Styger, T., Gabioud, B., Keller, E. (1993). Méthodes informatiques pour l'analyse de paramètres primaires en parole pathologique. *C.A.L.A.P., 12*.

METHODES INFORMATIQUES POUR L'ANALYSE DE PARAMETRES PRIMAIRES EN PAROLE PATHOLOGIQUE.

Thomas STYGER, Bernard GABIOUD, Eric KELLER.

I. INTRODUCTION.

Les perturbations de la parole (les “dysarthries”) sont parmi les séquelles les plus graves des lésions neurologiques occasionnées par des accidents cérébrovasculaires, des traumatismes crâniens etc… Dans ce domaine, chercheurs et cliniciens ont besoin d'un instrument simple et fiable pour l'évaluation des troubles de la parole. Une évaluation acoustique présente ces avantages; elle est simple à mettre en oeuvre, les coûts du système sont peu élevés et la gêne du patient réduite à un strict minimum (KELLER, 1992). L'analyse acoustique fournit des informations d'une richesse surprenante; elle est souvent en mesure de distinguer entre différents syndromes et permet d'identifier la sévérité de la perturbation. La difficulté majeure en analyse acoustique des pathologies de la parole se situe au niveau du volume des données à traiter. Pour ces raisons, un traitement automatisé des données s'impose.

Les critères suivants s'avèrent décisifs lors de l'évaluation des pathologies (KELLER, 1991; CANTER, 1985; DARLEY et al., 1969) : (i) la distinction entre segments voisés, non-voisés et fricatifs, (ii) l'organisation temporelle de la parole (i.e. la durée des silences), (iii) la distinction des différents niveaux d'amplitude ainsi que les variations fort rapides que peut subir la fréquence fondamentale, (iv) l'évolution temporelle de l'ensemble de ces paramètres. Le développement de mesures automatiques de ces événements représente un objectif important dans le cadre d'un examen approfondi des perturbations de la parole.

La mesure automatique de l'amplitude et de son évolution ne pose pas de problème, alors que l'analyse de la fréquence et des autres paramètres (voisement, silences et friction) se heurte à certaines difficultés. En particulier, l'analyse spectrale par transformées de Fourier est mal

adaptée aux modifications fort rapides que peut subir la voix pathologique.

Les méthodes que nous préconisons se basent sur une analyse temporelle du signal. Une inspection minutieuse de la structure temporelle (forme d'onde), selon un certain nombre de critères, permet une segmentation primaire fiable et précise. Quatre algorithmes (ou “senseurs”) sont présentés ici :

1) La segmentation Silence / Parole. Identifie la présence ou l'absence du signal vocal, ainsi que les silences occlusifs.
2) La segmentation Fricatif / Non-fricatif. Identifie la présence d'un bruit de friction dans le signal vocal.
3) La segmentation Voisé / Non-voisé. Permet de détecter la présence d'une excitation glottique.
4) La mesure de la fréquence fondamentale.

Dans le cadre du présent travail, ces senseurs ont tout d'abord été développés pour des signaux de parole normale. Leur performance a ensuite été examinée sur un corpus de signaux pathologiques.

II. ALGORITHMES DE SEGMENTATION.

Après avoir éliminé la composante continue, le signal est recodé selon les passages par zéro de sa dérivée. Nous retenons donc les instants k_i, k_{i+1} …, ainsi que les amplitudes a(i) associées aux extremums.

Ce prétraitement est avantageux, car il permet de réduire la quantité d'informations à traiter et le signal se présente sous une forme plus simple à manipuler (BAUDRI, 1978).

1) Le senseur Silence / Parole.

Le senseur d'identification des segments Silence et Parole se base sur la comparaison des amplitudes du signal avec le niveau de bruit. Cependant le senseur ne fait pas intervenir un traditionnel seuil d'énergie ou d'amplitude moyenne, mais un codage particulier des extremums.

a) Codage des extremums du signal.

Le niveau de bruit L_n est calculé en prenant la moyenne des valeurs absolues des amplitudes a(i) sur une portion de silence plus la moitié de l'écart type :

$$L_n = \overline{|a(i)|} + \frac{\sigma}{2}$$

Les extremums du signal sont ensuite classés en tenant compte de leur voisinage immédiat (Figure 1) : la classe AI (“Amplitude Inférieure”) correspond au type Silence et la classe AS (“Amplitude Supérieure”) au type Parole.

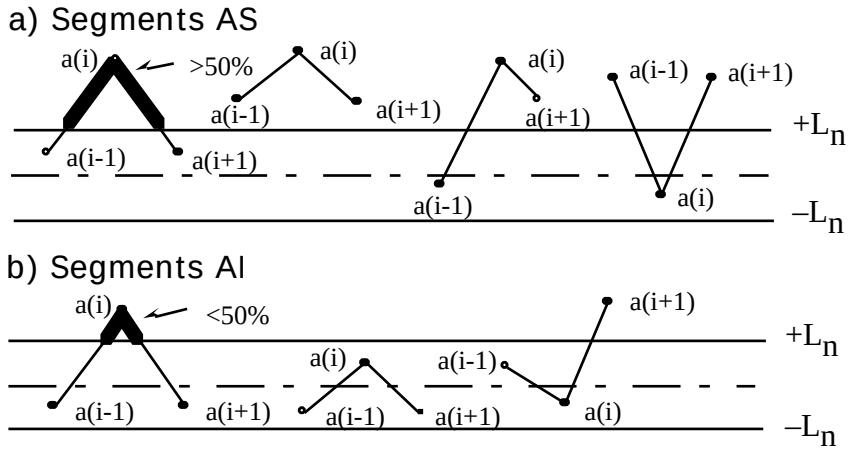

Figure 1. Codage des extremums selon le niveau de bruit (L_n).
Le premier extremum a(i) de la partie a) est considéré de type AS car >50% de sa durée se trouve dans le domaine $>L_n$. Cette condition n'est pas satisfaite dans le premier extremum de b).

Le résultat de la classification des segments est représenté par la figure 2.a. L'identification des éléments AI et AS fait ressortir des tendances de type silence (succession de segments de type AI de durée longue, entrecoupés de segments AS de durée plus courte), et des tendances de type parole (présence de segments de type AS plus longs). Par une élimination judicieuse des éléments les plus courts, nous déterminons le début et la fin des segments de parole. Le détail de l'algorithme est donné par l'organigramme de la figure 4.

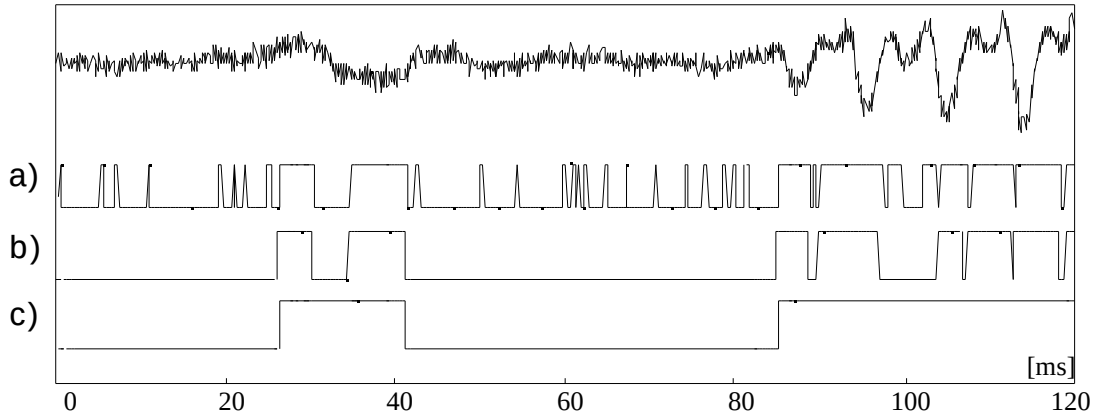

Figure 2. Détermination du début du segment de parole. Début de la barre de voisement de l'occlusive /d/. Le codage des éléments AS et AI est représenté par les signaux binaires a), b) et c). Un niveau haut correspond à AS, un niveau bas à AI. a) Codage initial des éléments AI et AS. b) Résultat après élimination des

AS plus brefs que 1/4 de période de voisement. c) Résultat après élimination des AI plus brefs qu'une période de voisement.

b) Détermination des segments Silence et Parole.

Pour commencer, les éléments AS d'une durée plus courte qu'un quart de la période de voisement sont éliminés (figure 2.b). Ce critère est basé sur l'observation de portions de signaux de faible amplitude (début ou barre de voisement). Ensuite, on élimine les segments de type AI dont la durée ne dépasse pas une période de voisement (figure 2.c).

La procédure décrite est bien adaptée aux signaux voisés, mais peut présenter des erreurs dans le cas des fricatives ou des barres d'explosion des occlusives de faible amplitude. De ce fait un second procédé est utilisé en parallèle, basé sur le niveau L_d de la première dérivée du signal et qui a pour effet de mieux faire apparaître les composantes de haute fréquence (HESS, 1976) :

$$L_d = \frac{1}{M} \sum_{k=N}^{N+M} |x(k+1) - x(k)|$$

où M est la durée de la fenêtre d'analyse, N le point de départ et x(k) les échantillons du signal. Les fricatives étant caractérisées par une distribution d'énergie plus importante dans les hautes fréquences, L_d est plus élevé (figure 3). Un seuil sur L_d permet d'assurer l'inclusion des frictions de faible amplitude dans les segments de type Parole.

