

Synthèse stochastique et modélisation acoustique

Analyse d'un générateur procédural de tempête côtière

Traitement du signal audio & synthèse procédurale

Août 2026

Résumé

Cet article présente l'analyse mathématique d'un générateur audio procédural produisant une boucle de 75 secondes évoquant une tempête côtière. Initialement conçu pour simuler une pluie douce, les paramètres du modèle engendrent un artefact acoustique remarquable : l'immersion dans un environnement maritime avec ressac, vent et éclaboussures. Nous formalisons l'intégralité de la chaîne de traitement du signal : façonnage spectral stochastique, modulation bipolaire périodique, synthèse stéréophonique Mid/Side, synthèse granulaire à phase directe, filtrage résonant, et chaîne finale de normalisation RMS et de réduction conditionnelle du gain.

1 Façonnage Spectral Stochastique

1.1 Génération du bruit gaussien complexe

Le lit sonore de la tempête est généré dans le domaine fréquentiel. On initialise un spectre de bruit blanc gaussien complexe sur $M = \frac{N}{2} + 1$ bins de fréquence :

$$X[k] = \mathcal{N}(0, 1) + i\mathcal{N}(0, 1), \quad \forall k \in [0, M - 1] \quad (1)$$

où $\mathcal{N}(0, 1)$ désigne la distribution normale standard. Ce spectre est ensuite sculpté par une fonction de transfert composite avant transformation inverse.

1.2 Courbes de magnitude de type Butterworth

Le façonnage spectral utilise des courbes de magnitude de type Butterworth appliquées directement aux bins FFT, et non comme filtres RII temporels. Le passe-haut est d'ordre 1 et le passe-bas d'ordre 2 :

$$H_{hp}(f) = \frac{1}{\sqrt{1 + \left(\frac{f_{hp}}{f}\right)^2}} \quad (2)$$

$$H_{lp}(f) = \frac{1}{\sqrt{1 + \left(\frac{f}{f_{lp}}\right)^4}} \quad (3)$$

Pour un filtre de Butterworth d'ordre N , l'exposant au dénominateur est $2N$. Ces courbes sont appliquées dans le domaine fréquentiel, ce qui garantit une réponse de phase nulle (filtrage zéro phase).

1.3 Tilt spectral et renforcements gaussiens

Trois termes de façonnage supplémentaires sont appliqués :

Tilt spectral. Un tilt en puissance modifie la pente générale du spectre :

$$H_{tilt}(f) = \left(\frac{f_{clamp}}{f_{ref}} \right)^\alpha \quad (4)$$

où f_{clamp} est bornée entre 60 Hz et 8 kHz, $f_{ref} = 1000$ Hz, et α est un exposant négatif typiquement compris entre $-0,15$ et $-0,90$ selon la couche.

Renforcement du corps grave. Un renforcement gaussien en fréquence logarithmique centre l'énergie autour des médiums-graves :

$$H_{body}(f) = 1 + A_{body} \exp \left(-\frac{1}{2} \left(\frac{\ln(f/f_{body})}{\sigma_{body}} \right)^2 \right) \quad (5)$$

où A_{body} contrôle l'amplitude du renforcement, f_{body} la fréquence centrale (typiquement 190–700 Hz), et σ_{body} contrôle la largeur dans le domaine logarithmique de fréquence. Avec $\sigma_{body} = 0,7$, la largeur correspondante est d'environ $0,7/\ln 2 \approx 1,01$ octave.

Renforcement de présence. Un second renforcement gaussien, plus large et plus discret, ajoute une légère présence autour de 3,5 kHz :

$$H_{presence}(f) = 1 + A_{pres} \exp \left(-\frac{1}{2} \left(\frac{\ln(f/3500)}{\sigma_{pres}} \right)^2 \right) \quad (6)$$

avec $A_{pres} = 0,02$, soit un renforcement de $+2\%$.

1.4 Fonction de transfert composite

La fonction de transfert totale est le produit des cinq termes :

$$H(f) = H_{hp}(f) \cdot H_{lp}(f) \cdot H_{tilt}(f) \cdot H_{body}(f) \cdot H_{presence}(f) \quad (7)$$

Le signal temporel est obtenu par transformation de Fourier inverse rapide :

$$x[n] = \text{IFFT} \{ X[k] \cdot H[k] \} \quad (8)$$

Suppression de la composante continue. Après synthèse, la valeur moyenne de chaque couche est soustraite :

$$x[n] \leftarrow x[n] - \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad (9)$$

Normalisation RMS par couche. Chaque couche est normalisée par son efficacité (RMS) avant application du gain de niveau :

$$x[n] \leftarrow \frac{x[n]}{\sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x[n]^2}} \cdot L \quad (10)$$

où L est le niveau de la couche.

