

Introduction

Dans ce mémoire, nous allons essayer de répondre à la question :

Comment pourrions-nous générer un environnement réaliste en fonction d'un flux audio ?

Aujourd'hui, notamment dans le domaine du jeu vidéo, il existe deux manières principales d'exploiter un flux audio : la génération d'un environnement abstrait et les jeux de rythme. Nos recherches ont été dirigées vers la génération d'un environnement avec pour mot-clé "réaliste". Ainsi nous posons les contraintes d'un environnement réaliste qui implique entre autres plusieurs complications telles que : la génération d'un univers cohérent ou bien la compréhension émotionnelle d'une mélodie pour un utilisateur... Nous porterons tout au long de ce mémoire des solutions qui aboutiront à résoudre ces contraintes.

Le sujet impose une certaine connaissance de matières externes à l'informatique afin de pouvoir comprendre au mieux les tenants et aboutissants. Pour ce faire, nous proposerons une introduction générale en matière d'analyse audio, de créations procédurales et un aperçu de la médecine neurologique, ainsi qu'un aperçu de l'existant. Cette première partie nous permettra de poser les bases de réflexions sur lesquelles s'appuyer pour construire notre théorie. Dans une seconde partie, nous pourrons alors développer les idées, recherches, expérimentations, ... afin d'expliquer de manière plus approfondie les solutions que nous proposons en matière d'analyse musicale, de générations procédurales et de calibrage utilisateur. Nous suggérerons ensuite des pistes d'implémentation et des sources d'optimisation avant de conclure.

Analyse de l'existant

Design sonore

Qu'est-ce que le Design sonore ?

" Le design sonore ou la conception sonore est l'art d'utiliser des éléments sonores afin d'obtenir un effet désiré."

- Wikipédia

Nous croisons son résultat tous les jours : lorsque nous regardons un film, écoutons une musique, allons au théâtre et, bien entendu, dans les jeux vidéo. Le travail du designer sonore consiste à arranger la musique, à la créer, à la nettoyer, mais aussi à la faire coller à un environnement donné. Par exemple, lors d'une scène triste dans un film, vous ne voudriez pas entendre une musique d'action derrière. Le designer sonore doit s'adapter au ressenti émotionnel dégagé par un lieu, un événement, un moment, ... pour y mettre au mieux en relation ces sonorités et ainsi pouvoir y dégager l'émotion qu'il souhaite transmettre.

Bien entendu chaque être humain perçoit les choses différemment, et pourtant le designer sonore doit s'appuyer sur des codes de conventions communes pour sa création.

Ces codes, quels sont-ils ?

Le rythme de la musique doit être adapté au rythme de l'action. Prenons une course poursuite par exemple : le nombre d'actions à la seconde est très important c'est pourquoi il faudra adapter la pulsation de la chanson en fonction de cette fréquence.

La « teinte » de la musique doit coller avec l'environnement visuel. L'effet sonore dans une église ne sera pas le même que celui dans un théâtre, une arène ou au grand air, nous y entendrons plutôt des bruits avec échos.

Le choix des instruments est également important, les timbres de ces derniers viendront imprégner des informations liées à la culture. De plus, dans un contexte

historique, par exemple au moyen âge, il n'est pas naturel d'y entendre des instruments récents qui provoqueront ainsi un anachronisme et donc potentiellement aller à l'encontre de l'ambiance souhaitée.

Il existe d'autres codes mais ceci nous permet de définir une base commune de réflexion. Il existe une sorte de consentement implicite mutuel entre les individus pour ces règles. Nous pouvons nous poser la question de savoir d'où viennent ces règles ?

La première réponse serait de dire que nous avons été éduqués comme cela. Effectivement, ces règles existent depuis assez longtemps pour que nous ayons toujours vécu avec sans pour autant que personne ne nous les ai énoncées. La plupart des films, pièces de théâtre, jeux et musiques les respectent et donc tout au long de notre éducation musicale, nous les avons assimilées implicitement. Nous « baignons dedans » et même si, à première vue, un individu dit ne pas faire attention aux musiques de film, il les entend quand même et son cerveau les assimile automatiquement, tout comme plein d'autres détails.