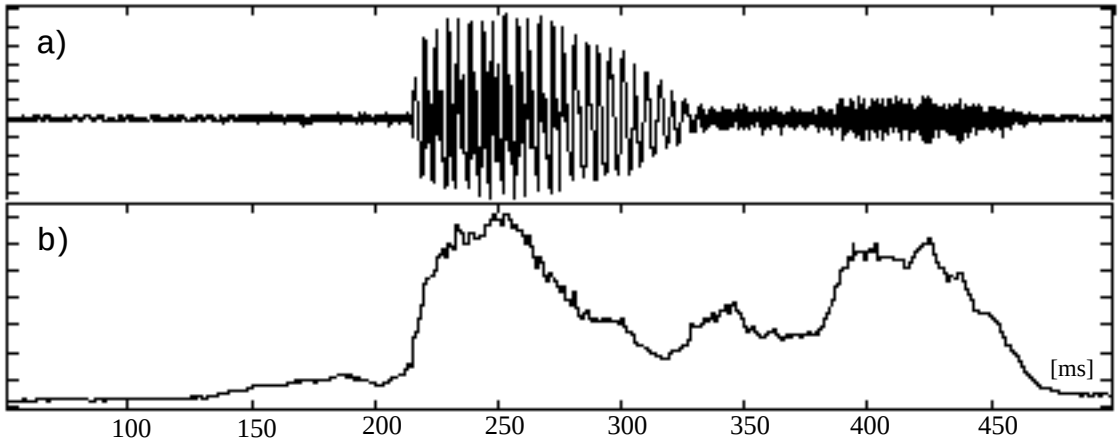

Figure 3. Enveloppe de la première dérivée du signal. a) Signal, les phonèmes /f ɛ s/ du mot “festin”. b) Enveloppe moyenne de la première dérivée du signal.

2) Le senseur de friction.

La technique appliquée pour la détermination des segments fricatifs du signal de parole se base principalement sur une statistique du nombre de passages par zéro de la dérivée du signal (SCARR, 1968; ITO et al., 1971).

La méthode que nous décrivons a été optimisée pour identifier une faible friction même lorsqu'elle se superpose à un voisement.

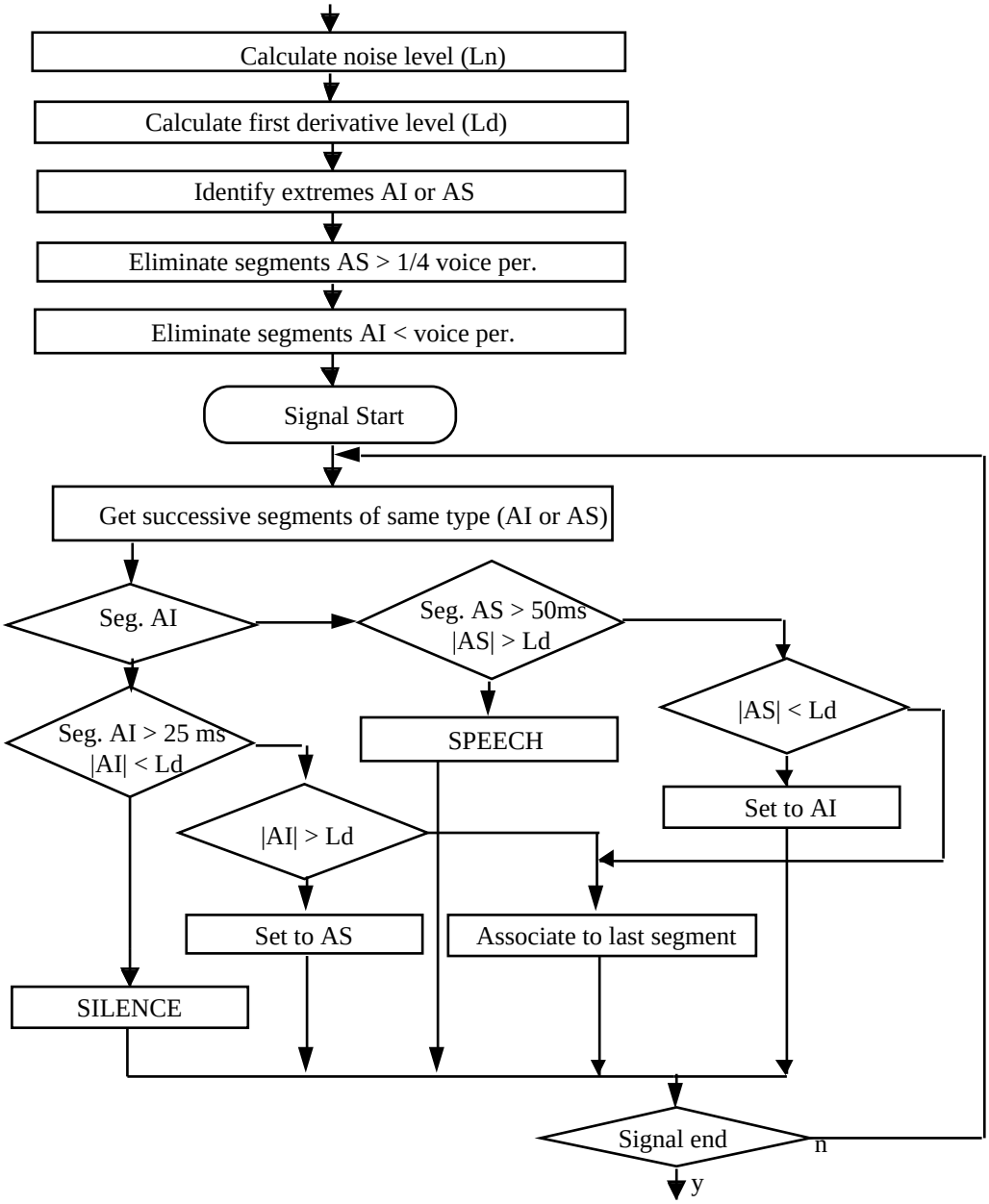

Figure 4. Organigramme : Principe de la détermination des segments Silence / Parole.

a) Motivations et principe.

Les voyelles, les consonnes liquides et nasales se distinguent par une énergie concentrée dans le bas du spectre. Les formants supérieurs à F_2 sont négligeables par leur faible énergie. En revanche, les fricatives ne présentent que peu (barre de voisement) ou pas d'énergie en dessous de 2 kHz, avec un premier maximum se situant en général aux alentours de 3 à 6 kHz. Ces différences au niveau du spectre en énergie peuvent également

être grossièrement identifiées sur la forme d'onde temporelle du signal. La figure 5 présente l'exemple d'une fricative et d'une voyelle.

Un bruit de constriction se caractérise par la présence d'oscillations rapides dont l'amplitude relative varie peu. Sur la figure 5.a on observe une succession de segments de droite dont la longueur ne varie pas de façon excessive d'un segment au suivant. Remarquons que la représentation en segments est due au codage des extremums du signal, qui dans ce cas est très avantageuse.

La situation d'une voyelle est différente. Le contenu en basses fréquences se caractérise par une forme d'onde variant lentement et principalement identifiée par des segments longs, alors que la présence de segments plus courts est caractéristique des composantes hautes fréquences de plus faible énergie. Ces segments de faible amplitude sont repérés sur la figure 5.b.

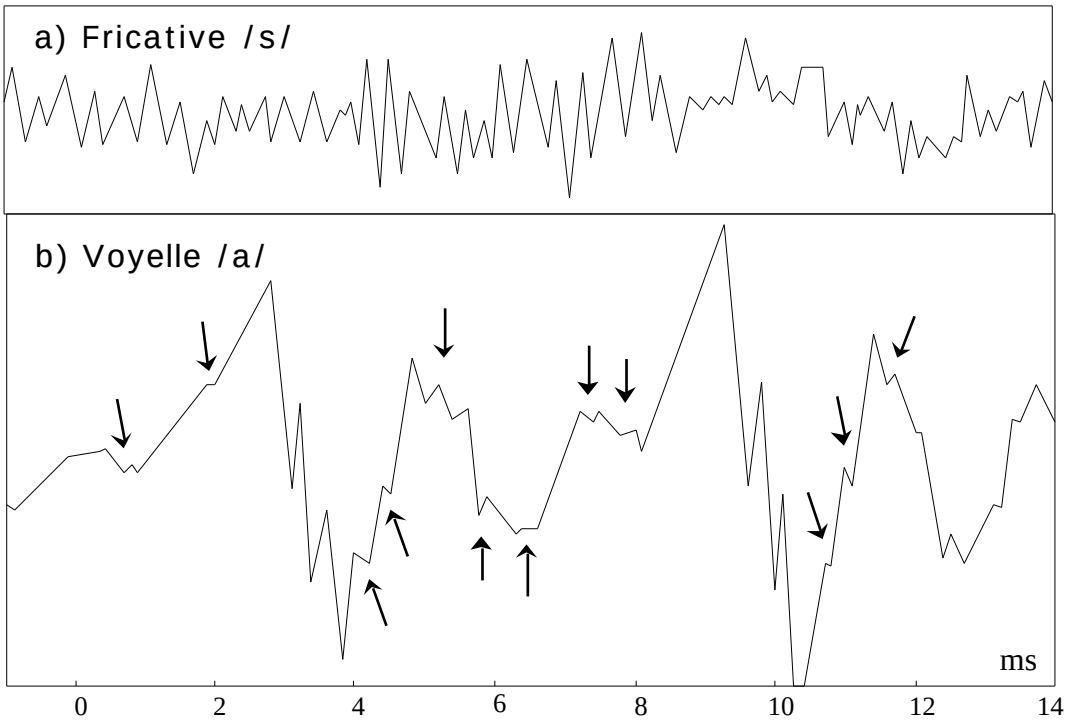

Figure 5. Structure temporelle d'une fricative /s/ et d'une voyelle /a/. Les sons bruités sont représentés par une succession de segments d'amplitude relative voisine, alors que les sons non fricatifs comprennent des segments longs et courts.

Le principe de l'algorithme consiste à éliminer ces segments courts lorsque ceux-ci sont identifiés parmi des segments plus longs dans les sons vocaliques. De cette façon il est possible de s'affranchir des composantes spectrales hautes fréquences sans pour autant éliminer le bruit de friction.

b) Implantation.

Le senseur de friction comporte 4 étapes:

i) Elimination des segments courts, suivant le critère :

$$\frac{|a(i+1) - a(i)|}{|a(i) - a(i-1)|} < Ls$$

où les a(i) sont les amplitudes des extremums locaux du signal.

Lorsque cette relation est vérifiée, le segment compris entre a(i) et a(i+1) est invalidé et la recherche continue jusqu'à ce que la taille relative du segment [a(i),a(i+1)] dépasse le seuil. La figure 6 illustre la procédure.

