HAL open science

Contrôle intuitif d'un synthétiseur d'environnements sonores spatialisés

Charles Verron, Mitsuko Aramaki, Richard Kronland-Martinet, Grégory Pallone

To cite this version:

Charles Verron, Mitsuko Aramaki, Richard Kronland-Martinet, Grégory Pallone. Contrôle intuitif d'un synthétiseur d'environnements sonores spatialisés. 10ème Congrès Français d'Acoustique, Apr 2010, Lyon, France. hal-00546833

HAL Id: hal-00546833

https://hal.science/hal-00546833

Submitted on 14 Dec 2010

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.

10ème **C**ongrès **F**rançais d'**A**coustique

Lyon, 12-16 Avril 2010

Contrôle Intuitif d'un Synthétiseur d'Environnements Sonores Spatialisés

Charles Verron[1], Mitsuko Aramaki[2], Richard Kronland-Martinet[1], Grégory Pallone[3]

[1] CNRS-LMA, 31 Chemin Joseph Aiguier, F-13402 Marseille cedex, {verron,kronland}@lma.cnrs-mrs.fr

[2] CNRS-INCM, 31 Chemin Joseph Aiguier, F-13402 Marseille cedex, aramaki@incm.cnrs-mrs.fr

[3] Orange Labs, Avenue Pierre Marzin, F-22307 Lannion cedex, gregory.pallone@orange-ftgroup.com

La création d'ambiances sonores 3D interactives constitue un enjeu majeur dans le domaine de la Réalité Virtuelle, en particulier, pour des applications nécessitant un rendu sonore temps-réel (e.g., bruitage, jeux vidéo, infrastructure CAVE...). Dans ce contexte, un synthétiseur spatialisé d'environnements sonores 3D compatible avec les formats audio existants (multicanal, Ambisonics, binaural) a été mis en œuvre. Une stratégie de contrôle intuitif de ce synthétiseur, basée sur une organisation hiérarchique des sons, est proposée. Ainsi, les scènes auditives sont construites en combinant des sources élémentaires catégorisées en Solides (impact...), Liquides (pluie, vague...) et Aérodynamiques (vent, feu...) et chaque source est modélisée comme une combinaison de cinq "atomes" signifiants du point de vue signal. Contrairement aux procédés communément utilisés, ce synthétiseur présente l'avantage de combiner modules de synthèse et de spatialisation au même niveau de la génération du son, offrant ainsi de nouvelles possibilités de contrôle simultané du timbre et de la position spatiale des sources via des descripteurs haut-niveau (e.g., taille, intensité, largeur de source...). Grâce à son faible coût de calcul, cette architecture permet de créer en temps réel des scènes complexes 3D comportant un grand nombre de sources.

1 Introduction

Aujourd'hui, les techniques utilisées pour le rendu d'environnements sonores sont basées principalement sur l'utilisation de sons pré-enregistrés (tables d'onde). De larges banques de sons sont collectées pour représenter l'ensemble des évènements sonores du monde virtuel. Cette méthode de sonification est actuellement largement utilisée mais présente des limitations. Tout d'abord, il est nécessaire d'enregistrer les sons et de les stocker, ce qui peut représenter de grosses quantités de mémoire à allouer et des temps de chargement longs. Mais surtout, l'interactivité est par nature limitée avec des sources pré-enregistrées.

Nous explorons ici une approche différente de celles basées sur les sons enregistrés pour sonifier un monde virtuel. Nous nous intéressons à la synthèse sonore temps-réel, qui consiste à calculer dynamiquement le son à générer, via des algorithmes informatiques dédiés. Nous souhaitons simuler par synthèse les événements de la vie de tous les jours, autres que la musique et la parole, regroupés sous le terme de sons d'environnement. L'avantage de cette approche réside dans les possibilités de contrôle accrues par rapport au simple déclenchement de tables d'onde.

Nous avons réalisé un synthétiseur de sons d'environnement intégrant en son coeur des modules de spatialisation pour une restitution sonore multicanale [1]. La combinaison de la synthèse et de la spatialisation nous a permis d'utiliser des propriétés spécifiques du modèle additif [2] pour positionner les sources de manière efficace et simuler leur extension spatiale dans la scène. Ce synthétiseur permet la création de sources sonores diverses, regroupées dans les trois catégories de sons d'environnement proposées par Gaver [3] : les sons de solides en vibration, les sons aérodynamiques et les sons de liquides. Nous avons également proposé une stratégie pour le contrôle du synthétiseur, basé sur une décomposition atomique des sons d'environnement [4].

