



**HAL**  
open science

## L'évaluation acoustique

Alain Ghio

► **To cite this version:**

Alain Ghio. L'évaluation acoustique. Auzou P.; Rolland V.; Pinto S.; Ozsancak C. Les dysarthries, Solal, 2007. hal-01616693

**HAL Id: hal-01616693**

**<https://hal.science/hal-01616693v1>**

Submitted on 17 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# L'évaluation acoustique

A. Ghio - Ingénieur de recherche

Laboratoire Parole et Langage, CNRS UMR 6057  
Université de Provence  
29, avenue Robert Schuman  
13621 Aix-en-Provence, France

Courriel : alain.ghio@jpl.univ-aix.fr

## I. Principes et présupposés de l'évaluation acoustique

### 1. Les concepts liés à l'évaluation

Comme le définit Condamines (1985), « qui dit évaluation, dit chiffrage, donc passage de la comparaison qualitative à la comparaison quantitative ». Indubitablement, l'évaluation est intimement liée à la mesure, opération qui consiste à donner une valeur à une observation. De façon sous-jacente, la mesure fait intervenir une notion de modèle qui est une représentation abstraite, simplifiée, d'un phénomène et qui peut se ramener à des paramètres et des grandeurs. Par exemple, dans le cas d'une voiture, le modèle correspondant à cet objet peut être résumé par un ensemble de grandeurs définies : longueur, largeur, hauteur, masse, puissance, couleur... La mesure ou, plus généralement, la métrologie couvre les méthodes et techniques qui permettent de paramétrer un modèle destiné à représenter la réalité. Une fois cette modélisation établie, l'objet et son modèle peuvent être étudiés, manipulés, comparés.

Des modèles de production de la parole existent depuis longtemps. Le lecteur pourra se référer aux chapitres correspondants dans cet ouvrage. Par contre, des modèles rendant compte de la physiopathologie de la production de parole restent parcellaires, incomplets et en tout cas, insuffisamment annoncés pour permettre de clairement établir les mesures utiles à l'évaluation. Il n'est pas certain que l'orthophoniste, l'ORL ou le neurologue attendent la même chose d'une évaluation car la voix et la parole ne sont pas, pour eux, le même objet et donc, ils ne font pas nécessairement référence à un même modèle sous-jacent. Pour l'orthophoniste, dont le rôle est la

rééducation, chaque patient est son propre « modèle » et quelques mesures longitudinales bien choisies, liées au problème du patient, peuvent suffire à l'évaluation. De son côté, l'ORL peut vouloir effectuer une évaluation du grade de la dysphonie de façon à orienter la thérapie en conséquence ou encore pour évaluer a posteriori l'impact d'un acte chirurgical. Il s'agit plus alors d'une tâche de classification de groupe. Enfin, le neurologue peut être amené à entreprendre une évaluation des troubles de la parole, non pas comme une finalité en soi, mais comme un révélateur ou un marqueur d'une dégénérescence neurologique ou encore pour mesurer l'effet d'une pharmacopée ou d'une stimulation électrophysiologique. L'évaluation acoustique de la parole devient ainsi un moyen très indirect d'appréhender son modèle de dysfonctionnement neurologique général.

### 2. L'acoustique est une projection non linéaire de la biomécanique

Dans le processus de communication parlée, le canal acoustique est le support essentiel porteur d'information. Mais il ne faut pas oublier qu'il ne s'agit que d'un support, que d'une projection résultant des processus cognitifs et des mouvements articulatoires du locuteur. Si l'objectif de l'analyse acoustique est de quantifier la « qualité » du résultat sonore pris comme élément principal de la communication parlée, notamment en testant, par exemple, l'intelligibilité ou plus largement le processus de transmission d'information, la méthode est directe. Si, par contre, par l'analyse acoustique, l'observateur cherche un marqueur sensible des mécanismes biomécaniques, articulatoires ou cognitifs, il faut rester prudent car le lien articulation/acoustique n'est pas une simple relation bijective. On pourrait même faire référence à l'allégorie de la caverne de Platon dans laquelle

l'acoustique serait l'ombre projetée sur les murs de la grotte des phénomènes biomécaniques de production de parole. Or, il est bien connu que cette projection n'est qu'un reflet déformé de la réalité.

### 3. Intérêt et limites de l'évaluation acoustique

L'analyse acoustique permet de faire des mesures, obtenir des indices, dégager des tendances, et ce, de façon quantitative et objective. On l'oppose souvent au jugement perceptif, éminemment subjectif, variable d'un auditeur à l'autre ou changeant selon le contexte pour un même auditeur. Dans son ouvrage de référence, Baken (1987) affirme que « de telles méthodes d'évaluation (perceptives) sont inadéquates... incapables de mettre en évidence de petites différences ». Il cite ainsi une liste de références ayant montré que « les jugements perceptifs de hauteur sont non fiables ». De même, Morsomme et Estienne (2006) confirment que si « notre premier outil d'évaluation acoustique est l'oreille humaine... l'audition est insuffisante... (car) elle traite le signal acoustique dans sa globalité. L'oreille peut donc être dupée ». Aussi, l'analyse acoustique s'avère finalement comme un moyen plus crédible d'évaluer et mesurer les dysfonctionnements de la voix et de la parole. Cependant, ces techniques ne sont pas non plus infaillibles car, comme tout procédé de mesure, elles sont sujettes à des erreurs, à des problèmes de non reproductibilité, à de la matière première inadaptée. En effet, pour le moment, la plupart des études acoustiques sur les dysfonctionnements laryngés s'appuient sur la production de voyelles tenues, exercice qui permet d'obtenir une fiabilité de mesure mais qui s'éloigne des situations naturelles de communication.

## II. Précautions techniques nécessaires à l'évaluation acoustique

### 1. Les conditions d'enregistrement nécessaires à l'évaluation acoustique

L'analyse acoustique de la voix et de la parole nécessite une bonne qualité du signal sonore, particulièrement dans l'étude des dysfonctionnements où il s'avère parfois nécessaire de mesurer de fines nuances qui ne doivent pas être entachées de bruit. Actuellement, la diffusion très large d'ordinateurs

équipés en série d'entrées audio autorisant l'enregistrement sonore sans difficulté ne doit pas occulter le fait que certaines précautions doivent être prises. Typiquement, il est déconseillé d'utiliser un microphone bas de gamme connecté directement sur l'entrée micro de l'ordinateur : le résultat pourra être convenable à l'oreille mais les logiciels d'analyse auront des difficultés à fournir des mesures correctes sur des données altérées (Fig. 1). Il est préférable, d'une part d'utiliser des microphones de qualité, le plus souvent électrostatiques. D'autre part, il est conseillé de connecter le microphone sur un préamplificateur externe et non directement à l'ordinateur. En effet, ces dispositifs périphériques (p. ex. : Edirol USB Audio Capture UA25) génèrent beaucoup moins de bruit et, comme la plupart intègrent à présent les convertisseurs analogique-numérique, l'expérimentateur récupère alors directement le signal numérisé, évitant ainsi de multiples manipulations génératrices de problèmes.