2 Modulation Périodique Bipolaire

2.1 Construction du modulateur

Pour simuler les variations macroscopiques d'intensité (rafales de vent, houle), on construit un modulateur comme somme de sinusoïdes :

$$g(t) = \sum_{k=1}^K \frac{1}{c_k^{0.85}} \sin \left(2\pi c_k \frac{t}{T} + \phi_k \right) \quad (11)$$

où :

- $c_k \in \mathbb{Z}^+$ est un nombre entier de cycles sur la durée T ,
- $\phi_k \sim \mathcal{U}(0, 2\pi)$ est une phase aléatoire,
- l'amplitude décroît avec la fréquence selon $c_k^{-0.85}$.

La contrainte d'intégrité des cycles garantit que $g(0) = g(T)$, assurant une boucle parfaite.

2.2 Normalisation par le pic

Le modulateur est normalisé linéairement par son pic absolu :

$$G(t) = \frac{g(t)}{\max |g(t)|} \in [-1, 1] \quad (12)$$

Absence de conversion en décibels. Contrairement à une modulation d'amplitude conventionnelle, il n'y a **pas** de conversion logarithmique $10^{D \cdot g(t)/20}$. Le modulateur reste bipolaire, pouvant prendre des valeurs négatives.

2.3 Effet de modulation bipolaire

La multiplication du bruit par le modulateur bipolaire produit :

$$y(t) = x(t) \cdot G(t) \quad (13)$$

Modulation en anneau. Contrairement à une multiplication par une constante, la modulation par un signal variant dans le temps crée un produit de convolution dans le domaine fréquentiel :

$$Y(f) = X(f) * G(f) \quad (14)$$

Il s'agit donc d'une modulation en anneau lente, générant des bandes latérales spectrales, et non d'une simple inversion de phase.

Quasi-silences. Lorsque $|G(t)| \approx 0$, l'amplitude est fortement atténuée. C'est le mécanisme principal créant les respirations dramatiques et les quasi-silences caractéristiques.

Interférences entre couches. Lorsque des couches indépendamment modulées sont sommées :

$$m(t) = \sum_i x_i(t) \cdot G_i(t) \quad (15)$$

des interférences constructives et destructrices partielles se produisent. Toutefois, le mécanisme acoustique dominant reste la variation d'amplitude $|G_i(t)|$ plutôt que les inversions de polarité elles-mêmes.

2.4 Effet acoustique résultant

Cette modulation bipolaire inhabituelle crée les effets caractéristiques de la tempête côtière :

- une **respiration dramatique** avec des montées et descentes marquées,
- des **quasi-silences** produits par les passages du modulateur près de zéro,
- des **effets de ressac** produits par les modulations lentes et indépendantes des différentes couches.

3 Synthèse Stéréophonique Mid/Side

3.1 Principe

Les couches de bruit façonnées indépendamment sont combinées en composantes Mid (M) et Side (S). Le canal Side est atténué globalement par un facteur $\gamma = 0,52$ pour préserver la compatibilité mono.

3.2 Reconstruction stéréo

Le décodage M/S produit les canaux gauche et droit :

$$L(t) = M(t) + S(t) \quad (16)$$

$$R(t) = M(t) - S(t) \quad (17)$$

Cette architecture crée un lit stéréo large tout en maintenant une somme mono cohérente :

$$L(t) + R(t) = 2M(t) \quad (18)$$

4 Synthèse Granulaire des Impacts

4.1 Génération des grains

Les impacts (gouttes ou éclaboussures) sont générés comme des grains courts avec des paramètres aléatoires :

- durée T_d ,
- fréquence fondamentale f_0 ,
- amplitude A ,
- catégorie (petit ou grand impact).

Les grands impacts utilisent des durées plus longues, des fréquences plus graves et des amplitudes plus élevées.