L'exemple le plus parlant nous vient des enfants en bas âges (3-5ans). Au départ, les enfants ne distinguent pas les différences musicales, ils ne feront pas la distinction entre une musique rapide ou lente. Leur éducation musicale en étant qu'à son balbutiement, ils ne seraient pas gênés par une musique dansante pendant une scène triste dans un film. Cependant, même seuls, ils vont apprendre sans que l'on ait besoin de leur donner les codes. Cela donne un lien vers la deuxième réponse : la cohérence vis-à-vis du monde.

Dans la vie de tous les jours, nous entendons des milliers de bruits. Nos oreilles captent une « musique permanente ». Que cela soit le bruit du vent et des oiseaux chantant (auxquels nous ne prêtons pas forcément attention), ou les bruits de la ville. Nous sommes constamment entouré de sons et notre cerveau assimile et associe encore une fois ces bruits à un environnement donné. Nous allons alors associer un type de bruit, une impression générale à une vision. C'est pourquoi une musique lente, avec des bruits aigus (dans un certains spectre) comme des musiques celtes nous feront penser à la nature. Alors qu'une musique rapide et plus électrique sera associé à un milieu urbain, qui correspond à notre état d'esprit et aux images

assimilées par notre cerveau. On dit que la musique fait voyager, c'est en partie dû à cette association mais aussi à la culture.

Intuitivement, nous saurons reconnaître (même sans en avoir beaucoup écouté) une musique venant des pays africains, asiatiques ou d'autres continents. Les cultures des différentes civilisations s'entendent à travers la musique. Ceci est dû à l'éducation, à l'environnement visuel mais aussi aux habitudes des civilisations et à leurs codes musicaux. La culture asiatique utilise beaucoup d'instruments que nous n'utilisons pas en Europe de l'ouest, tout simplement car notre culture ne nous a pas habitués à ça. C'est pourquoi la plupart des films asiatiques sont remastérisés pour correspondre aux codes du pays d'exportation avant de sortir chez lui.

Tous ces éléments nous ont permis de dresser une liste de lois implicites autour des codes de la musique. Bien entendu, celles-ci dépendent de la culture, c'est pourquoi les règles citées ci-dessus s'appliquent à toutes les cultures, elles sont le reflet du monde dans lequel nous vivons.

Technique d'analyse existante

Le son étant un art faisant appel, comme tout art, à la sensibilité de l'être humain, il est de fait, subjectif. C'est pourquoi le travail d'un designer sonore est requis pour ce genre de production et qu'il est difficile de l'automatiser. Cependant, le son est avant tout un signal. Pour rappel, le son est créé par des vibrations à différentes fréquences. Nous allons distinguer 3 catégories de sons, les infra-sons (en dessous de 20Hz), les sons audibles humainement (entre 20Hz et 20 000Hz) et les ultra-sons (au-dessus de 20 000Hz).

Figure 1 a) Domaine des fréquences audibles pour l'oreille humaine.

b) Limites supérieures des fréquences audibles chez l'homme et chez certains animaux.

Nous intéressants à la musique, nous ne traiterons que les sons audibles. Cependant, avec un ordinateur nous pourrions traiter de la même manière les infra et ultra-sons. D'un point de vue informatique, le son sera présenté sous la forme d'un tableau à 2 dimensions mettant en parallèle la fréquence et la nuance relative. Pour exemple prenons un son constant à 1 000 Hz, nous pourrions exprimer sa puissance entre 0 (aucun son) et 1 (puissance de sortie dans les enceintes du son). Donc à 0.2 le son sera le même qu'à 0.7 mais moins audible. La combinaison de ces nuances et de ces fréquences sur le spectre audible permet de concevoir des musiques.

