Une ou deux composantes ? La réponse de la diffusion en ondelettes

Vincent LOSTANLEN

Music and Audio Research Lab, Université de New York
35 West Fourth Street, New York, États-Unis
vincent.lostanlen@nyu.edu

Résumé – Dans le contexte d'une modélisation biologiquement plausible de l'écoute artificielle, on s'intéresse à la représentation d'un signal stationnaire multicomposantes par un réseau de diffusion en ondelettes. D'abord, nous montrons que la renormalisation des coefficients du second ordre par ceux du premier ordre donne un critère numérique simple pour établir si deux composantes de fréquences voisines interfèrent psychoacoustiquement. Ensuite, on généralise le cadre théorique "une ou deux composantes" à trois sinusoïdes ou plus, et montrons en particulier qu'un réseau de diffusion en ondelettes de profondeur $M = \log_2 N$ suffit à caractériser l'amplitude relative des N premiers termes d'une série de Fourier, tout en bénéficiant d'une propriété d'invariance à la transposition fréquentielle ainsi qu'au déphasage de chaque composante.

Abstract – With the aim of constructing a biologically plausible model of machine listening, we study the representation of a multicomponent stationary signal by a wavelet scattering network. First, we show that renormalizing second-order nodes by their first-order parents gives a simple numerical criterion to establish whether two neighboring components will interfere psychoacoustically. Secondly, we generalize the "one or two components" framework to three sine waves or more, and show that a network of depth $M = \log_2 N$ suffices to characterize the relative amplitudes of the first N terms in a Fourier series, while enjoying properties of invariance to frequency transposition and component-wise phase shifts.

1 Introduction

En physiologie de l'audition, les cellules ciliées de notre cochlée jouent le rôle de filtres à facteur de qualité constant, dont la bande passante est large d'environ un quart d'octave. Ainsi, étant données deux sinusoïdes $t \mapsto \boldsymbol{x_1}(t) = a_1 \cos(f_1 t + \varphi_1)$ et $t \mapsto \boldsymbol{x_2}(t) = a_2 \cos(f_2 t + \varphi_2)$ de fréquences respectives $f_1 > 0$ et $f_2 > 0$, nous percevons leur somme comme un accord de deux sons purs à condition que $\boldsymbol{x_1}$ et $\boldsymbol{x_2}$ appartiennent à des bandes critiques distinctes. En revanche, si $a_2 \ll a_1$ ou $f_2 \approx f_1$, alors la composante $\boldsymbol{x_2}$ est dite *masquée* par $\boldsymbol{x_1}$. Au lieu de sons purs, nous entendons un "battement" : une onde localement sinusoïdale dont la fréquence porteuse est $\frac{1}{2}(f_1 + f_2)$ et la fréquence de modulation est $\frac{1}{2}|f_1 - f_2|$. La perception de cette modulation d'amplitude met en jeu des processus cognitifs ultérieurs à la cochlée, et notamment le cortex auditif primaire.

Or, la diffusion en ondelettes (*wavelet scattering*) est un opérateur qui alterne, en cascade, un banc de filtres passe-bandes analytiques à facteur de qualité constant Q, et l'application en tout point du module complexe [8]. Dès lors, ses deux premiers niveaux de profondeur modélisent respectivement les propriétés psychoacoustiques de la cochlée et du cortex auditif primaire. Cet opérateur est employé à des fins de reconnaissance de la parole [1], de sons environnementaux [6], de structures musicales répétées [4], et de modes de jeux étendus [5]. La diffusion en ondelettes bénéficie donc, simultanément, d'une assise mathématique solide, d'applications innovantes en ingénierie informatique, et de liens privilégiés avec la neurophysiologie.

