

Une technique d'assistance par le bruit pour aider l'opérateur de Teager-Kaiser à suivre une composante fréquentielle perturbée

Stéphane ROSSIGNOL

CentraleSupélec – Campus de Metz, Équipe IMS
2 rue Édouard Belin, F-57070 Metz, France
Stephane.Rossignol@centralesupelec.fr

Résumé – Cet article propose une technique pour améliorer l'opérateur de Teager-Kaiser (OTK) utilisé pour le suivi d'une composante fréquentielle perturbée par un bruit additif. Les méthodes basées sur l'OTK classiques sont très peu robustes aux diverses perturbations possibles. Une méthode d'assistance par le bruit est mise en place. Nous montrons qu'elle permet de statistiquement améliorer cette robustesse.

Abstract – This paper deals with the improvement of the Teager-Kaiser energy algorithm (TKEA), designed to track a frequency component corrupted by an additive noise. Methods based on the original TKEA are not robust at all when corrupted by such perturbations. It is proposed here a method based on the Noise Assisted Data Analysis (NADA) paradigm. It is shown that it allows to significantly increase this robustness.

1 Introduction

Dans [1] est montré dans le cas de signaux musicaux que des méthodes basées sur l'OTK permettent de suivre efficacement le trajet en fréquence (f_0) d'une sinusoïde dans le cas où le signal n'est pas bruité et n'est pas gêné par les harmoniques de cette sinusoïde (de fréquences $2f_0, 3f_0\dots$). Ces méthodes reposant sur l'OTK se montrent alors plus performantes que la Transformée de Fourier à Court Terme (TFCT) et les méthodes basées sur le filtrage de Hilbert (FH), aussi bien quand il s'agit de suivre une modulation de fréquence (vibrato) et/ou un saut de fréquence (changement de note). Par contre, dès qu'un bruit additif est présent, même de faible variance, ou dès que les harmoniques sont présents, même d'amplitudes très petites, les techniques basées sur l'OTK s'effondrent complètement.

La particularité de l'OTK est qu'il n'a besoin que de quelques échantillons du signal pour estimer la fréquence. Typiquement, seulement 4 échantillons sont requis, alors que TFCT et FH requièrent un découpage en trames comptant quelques centaines d'échantillons (en traitement des sons, des trames de 30 ms sont communément utilisées, ce qui correspond à 240 échantillons pour une fréquence d'échantillonnage de 8 kHz). Les petites trames utilisées par l'OTK expliquent pourquoi grâce à lui il est possible de suivre plus précisément qu'avec les deux autres méthodes les détails des variations de fréquence du signal.

L'estimateur de fréquence basé sur l'OTK étudié dans cet article est DESA-1a (Discrete Energy Separation Algorithm; détails donnés en 2.1). L'idée est de coupler plusieurs estimations successives (de faire donc l'hypothèse que le signal est pseudo-stationnaire sur plus que les 4 échantillons nécessaires à la méthode classique), afin de réduire l'influence des paramètres perturbateurs et ainsi d'augmenter la robustesse de la

méthode, tout en considérant une taille de trames plus petite que celles requises par les deux autres méthodes (soit, la TFCT et le FH), ce afin de ne pas perdre les qualités intrinsèques des estimateurs basés sur l'OTK. Pour ce faire, des techniques d'assistance par le bruit doivent être mises en place. Ces méthodes appelées « Noise Assisted Data Analysis » (NADA), en anglais, sont de plus en plus utilisées entre autres en analyse spectrale.

2 Description de la technique

2.1 L'opérateur de Teager-Kaiser (OTK)

Une rapide description de la méthode est donnée ci-dessous. La littérature au sujet de l'OTK est abondante. Voir par exemple [2], [3] (articles de revue sur l'OTK), [4] (article de revue plus petit que les précédents; offrant de plus des développements théoriques), [5] (critique de l'OTK et des méthodes basées sur le signal analytique, c'est-à-dire sur le FH), [6] (méthodes basées sur le signal analytique; offrant de plus un grand nombre de développements théoriques). Le principe est rapidement décrit ici. Soit un signal AM-FM discret modélisé ainsi dans le cas général :

$$x[n] = a[n] \cos \left[\Omega_c n + \Omega_m \int_0^n q(m) dm + \Theta \right].$$

Sa fréquence instantanée est : $\Omega_i[n] = \Omega_c + \Omega_m q[n]$.