Cependant, tous les sons ne sont pas perçus de la même façon par l'oreille humaine. Prenons par exemple les basses (entre 20 Hz et 160 Hz voir 312 Hz en prenant large), ces dernières produisent plus de bruits audibles par l'oreille. C'est pourquoi produire un son à 20 Hz et d'une puissance 0.5 sera plus audible qu'un son à 2 000 Hz à la même puissance. Ceci a un impact non négligeable sur l'analyse sonore. Intuitivement, l'homme peut dire si une musique a une sonorité grave ou aiguë et, ceci étant subjectif, tout le monde peut ne pas être en accord avec lui. Mais la machine, ne disposant pas de sensibilité mais simplement d'informations, ne peut pas déterminer la teinte d'une musique. Il faudrait pour cela définir pour chaque fréquence un taux de conversion pour niveler une musique.

La plupart des oreilles expérimentées arrivent à reconnaître les instruments dans une musique. Bien souvent parce que leur utilisation est soit différée dans le temps (quelques secondes voir millisecondes) ou bien parce qu'elles n'ont pas la même sonorité (exemple : batterie et clavier ne produisent pas des sons dans les mêmes fréquences). Pour l'ordinateur, il est impossible de les distinguer sans une intervention humaine. Cette limite est due au fait que les fréquences des instruments peuvent se superposer et donc nous ne pouvons assurer que sur une plage définie de fréquences, celles-ci feront toujours référence au même instrument. Mais dans le cas où nous arriverions à séparer les instruments, les notes pourraient être retrouvées grâce à la fréquence fondamentale. La fréquence fondamentale est définie par l'instrument, cela sera le multiple sur lesquels les notes seront calquées.

Prenons pour exemple une corde de piano qui a pour fréquence fondamentale 440Hz, toutes ses notes seront exprimées en multiple de 440Hz.

Maintenant que nous avons pu voir les limites de l'analyse nous pouvons déterminer que la seule façon d'analyser une musique est sur ces données pures. Une des analyses les plus communes est l'analyse de la pulsation.

La pulsation définit en partie le rythme et il est souvent donné principalement (mais pas que) par un instrument à percussions. La pulsation peut être analysée en regardant l'énergie dissipée au cours du temps par une musique sur une bande.

Les filtres « passe haut », « passe bas » et « passe bande ». Ils permettent de filtrer une partie de la musique et donc d'en extraire une partition de la fréquence. Cet outil sert notamment pour analyser la pulsation.

Une autre analyse possible est aussi le développement d'intensité dans le temps. Prenons une ou plusieurs fréquences données et regardons leurs évolutions d'intensité dans le temps.

Analyse de l'harmonique en fonction de la gamme utilisée pour la détection de suites de notes caractéristiques d'un genre de musique ou d'une tonalité.

Ces éléments permettent de façon simple ou combinées, de déterminer un bon nombre de paramètres pour analyser la musique et la classer. Nous pouvons définir le genre, le rythme, et filtrer certaines fréquences. Nous développerons cette partie plus tard mais, à titre d'exemple, nous pourrions imaginer que dans un environnement désertique, une tempête fait rage et la puissance de cette tempête serait fonction du beat de la musique courante.

Level Design

En jeu vidéo, le level design est l'étude et la construction de la structure d'un niveau de jeu par un Level designer. Il doit définir ou s'adapter à la mécanique de jeu que nous appellerons plus communément gameplay. Cette définition est spécifique au jeu vidéo, cependant, il est possible de l'appliquer dans d'autres domaines. Au théâtre, le metteur en scène avec le responsable des décors fait lui aussi du level design, il pose les bases d'un environnement, il en va de même pour le cinéma. Au

final, nous pourrions définir le level design comme la création de l'environnement visuel pour le spectateur.

Dans notre cas nous allons nous intéresser aux procédés déterminants dans la création d'un environnement pour le ressenti émotionnel du joueur. Ici, il ne s'agit pas de définir un level design qui va impacter la manière de jouer du joueur mais un level design qui va impacter la vision émotionnelle du joueur. Nous serons donc plus proches de la mise en scènes telle qu'elle pourrait être créée dans le cinéma ou le théâtre que dans le jeu vidéo. C'est pourquoi nous nous baserons sur le travail d'un scénographe pour définir celui du level designer.