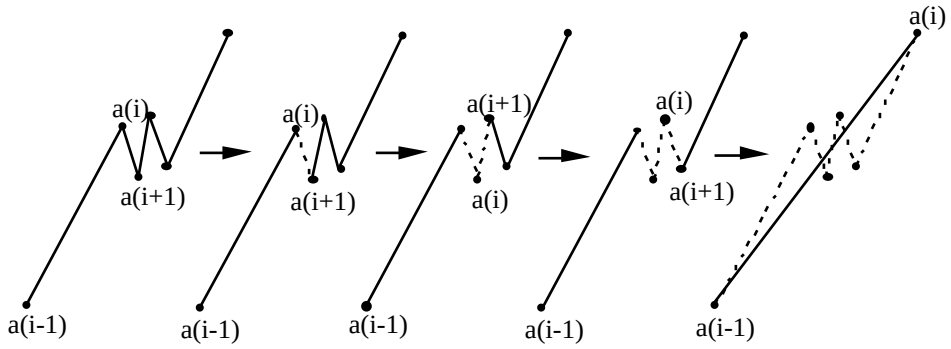

Figure 6 Principe de l'élimination des segments courts. Les segments du signal sont éliminés dont la longueur relative au segment précédent est inférieure au seuil L_S .

ii) Statistique du nombre de passages par zéro de la dérivée du signal. Pour obtenir une bonne précision sur l'identification des transitions, l'utilisation d'une fenêtre de courte durée est nécessaire (1 à 2 ms).

iii) Lissage de la statistique. L'utilisation d'un filtre médian moyenné (RABINER et al., 1975; GALLAGHER, 1981) permet d'éliminer le bruit introduit dans l'étape ii), sans pour autant adoucir les transitions.

iv) Les segments Fricatifs et Non-fricatifs sont ensuite déterminés par le positionnement d'un seuil de décision.

3) Le senseur de voisement.

La détection de voisement s'effectue selon plusieurs principes, suivant l'amplitude du signal. Dans le cas des signaux de forte amplitude (typiquement cinq fois le niveau de bruit) il est intéressant de se baser sur une mesure de passages par zéro du signal codé par les milieux des segments (BAUDRI, 1978). Ce codage consiste à repérer chaque milieu de segment compris entre deux extremums a(i) et a(i+1) et construire un nouveau signal passant par ces points. C'est une estimation grossière du contenu des basses fréquences du signal, qui se caractérise par peu de passages par zéro dans le cas de signaux voisés. La figure 7 présente un tel codage pour une fricative et une voyelle.

La technique utilisée pour l'extraction des segments Voisé / Non-voisé dans les portions de signal de forte amplitude est illustrée par l'organigramme de la figure 9. Avant de coder le signal par les milieux des segments, on opère d'abord une élimination des segments courts, telle qu'elle a été décrite sous II.2. Elle a pour effet d'éliminer les bruit de haute fréquence et de faible amplitude relative. Après codage, une statistique du nombre de passages par zéro du signal codée par les milieux des segments est effectuée sur une fenêtre de courte durée (5 ms), afin de conserver une bonne précision quant aux instants de transition. Les segments du signal dépassant un niveau d'énergie L_e sont alors identifiés comme VO (voisé) ou NV (non-voisé) selon un seuil de décision Z_{nv} du nombre de passages par zéro. Les segments de parole dont le niveau est inférieur à L_e ne sont pas classés à ce moment.

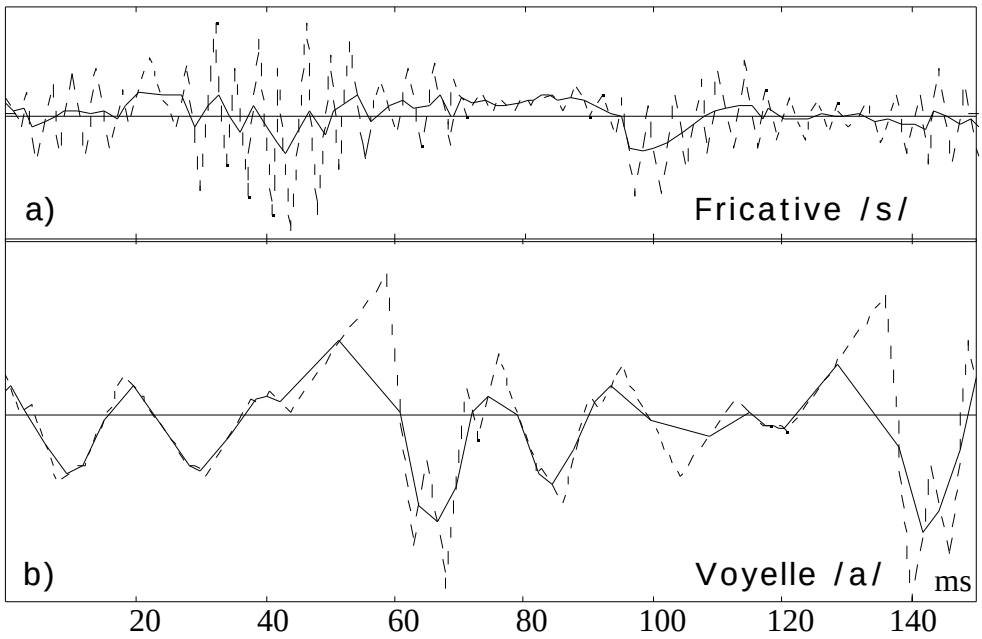

Figure 7. Codage du signal par les milieux des segments. Le signal représenté en traits interrompus est le signal original codé selon ses extremums. La représentation en traits pleins est le signal codé. a) Fricative /s/. b) Voyelle /a/.

La statistique étant calculée sur une faible durée, certains segments sont identifiés de façon erronée et doivent être corrigés. Ceci est généralement le cas pour les fricatives voisées fortement bruitées (ex. figure 8). La procédure appliquée se base sur un critère de durées cumulées. Entre deux segments identifiés correctement, la somme des segments alternés VO et NV est calculée :

$$S_{VO} = \sum |VO|$$
$$S_{NV} = \sum |NV|$$

Lorsque la somme des VO (S_{VO}) est supérieure à S_{NV}, alors tout le segment est déclaré VO, dans le cas contraire il sera déclaré NV.

Les portions de signal de petites amplitudes pour lesquels la méthode ci-dessus ne fonctionne pas sont représentatives des fricatives faibles, des barres de voisement des occlusives voisées, et les barres d'explosion et temps d'établissement du voisement des occlusives sourdes (VOT) qui ne sont pas considérées comme voisées si leur structure ne présente pas une structure régulière. La distinction Voisé / Non-voisé s'effectue dans ce cas par une statistique des passages par zéro du signal, permettant de classer les sons bruités comme Non-voisé (ATAL, 1975), ainsi qu'une mesure de régularité, basé sur la détection des début de cycle de voisement, qui sera décrite à la prochaine section.

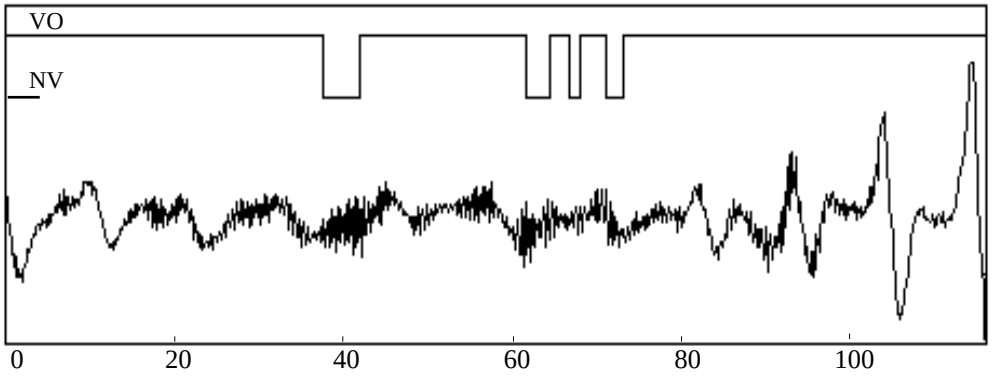

Figure 8. Identification du voisement. Signal supérieur : voisement / non voisement. Signal inférieur : phonème /g/. La présence du bruit de friction fait apparaître de petites erreurs (segments de type NV, dont la durée est petite) et qui doivent être corrigées.

La statistique du nombre de passage par zéro s'effectue à nouveau après élimination des segments courts (II.2) et sur une fenêtre de faible durée (2 ms). La décision de non voisement est prise à l'aide d'un critère de durée identique au cas précédent. Les segments qui ne sont pas identifiés comme non voisés sont traités par la prochaine étape de décision.

Pour terminer, un test de consistance est effectué sur les résultats issus de chaque méthode basé sur des critères de durée minimale de chaque segment.

4) Extraction de la fréquence fondamentale.

Notre algorithme suit les méthodes classiques d'extraction de la fréquence fondamentale par analyse de la structure temporelle du signal (HESS, 1983). Il subdivise en trois parties majeures. La première est un prétraitement, opérant un filtrage passe-bas, destiné à éliminer les composantes de bruit. La seconde est l'analyse de la structure temporelle proprement dite, ayant pour but de déterminer les pics caractéristiques du

début de chaque cycle de voisement par un certain nombre d'identificateurs. La dernière partie est le post-traitement constitué d'une procédure de correction d'erreurs basée sur un critère de régularité locale, ayant pour mission de rajouter, de déplacer ou d'enlever les identificateurs incorrects, du calcul de F_0 proprement dit et d'un lissage final.