L'OTK vaut : $\Psi[x[n]] = x^2[n] - x[n-1]x[n+1]$.

Alors, en faisant l'hypothèse forte supplémentaire suivante : si on admet que les signaux $a[n]$ et $\Omega_i[n]$ ne varient pas trop vite ou de manière trop importante, on peut estimer $a[n]$ et $\Omega_i[n]$ en utilisant seulement 4 échantillons successifs du signal.

L'estimateur appelé DESA-1a s'écrit finalement :

$$\begin{cases} \hat{\Omega}_i[n] \simeq \arccos\left(1 - \frac{\Psi[x[n] - x[n-1]]}{2\Psi[x[n]]}\right) \\ \hat{a}[n] \simeq \sqrt{\frac{\Psi[x[n]]}{1 - \left(1 - \frac{\Psi[x[n] - x[n-1]]}{2\Psi[x[n]]}\right)^2}} \end{cases}$$

Il utilise les deux expressions classiques de l'OTK, $\Psi[x[n]]$ et $\Psi[x[n] - x[n-1]]$, appelées Ψ_1 et Ψ_2 dans la suite.

Dans cet article, on ne s'intéresse qu'à la fréquence instantanée. On ne va pas considérer le cas où une modulation d'amplitude (trémolo) vient perturber l'estimateur.

2.2 Techniques d'assistance par le bruit

Les méthodes d'assistance par le bruit (ou NADA) ont été développées notamment dans le cadre de l'EMD (Empirical Mode Decomposition). Voir [7], où une bibliographie concernant les méthodes NADA, est fournie ; ou encore [8] pour des développements théoriques.

Bien sûr, rien ne permet de déterminer qu'une estimation de $\hat{\Omega}_i[n]$, prise individuellement, est mauvaise (soit, très éloignée de la vraie fréquence instantanée du signal). Il est assez aisé de montrer que la moyenne de l'OTK en présence d'un bruit blanc de moyenne nulle indépendant du signal n'est pas la valeur recherchée par l'OTK. En effet, on a, si $X[n] = x[n] + b[n]$, avec $b : \mathcal{N}(0, \sigma^2) : E[\Psi[X[n]]] = \Psi[x[n]] + \sigma^2$. Il est facile de montrer aussi que : $E[\Psi[X[n] - X[n-1]]] = \Psi[x[n] - x[n-1]] + 2\sigma^2$. De ce fait, moyenniser les estimations de $\hat{\Omega}_i[n]$ sur plusieurs trames de 4 échantillons successives ne permet pas, en général, de trouver la bonne fréquence instantanée. L'estimateur en fait est biaisé (la preuve théorique de ceci n'est pas simple à dériver, mais une preuve expérimentale le montre). À ce niveau les méthodes basées sur l'assistance par le bruit doivent intervenir. Ces méthodes sont développées depuis quelques années dans divers domaines du traitement du signal.

Les étapes de la stratégie mise en place dans le cadre de notre problème sont données ci-dessous.

[0.] Tirage d'un signal à analyser. Dans les sections suivantes, est décrit le signal simple de test utilisé. Lors de cette étape, certains paramètres doivent être tirés aléatoirement à chaque fois ; pour le signal de test, c'est la phase de la sinusoïde et le bruit perturbateur (de moyenne nulle et de variance σ^2 fixe ; il faut cependant noter que l'influence de σ^2 sur les performances de la méthode devra être mesurée elle aussi).

[1.] Ajout d'un bruit blanc artificiel, assistant, à la moyenne (nulle ici) et à la variance σ_a^2 parfaitement contrôlées, et tiré aléatoirement, au signal X original.