Les composantes scénographiques.

Prenons la définition encyclopédique du terme scénographie :

" C'est l'art de concevoir l'aménagement expressif d'un espace complexe, au moyen d'un ensemble d'éléments picturaux, plastiques et techniques qui permettent l'élaboration d'une mise en scène. "

- Larousse

Ainsi nous pourrions noter, dans un premier temps, que le scénographe a la possibilité de manipuler cinq composantes techniques de la manière la plus cohérente possible afin d'en dégager l'impression émotionnelle recherchée. Il compose avec des volumes, des objets, des couleurs, des lumières et des textures.

Le décor

Les Volumes

Avant toute chose, il y a la définition des volumes. C'est ce qui va permettre de cadrer les proportions de toute chose dans un espace donné. Ceux-ci sont donc basés sur la différence entre la vision de l'utilisateur/spectateur et les limites de l'espace auxquelles ce dernier est confiné.

Les objets

"Objets" est un terme général qui désigne toute chose contenue dans un volume. La taille de ces derniers correspond de manière proportionnelle, bien entendu, au volume auquel ils appartiennent. Les objets, de par leur présence ou absence, parlent à leur manière. Aucun objet n'est jamais muet, mais il faut parfois une analyse assez minutieuse pour en décrypter tout le message particulier. Les objets symbolisent donc l'essence même de la scène, ce sont eux qui économisent le narratif propre aux différents genres littéraires ou bien ici, dans le cas d'un jeu vidéo, à un scénario. Un agencement d'objets et d'éléments sont indépendants les uns des autres, mais constituant d'un tout, souvent éphémère, qui permet en un regard de découvrir le lieu et la temporalité de l'action.

Les couleurs

Les couleurs sont d'une complexité lorsqu'il s'agit de comprendre ce qu'elles représentent. Mais une fois admises, elles représentent un suprême potentiel de transmission de message. Dans un premier temps, il faut considérer que la couleur n'existe que dans notre cerveau. Il ne s'agit en fait que de fréquences qui vont inconditionnellement agir sur l'ensemble des êtres humains de la même manière. En effet, les analogies de structure de nos rétines, de nos systèmes nerveux vont répondre de manière plus ou moins automatique (par réflexe) à ces fréquences et ainsi agir, entre autre, sur notre rythme cardiaque et d'une manière plus générale, sur notre état d'esprit. C'est donc un bon moyen de conditionnement certes, mais il existe aussi des adaptations personnelles qui expliquent des préférences individuelles. En effet, dans un second temps, les couleurs permettent de véhiculer un message en fonction de la culture, l'éducation et l'expérience du monde et la nature qui l'entoure. Le transfert de l'image photographiée par l'œil, emprunte de nombreuses voies jusqu'à être transmise au cortex. Celui-ci analyse l'image et utilise les zones frontales associatives qui, en fonction des acquis mémorisés, vont tirer les conséquences pour alimenter, en réponse, les impressions qui lui semblent subjectivement associées.

Les lumières

La lumière est un atout majeur pour déterminer l'environnement dans lequel nous voulons plonger le joueur. D'un coup d'œil, le spectateur peut identifier

l'environnement dans lequel il se trouve. Est-il inquiétant ou bien réaliste. Dans quel moment de la journée nous trouvons nous. En intérieur ou bien en extérieur. La lumière permet de passer rapidement toutes ces informations. Pour prendre un exemple, en intérieur la lumière sera plus tamisée et plus artificielle. Elle ne viendra pas du soleil et donc surement de plusieurs sources différentes, c'est pourquoi les ombres seront multipliées et les températures (Kelvin) autres que celles du soleil. Cela peut renforcer un sentiment de malaise ou bien au contraire un sentiment de réconfort selon le type de source lumineuse.

