en compte aucune bosse spectrale lorsque le peigne est calé sur $F_0/2$, comme le montre la figure (cas M1,2). Il en va de même pour $F_0/4$, $F_0/6$ etc.

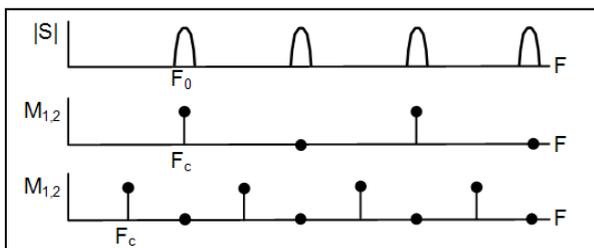


Figure 4: peigne à dents manquantes d'ordre 2, calé sur F_0 et sur $F_0/2$

Ce peigne supprime donc les sous-harmoniques indicés $(p,2k)$ de la fonction de pitch, comme le montre la figure 5 obtenue à partir du même son que précédemment (k est un entier positif).

De la même manière, le peigne à dents manquantes pdm3, peigne uniforme infini dont on aurait enlevé les dents 3, 6, 9... $3k$ supprimerait les pics $(1,3)$, $(1,6)$... $(p,3k)$.

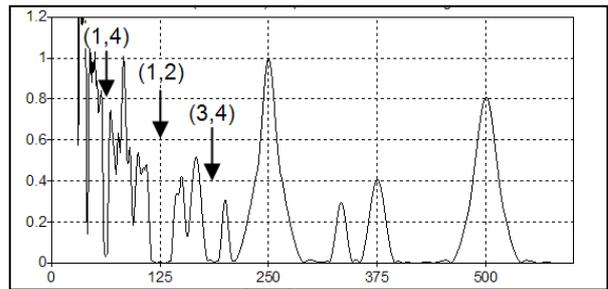


Figure 5: fonction de pitch obtenue avec le peigne impair sur un train d'impulsions de $F_0=250$ Hz. Les pics parasites $(1,2)$, $(1,4)$, $(3,4)$... $(p,2k)$ sont supprimés.

4.4 Peignes à dents négatives (pdn)

Dans le peigne uniforme infini à dents négatives (ou alternées) on ajoute des dents négatives entre les dents positives. Par convention la fréquence F_c du peigne est celle de la première dent positive. La figure 6 montre le pdn d'ordre 2, noté $A_{2,1}$. Lorsque F_c est calé sur F_0 la valeur de la FP est la même que celle du simple peigne uniforme infini puisque les dents négatives sont alignées avec des valeurs nulles du spectre. Mais lorsqu'il est calé sur $F_0/2$ l'effet des bosses paires et impaires s'annule globalement et le pic parasite correspondant $(2,1)$ est supprimé.

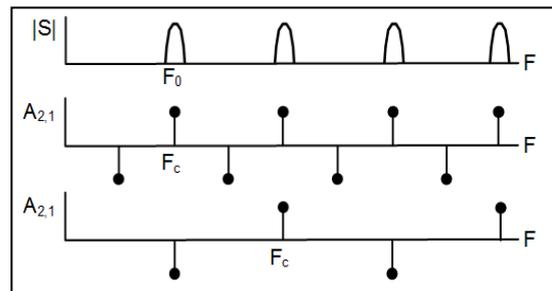


Figure 6: peigne à dents négatives d'ordre 2, calé sur F_0 et sur $2F_0$

La figure 7 montre la FP obtenue en appliquant ce peigne au même son que dans les exemples précédents. On voit que non seulement le pic d'octave $(2,1)$ est supprimé, mais qu'il en va de même pour tous les pics $(2p,q)$.

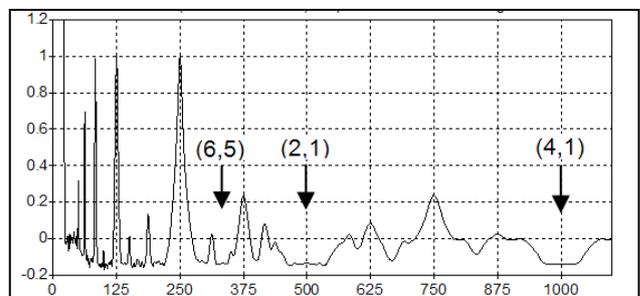


Figure 7: FP obtenue avec le son à 250 Hz analysé par un pdn d'ordre 2

Comme le précédent, ce type de peigne se généralise aisément: en intercalant deux dents négatives d'amplitude 0,5 régulièrement espacées entre deux dents positives (p_{dn}3) les pics parasites (3p, q) sont supprimés.