Le présent papier vient compléter ces travaux. Il est articulé en deux parties. Tout d'abord nous décrivons la décomposition atomique de sources représentatives des trois catégories de sons d'environnement. Ensuite nous définissons un ensemble de contrôles intuitifs pour manipuler l'extension spatiale et le timbre de ces sources. Ce travail a été intégré au synthétiseur temps-réel qui permet désormais de générer des sources sonores diverses (pluie, vent, feu, vagues, impacts...), de les contrôler via des paramètres haut-niveau, et de les manipuler dans un espace virtuel immersif.

2 Décomposition atomique de sons d'environnement

Les études qui ont été proposées pour la synthèse des sons d'environnement ont permis de dégager des modèles de synthèse, basés aussi bien sur la physique des sources que sur l'analyse de signaux enregistrés [5, 6, 7, 8]. Dans [1] nous avons présenté une approche générale basée sur le modèle additif [2] pour reproduire des sources environnementales diverses par synthèse additive. Le moteur de synthèse proposé permet de générer ces sources sous la forme de sinusoïdes et de bruit filtré variant au cours du temps. La manipulation directe des paramètres bas-niveau de la synthèse additive (i.e., fréquences et amplitudes des sinusoïdes, enveloppe spectrale du bruit) n'est pas toujours simple.

Pour cette raison, nous avons défini cinq atomes paramétriques, qui se veulent plus intuitifs que les paramètres bas-niveau de la synthèse additive et offrent un premier niveau de contrôle au synthétiseur [4]. Le paramétrage des atomes et leur agencement temporel pour reproduire une source donnée est réalisé à partir de modèles présentés dans la littérature, et de manière empirique, en analysant des spectrogrammes de sources naturelles et en évaluant le résultat perceptif obtenu par resynthèse. Nous présentons ici comment combiner les cinq atomes pour former différentes sources environnementales : des sons d'impact, du vent, des "whooshs" (i.e., souffles d'air produits par un object fendant l'air), du feu, de la pluie et des vagues.

Impacts Utilisant le modèle décrit dans [9], les sons d'impact sont réalisés à partir de deux atomes :

- L'**impact modal**, qui est une somme de M sinusoïdes décroissant exponentiellement d'amplitude a_m, de fréquences f_m, de phases Φ_m et de coefficient d'amortissement α_m :

$$x(t) = \sum_{m=1}^{M} a_m \cos\big(2\pi f_m t + \Phi_m\big) e^{-\alpha_m t}$$

- L'**impact bruité**, qui est une somme de B sous-bandes de bruit $s_b(t)$ décroissant exponentiellement, d'amplitudes a_b et de coefficient d'amortissement α_b :

$$x(t) = \sum_{b=1}^{B} a_b s_b(t) e^{-\alpha_b t}$$

Vent et whooshs Partant du modèle de synthèse décrit dans [8], les sons de vent et de whooshs sont réalisés par superposition d'atomes **bruit bande-étroite**, qui sont des bruits filtrés par un filtre variant au cours du temps, dont l'enveloppe spectrale au temps t est donnée par :

$$X(f) = \begin{cases} a(t) & \text{si } |f - f_0(t)| < \frac{B(t)}{2} \\ a(t) e^{-\alpha(t)\left(|f - f_0(t)| - \frac{B(t)}{2}\right)} & \text{si } |f - f_0(t)| > \frac{B(t)}{2} \end{cases}$$

avec $f_0(t)$ la fréquence centrale, $a(t)$ le gain du filtre, $B(t)$ la bande passante et $\alpha(t)$ la pente du filtre.

Feu Nous synthétisons le feu comme une somme de crépitements, de sifflements et de bruits de flammes, conformément au modèle décrit dans [8]. La synthèse nécessite la combinaison de trois atomes :

- L'**impact bruité** pour les crépitements.
- Le **bruit large-bande** pour le sons des flammes (dans les basses fréquences) et pour les sifflements (dans les hautes fréquences). Il s'agit d'un bruit dont l'enveloppe spectrale est définie par morceau sur 32 sous-bandes équi-réparties sur une échelle ERB (Equivalent Rectangular Bandwidth). Chaque sous-bande $s_b(t)$ a une amplitude variable dans le temps $a_b(t)$:

$$x(t) = \sum_{b=1}^{32} a_b(t) s_b(t)$$

- Le **bruit bande-étroite** pour les sifflements résonants.