Les textures

Une des composantes qui apportent le plus de crédibilité. Les textures sont appliquées aux objets en même temps que le choix des couleurs et de l'éclairage, ils sont tous les trois, étroitement liés. Après définition géométrique des éléments d'une scène, des propriétés s'appliquent aux objets qui leurs confèrent la rugosité, la couleur, la réflexion etc. Ces ensembles de propriétés sont inclus dans ce qui est techniquement connu sous le nom "Matériel". Le matériel comprend donc différents concepts associés avec le comportement de la lumière sur l'objet.

Cet ensemble de composantes va ainsi définir le lieu et l'environnement dans lequel se passe l'action, c'est ce que nous appelons le décor. Le décor bien entendu est une composante essentielle de l'identification. C'est pourquoi le décor nous donne des éléments visuels qui nous permettent d'identifier notre environnement, cependant, ces éléments doivent être cohérents entre eux pour définir un environnement. Une télévision n'aurait pas sa place dans un environnement médiéval par exemple. Et pourtant, ce décor permet un certain degré de liberté. Il serait possible de trouver une télévision dans un château, si le joueur explore à notre époque des « ruines ». Cela permet de créer une certaine ambiance dans un contexte établi. Dans un seconds temps, le scénographe va travailler sur : La mise en scène.

La mise en scène

La mise en scène peut se décomposer en deux parties, l'agencement scénique des décors et la cohérence de ceux-ci. Ces deux éléments permettent de définir l'univers. Prenons par exemple un cactus dans un désert. Jusqu'ici, rien de spécial, il est normal de trouver ces éléments ensembles. Maintenant prenons un cactus qui se

déplace dans un désert. Ici, le spectateur sait directement que l'environnement se place dans un contexte onirique, surréaliste ou comique. Le décor et/ou la lumière ne lui apprennent rien là-dessus, c'est bien la mise en scène qui s'en occupe et plus particulièrement la cohérence par rapport à l'environnement habituel que nous connaissons tous. Prenons maintenant comme exemple une salle, comme une salle à manger. Si tous les éléments sont au sol, le contexte semble être réel, mais sinon plaçons ces éléments sur le plafond, le contexte se modifie. L'agencement permet de définir ce contexte. Il peut parfois rejoindre la cohérence.

Espace scénique

L'espace scénique concerne pour nous l'endroit où le joueur pourra ou non se déplacer. Pourra-t-il traverser les murs et voler ? Les réponses à ces choix impactent la vision que le joueur aura de lui-même et donc l'appréhension de son environnement. Est-il humain ou fantôme ? Dans le deuxième cas nous nous trouvons dans un environnement surréaliste.

Ces composantes vont nous aider à définir l'environnement du joueur et donc l'impact des choix que nous faisons sur lui. L'émotion ressentie par le joueur ne sera pas la même dans un environnement réel que dans un environnement imaginaire.

Qu'est-ce qu'un environnement virtuel ?

Un environnement virtuel est une représentation d'une mise en scène, d'un décor de façon informatique. Principalement, il s'agit du croisement d'un jeu vidéo et d'un réseau social. Il s'agit d'immerger le spectateur dans un environnement réel et lui laisser le libre choix de ce qu'il va en faire. L'environnement virtuel le plus connu est le « jeu » : Second Life. Ici, le joueur est représenté par un avatar. Ce jeu pourrait être vu comme un simulateur de vie où le joueur peut effectuer une partie des actions qu'il ferait dans la réalité, comme par exemple travailler, aller boire un verre avec ses « amis », etc... C'est ici que nous voyons la dimension réseau social arriver. Le but est d'interagir avec les autres joueurs du monde. Outre cela, le joueur peut être amené à visiter différentes régions comme dans un jeu MMORPG (Massivement Multi-joueurs Online Role Play Game).

Le but d'un environnement virtuel est de reproduire la réalité. Comme vu précédemment, les contextes doivent donc être réels au niveau des composantes scénographiques.