Cet article caractérise la réponse théorique de la diffusion en ondelettes au signal $\boldsymbol{x}(t) = \boldsymbol{x_1}(t) + \boldsymbol{x_2}(t)$. À ce titre, il s'inscrit méthodologiquement dans une lignée de travaux antérieurs en traitement du signal non stationnaire, visant tous à étudier le comportement de tel ou tel opérateur convolutionnel non linéaire en fonction des grandeurs relatives $\frac{a_2}{a_1}$, $\frac{f_2}{f_1}$, et $(\varphi_2 - \varphi_1)$. Trois de ces opérateurs, qui peuvent mis en regard de la diffusion en ondelettes, sont la décomposition modale empirique [9], le *synchrosqueezing* [11], et l'analyse spectrale singulière [3].

Premièrement, nous établissons un critère numérique, appelé *coefficient de masquage*, fondé sur la diffusion en ondelettes d'ordre deux, et tendant vers 0 pour $a_2 \to 0$ ou $|f_2 - f_1| \to 0$ et proche de 1 pour $a_1 \approx a_2$ et $0 < |f_2 - f_1| \ll f_1$. Deuxièmement, nous généralisons le cadre théorique "une ou deux composantes" au cas de $N > 2$ composantes $\boldsymbol{x_1}, \ldots, \boldsymbol{x_N}$, et étudions l'apport respectif de chaque ordre dans la diffusion en ondelettes de $\boldsymbol{x} = \sum_n \boldsymbol{x_n}$. En particulier, nous démontrons qu'un réseau de profondeur $M = \log_2 N$ caractérise les N premiers termes de la série de Fourier $\boldsymbol{x}(t) = \sum_n^N a_n \cos(n f_1 t + \varphi_n)$ par leurs amplitudes et fréquences respectives tout en étant invariant à la phase relative φ_n de chacune de ses composantes sinusoïdales.

Ce travail est financé par la bourse ERC InvariantClass 320959. Le code source reproduisant nos expériences et figures est en libre accès a l'adresse www.github.com/lostanlen/scattering.m

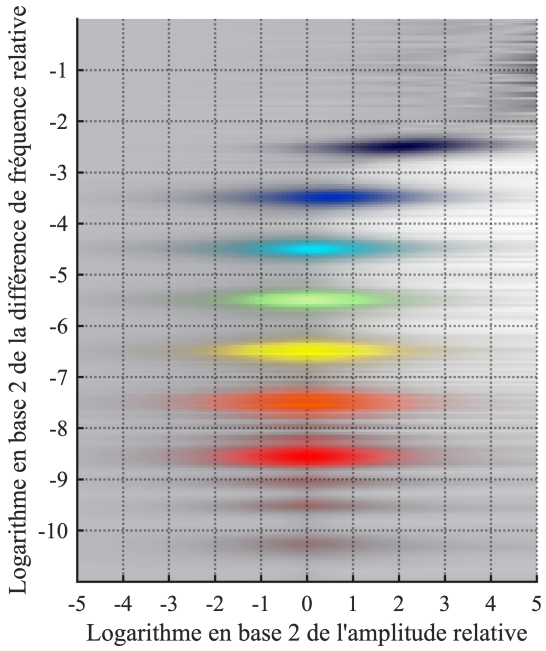

FIGURE 1 – Carte de chaleur du coefficient de masquage lors de la diffusion en ondelettes de deux composantes sinusoïdales $\boldsymbol{x_1}$ et $\boldsymbol{x_2}$, mesuré autour de la fréquence f_1, et évoluant en fonction de l'amplitude relative $\frac{a_2}{a_1}$ et de la différence de fréquence relative $\frac{|f_2-f_1|}{f_1}$. La couleur de chaque tache de diffusion indique la la résolution d'analyse λ_2 de la deuxième strate du réseau. Les ondelettes de diffusion ont un profil asymétrique (gammatone) et un facteur de qualité $Q = 4$. Elles couvrent un intervalle total de neuf octaves en-deçà de f_1. Par souci de clarté, on affiche une seule interférence par octave au lieu de quatre.