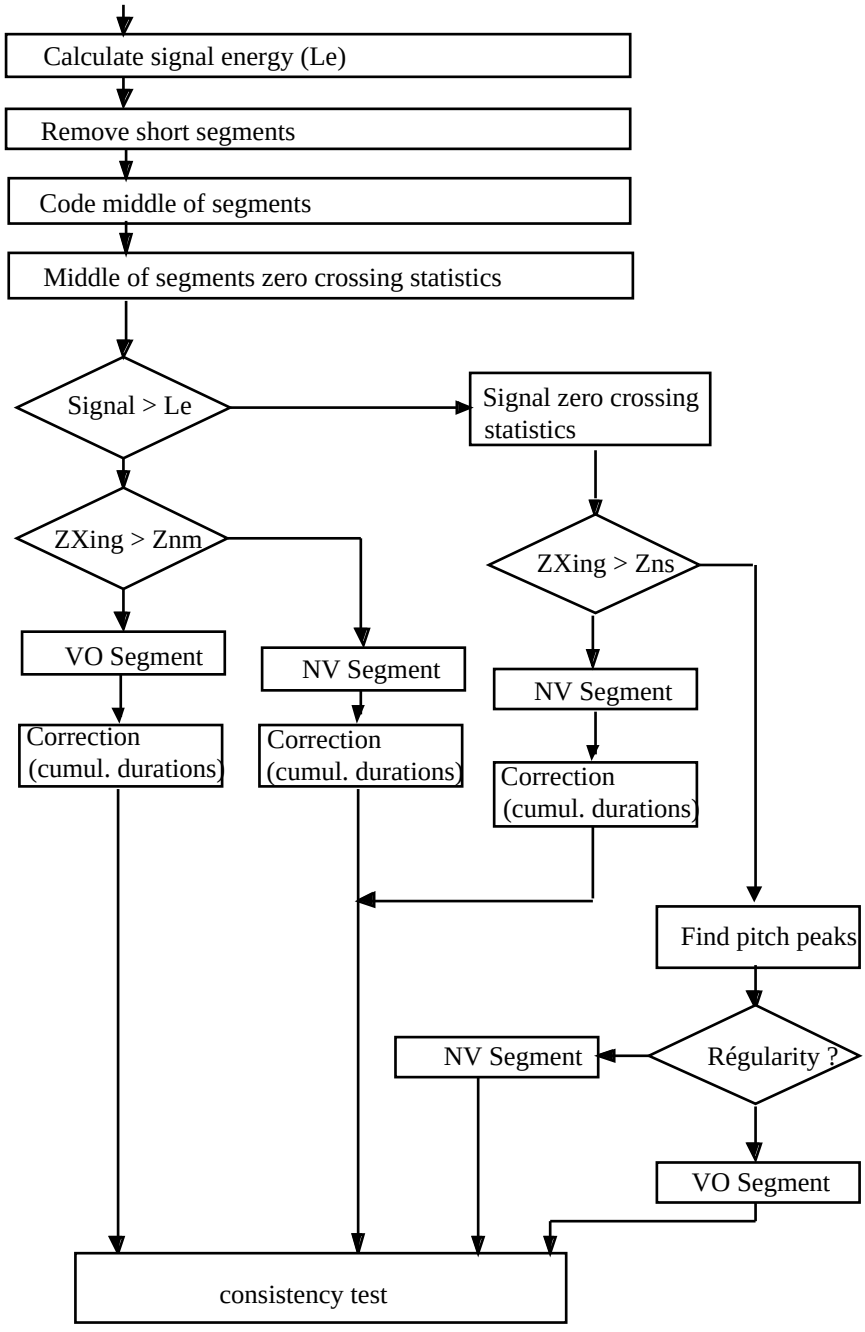

Figure 9. Organigramme : Principe de la détermination des segments Voisé / Non-voisé.

Les méthodes d'analyses temporelles s'avèrent avantageuses dans le cas des signaux pathologiques car l'identification précise de chaque période de voisement permet de suivre le fondamental lorsque celui-ci est sujet à de fortes variations.

a) Détermination des débuts de cycle de voisement.

Les méthodes de détection des débuts de cycle de voisement (DCV) se basent sur la modélisation du conduit vocal en termes de système linéaire passif (FANT, 1960), dont la réponse impulsionnelle est approximée par une somme de sinusoïdes modulées par une exponentielle décroissante. Par conséquent les instants d'ouverture de la glotte se caractérisent par de grandes amplitudes. L'approche que nous avons adoptée utilise deux critères. Le premier est une identification des pics de grande amplitude (GOLD, 1962; GOLD et al., 1969), dépassant un certain seuil. Le second consiste à calculer les cycles d'excursion maximaux (MILLER, 1974), c'est-à-dire la différence d'amplitude entre un pic positif et négatif. Le principe de l'identification est illustré par l'organigramme de la figure 10.

Les pics de début de cycle de voisement que nous sélectionnons sont les pics négatifs. On recherche d'abord un pic d'amplitude positive, supérieur au seuil L_1. Ce seuil, afin d'obtenir une bonne identification, est adaptatif et calculé en chaque instant en tenant compte d'une mesure de l'enveloppe du signal. A l'intérieur d'un intervalle délimité par le pic positif précédent et un autre pic positif dépassant le seuil L_1, le plus négatif et inférieur à un seuil L_2 (également adaptatif) est déterminé. Ce dernier est sélectionné comme identificateurs de fréquence fondamentale (F_0) et le cycle d'excursion est calculé.

La figure 11.b montre un exemple d'une telle identification. Outre les pics de DCV corrects, un certains nombre de pics secondaires ont été identifiés. Cependant leur cycle d'excursion est significativement plus petit et une procédure de correction d'erreurs appropriée permettra de les éliminer.

b) Correction d'erreurs et détermination de F_0.

La procédure de correction d'erreurs permet, outre d'éliminer les identificateurs des oscillations secondaires, de corriger les trous, lorsque un identificateurs est manquant, ainsi que les glissements.

Le principe se base sur un critère de régularité et d'amplitude des cycles d'excursions (HESS, 1983; GOLD, 1962; MILLER, 1974). Le critère de déviation d'une période du fondamental à une autre est relativement large, permettant de détecter des transitions abruptes. Cependant l'application de ce seul critère mènerait à une mesure imprécise du F_0 et

c'est précisément le deuxième critère sur les cycles d'excursion maximaux qui nous permet d'identifier les marqueurs corrects.

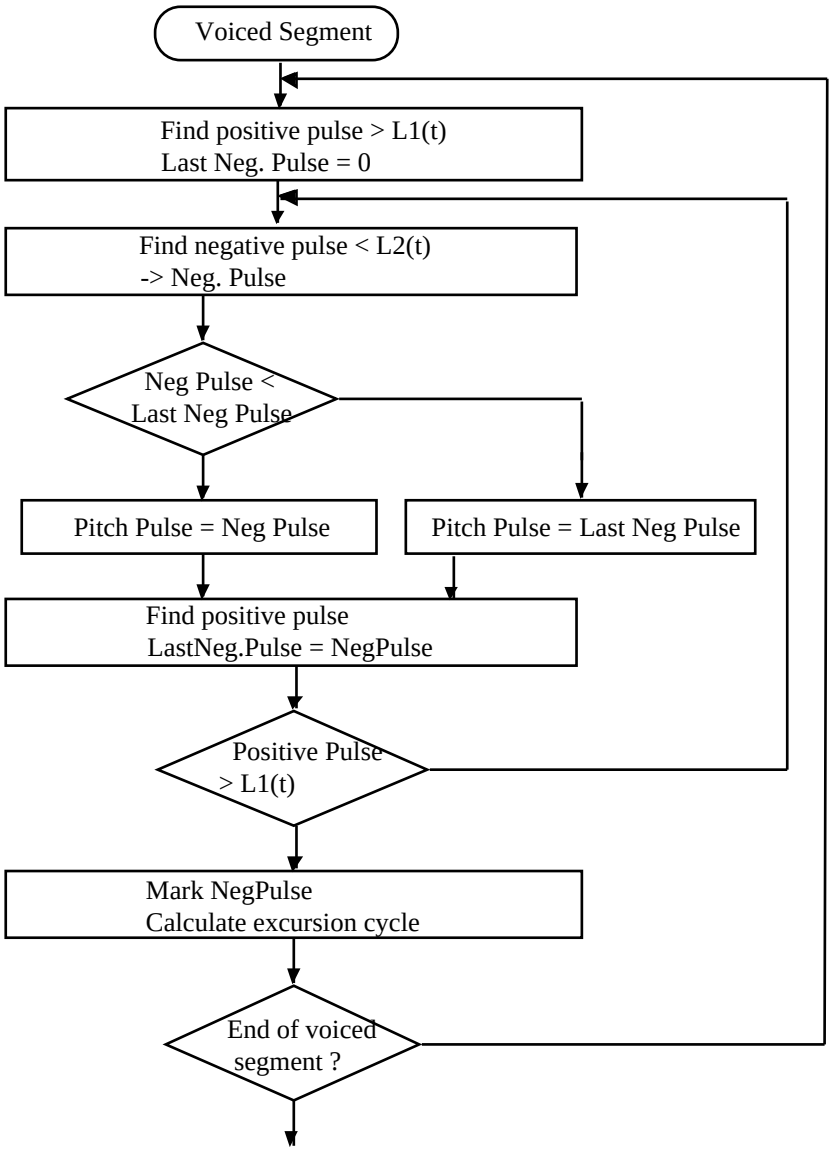

Figure 10. Organigramme : Principe de la détermination des pics de début de cycle de voisement.

L'application de ces principes est représentée sur l'organigramme de la figure 12.

On recherche pour commencer 5 identificateurs satisfaisant au critère de régularité, de sorte à pouvoir estimer de façon sûre la période instantanée du fondamental. La recherche se poursuit par l'identification d'un marqueur probable dans l'intervalle de variation toléré de T_0. S'il existe plusieurs candidats on choisit celui présentant le cycle d'excursion

maximal. Les éventuels trous (marqueurs manquants) sont également corrigés en insérant des identificateurs artificiels.

Lorsque les conditions ne sont plus vérifiées, la recherche est reprise dans une portion stable du signal en recherchant à nouveau 5 marqueurs. Ce type de situation survient en particulier lorsque la structure du signal est modifiée (ex. transition entre deux phonèmes). La figure 11.c montre un exemple où une telle discontinuité fait apparaître un glissement (deux marqueurs proches) dans l'identification des débuts de cycle de voisement. De telles erreurs sont éliminées lors du calcul de F_0, en opérant une interpolation sur la valeur précédente de la fondamentale.