### 2. Le problème du calibrage de l'intensité

Le sonomètre reste l'instrument idéal pour mesurer l'intensité sonore car il fournit une mesure SPL calibrée (voir dans cet ouvrage le chapitre « Onde sonore »). Il permet de faire des études comparatives fiables sur le niveau sonore. Ainsi, par exemple, il est possible de mesurer le fait que tel locuteur a augmenté de 6 dB sa puissance sonore en 3 mois de rééducation. A partir du moment où le signal sonore est enregistré sur un quelconque support (magnétique, informatique...) et/ou avec un microphone dont on ne connaît pas la sensibilité, ce type d'analyse comparative devient délicat. En effet, si les caractéristiques de la chaîne d'enregistrement constituée par le microphone, le préamplificateur et le convertisseur numérique ne sont pas parfaitement connues, il est impossible de connaître la valeur SPL finale de l'enregistrement, en particulier s'il est stocké sous forme numérique en fichier (p. ex. : fichier wav). En effet, l'amplitude observée du signal dépend du microphone utilisé, du volume d'enregistrement utilisé et de divers autres réglages. Autrement dit, on peut obtenir pour un locuteur à voix forte un fichier audio avec de faibles valeurs d'amplitude tout simplement parce qu'on a utilisé un microphone à faible sensibilité, placé loin du locuteur, avec un faible volume d'enregistrement. Inversement, un locuteur hypophonique peut fournir un fichier audio avec de fortes valeurs d'amplitude si a été utilisé un microphone sensible placé à 10 cm du locuteur avec un enregistreur à fort gain.

Hormis le sonomètre, il existe deux moyens d'obtenir des données sonores permettant la mesure calibrée en intensité. La première consiste à maîtriser

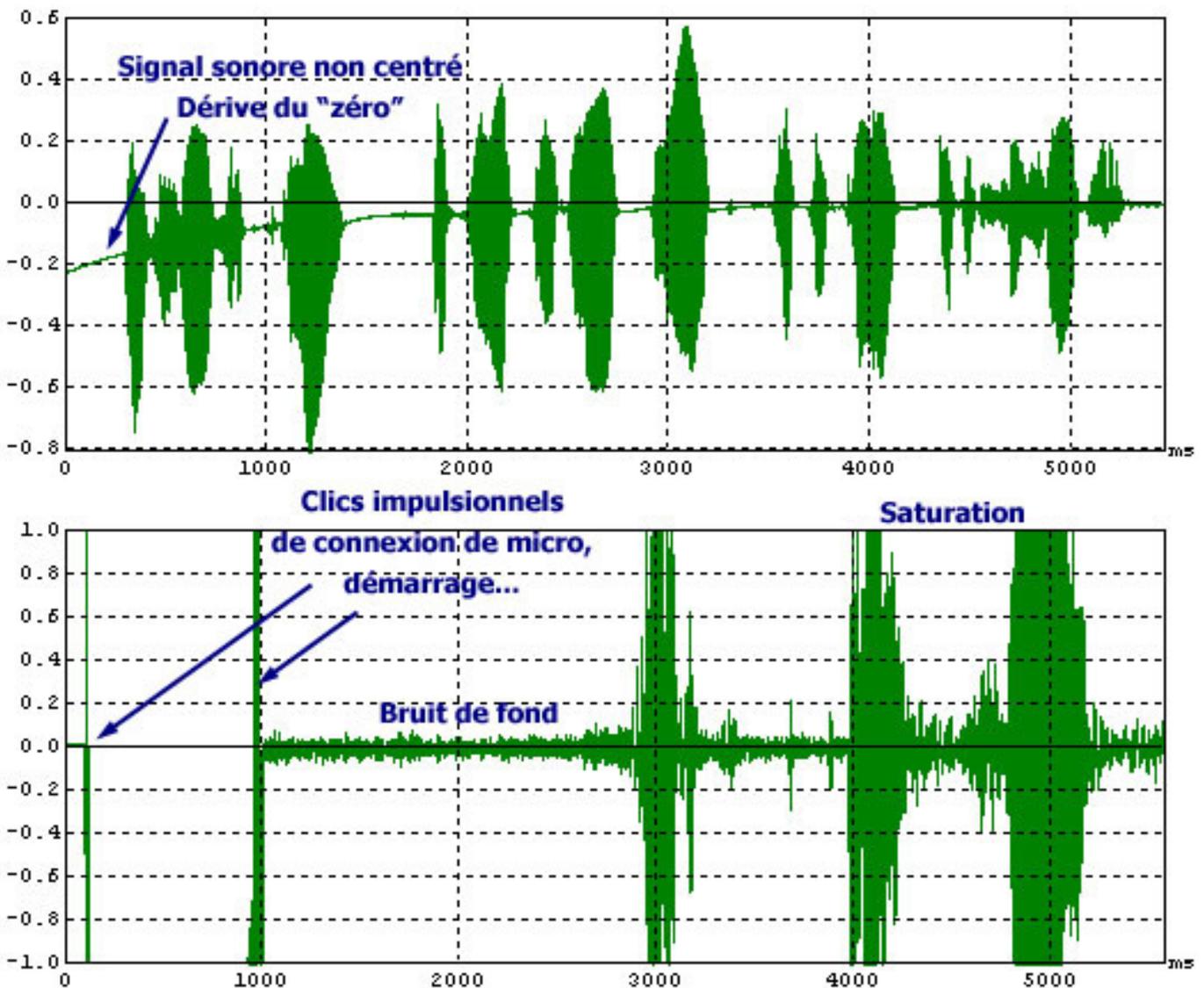


Fig. 1. Signaux de parole enregistrés sans précaution sur des dispositifs de mauvaise qualité. Une analyse acoustique fiable ne sera pas possible sur de telles données. En haut, on remarque un offset initial (le signal n'est pas centré sur zéro) et une dérive progressive. En bas, on observe des clics au démarrage, un bruit de fond résiduel important et la saturation du signal dans les parties énergétiques.

parfaitement toute la chaîne d'acquisition. C'est le cas dans le dispositif EVA2 fabriqué par SQLab-LPL (Teston et Galindo, 1995) où un microphone précis est connecté à un préamplificateur réglé pour fournir 10 V crête à 110 dB SPL à 30 cm. Le signal est ensuite transmis via un canal d'acquisition indépendant des boutons de réglage de volume et la numérisation est effectuée par une carte fournissant un signal codé sur 16 bits pleine échelle à 10 V. Dans ce cas-là, il existe un lien direct entre l'amplitude du signal numérique en fin de chaîne et l'amplitude acoustique réelle du départ.

La deuxième solution consiste à utiliser une source sonore étalon, ceci avant chaque enregistrement. On peut utiliser un «buzzer» dont l'intensité est connue

(p. ex. : 83 dB à 30 cm). En émettant un bip en début d'enregistrement, on peut alors en déduire, par comparaison, l'intensité du signal de parole (p. ex. : -10 dB par rapport au buzzer signifiera 73 dB). Si aucune précaution de calibrage n'a été prévue par l'expérimentateur, il est toutefois possible de faire des mesures d'intensité en valeurs relatives sur le même enregistrement. Autrement dit, il ne faut pas s'attacher à la valeur absolue d'intensité affichée mais par contre, si l'on mesure une baisse de 10 dB, par exemple, lors d'une tâche de répétition, cette diminution de 10 dB est effective. Dans tous les cas, il faut alors être certain que l'enregistreur ne possède pas un circuit électronique de contrôle automatique de volume d'enregistrement dont la particularité est d'écraser la dynamique naturelle du signal de parole.