Connaissance de la médecine neurologique

Si nous savons aujourd'hui que le cerveau joue bel et bien une part importante dans nos actions, nos réactions, nos émotions, etc... Il se trouve être encore bien complexe. Et pour cause :

"Si le cerveau était suffisamment simple pour que nous le comprenions, nous serions si simples d'esprit que nous ne le comprendrions pas."

- Lyall Watson

Nous nous devons de le comprendre un minima afin de pouvoir agir aux mieux avec ce qu'il permet. Pour cela, nous nous référons principalement, aux études du psychologue Abraham Maslow publiées dans la seconde édition en 1970 de son livre "Motivation and Personality", ainsi qu'aux observations faites au travers des différentes techniques et outils d'imagerie cérébrale apparus dans les années 1950 et qui ont provoqués un progrès significatif des sciences cognitives dans les années 1990.

Roland Jouvent, professeur en neuropsychologie, insiste beaucoup sur la cohésion qui existe entre effectuer une action et seulement l'imaginer. En effet, à la direction du "Centre émotion" du CNRS à la Salpêtrière, il a pu mettre en avant, au travers d'IRM, que les régions du cerveau activées sont les mêmes : les aires pré motrices et plus ou moins à l'identique, le cortex moteur. Imaginez-vous prendre une douche chaude. Cela déclenche les souvenirs de nos sens et ainsi le système limbique réagit de la même manière que lorsque nous prenons réellement cette douche. Plus la simulation sera précise, plus l'imagination perceptive du corps sera importante. À noter que l'excès de simulation peut entraver le passage à l'action, qui implique effort, résistance, douleur.

Pour simplifier, nous pouvons retenir que l'imagination n'est pas une aptitude abstraite de l'esprit, mais une action spécifique de notre organisme, non négligeable et innée.

En effet, cette aptitude à pouvoir simuler peut-être provoquée par la visualisation d'actions d'autrui : regarder un congénère active nos neurones miroirs dans le cortex moteur et pré moteur. Ils permettent l'apprentissage par imitation. Ceci est également vrai pour nos émotions : voir une expression sur la figure d'un semblable provoque les sentiments aussi bien que lorsque nous le percevons directement nous-mêmes.

Le système d'alerte

Nous pouvons tout d'abord comprendre (notamment au travers de la célèbre "pyramide de Maslow") que le cerveau fonctionne par priorité. Sa première des priorités est bien entendu de protéger l'harmonie du fonctionnement de l'organisme sans tenir compte de tous les éléments et toutes les contraintes qui lui sont extérieurs, nous nommons cela l'homéostasie.

Le cerveau use une petite partie (le tronc cérébral) afin de prémunir des sensations primaires telles que la soif et la faim, et ainsi faire évoluer l'organisme vers une réaction adéquate.

Lorsque le cerveau (le thalamus sensoriel) perçoit une douleur ou un danger il court-circuit ses procédures vers l'amygdale, qui met le corps en état d'alerte (accélération du rythme cardiaque), voire déclenche un réflexe inné ou acquis. L'amygdale dispose de ses propres souvenirs implicites, avec des estimations catégoriques : positives, négatives. Cela permet d'avoir des réponses motrices accrues : moins de 200 ms. Mais le thalamus envoie également des signaux stimulant vers le cortex pour y être traités de manière plus "réfléchie" usant d'une mémoire explicite cette fois-ci. Cela permet de discerner l'environnement et les objets puis de les conceptualiser avant de les contextualiser.

Ce système dépend intégralement du circuit de la punition (ou PVS) et aboutit à la sécrétion d'adrénaline et glucocorticoïdes.

Le système de récompense

Ce système est bien difficile à comprendre tant il dépend de différentes parties du cerveau. Lorsque le cortex décèle une jouissance envisageable (nous ne sommes donc plus dans une réaction primaire), de la dopamine est dégagée afin de déclencher certaines des réponses : végétatives, cognitives et la sensation de plaisir.

Ce système permet de mémoriser des situations plaisantes pour les rechercher à nouveau.