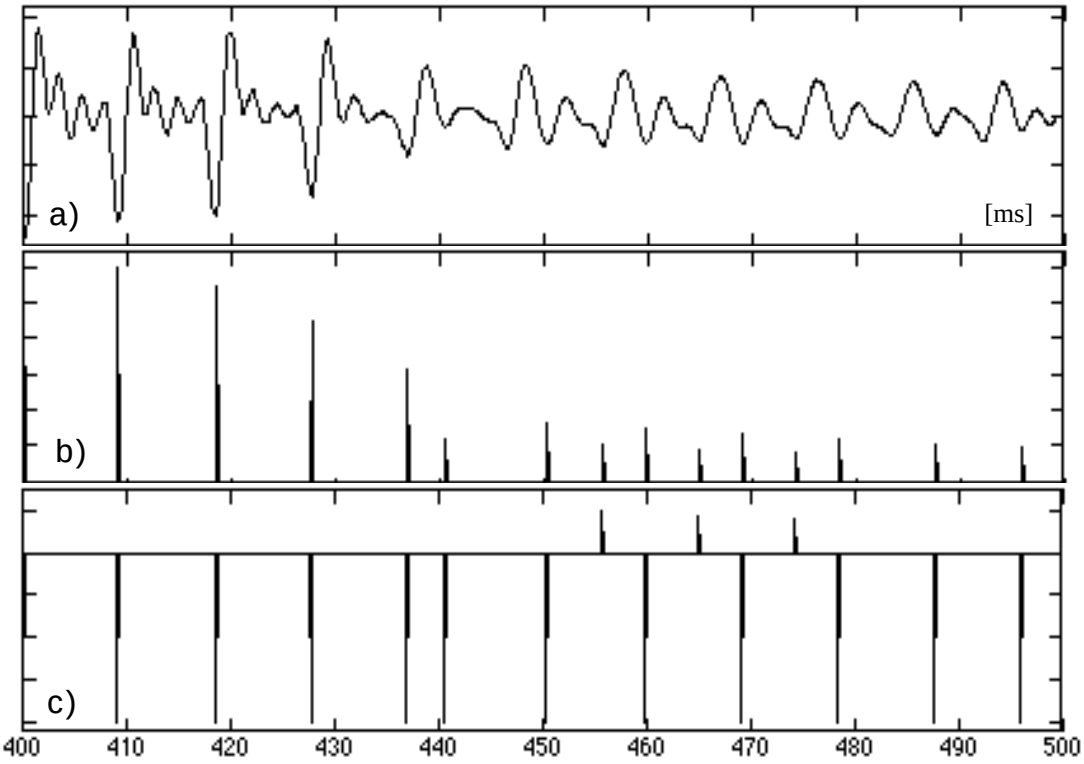

Figure 11. Identification des pics de DCV. a) Signal. b) Pics de DCV identifiés au terme de l'analyse structurelle. L'amplitude de chaque marqueur correspond au cycle d'excursion calculé. c) Identification des marqueurs de F_0 après la procédure de correction des erreurs. Les marqueurs sélectionnés sont représentés par les amplitudes négatives. A noter le glissement introduit vers 440 ms, dû à une modification de la structure du signal.

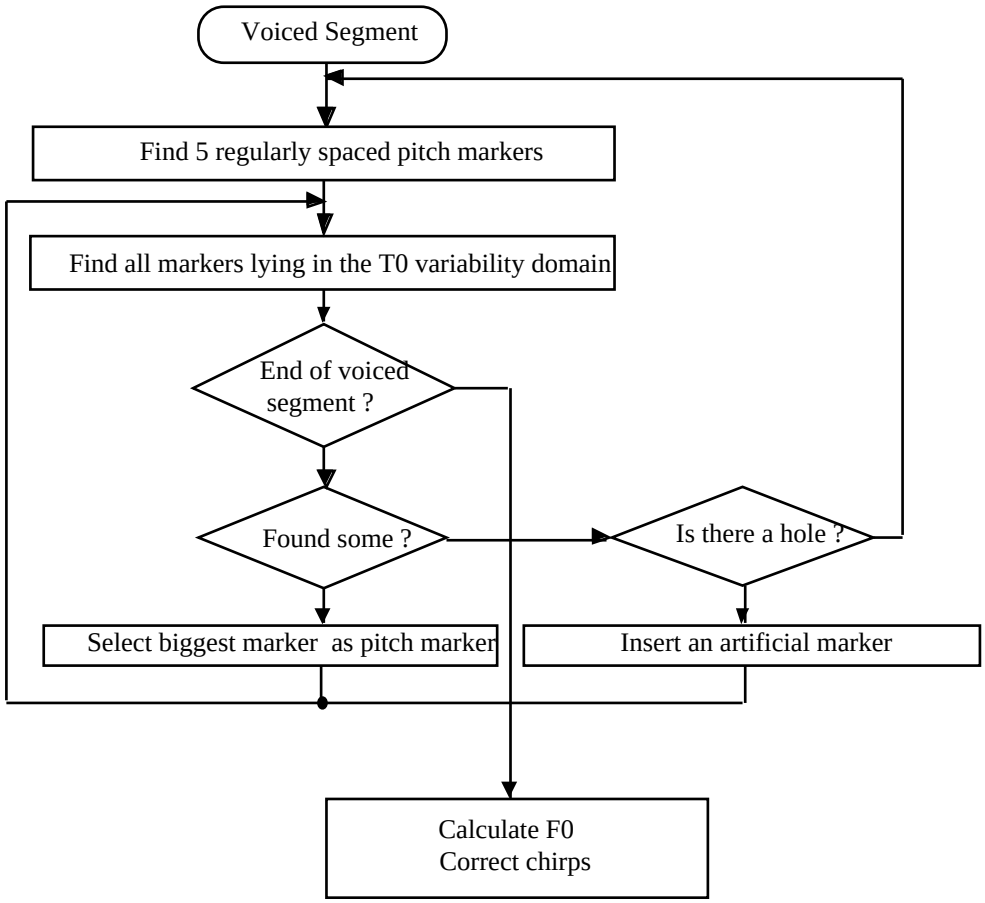

Figure 12. Organigramme : Correction des erreurs et calcul de la fréquence fondamentale.

III. EVALUATION DES SENSEURS.

1) Description des tests.

Les senseurs présentés ont été d'abord appliqués et testés sur des signaux de parole normale, puis dans le contexte de la parole pathologique. Les signaux pathologiques proviennent de patients qui ont été diagnostiqués comme étant affectés par des lésions ponto-cérébelleuses. La plupart d'entre eux souffraient d'une ataxie de Friedreich.

Nous ne présenterons ici que les résultats obtenus dans le cas pathologique. Dans chaque cas l'ensemble testé se composait de 40 signaux provenant exclusivement de locuteurs français, de sexe féminin (7 locuteurs) et masculin (5 locuteurs). Ces signaux se composent de de mots et de phrases et le débit d'élocution est de l'ordre de 6 à 10 phonèmes par seconde. Les signaux furent échantillonnés à 12 kHz et quantifiés sur 16 bits à l'aide d'une carte Audiomedia sur un ordinateur Apple Macintosh.

Deux aspects ont été retenus dans la procédure de test : (i) La fiabilité. (ii) La précision quant à laquelle les instants de transition peuvent être identifiés. Le test de fiabilité vise à déterminer si l'algorithme est en mesure de détecter la présence d'un segment particulier. Dans ce cas le senseur fournit un résultat correct lorsqu'il identifie les transitions d'un segment vers un autre (par ex. fin de silence / début de parole) et qu'il ne détecte pas des transitions supplémentaires ne devant pas faire partie du signal.

Il est à noter que les méthodes utilisées sont particulièrement sensibles aux conditions d'enregistrement. Le niveau de bruit de fond ainsi que des taux d'échantillonnage différents de ceux utilisés ici nécessitent une nouvelle adaptation des paramètres.

2) Senseur Silence/Parole.

La figure 13 présente un exemple d'identification de segments Silence / Parole.

Parmi les 40 signaux testés nous avons dénombré manuellement 154 transitions Silence/Parole. Le senseur est en mesure d'identifier 95% de ces transitions de façon correcte. Les segments sont repérés correctement et dans leur intégralité. La distinction entre silences en début et en fin de signal, et silences occlusifs se fait de façon certaine dans ces cas.

Les erreurs apparaissent principalement dans le cas des silences occlusifs très courts. Le critère de durée minimale de ces segments est dans ce cas trop grand et ne permet pas leur identification. Cette situation est caractéristique des signaux pour lesquels les critères de silence, sur lesquels se base notre algorithme, ne sont pas suffisamment bien marqués pour pouvoir être identifiés correctement. D'autre part il peut arriver, dans le cas de signaux de très faible amplitude, que l'on identifie un segment de silence lorsque celui-ci n'existe pas.

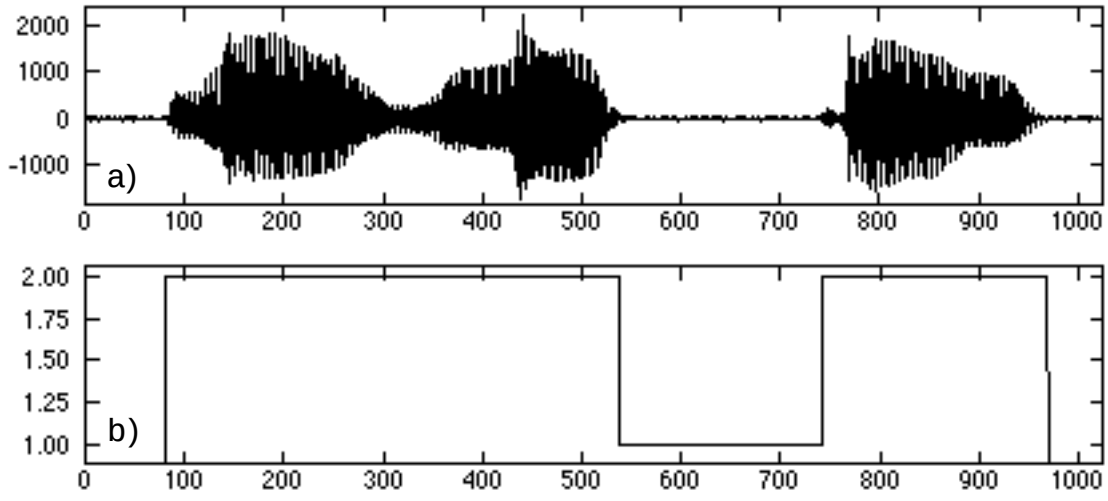

Figure 13. Extraction des segments Silence et Parole. a) Signal : “l'enveloppe”, locuteur féminin. b) Identification des segments Silence (niveau haut) et Parole (niveau bas). La précision des instants de transition est de l'ordre de 3 millisecondes.