2 Interférométrie en ondelettes

Soit $\boldsymbol{\psi} \in \mathbf{L}^2(\mathbb{R}, \mathbb{C})$ un filtre analytique de moyenne nulle, de fréquence centrale 1, et de bande passante efficace $1/Q$. On définit un banc de filtres passe-bandes à facteur de qualité Q constant comme la famille d'ondelettes $\boldsymbol{\psi_\lambda} : t \mapsto \lambda\boldsymbol{\psi}(\lambda t)$ de fréquences centrales λ, de bandes passantes efficaces λ/Q, et d'échelles temporelles efficaces $2\pi Q/\lambda$. En pratique, la variable fréquentielle λ est discrétisée selon une série géométrique de raison $2^{1/Q}$. Autrement dit, tout signal continu à bande limitée $[f_{\min}, f_{\max}]$ fait résonner au plus un nombre $Q\log_2(\frac{f_{\max}}{f_{\min}})$ d'ondelettes $\boldsymbol{\psi_\lambda}$. On définit le scalogramme comme le carré du module complexe de la transformée en ondelettes continue

$$\begin{aligned}\mathbf{U_1}\boldsymbol{x}(t,\lambda_1) &= \left|\boldsymbol{x} * \boldsymbol{\psi_{\lambda_1}}\right|^2(t) \\ &= \left|\int_{\mathbb{R}} \boldsymbol{x}(t')\boldsymbol{\psi_{\lambda_1}}(t-t')\,\mathrm{d}t'\right|^2. \end{aligned} \tag{1}$$

De même, on définit un second niveau de transformation non linéaire comme le "scalogramme du scalogramme"

$$\mathbf{U_2}\boldsymbol{x}(t,\lambda_1,\lambda_2) = \left|\left|\boldsymbol{x} * \boldsymbol{\psi_{\lambda_1}}\right|^2 * \boldsymbol{\psi_{\lambda_2}}\right|^2(t). \tag{2}$$

Une telle construction peut être itérée *ad infinitum* en effectuant, pour tout entier naturel m, une "diffusion" (en anglais *scattering*) du signal temporel multivarié $\mathbf{U_m}$ dans chacune des sous-bandes $\lambda_m < \lambda_{m-1}$:

$$\begin{aligned}&\mathbf{U}_{m+1}\boldsymbol{x}(t,\lambda_1\dots\lambda_{m+1}) = \\ &\qquad\left|\mathbf{U}_{m+1}\boldsymbol{x}(t,\lambda_1\dots\lambda_m) * \boldsymbol{\psi}_{\lambda_m}\right|^2(t,\lambda_1\dots\lambda_m).\end{aligned} \tag{3}$$

Puisque chaque strate m du réseau s'exprime comme la composition d'un système linéaire invariant (la transformée en ondelettes continue) et d'une opération ponctuelle (le module complexe), on constate, par récurrence sur m, que tout tenseur $\mathbf{U}_m\boldsymbol{x}$ est équivariant à l'action du délai $\mathcal{T}_\tau : \boldsymbol{x} \mapsto \boldsymbol{x}(t-\tau)$. Afin de remplacer cette propriété d'équivariance par une propriété d'invariance, on intègre $\mathbf{U_m}$ sur une durée pré-établie T, conduisant ainsi à la diffusion invariante (*invariant scattering*)

$$\mathbf{S}_m\boldsymbol{x}(t,\lambda) = \int_{\mathbb{R}} \mathbf{U}_m(t',\lambda)\boldsymbol{\phi}_T(t-t')\,\mathrm{d}t', \tag{4}$$

où le m-uplet $\lambda = (\lambda_1\dots\lambda_m)$ est un chemin de diffusion (*scattering path*) et la fonction $\boldsymbol{\phi}_T$ un filtre passe-bas symétrique et réel de support temporel efficace T.