Pluie Nous synthétisons un son de pluie comme une somme de gouttes qui tombent dans de l'eau, ou sur un matériau rigide, et d'un bruit d'ambiance qui simule un très grand nombre de gouttes. Utilisant les modèles décrits dans [7, 8], la synthèse est réalisée à partir de trois atomes :

- L'**impact bruité** pour les gouttes tombant sur des feuilles ou une surface rigide.
- L'**impact chirpé** pour les gouttes tombant dans un liquide. Il s'agit d'une sinusoïde d'amplitude initiale a, de phase à l'origine ϕ et de fréquence $f(t)$ variant dans le temps. L'amplitude suit une décroissance exponentielle de coefficient d'amortissement α :

$$x(t) = a \cos\Big(2\pi \int_0^t f(\nu) d\nu + \phi\Big) e^{-\alpha t}$$

- Le **bruit large-bande** pour le bruit diffus d'ambiance représentant un très grand nombre de gouttes.

Vagues Les sons de vagues sont réalisés à partir de l'atome **bruit large-bande**, les amplitudes dans les 32 sous-bandes étant contrôlées avec la méthode décrite dans [10]. La vague est découpée en cinq étapes successives : la crête de la vague se détache (étape 1), elle entre en collision avec la surface de l'eau se trouvant dessous (étape 2), la surface se stabilise (étape 3), l'eau s'avance sur la plage (étape 4) et se retire (étape 5) avant le déferlement d'une nouvelle vague. Les paramètres de synthèse utilisés sont [10] :

- la durée des cinq phases de la vague ;
- les amplitudes de chaque phase ;
- la brillance de chaque phase ;
- une couleur spectrale fixe définie sur 32 sous-bandes ERB.

3 Contrôles haut-niveau

Dans cette partie nous proposons un ensemble de contrôles haut-niveau pour simuler l'extension spatiale et manipuler le timbre des sources environnementales décrites précédemment. Ces contrôles sont liés à la nature physique des sources et constituent des exemples de mapping pour gérer l'ensemble des atomes du synthétiseur de manière intuitive.

3.1 Extension spatiale

Pour obtenir une source élargie, nous avons proposé de dupliquer la source originale en huit points de l'espace, chaque source ponctuelle étant une version décorrélée obtenue à partir des mêmes paramètres de synthèse [1]. Un test perceptif a permis d'évaluer les possibilités offertes par cette approche. La méthode donne de bons résultats, principalement pour les sons bruités et relativement stationnaires, tels que le vent ou les vagues.

Dans le cadre de notre décomposition atomique des sons d'environnement, les atomes impulsifs correspondent le plus souvent à des sons brefs bien localisés spatialement (par exemple un crépitement de feu). La méthode d'extension spatiale par décorrélation et duplication fonctionne pour ce type d'éléments. Elle peut

être utilisée pour transformer un crépitement isolé en une source large et diffuse. Cependant cela ne correspond pas toujours à une situation naturelle. On s'attend plutôt, dans le cas du feu, à percevoir un ensemble de crépitements sous la forme de sources ponctuelles localisées. Il semble alors plus judicieux de créer l'effet d'élargissement en répartissant les atomes impulsifs autour de l'auditeur.

Nous avons choisi de combiner ces deux approches (duplication et répartition des atomes) pour élargir les sources environnementales du synthétiseur. Pour la pluie, les gouttes d'eau (impacts bruités et chirpés) sont spatialisées indépendamment autour de l'auditeur, et le bruit de fond (bruit large-bande) est dupliqué dans les sources secondaires. Pour le feu, les crépitements (impacts bruités) sont repartis spatialement ; le bruit des flammes (bruit large-bande) et les sifflements (bruits large-bande et bande-étroite) sont dupliqués. Le vent et les whooshs sont constitués uniquement de bruits bande-étroite, qui sont dupliqués dans les sources secondaires. Les vagues, constituées uniquement de bruit large-bande, sont également dupliquées pour être élargies. Enfin, les sons d'impact peuvent être élargis individuellement par duplication des composantes modales, si celles-ci sont assez nombreuses, ou répartis autour de l'auditeur dans le cas d'impacts multiples (pour simuler une "pluie d'impacts" étalée spatialement).