Ainsi, les faibles amplitudes sont fortement rehaussées et, inversement, les fortes valeurs sont limitées. Cette technologie est quasi systématiquement utilisée dans les studios professionnels de radiophonie afin d'obtenir au final un signal sonore ayant approximativement une intensité sonore moyenne constante, évitant à l'auditeur de modifier sans cesse son volume d'écoute. Elle est aussi utilisée dans des enregistreurs portables de type minidisque pour lesquels il est alors nécessaire de désactiver ce réglage automatique. Bien évidemment, sans information sur la méthode d'enregistrement, il convient de rester prudent sur les mesures d'intensité.

### 3. La délicate tâche de détection de la fréquence fondamentale

La majeure partie des indices exploités dans l'évaluation acoustique de la voix ou de la mélodie dérive de la mesure de fréquence fondamentale. Il est possible de faire des mesures manuelles de F0, méthode exposée dans le chapitre « Onde sonore » de cet ouvrage et proposée dans (Baken, 1987). Toutefois, nous n'aborderons dans cette partie que le calcul automatique de cette grandeur très importante, des détails pouvant être consultés dans (Hess, 1983).

#### a. Les étapes et les différentes techniques de détection de la F0

L'étape préliminaire à la mesure de la F0 consiste à repérer le voisement dans le continuum acoustique. Cette détermination des segments voisés/non voisés n'est pas triviale, surtout dans le cas de dysfonctionnement où les cordes vocales peuvent « décrocher » ou se remettre en vibration très accidentellement. La plupart des algorithmes de détection explicite de voisé/non voisé sont fondés sur un ensemble de paramètres physiques tels que l'énergie, le taux de passage par zéro, le 1<sup>er</sup> coefficient LPC, le facteur de balance grave/aigu... dont les valeurs sont comparées à des seuils plus ou moins dynamiques. Par exemple, si une portion de signal a une faible énergie et un taux de passage par zéro important, le segment est déclaré non voisé. Des mécanismes de suivi temporel permettent ensuite de statuer sur le voisement ou pas des différents segments de parole.

Sur les parties détectées comme voisées, il est alors possible de mesurer la F0. Il existe différentes techniques pour cela (Hess, 1983). On peut citer les techniques temporelles d'autocorrélation, d'AMDF (Average Magnitude Difference Function) ou encore les techniques spectrales d'analyse harmonique

(technique du peigne) ou du cepstre. La plupart de ces méthodes sont fondées sur un calcul par trame, c'est-à-dire que le signal de parole est découpé en morceaux et une valeur de F0 est fournie sur chaque section de durée fixe comme, par exemple, toutes les 10 ms (trame centiseconde). Ce principe est suffisant pour ensuite obtenir des courbes mélodiques rendant compte des montées ou descentes de la F0. Par contre, un tel procédé est complètement inadapté pour évaluer la stabilité à court terme du vibrateur laryngé. En effet, pour cela, il est nécessaire d'utiliser un détecteur de cycle fondé, par exemple, sur du passage par zéro ou de la détection de pic. L'objectif est alors de « marquer » très précisément le début et la fin de chaque cycle vibratoire de façon à en calculer les variations instantanées. L'utilisation de méthodes mixtes avec des algorithmes à plusieurs passes apparaît désormais comme un gage de robustesse. Dans ce type d'approche, une détection simultanée par AMDF, peigne et cepstre suivie d'un algorithme de vote permet d'obtenir une F0 par trame fiable. On peut alors, dans un deuxième temps, déclencher une détection cycle à cycle rendue robuste par la détermination préliminaire de la F0 sur la trame. La 2<sup>e</sup> passe consiste alors à affiner la détection et fournir très précisément le début et la fin de chaque cycle comme préconisé par Titze et Liang (1993).

#### b. Les problèmes de détection de la F0 : erreurs de voisé/non voisé et sauts d'octave

La plupart des problèmes de détection de F0 portent sur 2 catégories : le voisement et les sauts d'octave. La décision voisé/non voisé est un compromis délicat : soit le détecteur élimine trop de parties voisées en les considérant comme non voisées et par conséquent, la F0 n'est pas calculée sur ces parties intéressantes ; soit le détecteur favorise trop le voisement et des parties non voisées sont soumises à un calcul de F0 qui se solde par des valeurs aberrantes car portant sur du bruit a périodique. Il peut être intéressant d'introduire une contre-réaction sur la décision voisé/non voisé suite au calcul de F0 en cas de valeurs aberrantes et déclarer non voisées des suites anarchiques de F0 dues, par exemple, à un calcul sur du bruit. Toutefois, cette solution peut s'avérer malvenue notamment dans le cas de la parole pathologique. Dans ce cas précis, il est préférable d'introduire un réglage semi-manuel laissé à l'expertise de l'observateur.

Le deuxième problème fréquent dans la détection de F0 est le saut d'octave. En effet, dans les techniques temporelles, si une périodicité est observée pour un temps  $\Delta t$ , on considérera qu'il s'agit alors de la période fondamentale  $T_0$ , inverse

de la F0 (voir chapitre « Onde sonore »). Mais 2T0, 3T0, 4T0 constituent aussi des candidats à la périodicité. Si l'algorithme, pour diverses raisons, sélectionne comme résultat  $T=2T0$ , cela entraîne une erreur de calcul de fréquence fondamentale égale à F0/2. Inversement, dans les techniques spectrales de type analyse harmonique (méthode du peigne), si un espacement  $\Delta f$  constant est observé entre deux raies spectrales, on considérera qu'il s'agit alors de la fréquence fondamentale F0, cette valeur étant l'écart systématique entre les harmoniques (voir chapitre 6). Mais 2F0, constitue aussi un candidat à l'intervalle spectral. Si l'algorithme, pour diverses raisons, sélectionne comme résultat  $F=2F0$ , cela entraîne une erreur de calcul de fréquence fondamentale.

### c. Garder l'œil critique

La détection de F0 n'est pas un processus sans faille comme peut l'être le calcul de l'intensité ou du spectrogramme. Il existe des seuils, des limites, des choix qui ont une incidence sur le résultat. Aussi, face à une courbe de F0, il faut toujours garder un œil critique. Tout saut brutal ou toute montée/descente tronquée doit être vue avec circonspection. En même temps, dans le cas de parole pathologique, ces « sauts » de F0 particulièrement visibles, par exemple, dans les diplophonies, font partie de la production vocale. L'œil expert reste donc important.

## III. Evaluation acoustique de la voix

Les travaux portant sur l'évaluation acoustique des troubles vocaux sont innombrables. La problématique du dysfonctionnement laryngé est plus simple que celle plus vaste de la dysarthrie car elle est plus restreinte. Nous commencerons donc par cette partie. Par contre, nous n'aborderons pas la problématique de la dysprosodie qui est étudiée dans un chapitre spécifique de cet ouvrage.