3 Coefficient de masquage

La convolution entre toute sinusoïde $\boldsymbol{x_n}$ et toute ondelette $\boldsymbol{\psi_{\lambda_1}}$ s'écrit comme une multiplication dans le domaine de Fourier. Puisque $\boldsymbol{\psi_{\lambda_1}}$ est analytique, seule la partie analytique $\boldsymbol{x_n^a} = \boldsymbol{x_n} + \mathrm{i}\mathcal{H}\{\mathbf{x_n}\} = a_n\exp(\mathrm{i}(f_nt+\varphi_n))$ du signal réel $\boldsymbol{x_n}$ est préservée :

$$\begin{aligned}&\left(\boldsymbol{x_n} * \boldsymbol{\psi_{\lambda_1}}\right)(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{\boldsymbol{x}}_{\boldsymbol{n}}(\omega)\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(\omega)\exp(\mathrm{i}\omega t)\,\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} a_n\mathrm{e}^{\mathrm{i}\varphi_n}\frac{\delta(\omega-f_n)+\delta(\omega+f_n)}{2}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(\omega)\exp(\mathrm{i}\omega t)\,\mathrm{d}\omega \\ &= \frac{1}{2}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(f_n)\exp\big(\mathrm{i}(f_nt+\varphi_n)\big) = \frac{1}{2}\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(f_n)\boldsymbol{x_n^a}(t).\end{aligned} \tag{5}$$

Par linéarité de la transformée en ondelettes continue, on décompose le cas à $N = 2$ composantes s'exprime comme une multiplication hétérodyne faisant intervenir le conjugué complexe :

$$\begin{aligned}&\left|(\boldsymbol{x_1}+\boldsymbol{x_2}) * \boldsymbol{\psi_{\lambda_1}}\right|^2(t) = \frac{1}{2}\left|\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(f_1)\boldsymbol{x_1^a}(t) + \widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(f_2)\boldsymbol{x_2^a}(t)\right|^2 \\ &= \Re\Big(\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(f_1)\widehat{\boldsymbol{\psi}}^*_{\boldsymbol{\lambda_1}}(f_2)\boldsymbol{x_1^a}(t)\boldsymbol{x_2^{a*}}(t)\Big) \\ &+ \frac{1}{2}\left|\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(f_1)\right|^2\left|\boldsymbol{x_1^a}(t)\right|^2 + \frac{1}{2}\left|\widehat{\boldsymbol{\psi}}_{\boldsymbol{\lambda_1}}(f_2)\right|^2\left|\boldsymbol{x_2^a}(t)\right|^2 \\ &= \Re\left(\widehat{\boldsymbol{\psi}}\Big(\frac{f_1}{\lambda_1}\Big)\widehat{\boldsymbol{\psi}}^*\Big(\frac{f_2}{\lambda_1}\Big)\right)a_1a_2\cos\big((f_2-f_1)t+(\varphi_2-\varphi_1)\big) \\ &+ \frac{1}{2}\left|\widehat{\boldsymbol{\psi}}(\frac{f_1}{\lambda_1})\right|^2a_1^2 + \frac{1}{2}\left|\widehat{\boldsymbol{\psi}}(\frac{f_2}{\lambda_1})\right|^2a_2^2.\end{aligned} \tag{6}$$

Puisque l'ondelette $\boldsymbol{\psi}$ est de moyenne nulle, les deux termes constants de l'équation ci-dessus, respectivement proportionels

à a_1^2 et a_2^2, sont absorbés par le réseau à l'ordre un, et disparaissent à partir de la deuxième strate. En revanche, le terme mixte, proportionnel à $a_1 a_2$, présente une fréquence fondamentale $\Delta f = |f_2 - f_1|$. Les auteurs d'une publication précédente ont remarqué que cette interférence produit un pic d'énergie au deuxième ordre de la diffusion pour $\lambda_1 = f_1$ et $\lambda_2 = |f_2 - f_1|$ [2]. En revanche, ils ne commentent pas la dépendance de ce pic en terme de l'amplitude relative $\frac{a_2}{a_1}$, du profil de l'ondelette $\boldsymbol{\psi}$, du facteur de qualité Q, et de l'échelle temporelle de stationnarité locale T. Cet article propose de remédier à ce manque, par une analyse plus exhaustive de ces différents paramètres.