Pour gagner en efficacité, les atomes impulsifs ne sont pas tous spatialisés indépendamment, mais regroupés dans huit sources secondaires. Cette méthode s'apparente au clustering spatial de sources proposé par Tsingos et al. dans [11]. Ainsi quel que soit le nombre d'impacts à synthétiser (il peut s'agir de plusieurs centaines dans le cas de gouttes de pluie ou de crépitements de feu) le nombre de sources à spatialiser est toujours limité à huit.

Pour conclure, l'effet d'extension spatiale des sources sonores intégré au synthétiseur combine deux approches : la duplication et la répartition des atomes. Les atomes bruités continus sont décorrélés et dupliqués, alors que les atomes impulsifs sont repartis dans huit sources secondaires autour de l'auditeur. Des tests d'écoute informels ont montré que cette approche combinée est pertinente pour élargir de manière réaliste des sources environnementales contenant du bruit stationnaire et des impulsions, comme du feu ou de la pluie. L'extension spatiale leur confère une dimension plus immersive qu'une simple source ponctuelle.

3.2 Sons de solides

Nous considérons ici les sons de solides en vibration, et plus spécifiquement les sons d'impact. Des contrôles haut-niveau ont été proposés dans [9]. L'utilisateur peut orienter la synthèse en fonction de propriétés de la source qu'il souhaite simuler, i.e., le matériau constituant l'objet, sa forme et sa taille, la force d'excitation, la dureté du maillet et la position de l'impact sur l'objet. Nous reprenons ici certains de ces contrôles haut-niveau, qui ont été implémentés dans le synthétiseur.

Force La force de l'impact est liée à la quantité d'énergie qui est communiquée au système résonant. Ce contrôle agit d'abord sur le volume du son synthétique. De plus, Aramaki et al. proposent de lier la force avec la brillance et le temps d'attaque du son de synthèse [9]. Un impact plus fort correspond à un son au contenu spectral plus riche en fréquences aiguës et au temps d'attaque plus court.

Matériau Notre perception du matériau d'un objet frappé est principalement lié à l'amortissement et au contenu spectral du son d'impact. Le contrôle du matériau proposé dans [9] exploite ces indices perceptifs. L'analyse de sons d'impact réels à permis d'extraire des profils d'amortissement et des profils spectraux typiques de différents matériaux, comme le bois, le verre et le métal. Aramaki et al. proposent ensuite de naviguer entre les matériaux en appliquant des techniques d'interpolation sur les paramètres de synthèse pour passer d'un profil à l'autre (morphing).

Forme de l'objet Dans [9] les auteurs proposent de relier la forme de l'objet à l'inharmonicité du son de synthèse. Nous avons repris cette méthode dans le synthétiseur. Trois prototypes de formes sont disponibles pour un objet, conduisant à des profils d'harmonicité différents : 1D (forme linéaire), 2D (membrane ou plaque) et 3D (forme complexe). A titre d'exemple, la simulation d'une forme linéaire conduit à un son harmonique alors qu'une plaque résonante produira un son inharmonique.

Point d'impact Lorsqu'un objet est excité à différents endroits, le son résultant connaît des variations de timbre. L'objet gardant toujours la même forme, ses fréquences propres sont conservées. Cependant, l'énergie donnée aux modes de résonance varie avec le point d'impact, celui-ci pouvant tomber sur un ventre ou sur un noeud pour un mode donné. Dans [9] les auteurs proposent de simuler ces variations de timbre en appliquant un filtrage sur le son généré, les creux et bosses du filtre variant selon le point d'excitation sur l'objet.

3.3 Sons aérodynamiques

3.3.1 Vent

Comme nous l'avons vu dans la première partie, le vent est constitué d'atomes bruit bande-étroite, dont les paramètres sont la largeur de bande, la fréquence centrale, la pente de l'enveloppe spectrale et l'enveloppe temporelle (gain variable au cours du temps). Des tests d'écoute informels ont montré que la superposition de trois atomes bruit bande-étroite donne de bons résultats perceptifs. Ici nous présentons deux contrôles haut-niveau, qui ont été intégrés au synthétiseur : la force et la "froidure" du vent.

Force Ce contrôle est basé sur les travaux de Farnell [8] qui propose de modifier conjointement l'amplitude et la fréquence de résonance des bruits filtrés en fonction d'un paramètre qui simule la vitesse du vent.