### 1. Les mesures de vibration laryngée

#### a. La fréquence fondamentale moyenne

La fréquence fondamentale F0 reste le meilleur indicateur des caractéristiques biomécaniques des cordes vocales. La F0 moyenne apporte une mesure globale de la hauteur de la voix du sujet (voix aiguë, grave...). Curieusement, dans la littérature, on trouve fréquemment des valeurs moyennes de F0 pour les hommes ou pour les femmes mais sans réelles mesures expérimentales. Nous avons donc repris les données de Peterson et Barney (1952) qui

portent sur 76 locuteurs américains (33 hommes, 28 femmes, 15 enfants) ayant prononcé 2 fois une liste de 10 mots contenant une voyelle sur laquelle ont été mesurées la F0 (Tableau 1) et les formants. Sauf écart très important (p. ex. : voix d'homme à 250 Hz) par rapport à ces données « normatives », il est difficile de corrélérer un dysfonctionnement et une simple valeur de F0 moyenne. Par contre, cette mesure peut être utile dans un suivi longitudinal du locuteur, sachant que, généralement, « la plupart des sujets qui souffrent d'une pathologie vocale ont une hauteur tonale aggravée » (Morsomme et Estienne, 2006).

Tableau 1. Moyenne et écart type de F0 par groupe de locuteurs à partir des données de Peterson et Barney (1952) disponibles sur le Web à l'adresse <http://prosodies.org/ling591/pb>.

Groupe	Hommes	Femmes	Enfants
F0 moyenne	131 Hz	223 Hz	264 Hz
Ecart-type F0	19 Hz	27 Hz	35 Hz

#### b. L'ambiguïté contradictoire des variations de F0

Les variations de F0 peuvent rendre compte du double phénomène de contrôle et de régulation du vibrateur laryngé. Dans un fonctionnement correct, la capacité de pouvoir faire varier la fréquence de vibration des cordes vocales est révélatrice d'une bonne maîtrise vocale. C'est particulièrement le cas chez le chanteur qui possède une importante étendue tonale. On mesure alors une adéquation entre variation importante de F0 et bon fonctionnement. Inversement, de fortes perturbations instantanées du cycle vibratoire se mesurent par des variations de F0 révélatrices d'une dérégulation du système phonatoire. On retrouve alors une corrélation opposée à la précédente entre variations et dysfonctionnement. En fait, seules les variations contrôlées sont synonymes de bonne maîtrise. D'où l'importance cruciale du contexte de production vocale et des consignes.

Une consigne de production de voyelle tenue permet de focaliser la mesure sur la stabilité et la régulation du vibrateur laryngé et d'écarter les variations contrôlées. C'est d'ailleurs l'un des intérêts de l'utilisation de voyelles tenues. Dans ce cas-là, où la stabilité est demandée, comme une sorte de posture statique avec recherche d'immobilité, toute variation est synonyme de dysfonctionnement. Inversement, dans de la lecture de texte offrant une possibilité d'expressivité, tâche nécessitant un dynamisme s'apparentant à de la marche, les changements lents traduisent une bonne capacité vocale alors que les variations à court terme, voire les

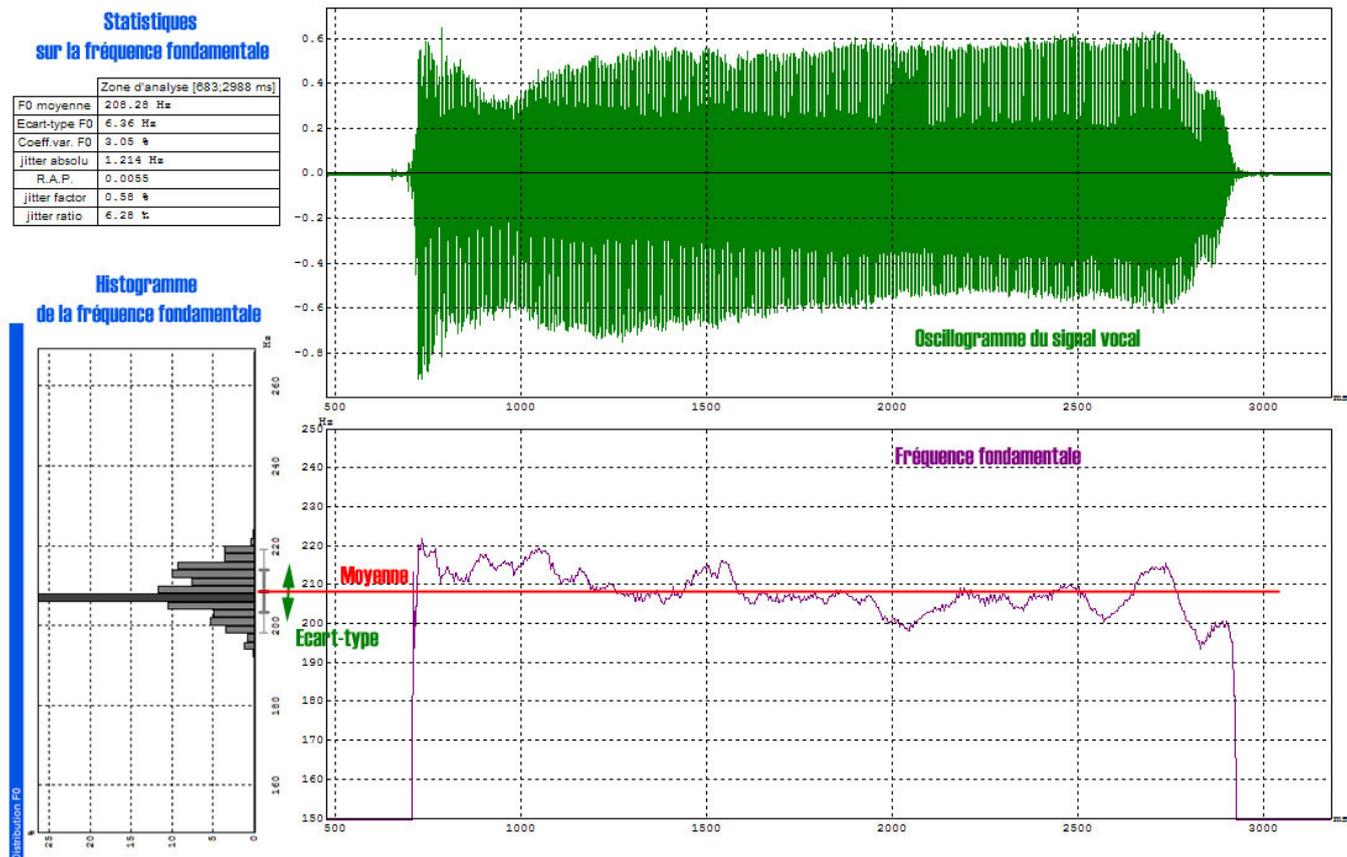


Fig. 2. Signal vocal sur un /a/ tenu d'un adolescent en cours de mue « difficile ». Décours temporel, histogramme et statistiques de la fréquence fondamentale.

On remarque les variations importantes de F0 alors que la consigne est à la stabilité. Le coefficient de variation F0 de 3 % mesure cette instabilité globale. Par contre, le faible jitter de 0.58 % traduit une relativement bonne stabilité instantanée.

dévoisements inattendus, révèlent le dysfonctionnement. D'où la nécessité d'introduire plusieurs paramètres.