4.2 Enveloppe asymétrique

L'enveloppe combine trois composants :

$$e(t) = \exp\left(-\frac{t}{\tau}\right) \cdot \left(1 - \frac{t}{T_d}\right)^3 \cdot \mathcal{A}(t)^2 \quad (19)$$

où le facteur de fondu linéaire $\mathcal{A}(t)$ est défini par :

$$\mathcal{A}(t) = \begin{cases} t/t_{attack} & \text{si } t < t_{attack} \\ (T_d - t)/t_{attack} & \text{si } t > T_d - t_{attack} \\ 1 & \text{sinon} \end{cases} \quad (20)$$

et est élevé au carré dans l'expression de l'enveloppe, produisant un fondu quadratique.

4.3 Glissade de phase directe

Calcul de la phase. La phase de l'oscillateur est calculée directement à partir du paramètre de fréquence descendant :

$$\phi(t) = 2\pi \cdot f(t) \cdot t \quad (21)$$

avec :

$$f(t) = f_0 \left(1 - \beta \frac{t}{T_d} \right), \quad \beta = 0,25 \quad (22)$$

Fréquence instantanée réelle. Il est crucial de noter que $f(t)$ n'est **pas** la fréquence instantanée de l'oscillateur. La fréquence instantanée est définie comme :

$$f_i(t) = \frac{1}{2\pi} \frac{d\phi}{dt} = f(t) + t \cdot f'(t) \quad (23)$$

Avec $f'(t) = -\beta f_0/T_d$, on obtient :

$$f_i(t) = f_0 \left(1 - 2\beta \frac{t}{T_d} \right) \quad (24)$$

La fréquence instantanée réelle descend donc **deux fois plus loin** que le paramètre $f(t)$ ne le suggère : de f_0 à $f_0(1 - 2\beta) = 0,5f_0$ au lieu de $0,75f_0$.

4.4 Composition du grain

Le grain brut combine une composante sinusoïdale et du bruit :

$$s(t) = (0,85 \sin(\phi(t)) + 0,15 w(t)) \cdot e(t) \quad (25)$$

où $w(t) \sim \mathcal{N}(0, 1)$ est un bruit blanc gaussien.

5 Filtrage Résonant Passe-Bande

5.1 Filtre biquad à gain de crête constant

Chaque grain traverse un filtre RII du second ordre de type passe-bande à gain de crête constant (formulation RBJ, 0 dB au pic). Les coefficients sont dérivés de :

$$\omega_0 = \frac{2\pi f_c}{f_s} \quad (26)$$

$$\alpha = \frac{\sin(\omega_0)}{2Q} \quad (27)$$

Les coefficients normalisés sont :

$$b_0 = \frac{\alpha}{1 + \alpha}, \quad b_1 = 0, \quad b_2 = \frac{-\alpha}{1 + \alpha} \quad (28)$$

$$a_1 = \frac{-2 \cos(\omega_0)}{1 + \alpha}, \quad a_2 = \frac{1 - \alpha}{1 + \alpha} \quad (29)$$

L'équation aux différences est :

$$y[n] = b_0 x[n] + b_1 x[n - 1] + b_2 x[n - 2] - a_1 y[n - 1] - a_2 y[n - 2] \quad (30)$$

5.2 Facteur de qualité

Le facteur de qualité Q contrôle la largeur de bande et la résonance temporelle :

- $Q = 8$ pour les grands impacts : bande étroite, résonance prononcée,
- $Q = 5$ pour les petits impacts : bande plus large, résonance plus douce.

Indépendance vis-à-vis de l'amplitude. Chaque grain étant normalisé par son pic **après** filtrage, le facteur Q affecte principalement la bande passante et le *ringing* temporel, plutôt que l'amplitude finale.

6 Normalisation par Grain

Le grain filtré est noté :

$$s_{\text{filt}}[n] = \text{Biquad}\{s[n]\} \quad (31)$$

Chaque grain filtré est normalisé par son pic avant application de l'amplitude aléatoire finale :

$$s_{\text{norm}}[n] = \frac{s_{\text{filt}}[n]}{\max |s_{\text{filt}}[n]|} \quad (32)$$

Cette normalisation maintient le niveau des impacts synthétisés prévisible et réduit le risque de transitoires excessifs lors du mixage. Les grains superposés et le lit de bruit peuvent néanmoins produire des pics globaux plus élevés, traités par l'étape finale de mastering.