[2.] Évaluation de $\hat{\Omega}_i[n]$ pour un nombre N d'instantanés n successifs. Ce nombre N donne la taille de la fenêtre d'analyse, qui doit rester bien sûr inférieure à celle utilisée pour la TFCT et le FH. On peut utiliser les expressions standards $\Psi[X[n]]$ et $\Psi[X[n] - X[n-1]]$, ou on peut utiliser une version modifiée de celles-ci, tenant compte du fait que la variance des bruits assistant est connue. On a, pour la première : $\Psi_1^m[x[n]] = x^2[n] - x[n-1]x[n+1] - \sigma_a^2$. La seconde, Ψ_2^m , est aisée à dériver.

[3.] Calcul de la moyenne de ces N mesures, pour obtenir $\overline{\hat{\Omega}_i[n]}$. On garde pour traitements ultérieurs cette moyenne si les N mesures sont toutes suffisamment proches de cette moyenne, c'est-à-dire si elles sont toutes à moins de b % de celle-ci.

[4.] Itération sur les points 1., 2. et 3. un nombre m de fois, qui peut être (et doit être sans doute) très grand. Sur ces m itérations, seulement $m_- \leq m$ sont gardées, selon la procédure décrite en 3. Si $m_- = 0$, il faut ajouter une étape supplémentaire : il faut augmenter b . Ceci est fait par pas de 1 %, jusqu'à ce que m_- soit différent de 0.

[5.] Calcul de la moyenne, sur les m_- valeurs gardées, des $\overline{\hat{\Omega}_i[n]}$; on l'appelle $\overline{\hat{\Omega}_i[n]}$ dans la suite : elle correspond à l'estimation finale de la fréquence instantanée $f_0[n]$.

[6.] On mesure si l'estimation obtenue est meilleure, ce qui veut dire ici plus proche du f_0 à obtenir, que l'estimation classique. Celle-ci est obtenue sans l'assistance du bruit, et en moyennant sur les N estimations. On obtient un binaire e valant 0 (échec de la méthode) ou 1 (succès de la méthode).

[7.] On recommande à l'étape 0. un grand nombre de fois (plusieurs milliers, pour chacun des tests présentés dans la suite), afin de pouvoir faire des statistiques sur e .

On voit qu'un grand nombre de paramètres peuvent entrer en ligne de compte et perturber les performances de la méthode : σ^2 , σ_a^2 , N , b , m , et si on utilise les OTK Ψ_1^m et Ψ_2^m plutôt que les opérateurs classiques Ψ_1 et Ψ_2 (voir l'introduction et l'étape 2.). Dans la suite de l'article, on essaie de caractériser les effets respectifs de ces paramètres.

3 Résultats – Sinusoïde de fréquence fixe

3.1 Modèle de signal

Ce travail s'inscrit principalement dans le cadre du traitement des sons (parole, musique, cris d'animaux, bruits divers) perçus par l'oreille humaine. On va donc considérer des valeurs de fréquence ayant sens pour ces domaines d'applications. On a de ce fait choisi comme fréquence d'échantillonnage 8 kHz (minimum requis) et comme fréquence de la sinusoïde 100 Hz (pitch typique d'une voix d'homme).

3.2 Un exemple

Un bruit blanc additif, perturbateur, vient s'ajouter à la sinusoïde. On le prend de variance σ^2 très petite : ça suffit pour dégrader de façon notable l'estimateur DESA-1a classique (voir [1] à ce sujet). Les valeurs utilisées pour les autres paramètres sont : $\sigma = 3.10^{-5}$: écart type du bruit ; $\sigma_a = \sigma$: écart type des bruits assistant ; $b = 5$: écart maximum (en pour-cent) entre chacune des N estimations et leur moyenne pour que celle-ci soit acceptée ; les expressions de l'OTK où σ_a^2 et $2\sigma_a^2$ sont enlevés sont utilisées (soit : Ψ_1^m et Ψ_2^m).