### c. Instabilité de la vibration

La dynamique ou l'instabilité laryngée se traduit par des variations de F0 au cours du temps. Elle peut se mesurer par l'écart type de la F0, qui correspond à l'ampleur en Hz des variations de F0 autour de la moyenne. Le coefficient de variation permet de relativiser l'écart type en le comparant à la F0 moyenne. Il correspond donc à l'ampleur en % des variations de F0 par rapport à la F0 moyenne. Ainsi, à un écart type de 4.9 Hz pour une F0 moyenne de 180 Hz correspond un coefficient de variation de 2.7 %, valeur importante sur une voyelle tenue. Le même écart type pour une F0 moyenne de 500 Hz fournit un coefficient de variation de 0.98 %, valeur beaucoup plus normale sur ce type de production. Le coefficient de variation de la F0 est donc le meilleur indice pour explorer globalement la dynamique ou l'instabilité du vibrateur laryngé (Fig. 2). Sur des voyelles tenues, il devient important en présence de vibrato pathologique, chevrottement, instabilité d'origine neurologique.

L'instabilité à court terme de la F0 se traduit par des variations de fréquence entre les cycles d'oscillation (Fig. 3). Elle se calcule par le JITTER absolu moyen qui est la moyenne de la différence de F0 entre deux cycles de vibration consécutifs. Ces variations sont calculées cycle à cycle, de manière très précise. Le JITTER factor permet de relativiser le jitter moyen en le comparant à la F0 moyenne. Ainsi, à un jitter moyen de 0.677 Hz pour une F0 moyenne de 180 Hz correspond un jitter factor de 0.38 %, valeur faible. Le même jitter moyen pour une F0 moyenne de 80 Hz fournit un jitter factor de 0.84 %, valeur plus problématique. Le jitter factor est donc un meilleur indice pour explorer la stabilité de la fréquence fondamentale à court terme. Il reste le meilleur corrélat acoustique de la raucité, bien que cette corrélation simple ne permette pas de rendre compte complètement de cette dimension vocale. Le RAP (Relative Average Perturbation) mesure la moyenne des variations de trois périodes consécutives rapportée à la période moyenne du signal observé. Le JITTER ratio mesure la moyenne des variations de période, inverse de la fréquence, entre deux cycles de vibration consécutifs et relativise cette valeur par la période moyenne du signal observé.

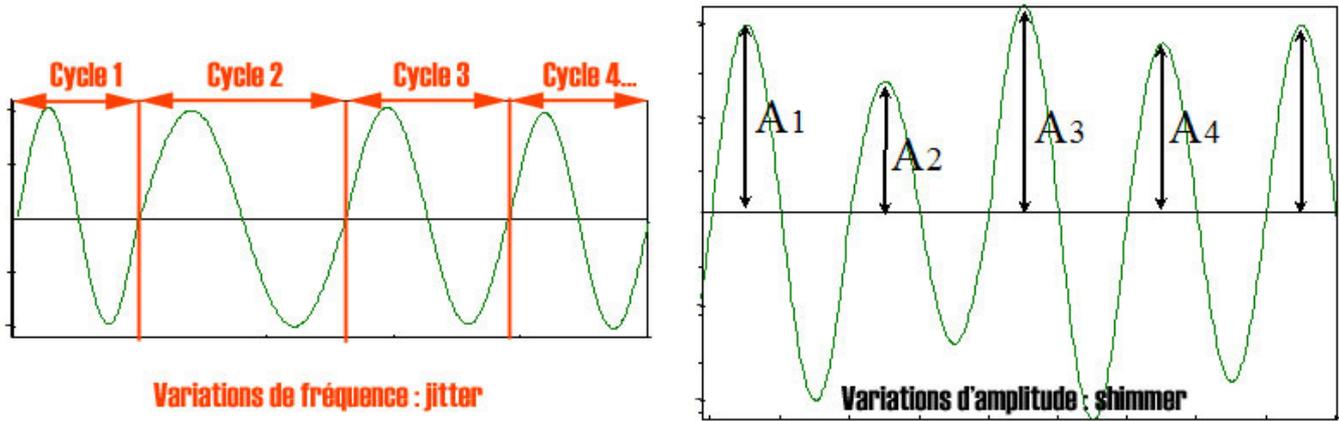


Fig. 3. Variations instantanées des caractéristiques de la vibration à gauche, l'écart de fréquence entre deux cycles consécutifs entraînera un jitter important ; à droite, l'écart d'amplitude entre deux cycles consécutifs entraînera un shimmer important.

Il faut signaler qu'un jitter élevé entraîne toujours un coefficient de variation important. La réciproque n'est pas vraie. En effet, de petites variations croissantes ou décroissantes de  $F_0$  entre les cycles ne fournira pas un jitter important mais pourra donner lieu à des variations globales importantes de  $F_0$  (exemple typique : un vibrato).

L'instabilité à court terme de l'amplitude de la vibration se traduit par des variations d'amplitude entre les cycles d'oscillation (Fig. 3). Elle se calcule par le SHIMMER moyen qui est la moyenne des

rapports d'amplitudes entre deux cycles de vibration consécutifs. Ce rapport est généralement exprimé en dB. Ces variations sont calculées cycle à cycle, de manière très précise. L'APQ (Amplitude Perturbation Quotient) mesure la moyenne des variations d'amplitude en tenant compte de 11 périodes consécutives, le tout rapporté à l'amplitude moyenne du signal observé.

Il existe de nombreux paramètres de ce type et il serait fastidieux de les présenter tous. Pour plus de détails, le lecteur pourra se référer à (Baken, 1987).

Tableau 2 . Principaux indices liés à la vibration laryngée et sa stabilité (Baken, 1987).

Moyenne de la $F_0$	moyenne (en Hz) = $\frac{1}{N} \sum_{i=1}^N F_{0i}$
Ecart type de la $F_0$	écart-type (en Hz) = $\sqrt{\frac{1}{N} \sum_{i=1}^N (F_{0i} - F_{0\text{ moy}})^2}$
Coefficient de variation	coefficient de variation (en %) = $100 \times \frac{\text{écart type}}{\text{moyenne}}$
Perturbation Cycle à Cycle	<u>jitter</u> moyen (en Hz) = $\frac{1}{N-1} \sum_{i=1}^{N-1}  F_{0i} - F_{0i+1} $
Perturbation Moyennée	Relative <u>Average</u> Perturbation (sans unité) = $\frac{1}{N-2} \sum_{i=2}^{N-1} \left  \frac{T_{0i-1} + T_{0i} + T_{0i+1}}{3} - T_{0i} \right $
Pourcentage	<u>Jitter factor</u> (en %) = $100 \times \frac{\text{jitter moyen (Hz)}}{F_{0\text{ moy}}}$
	<u>Jitter ratio</u> (en %) = $1000 \times \frac{\text{jitter moyen (ms)}}{T_{0\text{ moy}}}$
Perturbation Cycle à Cycle	<u>shimmer</u> moyen (en dB) = $\frac{1}{N-1} \sum_{i=1}^{N-1} \left  20 \cdot \log \left( \frac{A_i}{A_{i+1}} \right) \right $
Perturbation Moyennée	Amplitude Perturbation Quotient APQ = $\frac{1}{N-10} \sum_{i=6}^{N-5} \left  \frac{A_{i-5} + \dots + A_i + \dots + A_{i+5}}{11} - A_i \right $

## 2. Les mesures de signal/bruit

Les mesure de signal et de bruit explorent la présence (ou absence) de « bruits » au cours de la phonation. Ce bruit peut être dû à un écoulement turbulent causé par un mauvais accolement des cordes vocales. On parle alors de « bruit additif ». Toutefois, il faut signaler qu'une vibration très instable peut être elle aussi mise en évidence par cette analyse : l'écart de la  $F_0$  et de ses harmoniques par rapport à la valeur moyenne est détecté comme un « bruit » de modulation.