Cette vitesse est astucieusement générée par un oscillateur basse-fréquence auquel est ajouté un bruit passe-bas pour simuler des rafales et autres bourrasques. Soit a_0 la vitesse moyenne du vent, a_1 l'amplitude et f_0 la fréquence de la sinusoïde, et $b(t)$ les perturbations aléatoires qui simulent des rafales soudaines, la vitesse du vent est donnée par :

$$v(t) = a_0 + a_1 sin(2\pi f_0 t) + b(t)$$

La composante $b(t)$ est obtenue par filtrage passe-bas d'un bruit blanc, avec une fréquence de coupure très basse (autour de 1Hz). Le paramètre de vitesse du vent agit sur l'amplitude et la fréquence centrale des filtres passe-bande dans lesquels sont injectés du bruit blanc. Les fréquences centrales varient proportionnellement à la vitesse :

$$f_c(t) = k_0 v(t - \tau) + k_1$$

où k_0 et k_1 sont des constantes. L'introduction d'un retard τ différent pour chaque filtre permet de créer des variations entre les composantes bruitées qui augmentent le naturel du son. Dans le synthétiseur, les retards des trois atomes bruit bande-étroite formant le vent sont fixés à $\tau = 0$, $\tau = 1.74$ et $\tau = 0.14$ secondes respectivement.

Le contrôle haut niveau "force du vent" couple la vitesse moyenne a_0, l'amplitude de la sinusoïde a_1 et l'amplitude des perturbations $b(t)$. Soit F le contrôle de la force du vent, variant de 0 (pas de vent) à 1 (vent fort), le paramètre de vitesse du vent s'exprime alors :

$$v(t) = F\left(1 + \frac{1}{4} sin(2\pi f_0 t) + b(t)\right)$$

La vitesse agit ensuite sur la fréquence de coupure et l'enveloppe temporelle (gain) des atomes bruit bande-étroite, comme décrit précédemment.

Froidure Si l'enveloppe temporelle et la fréquence de coupure sont définies par la force du vent, il est nécessaire de spécifier les largeurs de bande, les pentes et les gains relatifs des trois atomes de bruit bande-étroite utilisés. Nous avons introduit un contrôle appelé "froidure du vent" qui agit sur ces paramètres. Le contrôle de froidure intervient également sur les valeurs de k_0 et k_1 qui régissent la relation entre vitesse du vent et fréquences centrales des atomes. Nous avons défini deux jeux de valeurs ("presets") pour ces cinq paramètres, correspondant à deux types de vent très différents. Le premier correspond à un vent résonnant, évoquant un vent glacial sur la banquise. Le second est un vent très bruité, qui évoque une tempête dans des arbres. Le contrôle de froidure réalise simplement une interpolation linéaire des paramètres pour passer progressivement du vent glacial à la tempête dans les arbres. Les valeurs intermédiaires de froidure permettent d'obtenir différents types de vent.

3.3.2 Whooshs

Tout comme le vent, les whooshs sont simulés par une somme de trois atomes bruit bande-étroite. Les amplitudes des atomes sont fixées à 0.1, 0.5 et 1, leurs largeurs de bande à 43Hz et leurs pentes à $\frac{100}{f_s}$, $\frac{1100}{f_s}$ et $\frac{500}{f_s}$ respectivement, f_s étant la fréquence d'échantillonnage. Ces valeurs donnent de bons résultats perceptifs pour simuler des souffles d'air correspondant à un objet (petit ou grand) fendant l'air. Les contrôles haut-niveau proposés sont la durée, la taille de l'objet, et l'étendue de son mouvement.

Durée Le contrôle de durée permet à l'utilisateur de spécifier les temps d'attaque t_1 et de relâchement t_2 du son. L'analyse temps-fréquence d'enregistrements de whooshs nous a conduit à définir une enveloppe d'amplitude $v(t)$ sous la forme de deux exponentielles, l'une croissante pour la phase d'attaque, l'autre décroissante pour la phase de relâchement :

$$v(t) = \begin{cases} e^{\alpha_1(t-t_1)} & \text{pour } 0 < t \leq t_1 \\ e^{-\alpha_2(t-t_1)} & \text{pour } t_1 < t \leq t_2 \end{cases}$$

Les coefficients d'amortissement α_1 et α_2 sont calculés pour former l'enveloppe de manière à ce qu'elle soit exponentiellement croissante de 10^{-4} à 1 en t_1 secondes, puis exponentiellement décroissante de 1 à 10^{-4} en t_2 secondes. Cette enveloppe agit sur le gain et la fréquence centrale des atomes de bruit bande-étroite constituant le whoosh. Ce procédé est similaire à celui présenté pour le paramètre de vitesse du vent.