Il est cependant à noter que la détection des silences occlusifs se fait uniquement lorsque ceux-ci sont franchement présents, en raison des critères de distinction que nous appliquons dans notre algorithme. Certains types de pathologies montrent un contrôle insuffisant dans la fermeture de la glotte (figure 14), au quel cas notre algorithme n'est pas en mesure de le détecter.

La précision est généralement bonne. Il s'agit cependant de distinguer plusieurs cas. Les segments de parole débutant avec du voisement ou une transition brusque sont repérés avec une précision de l'ordre de 3 ms. La précision de l'identification de la fin d'un segment de parole est parfois plus critique. En fin de phonation le relâchement lent des organes articulatoires est la cause d'une oscillation de faible amplitude, dont la durée s'étend sur plusieurs périodes de voisement. Il est souvent difficile de définir un instant précis de fermeture. Dans ce cas la précision se chiffre entre 5 et 7 ms. Retenons que la même différence de fiabilité a été observée dans des expériences de mesures manuelles, effectuées dans le laboratoire précédent du troisième auteur: pour le début de l’énoncé, la fiabilité inter-juges avec un critère de 1 ms était de l’ordre de 98%, tandis qu’elle ne chiffrait qu’aux alentours de 88-92% pour les fins d’énoncé.

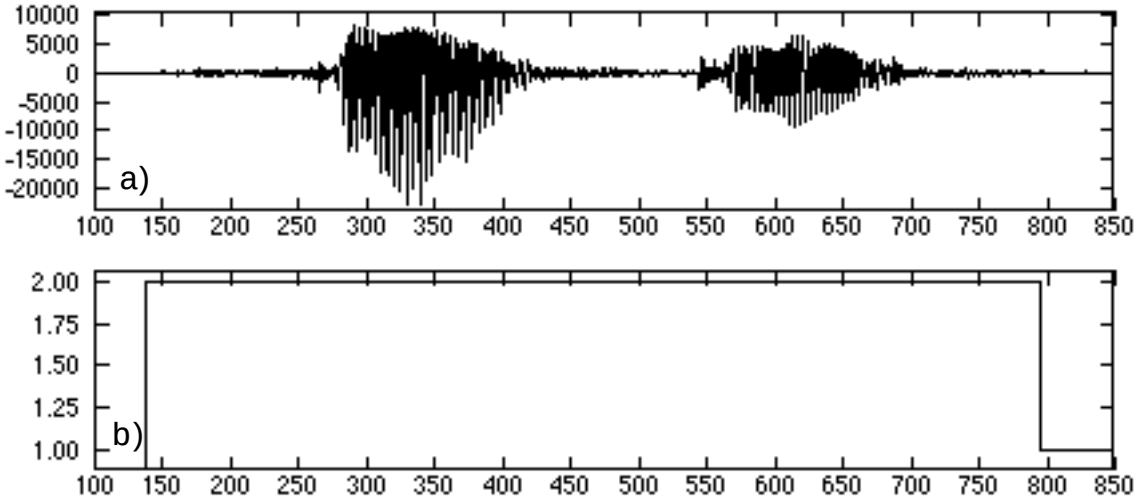

Figure 14. Occlusion non totale. a) Signal : “sa pipe”, locuteur féminin. b) Identification des segments Silence (niveau haut) et Parole (niveau bas). L'occlusion du /p/ n'est pas totale et le segment est identifié de type Parole. Perceptivement le phonème est plutôt perçu comme un /b/. On peut observer une très brève occlusion précédant immédiatement l'explosion, mais dont la durée est trop faible pour être détectée.

3) Senseur de friction.

Les phonèmes fricatifs et affriqués sont identifiés correctement dans l'ensemble des cas examinés. Les phonèmes critiques, tel que les fricatives voisées faiblement bruitées sont reconnues correctement (figure 16). En choisissant un ordre N du filtre de lissage adéquat il est également possible d'identifier des segments fricatifs de très courte durée. Les bruits de friction de courtes durées déterminés sur la barre d'explosion des occlusives sont également repérés correctement. La figure 17 démontre la puissance de cet algorithme. Moyennant un faible lissage de la statistique, il est possible d'identifier la barre d'explosion du phonème /g/, dont la durée s'étend sur 8 ms.

Il existe cependant quelques cas posant des problèmes. Les voyelles antérieures fermées /i/ /e/ et /ɛ/ sont parfois identifiées comme fricatives. Pour ces voyelles, la position antérieure de la langue, proche du palais peut créer un rétrécissement de la cavité buccale occasionnant ainsi l'apparition d'une turbulence aérodynamique caractéristique de la friction. Ces phonèmes présentent donc souvent les caractéristiques des fricatives. De même les barres de voisement (des occlusives sonores par exemple), dont le contenu fréquentiel est concentré dans les basses fréquences sont très sensibles à la présence de bruit parasite issu de l'enregistrement.

La précision à laquelle on détecte la transition entre un segment Fricatif et un segment Non-fricatif dépend des cas. Les passages francs entre un segment bruité et affriqué, tel que la barre d'explosion des occlusives, sont déterminés avec une précision que l'on peut estimer à environ 2 ms. Par contre lorsque la friction s'installe progressivement, cette transition n'est pas toujours bien marquée. En particulier lorsque le bruit de constriction est faible et se superpose à un voisement important, il se peut que son identification survienne avec un certain retard. Ceci est dû à l'algorithme d'élimination des segments courts qui supprime parfois certains segments courts de type fricatif lorsqu'ils sont précédés ou suivis d'un segment long.

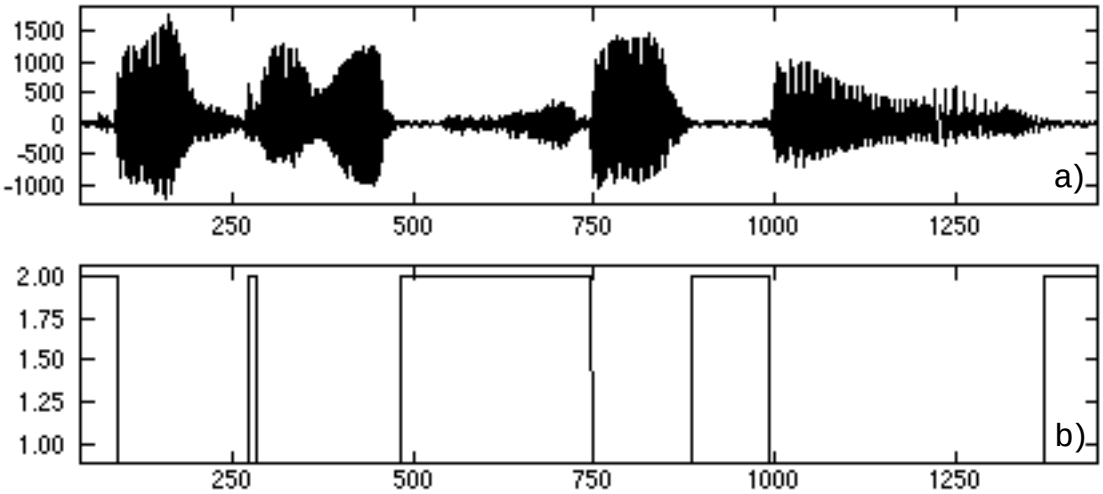

Figure 15. Identification des segments Fricatifs et Non-fricatifs. a) Signal : “ta grippe s'empire”, locuteur féminin. b) Identification des segments Fricatif (niveau haut) et Non-fricatif (niveau bas). Les phonèmes fricatifs ainsi que le bruit de friction sur l'explosion du /p/ sont identifiés avec une précision de l'ordre de 2 ms.

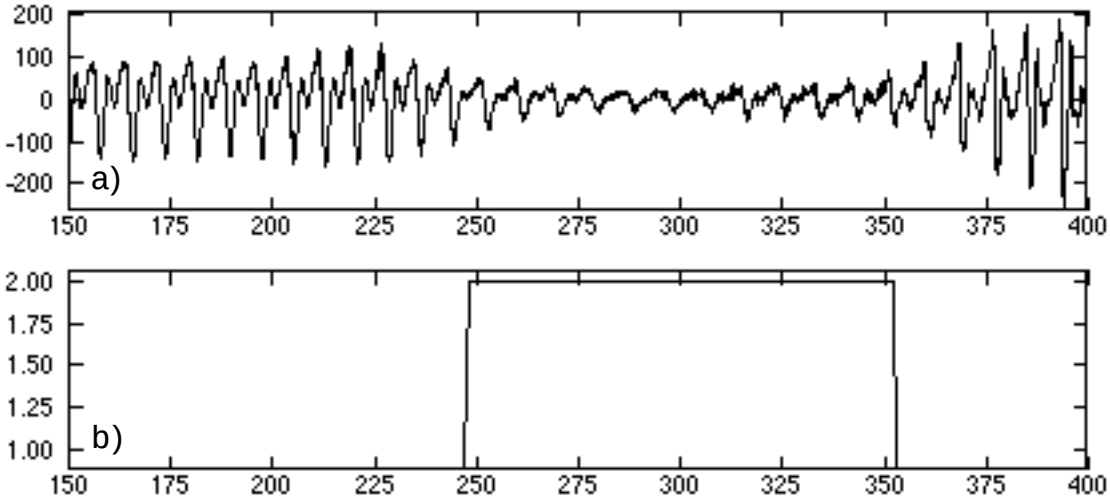

Figure 16. Identification des fricatives voisées. a) Signal : détail du phonème /v/ dans le mot “niveau”; locuteur masculin. b) Identification des segments Fricatif (niveau haut) et Non-fricatif (niveau bas). La faible friction superposée au voisement est identifiée correctement.