La figure 1 présente un exemple de ce qu'on obtient pour $N = 2$. À gauche, on donne la répartition des deux fréquences obtenues ; à droite, on donne les doublets d'estimations gardés :

ici, environ un tiers des doublets sont gardés.

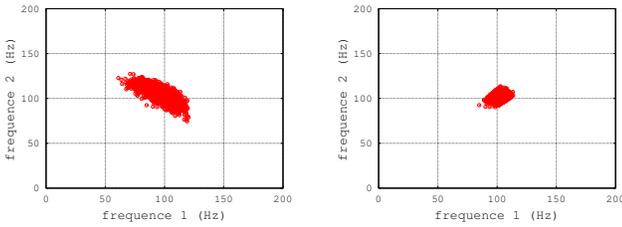


FIG. 1 – Répartition des $N = 2$ fréquences estimées, pour l'ensemble des m mesures et pour les $m_- \leq m$ gardées

Considérant l'histogramme des moyennes $\widehat{\Omega}_i[n]$ des doublets, on constate que son maximum (correspondant presque à $\widehat{\Omega}_i[n]$) tombe bien autour de la fréquence cherchée $f_0 = 100$ Hz. Il est heureux aussi de constater que les doublets se répartissent plutôt sur la diagonale perpendiculaire à la diagonale des $f_1 = f_2$: ainsi, moins de doublets sont acceptés, mais ceux qui le sont tombent sur des estimations plus justes de f_0 .

En fait, si on considère les deux estimations standards (c'est-à-dire sans les bruits assistant), on obtient 102.19 Hz et 92.30 Hz (on utilise Ψ_1^m et Ψ_2^m pour ces mesures aussi), ce qui donne pour moyenne 97.25 Hz ; notre méthode donne 97.78 Hz. Dans cet exemple spécifique, notre méthode améliore l'estimation de f_0 . Il faut à présent mesurer statistiquement cette amélioration, et déterminer les influences respectives des paramètres de la méthode.

3.3 Effets de N et m

Sur la figure 2/gauche, on montre l'influence de N , le nombre d'estimations, et de m , le nombre de bruits assistant utilisés. On voit que, de toute façon, dans plus de 50 pour-cent des cas on améliore la mesure de f_0 . On constate aussi que les paramètres N et m sont liés. On voit que plus N est grand, plus il faut que m le soit aussi pour que les résultats continuent de s'améliorer. En effet, plus N est grand, moins il y a d'ensembles d'estimations qui respectent le critère b et plus les statistiques sur e deviennent du coup incertaines. Par contre, le coup en temps de calcul peut devenir très prohibitif. Dans cet article, on considère qu'on a à notre disposition tout le temps de calcul qu'on pourrait requérir. On constate qu'avec $N = 5$ et m de l'ordre de 2000 ou plus, il est déjà possible d'obtenir dans 80 % des cas une meilleure estimation de f_0 avec la méthode décrite dans cet article qu'avec la méthode classique.

3.4 Effet de σ_a

Bien sûr, la variance σ^2 du bruit additif perturbateur n'est pas connue. Il s'agit ici de mesurer jusqu'à quel point on peut prendre une valeur pour σ_a^2 très différente de celle de σ^2 . Déjà, si σ_a^2 est trop petite, il est clair que les bruits assistant ne pourront pas faire leur office (ils seront noyés dans le bruit addi-

tif) ; par contre, si σ_a^2 est très grande, rien ne permet de dire qu'en augmentant N et/ou m (préférentiellement) on ne pourrait pas arriver à compenser le bruit additif. On garde les mêmes valeurs de paramètres que ci-dessus ; on fixe N à 5. Sur la figure 2/droite, on observe le comportement de la méthode. La variance du bruit additif est indiquée par un trait noir vertical.