Taille de l'objet Le paramètre $v(t)$ défini dans le paragraphe précédent contrôle également la fréquence de coupure des atomes du whoosh, via une relation linéaire :

$$f_c(t) = k_0 v(t) + 20 \tag{1}$$

Le contrôle "taille de l'objet" (noté $Taille$) spécifie la valeur de k_0 pour les trois atomes du whoosh :

$$k_0 = cst.2^{\frac{69-Taille}{12}} \tag{2}$$

où cst est égale à 900 pour le premier atome du whoosh, 1200 pour le deuxième et 1500 pour le troisième. $Taille$ varie entre 0 et 127 par souci de compatibilté avec la norme MIDI [12]. La relation spécifiée par les équations 1 et 2 est satisfaisante perceptivement : la brillance du son diminue lorsque la taille de l'objet augmente. Une étape de calibration serait néanmoins nécessaire pour déterminer plus finement la relation entre la taille de l'objet exprimée en centimètres par exemple, et la valeur de k_0 correspondante.

Mouvement Un son de whoosh étant provoqué par un objet en mouvement, il est naturel de faire varier la position de la source virtuelle autour de l'auditeur durant le son. C'est pourquoi nous proposons un contrôle appelé "mouvement" qui agit sur la position du whoosh au cours du temps, et assure un son rotatif autour de l'auditeur en augmentant linéairement l'azimut de la source virtuelle :

$$\theta(t) = \theta(t_0) + M\frac{t}{t_2}$$

M étant la valeur du contrôle de mouvement, qui peut varier de 0 (pour un son fixe) à 720 (pour deux tours complets).

3.3.3 Feu

Pour le feu, nous proposons dans le synthétiseur trois contrôles de haut-niveau, qui permettent d'obtenir différents types de sons allant d'un bruit de chaudière à un feu de forêt, en passant par un feu de cheminée. Ces contrôles sont la densité des crépitements, la taille des flammes et l'intensité globale.

Densité de crépitements Les impacts bruités simulant les crépitements sont déclenchés aléatoirement en suivant une loi de Poisson. A chaque trame, il y a une chance sur λ qu'un impact soit déclenché (λ étant le nombre moyen de crépitements qui tombent par unité de temps). Le gain et la densité λ des crépitements sont modifiables interactivement.

Taille des flammes La taille des flammes est simulée au travers du gain des quatre premières bandes ERB de l'atome bruit large-bande présent dans le feu. Le contrôle taille des flammes amplifie ou diminue l'énergie du son dans ces bandes de fréquence. Pour apporter de la "vie" au bruit des flammes, celui-ci est également modulé par un bruit très basse fréquence (fréquence de coupure autour de 1Hz).

Intensité Le paramètre intensité du feu est un contrôle global qui agit simultanément sur la densité et le gain des crépitements, ainsi que sur la taille des flammes et le gain des sifflements pour transformer un petit feu de cheminée en un brasier ardent. Le bruit basse fréquence modulant la taille des flammes est également relié linéairement au contrôle intensité du feu, pour générer des fluctuations d'amplitude plus grandes lorsque l'intensité augmente.

La figure 1 montre une représentation graphique du feu et illustre l'intérêt des contrôles haut-niveau dans un cadre audio-visuel : une forte interaction image/son peut être obtenue en contrôlant les deux modalités par les mêmes paramètres haut-niveau.

