

Caractérisation des singularités du signal de parole

A. BOUZID¹ et N. ELLOUZE²

¹ISET, Institut Supérieur des Etudes Technologiques de Sfax, Rte Mahdia Km 2.5, BP 88A, 3099 Elbustan Sfax, Tunisie

²ENIT, Ecole Nationale d'Ingénieurs de Tunis, BP. 37, Le Belvédère 1002 Tunis, Tunisie

bouzydacha@yahoo.fr, N.Ellouze@enit.rnu.tn

Résumé – Dans ce papier, nous présentons une méthode de caractérisation des singularités lissées détectées sur le signal de parole au voisinage des instants de fermeture de la glotte. Un modèle en dents de scie lissé par une gaussienne de variance σ^2 permet de simuler les discontinuités des tangentes observées sur le signal de parole lors de la fermeture de la glotte. La décroissance des modules maximaux de la transformée en ondelettes du signal de parole et du modèle lissé permettent de donner une estimation de la valeur de l'échelle de diffusion à travers la valeur de σ qui donne un modèle se rapprochant le plus du signal de parole.

Abstract – In this paper, we propose a model of a smoothed singularity detected in speech signal in the neighbourhood of glottal closure instant. It's about a signal having a fast and constant decay and presenting two derivative discontinuities smoothed by a Gaussian function having a small variance σ^2 . The determination of this variance is found by comparing the decay of wavelet modulus maxima (wavelet which is the n^{th} derivative of a Gaussian function) of speech signal and the smoothed model signal for different values of σ . The smoothed signal that gives the best fitting of the speech signal gives the estimation of σ .

1. Introduction

Les phénomènes de singularités localisées du signal de parole lors de la fermeture de la glotte (IFG) sont explorés par l'analyse de la décroissance des coefficients en ondelettes du signal aux fines échelles, ce qui permet de déterminer le coefficient de Lipschitz local α qui caractérise la singularité. La méthode commence par la détermination de la transformée en ondelettes à différentes échelles du signal. Le nombre de moments nuls de l'ondelette permet de caractériser la nature de la singularité. Aux faibles échelles, la décroissance des modules maximaux est donnée par le nombre de moments nuls de l'ondelette, alors qu'aux fortes échelles, la décroissance est caractérisée par le coefficient α . Toutefois, souvent le signal analysé présente un effet de lissage, dû aux prétraitements du signal, ce qui a pour effet de masquer les discontinuités et faire varier l'estimation de α . Par ailleurs lorsque le support d'une ondelette couvre à une échelle donnée plusieurs discontinuités, la mesure de α s'en trouve fortement dégradée.

Le paragraphe 2 présente des rappels sur la mesure de la régularité locale par les modules maximaux de la transformée en ondelettes et sur les singularités lissées. Le paragraphe 3 fait l'objet de la présentation d'un signal présentant des discontinuités de dérivée lissées, correspondant au modèle simplifié du comportement du signal de parole aux instants de fermeture de la glotte. Le paragraphe 4 est une application à la mesure de l'échelle de diffusion sur le signal de parole.

2. Mesure de la régularité locale

Une remarquable propriété de la transformée en ondelettes est de permettre de caractériser la régularité locale d'une fonction [1] [2], cette régularité est souvent mesurée par

l'exposant de Lipschitz α [5]. Jaffard a démontré que les propriétés de micro-localisation sont équivalentes à des conditions spécifiques de décroissance de l'amplitude de la transformée en ondelettes. En effet la décroissance de la transformée en ondelettes en fonction de l'échelle est liée à la régularité ponctuelle mesurée par α [3] [7]. Mesurer cette décroissance revient à faire un zoom sur les structures du signal avec une échelle qui tend vers zéro [6]. L'analyse de la régularité d'une fonction f en un point peut être délicate car f peut avoir des singularités de types différents agrégées au voisinage de ce point.