Le rapport Harmonique/Bruit (Harmonic to Noise Ratio, HNR) se calcule à partir d'une méthode temporelle décrite dans Yumoto et Gould (1982). 25 cycles consécutifs du signal voisé permettent d'établir une forme ondulatoire moyenne ; la soustraction entre le signal originel et la forme moyenne permet d'obtenir un résidu considéré comme le bruit. Si  $H$  est l'énergie de l'onde moyennée périodique et  $N$  l'énergie du bruit résiduel, le rapport  $H/N$  calculé régulièrement fournit la valeur du HNR éventuellement converti en dB.

Le Signal Ratio ( $S_r$ ) se calcule à partir d'une méthode spectrale décrite dans (Hiraoka *et al.*, 1984). On considère le signal vocal comme la somme de deux composantes : une composante périodique ( $F_0$  et harmoniques) et une composante de bruit (le reste). A partir d'un spectre effectué sur une portion de signal, l'énergie présente dans les raies est vue comme la composante périodique du signal. Le calcul de cette énergie dans les raies par rapport à l'énergie totale fournit un rapport appelé  $S_r$ . De plus, l'auteur préconise de faire le même type de calcul non pas sur tout le spectre mais uniquement sur des zones de moyennes et hautes fréquences comme, par exemple, au-dessus de 1 kHz. Dans ce cas-là, le Signal Ratio est particulièrement révélateur de dysphonie car il devient très faible chez ces patients dont le dysfonctionnement entraîne une composition spectrale appauvrie dans les aigus, caractéristique captée par le  $S_r$  calculé sans les graves.

L'énergie normalisée du bruit (NNE) se calcule à partir d'une méthode spectrale décrite dans Kasuya *et al.* (1986). Cette méthode est assez proche de la précédente excepté dans la formulation finale qui indique non pas le pourcentage de périodicité mais plutôt la part de bruit exprimée en dB par rapport au signal complet.

## 3. Les portraits de phase et les mesures de non-linéarité

La représentation du signal sous la forme d'un portrait de phase est un outil classique de la physique non linéaire. Dès la version 1 du dispositif EVA (Teston et Galindo, 1995), un diagramme à deux dimensions permettait de distinguer les instabilités aléatoires ou « organisées ». Le principe de ce tracé (Fig. 4) est de placer un point aux coordonnées  $\{x=F_0(t) ; y=F_0(t+\tau)\}$ ,  $\tau$  étant défini comme un délai temporel. Un ensemble de points peu espacés indique alors une bonne stabilité de la vibration laryngée car cela signifie que la valeur de  $F_0$  au temps ( $t$ ) est proche de celle au temps ( $t+\tau$ ), synonyme de constance. Par contre, la présence de points éloignés de la droite bissectrice  $F_0(t) = F_0(t+\tau)$  indique des variations instantanées de  $F_0$  (Fig. 4a). D'autre part, une distribution de points le long de la droite bissectrice témoigne d'une dérive à long terme (p. ex. : vibrato) mais d'une bonne stabilité à court terme (Fig. 4b). Enfin, un nuage dispersé met en évidence une vibration anarchique (Fig. 4c). Il est possible de passer à une représentation en trois dimensions en rajoutant la coordonnée  $z=F_0(t+2\tau)$  pour faire apparaître des trajectoires organisées en attracteur (Fig. 4d) ou bien des chevelus (Fig. 4e). L'analyse quantitative des attracteurs a fait l'objet de travaux dans le domaine. Le calcul des coefficients de Lyapunov a été proposé sur le plan théorique par Titze puis explicitement développé dans l'évaluation vocale par Yu *et al.* (2000). Le principe général est le calcul de la divergence de deux trajectoires voisines d'un portrait de phase. Toutefois, le Plus Grand Exposant de Lyapunov, bien que paramètre d'analyse très prometteur, reste peu utilisé du fait de la difficulté de relier cette mesure à une réalité physiologique.