4.5 Peigne à Suppression Harmonique (PSH)

Chacun des pdm et pdn présentés ci-dessus a comme caractéristique de supprimer une famille de pics parasites, tout en préservant le pic principal. Dans le Peigne à Suppression Harmonique PSH ([Lie08], [Sig09]) les peignes pdm et pdn sont utilisés conjointement (produit des FP correspondantes, ordres 2, 3, 5 et 7). Le résultat sur notre son de référence est présenté ci-dessous, figure 8, à comparer avec la figure 3. Les deux pics les plus dangereux (octave et sous-octave) se trouvent maintenant à -20 dB du pic principal. Ce surcroît d'émergence permet d'assurer une plus grande fiabilité à la détection du maximum dans le cas d'un signal monovoix, et de séparer plusieurs pics principaux dans le cas d'un mélange de plusieurs voix. La détection du voisement acoustique est elle aussi grandement facilitée.

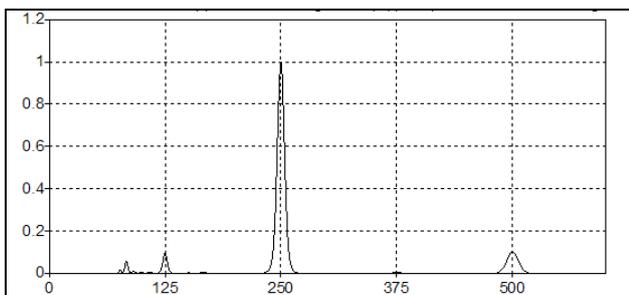


Figure 8: FP du son à 250 Hz obtenue par application conjointe des pdn et pdm d'ordres 2, 3, 5 et 7.

5. CONCLUSION

Depuis les années 70 Philippe Martin a mis au point et continuellement perfectionné une méthode originale d'analyse et de représentation de la fréquence fondamentale de la parole, outil essentiel pour ses recherches prosodiques. Sa méthode du peigne spectral donne un résultat immédiat et fiable, répondant à ses exigences expérimentales. Ayant à l'esprit un autre but, relevant de la perception de la parole, nous avons voulu en comprendre le fonctionnement en profondeur pour l'adapter à notre propre problème d'analyse de voix mélangées. Cette démarche nous a permis de constater l'efficacité de la méthode, et de proposer diverses extensions qui en conservent l'esprit et certaines qualités, en particulier la simplicité et la rapidité de calcul.

RÉFÉRENCES

- [Cam08] Camacho, A. and Harris, J.G. (2008), "A sawtooth waveform inspired pitch estimator for speech and music", *J. Acoust. Soc. Am.* Vol. 124, No 3, pp 1639-1652.
- [Hes83] Hess, W. (1983), *Pitch determination of speech signals*, Springer-Verlag.
- [Lie07] Liénard, J.S., Signol, F. and Barras, C. (2007), Speech fundamental frequency estimation using the Alternate Comb, *InterSpeech*, Anvers.
- [Lie08] Liénard, J.S., Barras, C. and Signol, F. (2008): "Using Sets of Combs to Control Pitch Estimation Errors", *ASA Proceedings Of Meetings on Acoustics*, Vol 4, 060005.
- [Mar73] Martin, Ph (1973), *Réalisation d'un analyseur de mélodie fonctionnant en temps réel en vue de la recherche phonétique et l'enseignement*, Dissertation, Université Libre de Bruxelles.
- [Mar81] Martin, Ph. (1981), "Extraction de la fréquence fondamentale par intercorrélation avec une fonction peigne", *12e Journées d'Etude sur la Parole*, Montréal, pp 221-232.
- [Mar82] Martin, Ph. (1982), "Comparison of pitch detection by cepstrum and spectral comb analysis", *Icassp*, Paris, pp 180-183.
- [Mar96] Martin, Ph. (1996), "WinPitch: un logiciel d'analyse temps réel de la fréquence fondamentale fonctionnant sous Windows", *21e Journées d'Etude sur la Parole*, Avignon.
- [Mar00] Martin, Ph. (2000), "Peigne et brosse pour F0: mesure de la fréquence fondamentale par alignement de spectres séquentiels", *23e Journées d'Etude sur la Parole*, Aussois, pp 245-248.
- [Mar08] Martin, Ph. (2008), "Crosscorrelation of adjacent spectra enhances fundamental frequency estimation", *InterSpeech*, Brisbane.
- [Mar12] Martin, Ph. (2012), "Multi methods pitch tracking", *ISCA Speech Prosody*, Shangaï.
- [Sch68] Schroeder, M.R. (1968), "Period histogram and product spectrum: new methods for fundamental frequency measurement", *J. Acoust. Soc. Am.* Vol. 24, pp 456-465.
- [Sig09] Signol, F. (2009), *Estimation de fréquences fondamentales multiples en vue de la séparation de signaux de parole mélangés dans un même canal*, thèse de doctorat, Univ. Paris XI.