Dans la définition originelle de la diffusion en ondelettes, la non-linéarité employée est l'amplitude complexe ($|z| = \sqrt{z\bar{z}}$) plutôt que la puissance ($|z|^2 = z\bar{z}$) , afin de garantir que chaque opérateur soit Lipschitz-contractant [8]. Néanmoins, pour simplifier les calculs et épargner une étape de linéarisation de la racine carrée, nous choisissons d'adopter une mesure de puissance plutôt que d'amplitude.

L'équation 6 montre que le premier ordre de diffusion transforme le signal à deux composantes $\boldsymbol{x} = \boldsymbol{x_1} + \boldsymbol{x_2}$ en un signal à une composante. Pour que cette composante soit non négligeable, trois conditions doivent être réunies. Premièrement, le produit $a_1 a_2$ doit être non négligeable devant chacun des carrés a_1^2 et a_2^2. Deuxièmenent, il doit exister une résolution d'analyse λ_1 contenant les deux fréquences f_1 et f_2 dans sa bande passante. Autrement dit, λ_1 doit satisfaire l'inégalité $|\frac{f_n}{\lambda_1} - 1| < \frac{1}{Q}$ pour $f_n = f_1$ et $f_n = f_2$ à la fois. Troisièmement, la différence de fréquence Δf doit appartenir à la bande passante d'une ondelette $\boldsymbol{\psi_{\lambda_2}}$ au sein de la deuxième strate du réseau de diffusion. Or, en pratique, afin d'assurer la localisation temporelle des coefficients et de restreindre chaque banc de filtres à un nombre fini d'octaves, la dilatation des ondelettes $\boldsymbol{\psi_{\lambda_m}}$ est bornée par la constante temporelle T. Par conséquent, il est nécessaire que la période $\frac{2\pi}{|f_2 - f_1|}$ du signal différentiel soit inférieure à la pseudo-période de l'ondelette de support efficace T, c'est-à-dire QT. On a donc nécessairement $|f_2 - f_1| < \frac{2\pi Q}{T}$.

Une manière simple de mesurer le degré d'interférence mutuelle des composantes $\boldsymbol{x_1}$ et $\boldsymbol{x_2}$ consiste à renormaliser les coefficients de diffusion du second ordre par ceux du premier ordre [1]. Une telle opération de post-traitement se rapproche conceptuellement de méthodes classiques de contrôle adaptatif du gain en reconnaissance de parole, et notamment de la normalisation d'énergie par canal (*per-channel energy normalization* ou PCEN) [7].

En accord avec la méthodologie "une ou deux composantes" [9], la figure 1 illustre la valeur de ce rapport d'énergie dans la bande de fréquence $\lambda_1 = f_1$, pour différentes valeurs de l'amplitude relative $\frac{a_2}{a_1}$ et de la différence de fréquence relative $\frac{|f_2 - f_1|}{f_1}$. On a fixé $f_2 < f_1$ sans perte de généralité. Comme attendu, on constate que, pour $a_2 \approx a_1$ et une différence de fréquence relative comprise entre $\frac{Q}{f_1 T}$ et $\frac{1}{Q}$, les ondelettes de la seconde strate $\boldsymbol{\psi_{\lambda_2}}$ résonnent avec le signal différentiel produit par l'interférence des composantes $\boldsymbol{x_1}$ et $\boldsymbol{x_2}$. La différence des logarithmes entre λ_1 et λ_2 caractérise l'intervalle musical formé par ces deux composantes. De plus, afin d'approximer la réponse des cellules ciliées de la cochlée, on emploie des ondelettes Gammatone [6]. Ce choix a pour conséquence de produire un masquage fréquentiel asymétrique, : puisque $f_2 < f_1$ par convention, il est possible, à a_1 et a_2 fixées, que $\boldsymbol{x_1}$ soit dans la bande critique de $\boldsymbol{x_2}$ sans que l'inverse soit nécessairement vrai. L'asymétrie est d'autant plus prononcée que la différence de fréquence est grande. Ce phénomène est en accord avec des résultats expérimentaux connus en psychoacoustique.