Il est prouvé que la régularité lipschitzienne locale de f à l'instant v dépend de la décroissance de $|Wf(u,s)|$ aux fines échelles dans le voisinage de v [3], [7]. Toutefois cette décroissance n'a pas besoin d'être mesurée directement dans le plan temps-échelle (u,s) . Elle peut en effet être contrôlée par les valeurs des maxima locaux de la transformée en ondelettes.

2.1 Moments nuls d'ondelette

Le nombre de moments nuls caractérisant l'ondelette d'analyse est important pour la mesure de la régularité locale d'un signal.

Une ondelette ψ à décroissance rapide, a n moments nuls si et seulement si il existe une fonction θ à décroissance rapide tel que

$$\psi(t) = (-1)^n \frac{d^n \theta}{dt^n} \quad (1)$$

Ce théorème montre qu'une ondelette à n moments nuls peut s'écrire comme la dérivée à l'ordre n d'une primitive θ [4]. Comme

$$wf(u,s) = f * \bar{\psi}_s(u) \quad (2)$$

avec

$$\bar{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{-t}{s}\right) \quad (3)$$

Ainsi la transformée en ondelettes s'exprime comme la dérivée $n^{\text{ième}}$ de la fonction lissée à chaque échelle s

$$wf(u, s) = s^n \frac{d^n}{du^n} (f * \bar{\theta}_s)(u) \quad (4)$$

$$\bar{\theta}_s = \frac{1}{\sqrt{s}} \theta\left(\frac{-t}{s}\right) \quad (5)$$

2.2 Mesure de la régularité

Jaffard a montré que pour une fonction f de carré sommable et uniformément Lipschitz $\alpha < n$ en v alors il existe un réel $A > 0$ tel que

$$\forall (u, s) \in \mathbb{R} \times \mathbb{R}^+ , |wf(u, s)| \leq A s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^\alpha\right) \quad (6)$$

Ceci permet d'écrire pour tout u appartenant au cône d'influence de v que

$$|wf(u, s)| \leq A' s^{\alpha+1/2} \quad (7)$$

Ce qui est équivalent à

$$\log_2 |wf(u, s)| \leq \log_2 A' + \left(\alpha + \frac{1}{2}\right) \log_2 s \quad (8)$$

Ce théorème prouve que la régularité lipschitzienne locale de f en v dépend de la décroissance aux fines échelles de $|Wf(u, s)|$ observée au voisinage de v . Cette décroissance n'a pas besoin d'être mesurée directement dans le plan temps-échelle (u, s) . La décroissance de $|Wf(u, s)|$ peut en effet être contrôlée par les valeurs de ses maxima locaux. Les singularités sont détectées en cherchant les abscisses où convergent les modules maximaux d'ondelettes aux fines échelles et dans ce cas le nombre de moments nuls, doit être supérieur à α qui caractérise le type de la singularité. Quand la transformée en ondelettes de f ne présente pas de maxima aux fines échelles alors f est localement régulière.

2.3 Singularités lissées

Un signal peut présenter d'importantes variations qui sont infiniment continûment différentiables. Au voisinage de ces transitions le signal peut être modélisé par le produit de convolution d'une fonction $f_0(t)$ possédant une singularité Lipschitz α par une fonction de lissage $g_\sigma(t)$.

$$f(t) = f_0 * g_\sigma(t) \quad (9)$$

Quand g_σ est une gaussienne $N(0, \sigma^2)$ exprimé par [4]

$$g_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right) \quad (10)$$

et l'ondelette mère donnée par la dérivée $n^{\text{ième}}$ d'une gaussienne $N(0, \beta^2)$, L'inégalité (7) devient

$$|wf(u, s)| \leq A s^{\alpha+1/2} \left(1 + \frac{\sigma^2}{\beta^2 s^2}\right)^{-(n-\alpha)/2} \quad (11)$$

Cette inégalité montre la relation entre la décroissance de la transformée en ondelettes et $a = \sigma/\beta$ l'échelle de diffusion.