Système d'inhibition de l'action.

Henri Laborit a mis au jour un troisième système dans les années 60. Si le cortex préfrontal a appris précédemment qu'une action est vaine, il va figer l'organisme. Cette troisième réaction, tente d'éviter par exemple l'agression. Lorsque le système d'inhibition de l'action est activé de manière répétée, il entraîne l'hypertension. Par extension empirique cela provoque la destruction du système immunitaire, des troubles de l'humeur, de la dépression et pour finir du suicide.

L'intérêt de ces recherches et la connaissance de ces trois systèmes permettent de mieux maîtriser certaines émotions que nous souhaiterions éveiller chez l'utilisateur. Ainsi nous pourrions par exemple faire appel à certaines émotions en choisissant de provoquer celle-ci par mimétisme, ou bien par rappel d'un moment vécu. Un autre cas qui pourrait être employé est l'apparition d'image subliminale pour ainsi travailler sur le système d'alerte et exciter l'amygdale. Les actions possibles sont donc multiples.

Projet existant liant son et graphisme

Les différents types

Les projets de jeu vidéo, liant le son et les graphismes, se font assez rares pour la simple et bonne raison que l'analyse d'un signal désordonné a du mal à produire un graphisme ordonné de façon automatique. La musique a un impact émotionnel qu'il est difficile de traduire pour un ordinateur. La machine ne peut pas ressentir des émotions, c'est pourquoi il est nécessaire d'avoir une intervention humaine. Cependant, certains s'y sont essayés. Bien souvent dans le domaine de l'abstrait ou alors dans un but rythmique de gameplay.

Les jeux abstraits

Dans ces jeux, le but est de générer un environnement virtuel en fonction de paramètres ou d'actions. Dans cette catégorie nous pourrions citer :

Panoramical : Jeu dans lequel il faut contrôler via des jauges, la puissance de certaines fréquences et qui génère plus ou moins un environnement.

Cet environnement n'a rien de réaliste, il relève plus du concept artistique que d'un environnement virtuel, de plus la seule interaction possible est celle de modifier les fréquences.

Les jeux basant leur gameplay sur la musique

Cette deuxième catégorie met en avant différents gameplay possibles en rapport avec la musique.

Guitar Hero : Jeu de rythme dans lequel nous possédons une guitare réelle avec des couleurs associées aux boutons, il s'agit d'appuyer sur les boutons au bon moment pour que la note se joue correctement. Ici, si vous vous trompez de note, la musique fait une fausse note et l'environnement (les spectateurs) réagissent en conséquence.

Flower : Jeu poétique qui prend place dans un environnement réel, une prairie. Vous jouez un étal de fleur volant au vent et qui doit collecter d'autres pétales auprès des fleurs. A chaque fois que vous ramassez un pétale cela joue un son, à vous de trouver le bon rythme pour produire une jolie musique.

Crypt of the Necrodancer : Ici nous nous rapprochons plus de notre concept qui est d'analyser la musique pour en tirer des évènements. Necrodancer vous laisse jouer avec vos propres musiques. Il va analyser ces dernières pour en déterminer le beat. Vous ne pouvez effectuer vos actions que sur un beat de la musique. Les ennemis et vous-même devez avancer au rythme de la musique. Il s'agit d'un jeu de type dungeons crawler ou le but est de se déplacer de salle en salle tout en éliminant les ennemis.

Réalisation

Traitement du son

Problématique : Comment est-il possible d'extraire le son et ces composantes fondamentales d'analyses ?

La phase pré-analytique

Définition

Pouvons-nous résumer une chanson à une simple suite de note de musique prises au hasard dans un tout ?