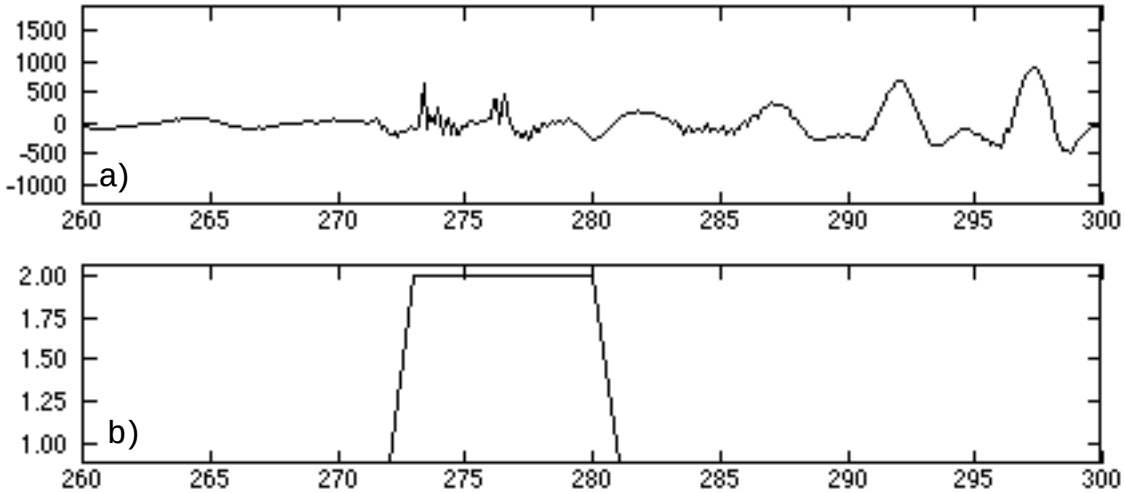

Figure 17. Identification de segments fricatifs de courte durée. a) Signal : détail de la barre d'explosion du phonème /g/ dans le mot “ta grippe s'empire” (exemple de la figure 16). b) Identification des segments Fricatif (niveau haut) et Non-fricatif (niveau bas). Le choix d'un ordre adéquat pour le filtre de lissage permet de détecter la barre d'explosion très courte du phonème /g/, de durée 8 ms.

4) Senseur de voisement.

L'ensemble du corpus des signaux pathologiques analysés présente 182 transitions Voisé / Non-voisé. Parmi ceux-ci le senseur était en mesure d'en identifier 90% de façon correcte. L'algorithme présente de bonnes performances pour diverses perturbations en parole pathologique du contrôle glottique, tel que les voix bitonales, les voix étranglées ou dans le cas du “pitch break”. Par contre les performances observées étaient plutôt médiocres dans les conditions de voix rauque (“raspy voice”), en particulier pour les signaux de faible amplitude. Ce type d'irrégularité dans la structure harmonique est difficilement identifiable par les méthodes utilisées pour les signaux de faible amplitude (II.3), c'est-à-dire ne pouvant appliquer la méthode de codage par les milieux des segments. D'autre part il est à noter que la méthode utilisée pour les signaux voisés de faibles amplitudes ne mesure que des oscillations régulières. Lorsque le voisement n'est pas suffisamment marqué, l'algorithme est mis en déroute.

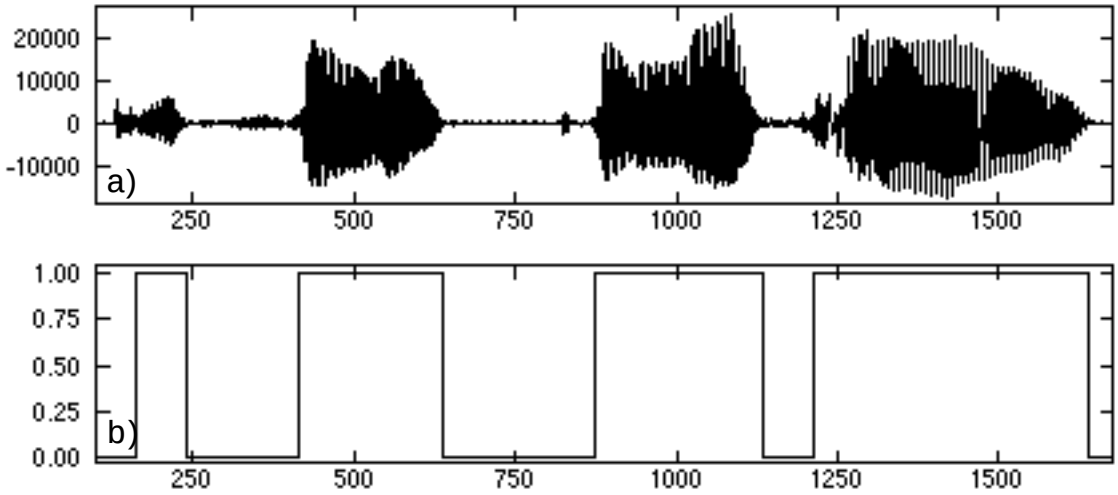

Figure 18. Identification des segments Voisé / Non-voisé. a) Signal : “A-t'-il téléphoné”, locuteur masculin. b) Identification des segments Voisé (niveau haut) et Non-voisé (niveau bas). L'identification du début et de la fin du voisement se fait au plus tard sur la deuxième période de voisement.

Un avantage notable de la méthode du codage des passages par zéro du signal codé par le milieu des segments est l'identification correcte du voisement lorsque le signal est fortement fricatif. Par exemple la figure 19 montre que le voisement est toujours identifiable dans une situation assez fortement bruitée.

Du point de vue précision, les transitions sont en général détectées sur la première ou la deuxième période de voisement. Pour les passages de faible amplitude les performances sur la précision des instants de transition, peuvent être légèrement dégradées, en fonction du type de pathologie rencontrée.

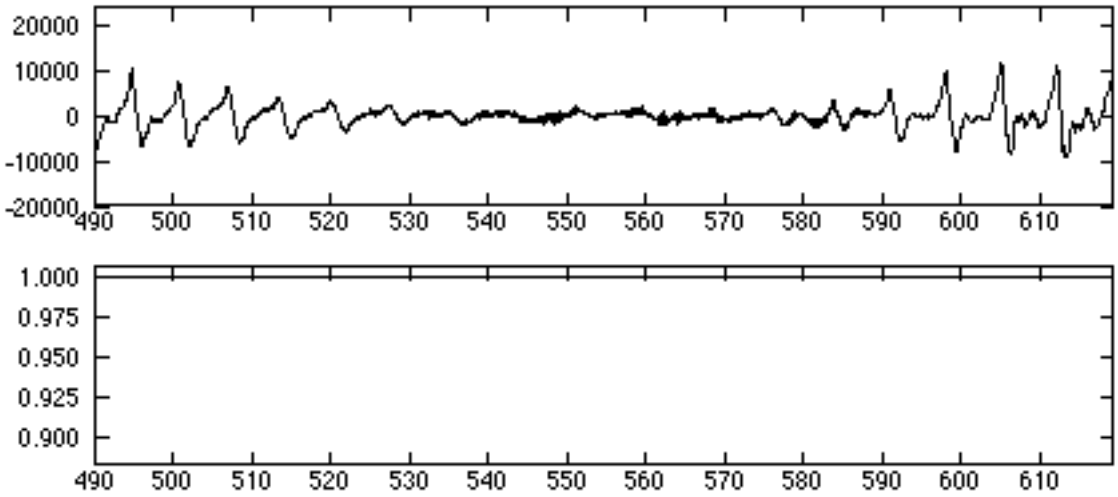

Figure 19. Identification des fricatives voisées. a) Signal : phonème /g/ du mot “donjon”, locuteur masculin. b) Identification des segments Voisé (niveau haut) et Non-voisé (niveau bas). La fricative voisée /g/ fortement bruitée est identifiée

correctement à l'aide de la méthode des passages par zéro du signal codé par les milieux des segments.

5) Extraction de la fréquence fondamentale.

L'algorithme d'extraction de la fréquence fondamentale a été testé selon deux critères. L'extraction est considérée comme étant correcte lorsque le senseur est en mesure de poursuivre le fondamental quand il peut être mesuré. D'autre part la mesure est considérée comme précise si l'erreur est inférieure de 20% par rapport à la valeur mesurée manuellement. Pour ces conditions de test, l'algorithme fournit des résultats corrects et précis pour les signaux de parole pathologique dans environ 85% des segments voisés. La figure 20 présente un exemple d'extraction.

Les erreurs d'imprécision sont principalement marquantes pour les phonèmes fricatifs voisés, où les pics de début de cycle de voisement ne peuvent être précisément identifiés en utilisant une méthode par analyse structurelle du signal. D'autre part certains glissements en fréquence peuvent également introduire des imprécisions. Ce défaut à déjà été souligné au chapitre II.4.b. où des modifications de la forme de l'onde peuvent présenter un décalage du pic le plus négatif. Dans ce cas le critère de régularité, que nous avons choisi de façon large, afin de pouvoir suivre des transitions rapides, ne peut détecter ce glissement et résultera en un saut brusque de la fondamentale. Le choix d'un critère de variation large permet cependant de suivre précisément les variations brusques et rapides de la fondamentale, pour lesquels l'algorithme présente de bon résultats. La figure 21 présente l'exemple d'un “pitch break”, dans lequel la fondamentale passe de 235 Hz à 365 Hz en l'espace de 20 ms.

Par contre dans les situations où la périodicité du signal n'est pas suffisamment marquée, tel que les voix bitonales ou présentant une diplophonie importantes, le senseur de pitch montrera de faibles performances.