En opérant une régression, on obtient l'approximation suivante pour les modules maximaux :

$$\log_2 |wf(u, s)| \approx \log_2(K) + \left(\alpha + \frac{1}{2}\right) \log_2 s - \frac{n-\alpha}{2} \log_2 \left(1 + \frac{a^2}{s^2}\right) \quad (12)$$

Les paramètres K , α et σ sont mesurées numériquement à partir de la décroissance des modules maximaux le long des courbes de maxima qui convergent vers v . Cette équation montre qu'aux fines échelles la pente de décroissance de $\log_2 |Wf(u, s)|$ en fonction de $\log_2(s)$ est donnée par $n+1/2$, alors que pour les fortes échelles la pente est donnée par $\alpha+1/2$.

3. Singularité du signal au voisinage des IFG

3.1 Présentation du modèle

Aux instants de fermeture de la glotte, le signal de parole comme le montre la figure 1, présente une décroissance rapide délimitée par deux discontinuités de tangentes. Ceci est modélisé par le signal en dents de scie. Le modèle proposé est formé d'un segment de support 10 points (fréquence d'échantillonnage de 20 kHz) et de pente négative encadrée par deux segments à pente positive et de support chacun de 20 points. Ce modèle à décroissance rapide possède quatre discontinuités sur sa dérivée. Le modèle est lissé par une fonction gaussienne de variances σ^2 . Le modèle ainsi que ses représentations lissées pour différentes valeurs de σ sont donnés en figure 2.

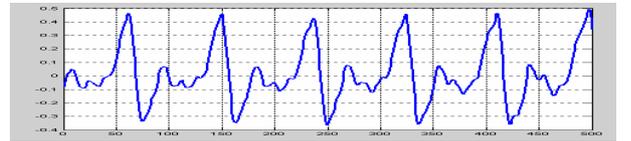


FIG. 1 : représentation du signal de parole : /o/ prononcé par une femme de la base de Keele

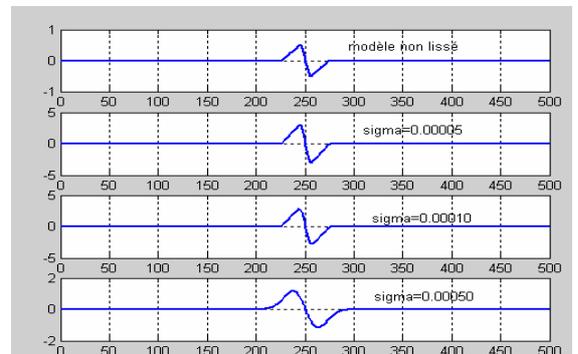


FIG. 2 : signal modèle non lissé suivi de ses représentations lissées pour 3 valeurs de $\sigma : 10^{-5}$, 10^{-4} et 510^{-4}

La transformée en ondelettes $Wf(u, s)$ du modèle est calculée avec une ondelette chapeau mexicain $\psi = \theta'$, θ étant une gaussienne $N(0, \beta^2)$, aux échelles allant de 2^{-4} à 2^0 .

Cette transformée est donnée en figure 3 pour le modèle lissé et pour la valeur de $\sigma = 5 \cdot 10^{-5}$. Les modules maximaux sont observés aux niveaux des discontinuités des tangentes.

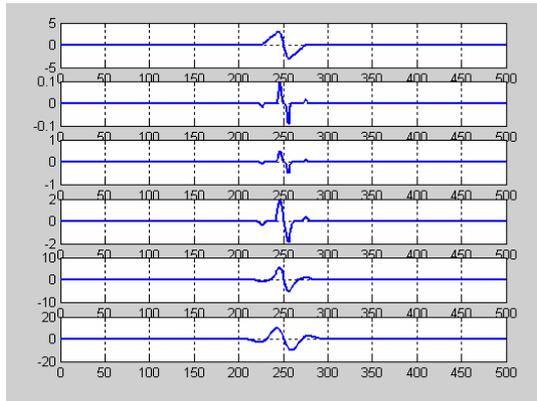


FIG. 3 : modules maximaux de la transformée en ondelettes (chapeau mexicain) du modèle lissé pour $\sigma=5 \cdot 10^{-5}$ et les échelles allant de 2^{-4} à 2^0

La décroissance de $\log_2 |Wf(u,s)|$ le long des courbes de maxima pour les modèles non lissé et lissés aux valeurs de $\sigma = 10^{-4}$ et $5 \cdot 10^{-5}$ est donnée en figure 5.