4 Trois composantes ou plus

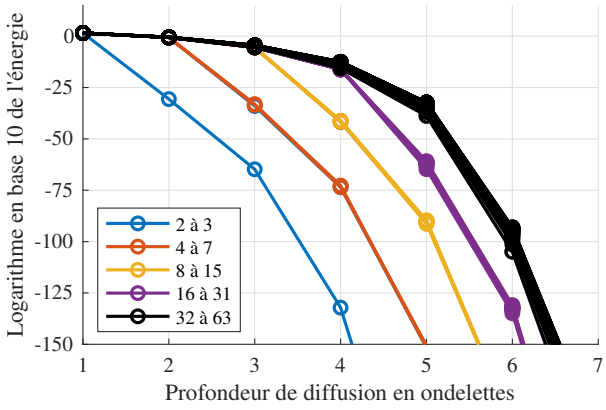

FIGURE 2 – Courbes de décroissance de l'énergie diffusée en fonction de la profondeur m du réseau de diffusion en ondelettes, pour des signaux à n composantes d'amplitude égale et de fréquences arithmétiquement espacées. La couleur de chaque courbe indique la partie entière de $\log_2 n$. Les ondelettes de diffusion ont un profil en sinus cardinal (ondelettes de Shannon) et un facteur de qualité $Q = 1$. À chaque niveau de profondeur, le banc de filtres couvre un intervalle total de sept octaves.

En acoustique musicale, les sons tonaux naturels sont rarement approximables comme un mélange de seulement deux composantes. Bien plus souvent, ils comportent dix composantes, voire plus, et couvrent simultanément plusieurs octaves. Dès lors, un calcul à l'ordre deux des coefficients de masquage n'apporte qu'une représentation grossière du contenu timbral de chaque bande critique. En effet, ces coefficient encodent les relations mutuelles entre paires de composantes, mais demeurent peu sensibles à la présence de structures plus globales dans l'enveloppe spectrale de la note de musique analysée.

Dans un contexte d'écoute artificielle, l'identification d'une source peut se formuler — en régime stationnaire, du moins — comme un problème inverse sur les paramètres physiques de l'équation des ondes, sous contrainte d'invariance à la fréquence fondamentale. S'agissant d'instruments de musique, on sait par exemple que le mode d'excitation (corde pincée ou frappée) influence l'exposant de décroissance de l'amplitude en fonction de la fréquence, tandis que les conditions aux bords (conduit fermé ou semi-ouvert) influence l'amplitude relative des harmoniques

pairs par rapport aux harmoniques impairs.

La construction d'un espace de similarité dans lequel les deux facteurs de variabilité susnommés se traduisent par des effets quasi linéaires, orthogonaux, invariants à la fréquence fondamentale, et robustes à la présence de bruit, est essentielle à l'amélioration des performances en apprentissage automatique, notamment supervisé, des sons musicaux. Pourtant, la difficulté de sa résolution réside dans son aspect multi-échelles. Si l'on restreint l'interférométrie à deux partiels adjacents en fréquence, il est difficile de distinguer les effets respectifs des conditions initiales (onde de forme rectangulaire ou triangulaire) et des conditions aux bords (onde à symétrie paire ou impaire) sur les paires d'amplitudes observées. En revanche, cette même distinction devient plus simple si l'on considère un quadruplet de partiels, comme par exemple les partiels 4 à 7 d'une série harmonique, occupant la troisième octave au-dessus de la fréquence fondamentale. À l'intérieur de cette octave, une variation dans l'exposant de décroissance du spectre se traduit par un déplacement d'énergie à large bande (petite échelle), des partiels 4 et 5 vers les partiels 6 et 7. À l'inverse, une variation dans l'amplitude relative des harmoniques impairs se traduit par un déplacement d'énergie à bande plus étroite (grande échelle), du partiel 4 vers le partiel 5 et du partiel 6 vers le partiel 7.