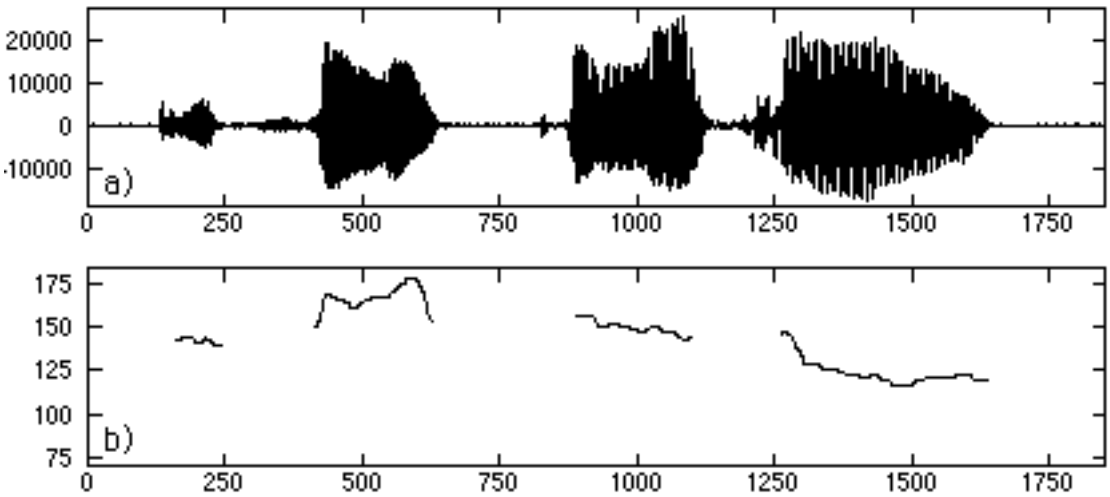

Figure 20. Mesure de la fréquence fondamentale. a) Signal : phrase “a-t-il téléphoné”, locuteur masculin. b) Fréquence fondamentale en Hz.

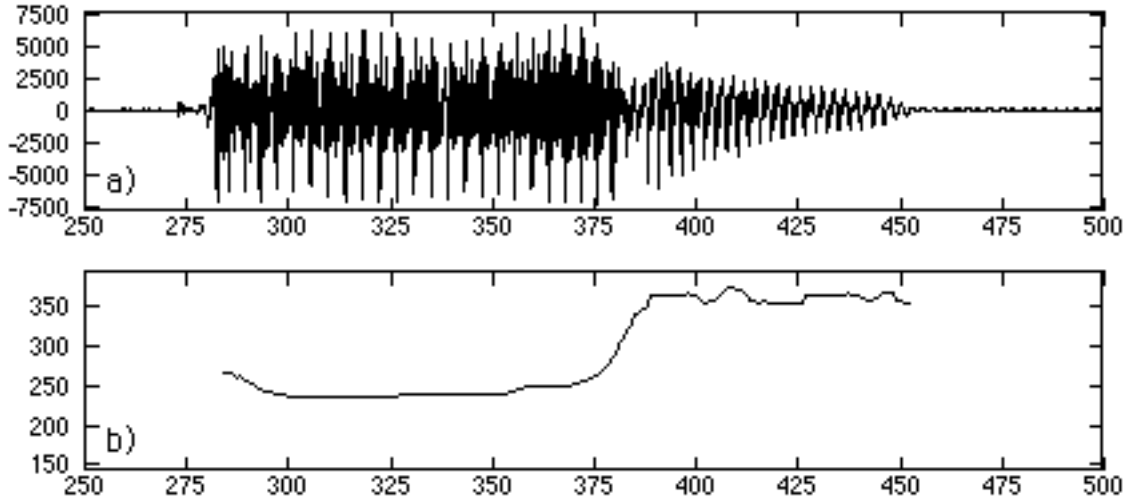

Figure 21. Pitch break. a) Signal : séquence “pa” répétée, locuteur féminin. b) Fréquence fondamentale en Hz. F_0 passe brusquement de 235 Hz à 365 Hz environ.

IV. CONCLUSION.

Ce travail avait pour but principal de combler une déficience actuelle quant aux algorithmes fondamentaux d'analyse en parole pathologique. Différents travaux ayant relié différents types de dysarthries à des défaillances dans les durées syllabiques et dans la fréquence fondamentale, nous avons élaboré quatre senseurs destinés à extraire les segments primaires du signal de la parole. Ce sont les segments Silence / Parole, Fricatif / Non-fricatif, Voisé / Non-voisé et la fréquence fondamentale. Nous avons montré que l'utilisation de techniques basées sur l'analyse de la forme structurelle du signal était en mesure de combler ce vide. Les

résultats obtenus montrent que ces méthodes représentent des outils puissants qui sont en mesure d'identifier de façon fiable et précise les différents segments primaires en parole pathologique.

Il est à souhaiter que d'autres recherches systématiques sur ces techniques et impliquant un large ensemble de syndromes différents, permette de développer un ensemble d'algorithmes robustes que chercheurs et cliniciens pourront utiliser pour évaluer les troubles de la parole.

REMERCIEMENTS.

Les auteurs désirent remercier monsieur le professeur Mario Rossi et son équipe, du Laboratoire d'Electromagnétisme et Acoustique de l'Ecole Polytechnique Fédérale de Lausanne (EPFL). Les patients ont été enregistrés avec la permission du Dr. M. Botez, Hôtel-Dieu de Montréal. Ce travail a été subventionné par un subside UNIL-EPFL.

********** J'ai changé tous les titres des publications anglaises de manière à mettre les mots lexicaux en majuscules, comme c'est la norme *********

REFÉRENCES.

********** MAJUSCULES SUR LE PROCHAIN TITRE *******

Atal, Bishnu S., Rabiner, Lawrence R., (1975), Voiced-unvoiced decision without pitch detection, JASA vol. 58 No. 4.

ATAL, Bishnu S., RABINER, Lawrence R., (1976), *A pattern recognition approach to voiced- unvoiced -silence classification with application to speech recognition*, IEEE ASSP-24 No. 3.

BAUDRI, Marc, (1978), *Analyse du signal vocal dans sa représentation amplitude-temps. Algorithmes de ségmentation et de reconnaissance de la parole*, Thèse Université Pierre et Marie Curie, Paris 6.

BENOIT, Christian, (1986), *Note on the use of correlation in speech timing*, JASA vol. 60 No. 12.

BOITE, René, KUNT, Murat, (1987), *Traitement de la parole*, Complément au traité d'électricité EPFL-PPR.

CALLIOPE, (1989), *La parole et son traitement automatique*, Masson.

CANTER, Gerald J., (1963), *Speech characteristics of patients with Parkinson's disease : I intensity, pitch and duration*, Journal of Speech and Hearing Disorders vol. 28.

CANTER, Gerald J., VAN LANCKER, Diana, (1985), *Disturbances of the temporal organisation of speech following bilateral thalmic surgery in a patient with Parkinson's disease.* Journal of Communication Disorders vol. 18.

DARLEY, Frederic L., ARONSON, Arnold E., BROWN, Joe R., (1969), *Clusters of deviant speech dimensions in the dysarthrias*, Journal of Speech and Hearing Research vol. 12.

FANT, Gunnar, (1960), *Acoustic theory of speech production*, Mouton the Hague.

GALLAGHER, Neal C., (1981), *A theoretical analysis of the properties of median filters*, IEEE ASSP-29 No. 6.

GOLD, Bernard (1962), *Computer program for pitch extraction,* JASA vol. 34 No. 7.
GOLD, Bernard, RABINER, Lawrence R., (1969), *Parallel processing for estimating pitch periods of speech in the time domain*, JASA vol. 46.
HESS, Wolfgang, (1976), *A pitch-synchronous digital feature extraction system for phonemic rcognition of speech*, IEEE ASSP-24 No. 1.
HESS, Wolfgang, (1983), *Pitch determination of speech signals: Algorithmes and devices*, Springer-Verlag.
ITO, Mabo R., DONALDSON, Robert W., (1971), *Zero crossing mesasurements for analysis and recognition of speech sounds*, IEEE AU-19 No. 3.
KELLER, Eric, *An expert system for the acoustic analysis of speech disorders*, in Ph. Martin (éd.), *Mélanges Léon: Hommages à Pierre Léon* (pp. 211-230). Toronto: Éditions Mélodie.
KELLER, Eric, VIGNEUX, Patrick, LAFRAMBOISE, Martine (1991), *Acoustic analysis of neurologically impaired speech*, British Journal of the Disorders of Communication, vol. 26.
LEHISTE, Ilse, (1965), *Some acoustic characteristics of dysartria*, Basel : Bibliotheca Phonetica.
MILLER, Norbert, (1974), *Pitch detection by data reduction*, IEEE Symp. Speech Recogn.
NODES, Thomas A., GALLAGHER, Neal C., (1982), *Median filters: Some modifications and their properties*, IEEE ASSP-30 No. 5.
RABINER, Lawrence.R., SAMBUR, Marvin R., SCHMIDT, Carolyn E., (1975), *Application of nonlinear smoothing algorithmes to speech processing,* IEEE ASSP-23 No. 6.
REDDY, Donald R., (1966), *Segmentation of speech sounds*, JASA vol. 40.
RODET, Xavier, (1977), *Analyse du signal vocal dans sa représentation amplitude-temps. Synthèse de la parole par règles,* Thèse Université Pierre et Marie Curie, Paris 6.
SCARR, Robert, (1968), *Zero crossing as a mean of obtaining spectral information in speech analysis*, IEEE AU-16 No. 2.
SIEGEL, Leah J., (1979), *A procedure for using pattern recognition classification techniques to obtain a voiced-unvoiced classifier*, IEEE ASSP-27 No. 1.
SIEGEL, Leah J., (1982), *Voiced/unvoiced/mixed excitation classification of speech*, IEEE ASSP-30 No 3.
STYGER, Thomas, (1992), *Reconnaisseur de durées syllabiques en parole pathologique*, Projet de diplôme, Ecole Polytechnique Fédérale de Lausanne, Laboratoire d'Electromagnétisme et Acoustique.