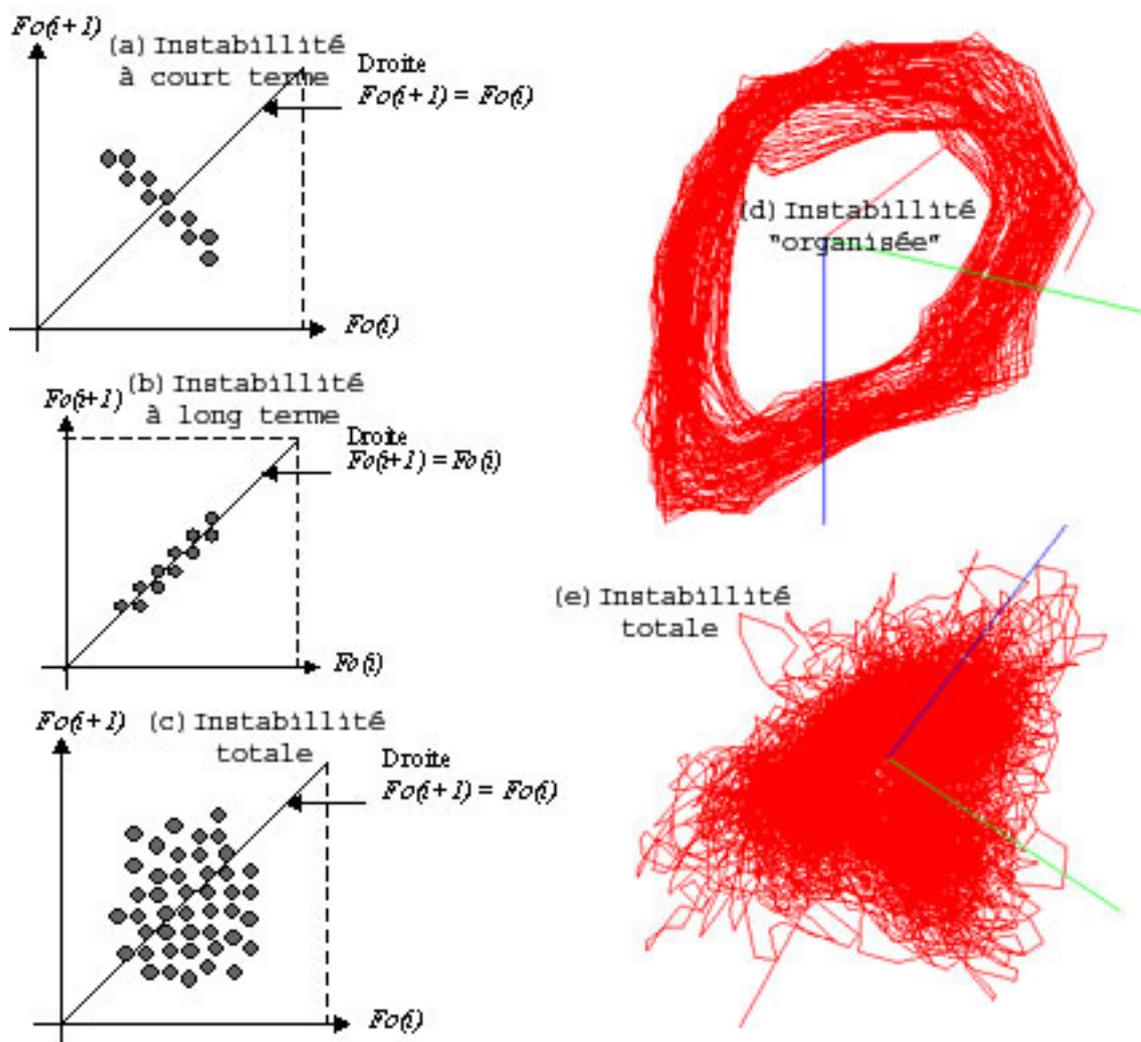


Fig. 4. Représentations en 2D de l'instabilité vocale (a, b, c) et portraits de phase en 3D (d, e).

## IV. Evaluation acoustique de la parole : aspects segmentaux

Les aspects segmentaux de la parole concernent les éléments phonémiques du message oral. La complexité et les phénomènes de variabilité liés au décodage phonétique empêche l'utilisation de techniques rapides ou semi-automatiques dans le cas de parole pathologique, d'où sa faible utilisation dans le cadre d'évaluation acoustique clinique.

### 1. La mesure de formants

Les formants correspondent aux zones fréquentielles mises en relief dans l'énoncé sonore. Les valeurs de ces pics spectraux sont dépendantes des configurations du conduit vocal et sont différentes selon la nature de la voyelle (voir chapitre « Résonance sonore »). De plus, le passage d'une consonne à une voyelle entraîne une migration des formants, phénomène bien connu sous le nom de transition formantique. Celles-ci

varient en fonction de la consonne et portent de l'information. Par conséquent, mesurer les valeurs des formants et analyser la forme des transitions est un indicateur des configurations articulaires. En première approximation, on peut même corrélérer le 1<sup>er</sup> formant au degré d'ouverture de la voyelle. De même, le 2<sup>e</sup> formant est un bon indicateur du lieu avant/arrière de l'articulation. Il faut toutefois tempérer cette conception car il existe une forte variabilité entre locuteurs ou en fonction du contexte linguistique et pragmatique.

Cette approche, fondée sur la mesure de formants, est régulièrement utilisée dans l'étude de la dysarthrie (Buder *et al.*, 1996 ; Weismer *et al.*, 2001 ; Baudelle *et al.*, 2003). Weismer *et al.* (2001) montrent notamment une réduction quantitative et mesurée de l'espace vocalique dans le cas de patients atteints de SLA. Toutefois, ce type d'expérimentation souffre de plusieurs difficultés. D'une part, les auteurs se heurtent aux phénomènes de variabilité qui rendent floue la différence entre inconstance « naturelle » des

formants et déviation «pathologique». Par exemple, la réduction de l'espace vocalique mesuré par Weismer *et al.* (2001) pour des SLA est aussi visible chez des locuteurs «normaux» mais dans des contextes d'élocution particuliers (Meunier *et al.*, 2006). D'autre part, une telle expérimentation est souvent longue et fastidieuse car elle nécessite, la plupart du temps, l'observation et la mesure manuelle des valeurs de formants, tâche bien connue en phonétique comme un travail de bénédictin difficilement réalisable en milieu clinique.

## 2. La mesure du VOT

Le VOT (voice onset time) est le délai temporel entre l'explosion d'une occlusive et le début du voisement (Fig. 5). En français, cette durée n'a pas vraiment de propriété linguistique et peut ainsi rendre compte de la capacité du locuteur à coordonner ses organes constricteurs et phonatoires. Ainsi, lors de la production de /pa/, le locuteur doit maîtriser le relâchement des lèvres responsables de l'occlusion du /p/ suivi d'une mise en mouvement des cordes vocales pour produire la voyelle. Or, une dérégulation d'origine neurologique peut entraîner un allongement ou un raccourcissement de cette durée si l'ordre moteur pour la vibration laryngée est

retardé ou anticipé par rapport au geste d'ouverture de la consonne. Des détails peuvent être consultés dans Morris (1989) qui a étudié ce paramètre pour différents types de dysarthriques et des aspects méthodologiques sont abordés dans Özsancak *et al.* (2001) pour le français.

## 3. Des approches issues des techniques de traitement automatique

Dans le domaine des sciences et technologies de l'information et de la communication, les dispositifs de traitement automatique de la parole sont à présent éprouvés et relativement robustes. Ces procédés sont ceux de la dictée vocale ou de l'identification vocale de locuteurs. Récemment est apparue une nouvelle approche pour tester ces techniques sur des locuteurs atteints de pathologies, en l'occurrence des dysphoniques, ceci dans un but d'évaluation objective du trouble de communication. Si on la compare à des méthodes instrumentales classiques, les originalités de cette approche basée sur une modélisation statistique, reposent sur sa capacité à analyser de la parole continue (et non des voyelles tenues) proche de l'élocution naturelle et sa capacité à traiter de grands corpus, permettant de mener des études à grande échelle et d'obtenir des informations statistiques significatives.