7 Panoramique à Puissance Égale

Chaque impact est positionné dans l'espace stéréo via un panoramique à puissance égale. Le gain final du grain est d'abord calculé :

$$g = A \cdot L_{\text{drop}} \quad (33)$$

où A est l'amplitude aléatoire de l'impact et L_{drop} le niveau global des impacts.

Les canaux gauche et droit sont ensuite :

$$L_{\text{grain}} = g \cdot \cos\left(\frac{p\pi}{2}\right) \cdot s_{\text{norm}} \quad (34)$$

$$R_{\text{grain}} = g \cdot \sin\left(\frac{p\pi}{2}\right) \cdot s_{\text{norm}} \quad (35)$$

où $p \in [0,18,0,82]$ est la position de panoramique.

Cette contrainte évite les positions extrêmes gauche/droite. La propriété de puissance égale garantit :

$$\cos^2\left(\frac{p\pi}{2}\right) + \sin^2\left(\frac{p\pi}{2}\right) = 1 \quad (36)$$

8 Bouclage Sans Couture

8.1 Périodicité du bruit façonné

Le bruit synthétisé par FFT est périodique sur la durée complète du tampon. Chaque composante de modulation complète un nombre entier de cycles sur cette même durée. Les deux se raccordent donc continûment à la frontière de boucle.

8.2 Enveloppement modulaire des impacts

Les impacts qui s'étendent au-delà de la fin du tampon sont explicitement enveloppés au début par indexation modulaire :

$$\text{idx} = (\text{start} + j) \mod N \quad (37)$$

Leurs queues continuent ainsi à travers la frontière de boucle plutôt que d'être tronquées.

9 Chaîne de Mastering

9.1 Suppression de la composante continue

La composante continue est supprimée à deux étapes :

1. après chaque synthèse de couche (dans le domaine temporel),
2. après mixage complet des canaux gauche et droit.

9.2 Normalisation RMS

Après mixage et suppression du DC, le signal stéréo complet est mis à l'échelle vers un niveau RMS cible :

$$\text{RMS} = \sqrt{\frac{1}{2N} \sum_{n=0}^{N-1} (L[n]^2 + R[n]^2)} \quad (38)$$

$$y[n] \leftarrow y[n] \cdot \frac{L_{\text{target}}}{\text{RMS}}, \quad L_{\text{target}} = 0,085 \quad (39)$$

9.3 Réduction conditionnelle de gain avec plafond de crête

Après normalisation RMS, le pic absolu maximal sur les deux canaux est mesuré :

$$P = \max_n \max(|L[n]|, |R[n]|) \quad (40)$$

Si $P > P_{\text{max}}$ avec $P_{\text{max}} = 0,92$, le tampon stéréo complet est réduit par un facteur de gain uniforme :

$$y[n] \leftarrow y[n] \cdot \frac{P_{\text{max}}}{P} \quad (41)$$

Distinction avec l'écrêtage et la limitation dynamique. Il ne s'agit pas d'un écrêtage (clipping) : aucune valeur n'est tronquée. Le signal est simplement mis à l'échelle globalement. Il ne s'agit pas non plus d'un limiteur dynamique : la réduction est uniforme sur tout le fichier. Comme cette étape intervient après la normalisation RMS, le signal final n'atteint le RMS cible que si le plafond n'est pas engagé.

9.4 Conversion PCM

Les échantillons flottants maîtrisés sont convertis en PCM signé 16 bits pour la sortie WAV :

$$s_{16}[n] = \text{clamp}(\text{round}(32767 \cdot y[n]), -32768, 32767) \quad (42)$$

10 Conclusion

La modélisation de cette chaîne DSP démontre comment l'interaction entre un bruit large bande façonné spectralement, une modulation bipolaire linéaire, et un filtrage résonant à fort facteur de qualité transforme une simple pluie procédurale en une simulation acoustique complexe de tempête maritime.

La clé de l'effet « tempête » réside en grande partie dans la modulation lente des couches spectrales. Le modulateur bipolaire agit comme une modulation en anneau lente ; ses passages près de zéro produisent de fortes atténuations, tandis que les modulations indépendantes des différentes couches créent une évolution complexe du spectre et de l'amplitude.

L'absence de conversion en décibels s'est révélée être l'ingrédient essentiel de l'identité sonore du générateur.