Les mesures des pentes sont paramétrées par le facteur de diffusion σ . Les courbes changent de pentes aux valeurs élevées des échelles, ceci est due à la proximité des discontinuités de dérivée, la mesure de α s'en trouve ainsi fortement dégradée aux fortes échelles, voir figure 6.

3.2 Application au signal de parole

La transformée en ondelettes (chapeau mexicain) est appliquée maintenant sur un signal de parole de type voisé, il s'agit de la voyelle /o/ d'un locuteur féminin, de longueur 25 ms, échantillonnée à 20 KHz, et extraite du mot « north » de la base de sons Keele.

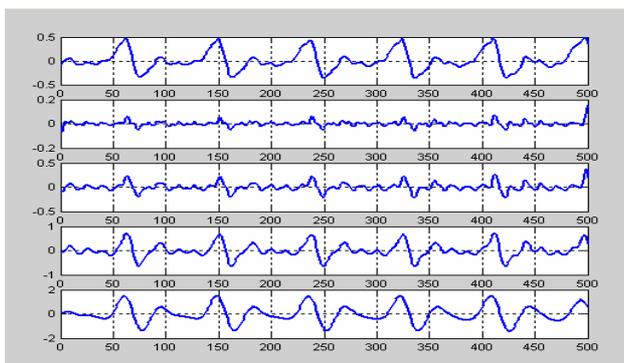


FIG. 4 : modules maximaux de la transformée en ondelettes (chapeau mexicain) de la voyelle /o/ d'un locuteur féminin pour les échelles allant de 2^{-4} à 2^0

La figure 4 montre le signal de parole suivi de ses transformées en ondelettes allant de l'échelle 2^{-4} à 2^0 .

Cette transformée en ondelettes présente deux lignes de modules maximaux qui convergent en deux points distincts correspondants aux deux points de discontinuités de tangentes qui encadrent la fermeture de la glotte.

D'après l'approximation 12, la décroissance de $\log_2 |Wf(u,s)|$ représentée en fonction de $\log_2(s)$ permet de donner aux faibles échelles la valeur $n+1/2$ et aux fortes échelles la valeur de $\alpha+1/2$.

La correspondance avec les mesures opérées sur le modèle lissé permet d'affecter au signal de parole une faible valeur du coefficient de diffusion, car la pente de décroissance de $\log_2 |Wf(u,s)|$ (voir figure 6) passe de 2.5 correspondant à $2+1/2$, 2 étant l'ordre de l'ondelette vers la valeur 1.5, α étant égale à 1, en passant par des valeurs intermédiaires.

Elle continue à décroître vers des valeurs inférieures à 1.5 car à cette échelle le support de l'ondelette couvre d'autres discontinuités du signal.

La mesure opérée sur une voix masculine donne un résultat similaire. L'échelle de diffusion étant relativement faible montre que le signal de parole est faiblement lissé.

La mesure opérée à l'aide d'ondelettes de moment 1 [8] ne révèle pas de maxima locaux dans la zone de décroissance rapide du signal entre les discontinuités de tangentes, alors que l'analyse par ondelette de moment 2 montre bien deux discontinuités de tangente associées à un léger lissage.