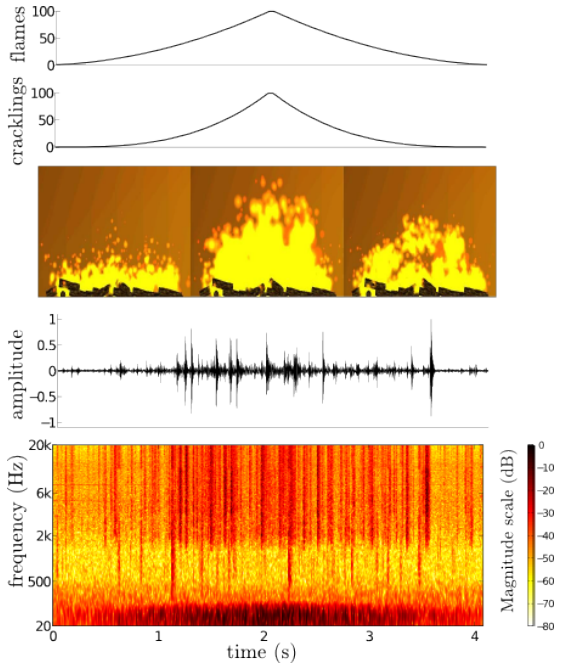

FIGURE 1 – *Exemples de contrôles haut-niveau du feu agissant simultanément sur le rendu visuel et sonore. L'utilisateur peut contrôler la taille des flammes et la densité des crépitements pour simuler différentes intensités de feu (image+son).*

3.4 Sons de liquides

3.4.1 Pluie

Les contrôles haut-niveau proposés pour le son de pluie sont la densité de gouttes, le gain du bruit d'ambiance et l'intensité globale. La figure 2 illustre ces possibilités de contrôle : on y voit la densité des gouttes et le volume du bruit d'ambiance augmenter au cours du temps pour simuler une pluie de plus en plus forte.

Densité de gouttes Tout comme pour les crépitements du feu, les gouttes (qu'il s'agisse de gouttes sur une surface rigide ou liquide) sont déclenchées aléatoirement en suivant une loi de Poisson. Il est possible pour l'utilisateur de spécifier la densité temporelle et le gain des gouttes qui tombent dans l'eau, de même pour les gouttes qui tombent sur une surface rigide.

Bruit d'ambiance Le bruit d'ambiance de la pluie, réalisé via un atome bruit large-bande, peut être contrôlé indépendamment des gouttes. Il est possible de spécifier son gain et sa couleur. Un preset de couleur est fourni à l'utilisateur. Il provient d'une analyse d'un son de pluie réelle, via un vocodeur à 32 canaux sur une échelle ERB, comme décrit dans [13].

Intensité Le contrôle de l'intensité de la pluie agit simultanément sur l'ensemble des paramètres précédents : le gain des gouttes sur les feuilles, le gain des gouttes dans l'eau, leurs densités, le gain du bruit d'ambiance. Ce contôle global permet de simuler une transition continue entre une pluie fine et une pluie forte. En pratique le mapping consiste en une relation linéaire ou exponentielle entre la valeur de l'intensité de la pluie et les paramètres de gain et de densité des gouttes. Cette stratégie procure une bonne adéquation entre le contrôle et le rendu sonore, mais une calibration plus ajustée doit encore être réalisée.

3.4.2 Vagues

Les contrôles haut-niveau que nous proposons sont la taille des vagues et la régularité de l'enchaînement temporel entre deux vagues successives.

Taille Nous avons spécifié les paramètres de synthèse (i.e., enveloppe temporelle, brillance, amplitude et couleur) pour deux vagues de référence associées à des tailles différentes : petite et grande. Pour synthétiser une vague de taille intermédiaire, nous réalisons une interpolation linéaire des paramètres de synthèse.

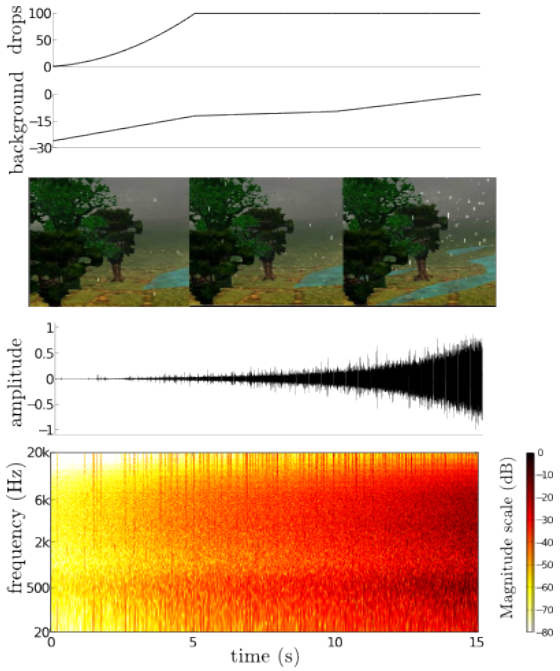

FIGURE 2 – *Exemples de contrôles de haut-niveau pour la pluie. La densité de gouttes et le volume du bruit d'ambiance peuvent être modifiés interactivement pour passer de façon continue de quelques gouttes à une grosse averse.*

Régularité Ce contrôle ajoute une valeur aléatoire aux durées des cinq phases du déferlement de la vague. Toutes les vagues déferlent de la même manière si la régularité est à 1. Sinon, les durées des cinq phases sont différentes d'une vague à l'autre. Cela permet d'éviter les répétitions et d'augmenter le réalisme de la scène de mer.