Les deux phénomènes décrits ci-dessus sont bien visibles. Une sous-estimation de la variance du bruit additif est plutôt néfaste pour la méthode ; une sur-estimation semble pouvoir être compensée en augmentant m (les limitations relevant dès lors de la puissance de calcul disponible). Notre méthode devient efficace dès que σ_a^2 devient supérieure à un tiers de σ^2 ; et elle le reste aisément pour des variances très grandes dès que m devient suffisamment grand (de l'ordre de 2000). Il semble y avoir un optimum entre $0.5\sigma^2$ et $2\sigma^2$, pour lequel on peut espérer dépasser 80 % de meilleures estimations de f_0 avec notre méthode qu'avec la méthode classique.

3.5 Effet de b

Plus b est petit, plus m_- est petit. De ce fait, il faut augmenter m pour assurer une bonne estimation de la fréquence instantanée. Par contre, plus b est petit, plus on peut espérer que notre méthode estime précisément cette fréquence. De la même façon, plus b est grand, plus l'estimation est éloignée de la vraie valeur, car des cas plus déviants sont acceptés. De ce fait, là aussi il faut augmenter m pour améliorer l'estimation. Il semble donc que, pour un m donné, on passe pas un b optimal.

On obtient la figure 3/gauche. Les phénomènes décrits ci-dessus ne sont pas forcément facilement observés, tout du moins pour les petits b : ceci est dû à la sous-étape supplémentaire présente pour l'étape 4. de la méthode. On voit que le choix fait dans les sections précédentes de prendre b égale à 5 est raisonnable (même si un peu petit, peut-être), tout du moins quand m est suffisamment grand (dès qu'égal à 1000 au plus).

3.6 Utiliser Ψ_1^m, Ψ_2^m au lieu de Ψ_1, Ψ_2

On garde les mêmes valeurs pour les paramètres que précédemment : $\sigma = 3.10^{-5}$, $\sigma_a = \sigma$, $b = 5$. Dans le tableau 1, est donnée la moyenne de e obtenue dans l'une et l'autre configurations, calculée sur 7000 essais, ce pour différents m . Pas d'effet vraiment notable n'est remarqué.

3.7 Effet de la variance du bruit additif

On donne les figures 3/droite et 4 pour $\sigma_a^2 = \sigma$. Bien sûr, en général, on ne connaît pas la variance du bruit additif, et donc il n'est pas aisé d'arriver à ça. On examine aussi l'écart entre la fréquence trouvée et celle à trouver. On voit clairement l'apport de la méthode proposée : elle donne dans plus de 70 % des cas de meilleurs résultats que la méthode classique, et surtout des résultats qui restent semble-t-il non-biaisés quelque soit la variance du bruit perturbateur, contrairement à la méthode classique, qui s'effondre : ce résultat en soit est extrêmement inté-

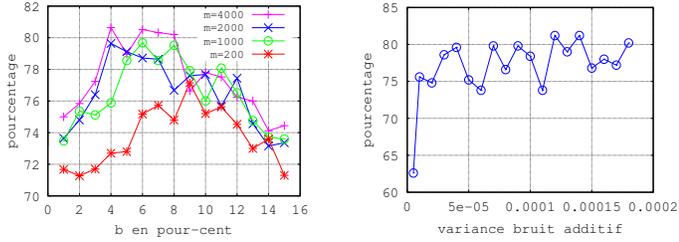
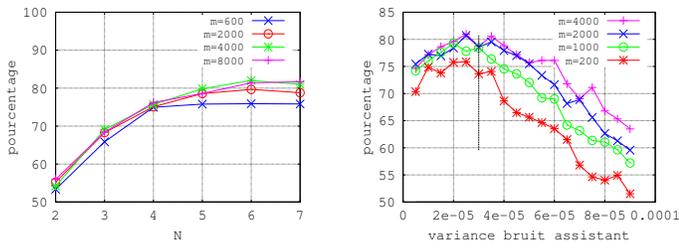


FIG. 3 – Pourcentage de e valant 1, indiquant que notre méthode améliore (étape 6.) la méthode classique. Gauche : b varie ; pour différents m . Droite : la variance du bruit perturbateur varie ; pour $m = 1000$

ressant et valide à lui tout seul notre méthode.