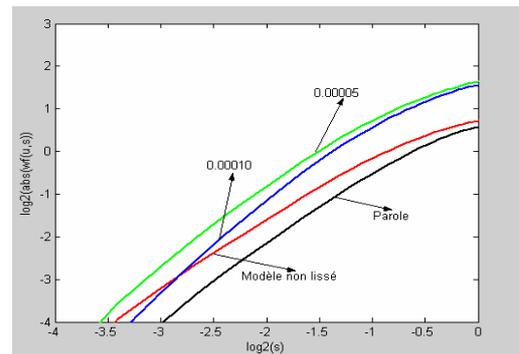


FIG. 5 : $\log_2 |Wf(u,s)|$ en fonction de $\log_2(s)$, la courbe la plus basse correspond à la voyelle /o/, la courbe supérieure concerne le modèle sans lissage suivi des modèles lissés pour deux valeurs de σ : $5 \cdot 10^{-5}$ et 10^{-4}

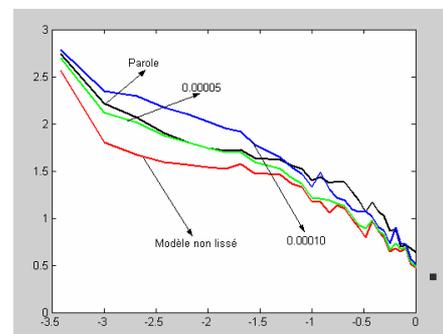


FIG. 6 : pentes calculées de $\log_2 |Wf(u,s)|$ en fonction de $\log_2(s)$ correspondantes au signal de parole, modèle non lissé et lissé pour les deux valeurs σ : $5 \cdot 10^{-5}$ et 10^{-4}

4. Conclusion

La caractérisation de la régularité du signal de parole aux instants de fermeture de la glotte est opérée sur un son de parole voisé. La mesure du coefficient de Lipschitz est utilisée pour caractériser la singularité, cette mesure est opérée à l'aide de la décroissance des modules maximaux de la transformée en ondelettes aux faibles échelles.

Un modèle simplifié de la décroissance du signal de parole à la fermeture de la glotte est utilisé. Ce modèle est lissé a priori pour la mesure du facteur de diffusion. La mesure de la décroissance des modules maximaux aux niveaux des singularités de tangente est opérée grâce à une ondelette à deux moments nuls (chapeau mexicain).

Ce modèle montre que l'échelle de diffusion du signal de parole est très faible ce qui permet d'atteindre rapidement la valeur du coefficient de Lipschitz égale à 1, toutefois la proximité des singularités est mise en valeur par une décroissance plus faible que 1.5.

Ceci permet de considérer que la décroissance rapide de la transformée en ondelettes du signal de parole à la fermeture de la glotte n'a pas pour origine une discontinuité du signal mais plutôt deux discontinuités de la dérivée.

Références

- [1] S. Mallat et A. Samarin. *Singularity Detection and processing with wavelets*. IEEE Transactions on Information Theory, Vol. 38, No 2, pp. 617-643, March 1992.
- [2] S. Kadambe and G. Faye Boudreaux-Bartels. *Application of the Wavelet Transform for Pitch Detection of Speech Signals*. IEEE Transactions on Information Theory, vol. 38, No2, pp. 917-924, March 1992.
- [3] M. Holshneider and P. Tchamitchian. *Régularité Locale de la Fonction non -Differentiable de Riemann*. in Les ondelettes en 1989, Lecture notes in Mathematics, P. G. Lemarie, Edition New York : Springer-Verlag, 1989.
- [4] S. Mallat. *Une Exploration des Signaux en Ondelettes*. Les Editions de l'Ecole Polytechnique, Paris, Juillet 2000.
- [5] S. Jaffard and Y. Meyer. *Wavelet Methods for Pointwise Regularity and Local Oscillations of Functions*. Tome 123. American Mathematical Society, Providence RI, 1996.
- [6] A. Witkin. *Scale Space Filtering*. Presented at Proc. Int. Joint conf. Artificial Intell, 1983.
- [7] S. Jaffard. *Exposants de Holder en des points donnés et coefficients d'ondelettes*. Notes au compte rendu de l'académie des sciences, France, vol. 308, ser. I, pp. 79-81, 1989.
- [8] A. Bouzid et N. Ellouze. *Caractérisation des singularités aux instants de fermeture de la glotte du signal de parole voisé par les modules maximaux de la transformée en ondelettes*. International Conference on Sciences of