4 Conclusion

Nous avons d'abord présenté une décomposition atomique adaptée à plusieurs types de sons d'environnement : des impacts, du vent, des whoosh, du feu, des gouttes, de la pluie et des vagues. Pour déterminer les combinaisons d'atomes appropriées, nous nous sommes inspirés de modèles existants (qu'ils soient basés sur la physique ou le signal) et avons examiné avec attention des représentations temps-fréquence d'enregistrements. A l'avenir, cette démarche devrait permettre de définir de nouveaux atomes, si les cinq proposés ne suffisent pas pour reproduire un son particulier, et de déterminer leur paramétrage pour reproduire d'autres sources environnementales.

Nous avons ensuite proposé un ensemble de contrôles haut-niveau pour gérer les atomes du synthétiseur de manière intuitive. L'utilisateur peut modifier interactivement la position et l'extension spatiale des sources sonores. Il dispose également d'un ensemble de moyens d'agir sur les propriétés physiques des sources, par exemple, l'intensité de la pluie, la densité de crépitements d'un feu ou la taille d'une vague. Les contrôles présentés ont été validés par des tests d'écoute informels. Une calibration plus rigoureuse serait nécessaire. Des exemples sonores sont disponibles en ligne [14].

Pour agir sur la scène sonore toute entière, nous envisageons de définir des contrôles manipulant plusieurs sources simultanément, par exemple les vagues, le vent et la pluie pour réaliser une scène de tempête de mer interactive. Toutes ces sources doivent être reliées d'une manière ou d'une autre dans le synthétiseur, pour assurer la cohérence de la scène globale. L'intégration de ce type de fonctionnalités est actuellement en cours d'investigation.

Références

[1] Verron C., Aramaki M., Kronland-Martinet R., Pallone G. "A 3D Immersive Synthesizer for Environmental Sounds". *accepted to IEEE Transactions on Audio, Speech, and Language Processing* .

[2] Serra X., Smith J.O. "Spectral modeling synthesis : A sound analysis/synthesis system based on a deterministic plus stochastic decomposition". *Computer Music Journal* 14 (4), (1990) 12–24.

[3] Gaver W.W. "What in the World Do We Hear ? An Ecological Approach to Auditory Event Perception". *Ecological Psychology* 5(1), (1993) 1–29.

[4] Verron C., Aramaki M., Kronland-Martinet R., Pallone G. "Controlling a Spatialized Environmental Sound Synthesizer". *Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (2009).

[5] Cook P.R. *Real Sound Synthesis for Interactive Applications*. A. K Peters Ltd. (2002).

[6] Dobashi Y., Yamamoto T., Nishita T. "Real-time rendering of aerodynamic sound using sound textures based on computational fluid dynamics". *ACM Transactions on Graphics (Proc. SIGGRAPH 2003)* 22 (3), (2003) 732–740.

[7] van den Doel K. "Physically-based Models for Liquid Sounds". *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display* (2004).

[8] Farnell A. *Designing sound, procedural audio for games and film*. Applied Scientific Press (2008).

[9] Aramaki M., Gondre C., Kronland-Martinet R., Voinier T., Ystad S. "Thinking The Sounds : an Intuitive Control of an Impact Sound Synthesizer". *Proceedings of ICAD 09 - 15th International Conference on Auditory Display* (2009).

[10] Marque E. "Synthèse de sons d'environnement pour la réalité virtuelle". Master's thesis, Master jeux et médias interactifs numériques (2009).

[11] N. Tsingos E.G., Drettakis G. "Perceptual audio rendering of complex virtual environments". *ACM Transactions on Graphics* 23(3), (2004) 249–258.

[12] *http ://www.midi.org/*.

[13] Verron C., Aramaki M., Kronland-Martinet R., Pallone G. "Analysis/Synthesis and Spatialization of Noisy Environmental Sounds". *Proceedings of ICAD 09 - 15th International Conference on Auditory Display* (2009).

[14] *www.lma.cnrs-mrs.fr/∼kronland/CFA/index.html*.