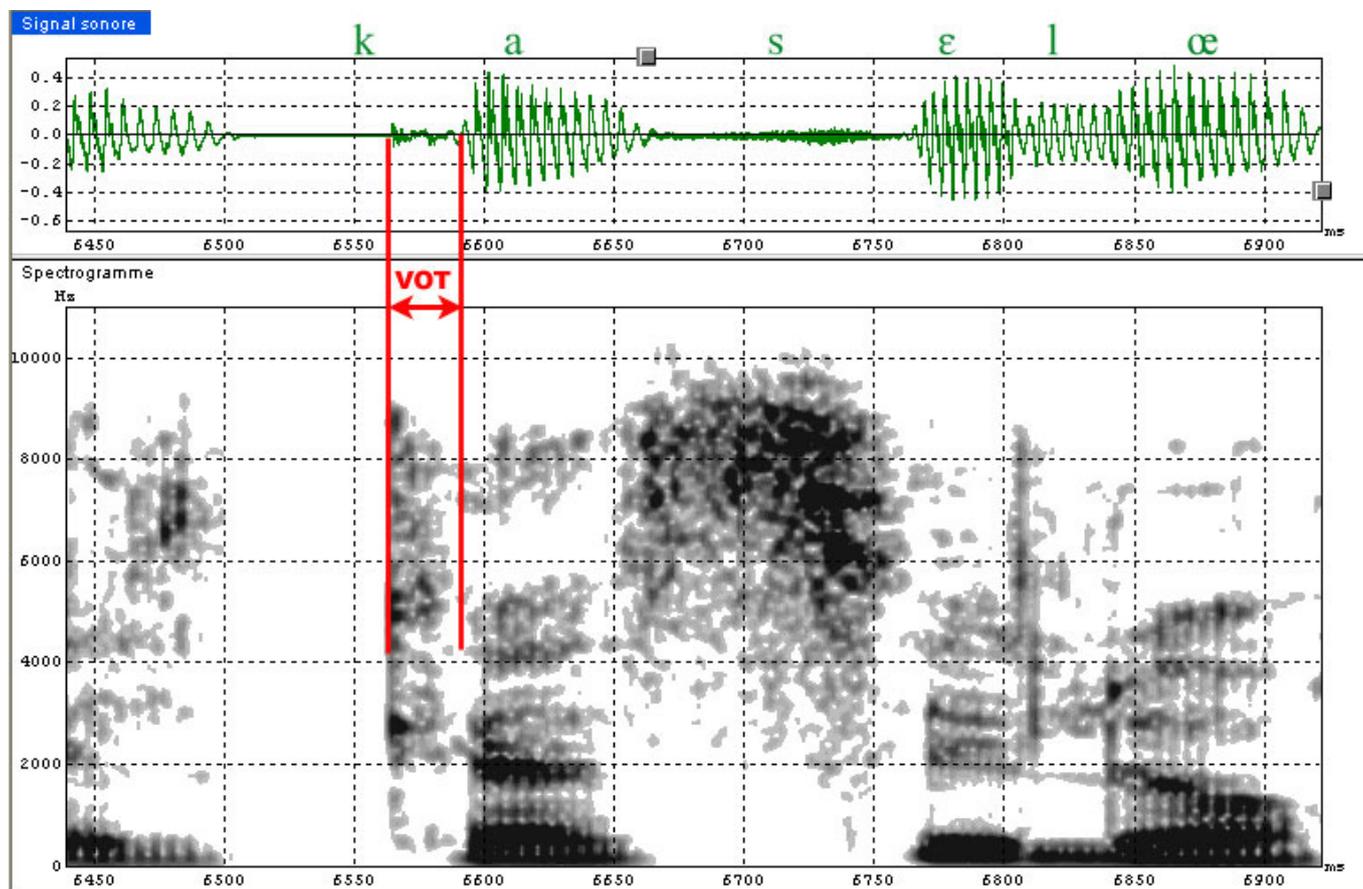


Fig. 5. Mesure du VOT (voice onset time) sur le /ka/ de l'énoncé « elles cassaient leur corde ».

Les systèmes conçus pour cette tâche s'appuient sur une première phase dite d'apprentissage où le dispositif apprend des caractéristiques à partir d'un ensemble très important d'exemples auquel il est confronté. Une fois cette phase effectuée, la machine confronte ses informations apprises aux données recueillies chez un nouveau patient. Des détails sont fournis dans Fredouille *et al.* (2005).

## V. Conclusions

Les caractéristiques multidimensionnelles de la parole imposent à la plupart des études une analyse simultanée de plusieurs paramètres. On définit ainsi l'analyse multiparamétrique. Le nombre de mesures possible est considérable et il y a peu de chances que le recueil en aveugle de différentes grandeurs permette, à terme, d'obtenir un protocole objectif d'évaluation de la dysarthrie prise dans son ensemble. Déjà l'estimation quantitative de la qualité vocale, problématique bien plus simple car plus restreinte, n'a toujours pas abouti à un consensus accepté par une majorité. De ce fait, sur un terrain bien plus vaste qui est celui de la dysarthrie, il nous semble plus raisonnable, dans l'état actuel des connaissances et des techniques, d'adopter une approche ciblée sur une facette du problème pour aboutir à une évaluation fiable, notamment en complétant l'analyse acoustique par d'autres techniques complémentaires comme l'analyse des mouvements ou les mesures aérodynamiques.

## Références

1. Baken JR. (1987). *Clinical Measurement of Speech and Voice*. Taylor & Francis Ltd. Londres.
2. Baudelle E, Vaissière J, Renard JL, Roubeau B, Chevré-Müller C. *Caractéristiques vocaliques intrinsèques et co-intrinsèques dans les dysarthries cérébelleuse et parkinsonienne*, *Folia Phoniatrica et Logopaedica* 2003, 55 : 137-146.
3. Buder EH, Kent RD, Kent JF, Milenkovic P, Workinger M. *FORMOFFA : an automated formant, moment, fundamental frequency, amplitude analysis of normal and disordered speech*, *Journ of Clinical linguistics & phonetics* 1996, 10, 31-54.
4. Condamines R. (1985). *Acoustique psycho-physique*. Ed. Masson. Paris.
5. Hess W. (1983). *Pitch Determination of Speech Signals, Algorithms and Devices*, Springer Series in Information Sciences.
6. Fredouille C, Pouchoulin G, Bonastre JF, Azzarello M, Giovanni A, Ghio A. *Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia)*. *Proceedings of European Conference on Speech Communication and Technology (Eurospeech 2005)*. Lisboa. 2005, pp 149-152.
7. Hiraoka N, Kitazoe Y, Ueta H. *Harmonic-intensity analysis of normal and hoarse voices*. *J Acoust Soc Am* 1984, 76 : 1648-1651.
8. Kasuya H, Ogawa S, Mashima K, Ebihara S. *Normalized noise energy as an acoustic measure to evaluate pathologic voice*. *J Acoust Soc Am* 1986, 80 : 1329-1334.
9. Meunier C, Espesser R, Frenck-Mestre C. *Aspects phonologique et dynamique de la distinctivité au sein des systèmes vocaliques: une étude inter-langue*. *Actes, Journées d'Etude sur la Parole (JEP) (26 : 2006 juin 12-16 : Dinard, France)*. 2006, pp 333-336.
10. Morris JR. *V.O.T. and dysarthria : a descriptive study*. *Journal of communication disorders*. 1989, 22 : 23-33
11. Morsomme D, Estienne F. *Bilan vocal*. In Estienne F, Piérart B (eds). *Les bilans de langage et de voix : fondements théoriques et pratiques*. Masson. Paris. 2006.
12. Özsancaç C, Auzou P, Jan M, Hannequin D. *Measurement of voice onset time in dysarthric patients : Methodological considerations*. *Folia phoniatrica et logopaedica*. 2001, 53 : 48-57.
13. Peterson GE, Barney HL. *Control methods used in a study of the vowels*. *J Acoust Soc Am* 1952, 24 : 175-184.
14. Teston B, Galindo B. *A diagnostic and rehabilitation aid workstation for speech and voice pathologies*. *Proceedings of European Conference on Speech Communication and Technology (Eurospeech 1995)* Madrid. 1995, pp 1883-1886.
15. Titze IR, Liang H. *Comparison of F0 Extraction Methods for High-Precision Voice Perturbation Measurements*. *Journal of Speech and Hearing Research* 1993, 36 : 1120-1133.
16. Weismer G, Jeng JY, Laures JS, Kent RD, Kent JF. *Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders*. *Folia Phoniatr Logop* 2001, 53 : 1-18.

17. Yu P, Ouaknine M, Giovanni A. *Clinical significance of calculating the coefficients of Lyapunov in the objective assessment of dysphonia. Rev Laryngol Otol Rhinol*, 2000, 121 : 301-305.
18. Yumot E, Gould W. *Harmonics-to-noise ratio as an index of the degree of hoarseness. J Acoust Soc Am* 1982, 71 : 1544-1550.