4 Conclusion

Nous avons montré qu'il est possible d'augmenter la robustesse des méthodes de suivi d'une composante fréquentielle perturbée fonctionnant grâce à l'OTK, qui est un opérateur à extrêmement court terme. Le réglage des paramètres ne semble pas critique, si on dispose d'une puissance de calcul suffisamment élevée ; si ce n'est pas le cas, il faut faire attention. On donne quelques perspectives ici. Il reste à explorer plus en détails les effets des divers paramètres. Il serait intéressant d'observer le comportement de la méthode en présence d'une modulation, d'une transition et d'harmoniques perturbateurs. Dans le cadre du traitement de la musique, la modulation d'amplitude correspond à un trémolo, or on sait que pour la voix chantée quand un vibrato (une modulation de fréquence) est présent, il y a forcément un trémolo. De plus, pour les sons, chaque transition en fréquence est accompagné d'une transition en amplitude. Il existe d'autres estimateurs que DESA-1a : il serait bon d'explorer les performances obtenus pour ces autres estimateurs. Le bruit additif et surtout la présence des harmoniques restent un problème. Il est supposé dans cet article que les amplitudes des harmoniques perturbateurs étaient petites devant celle de la sinusoïde à analyser. Bien évidemment, ce n'est en général pas le cas. Il faudrait trouver un moyen d'isoler chacun des harmoniques, et ce plus efficacement que grâce aux mé-

TAB. 1 – Moyenne (en %) de e pour les 2 types de l'OTK et pour différents m

	Avec Ψ_1 et Ψ_2	Avec Ψ_1^m et Ψ_2^m
$m = 200$	$m = 74.057$	$m = 73.057$
$m = 1000$	$m = 77.800$	$m = 77.386$
$m = 4000$	$m = 79.320$	$m = 79.740$

thodes décrites dans [1] (voir [9]). Le coût de calcul peut être problématique, mais la méthode est aisément parallélisable.

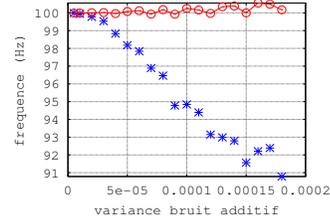


FIG. 4 – f_0 trouvée avec notre méthode (plein) et la méthode classique (étoiles) ; la variance du bruit additif varie ; pour $m = 1000$; en sachant que le f_0 à trouver est 100 Hz

Références

- [1] S. Rossignol, P. Desain, H. Honing, Refined Knowledge-Based f_0 Tracking : Comparing Three Extraction Methods, ICMC, 2001.
- [2] P. Maragos, Energy separation in signal modulations with application to speech analysis, IEEE Transactions on Signal Processing, v41n10, 1993.
- [3] D. Dimitriadis, P. Maragos, Continuous energy demodulation and application to speech analysis, Speech Communication, v48, 2006.
- [4] B. Boashash, Estimating and interpreting the instantaneous frequency of a signal, Proc. IEEE, v80n4, 1992.
- [5] D. Vakman, On the AS, the TK energy algorithm, and other methods for defining amplitude and frequency, IEEE Transactions on Signal Processing, v44n4, 1996.
- [6] A. L. Wang, Instantaneous and frequency-warped signal processing techniques for auditory source separation, PhD thesis, Stanford University, 1994.
- [7] Z. Wu, N. E. Huang, Ensemble Empirical Mode Decomposition : A Noise Assisted Data Analysis Method, Advances in Adaptive Data Analysis, 2009.
- [8] M. A. Colominas, G. Schlotthauer, M. E. Torres, P. Flandrin, Noise-assisted EMD methods in action, Advances in Adaptive Data Analysis, 2012.
- [9] A. Bouchikhi, A.-O. Boudraa, Multicomponent AM-FM Signals Analysis Based on EMD-B-Splines ESA, Signal processing, v92, 2014.