L'application alternée d'une transformée en ondelettes continue et du module complexe présente une forte ressemblance avec l'architecture des réseaux de neurones convolutifs. On parle donc de réseau de diffusion (*scattering network*), organisé selon différents niveaux (*layers*) de profondeur. Intuitivement, le passage d'un niveau au suivant double la multiplicité de l'interférence, exprimée en termes de nombre de composantes.

La figure 2 illustre cette dépendance logarithmique entre le nombre de composantes et la profondeur de diffusion maximale. On se place dans le cas particulier de la série de Fourier $\boldsymbol{x} : t \mapsto \sum_x^N \cos(nf_1 t)$, c'est-à-dire $a_n = a_1$, $f_n = nf_1$, et $\varphi_n = 0$ pour tout entier n. On étudie la norme ℓ^2 de chaque strate, obtenue en sommant les chemins $\lambda = (\lambda_1 \dots \lambda_M)$ de profondeur finie M. Pour N compris entre 2^M et 2^{M-1}, on constate que les M premières strates du réseau de diffusion parviennent à absorber la quasi-totalité de l'énergie totale du signal.

5 Conclusion

Dans cet article, on a abordé la question de comprendre le rôle de chaque strate dans un réseau de diffusion à partir d'une méthodologie de type "une ou deux composantes", particulièrement fructueuse en analyse temps-fréquence. On a montré qu'en présence d'une somme finie de composantes, le chemin de diffusion dans le réseau décroît en fréquence et en énergie. Dans le cas d'une série de Fourier, on a montré que l'énergie tombe subitement à zéro à partir d'une profondeur de l'ordre du logarithme du nombre de composantes. Ce résultat est un cas particulier du théorème de décroissance exponentielle des coefficients de diffusion [10], qui est généralement valable dans $\mathbf{L}^2$ mais exprimé en terme de borne supérieure d'énergie, non de borne supérieure de profondeur.

Références

[1] J. ANDÉN et S. MALLAT. « Deep scattering spectrum ». In : *IEEE Transactions on Signal Processing* 62.16 (2014), p. 4114–4128.

[2] J. ANDÉN et S. MALLAT. « Scattering representation of modulated sounds ». In : *Proceedings of the International Conference on Digital Audio Effects (DAF-x)*. 2012.

[3] J. HARMOUCHE et al. « Une ou deux composantes: la réponse de l'analyse spectrale singulière ». In : *Actes du colloque GRETSI*. 2015.

[4] V. LOSTANLEN. « Découverte de structures musicales en temps réel par la géométrie de l'information ». Mém.de mast. Ircam, 2013.

[5] V. LOSTANLEN, J. ANDÉN et M. LAGRANGE. « Extended playing techniques: The next milestone in musical instrument recognition ». In : *Proceedings of the International Conference in Digital Libraries for Musicology*. 2018.

[6] V. LOSTANLEN et al. « Relevance-based quantization of scattering features for unsupervised mining of environmental audio ». In : *EURASIP Journal on Audio, Speech, and Music Processing* 2018.1 (2018), p. 15.

[7] V. LOSTANLEN et al. « Per-Channel Energy Normalization: Why and How ». In : *IEEE Signal Processing Letters* 26.1 (2019), p. 39–43.

[8] S. MALLAT. « Group invariant scattering ». In : *Communications on Pure and Applied Mathematics* 65.10 (2012), p. 1331–1398.

[9] G. RILLING et P. FLANDRIN. « One or two frequencies? The empirical mode decomposition answers ». In : *IEEE Transactions on Signal Processing* 56.1 (2008), p. 85–95.

[10] I. WALDSPURGER. « Exponential decay of scattering coefficients ». In : *Proceedings of the International Conference on Sampling Theory and Applications (SampTA)*. IEEE. 2017, p. 143–146.

[11] H.-T. WU, P. FLANDRIN et I. DAUBECHIES. « One or two frequencies? The synchrosqueezing answers ». In : *Advances in Adaptive Data Analysis* 3.01–02 (2011), p. 29–39.