Commençons par définir ce qui résume une chanson. Distinguons deux points de réflexion :

Du point de vue humain

Du point de vue analytique

Du point de vue humain, ce qui résume une musique est ce qui va lui rappeler cette même musique. Prenons pour exemple : « Cette soirée-là » de Yannick, et « Cette

année-là » de Claude-François, la première étant une reprise de la seconde. Ici nous pouvons dire que pour les gens qui connaissent l'original de Claude François, la première chanson est un résumé de la seconde, car malgré des « instruments » différents la première reprend les mêmes notes et sonorités que la seconde et donc fait appel au souvenir de la chanson originale.

D'un point de vue analytique ces deux chansons ne seront pourtant pas semblables car les fréquences utilisées ne seront pas les mêmes (malgré le fait que les notes soient les mêmes, un "La" en guitare ne sera pas le même qu'un "La" en flutes). Nous pourrions toujours définir la suite de notes en prenant la fréquence fondamentale et en exprimant les notes en fonction de celle-ci. De façon analytique deux musiques seront semblables si elles utilisent les mêmes notes répétées (d'un point de vue légal une suite de 8 mêmes notes est considérée comme un plagia).

Ce qui définit une musique est donc une suite de notes consécutives. Maintenant, il nous faut savoir d'où nous pouvons extraire cette suite de notes ?

Lorsque que l'être humain se souvient d'une chanson c'est souvent par son refrain. Le refrain étant une suite de notes revenant à intervalle régulier sert souvent à caractériser une chanson car de par son aspect répétitif elle marque plus facilement l'utilisateur. Mais le refrain seul n'est pas suffisant, car il n'est pas toujours présent et selon la taille de la chanson peut avoir un impact moindre. C'est pourquoi la meilleure façon de caractériser une chanson est d'en déterminer la plus longue suite de notes répétées dans une chanson.

Transformation du son en spectre

Les fichiers sonores courants sont des fichiers binaires. L'analyse musicale passe par l'analyse d'un spectre sonore. C'est pourquoi il va nous falloir transformer ces fichiers audio en spectre audio.

Qu'est-ce qu'un spectre sonore ?

Le spectre sonore est un graphique tridimensionnel qui met en correspondance une fréquence avec une amplitude. Cette amplitude définit la puissance d'une fréquence à un moment donné. Nous avons donc l'évolution au cours du temps de la puissance d'une fréquence.

Attention : l'amplitude d'une fréquence ne définit pas sa puissance de sortie dans les enceintes (les décibels). Effectivement les fréquences étant différentes il faudra une amplitude plus ou moins élevée selon la fréquence.

Dans un premier temps il va nous falloir décompresser le fichier audio. Vu le nombre de fichiers existants, nous n'allons pas tout décrire leur système de décompression, cependant voici un exemple pour les fichiers MP3 (MPEG 1/2 audio layer 3 de son nom complet).

Le fichier MP3 contient un fichier d'entête qui permet de définir son contenu. Ce header a une taille fixe de : 32 bits soit 4 octets. Dans lesquels se trouvent différentes informations comme le taux d'échantillonnage (sample rate) qui définit la plage de fréquence audio (cf. spectre sonore), le type de sortie : mono, stéréo, dual Channel, etc... Mais aussi diverses informations sur la musique en elle-même : copyright, style de musique et auteur.

(Source : http://mpgedit.org/mpgedit/mpeg_format/mpeghdr.htm)

Pour la décompression nous allons surtout nous intéresser au taux d'échantillonnage et au débit (bit rate) qui nous permettront de définir le nombre de bits pour un état du spectre dans le temps.

La formule étant la suivante : (en fonction du layer défini dans le header)

Pour la couche I, les fichiers utilisent cette formule :

$$\text{FrameLengthInBytes} = 48 \times \text{BitRate} / (\text{SampleRate} + \text{Padding})$$

Pour les couches II et III, les fichiers utilisent cette formule :

$$\text{FrameLengthInBytes} = 144 \times \text{BitRate} / (\text{SampleRate} + \text{Padding})$$

Nous savons maintenant que nous pouvons lire le reste du fichier et découper par tranche de taille `FrameLengthInBytes` pour décoder notre musique. Nous savons que dans une « Frame » nous avons « `SampleRate` » fréquence soit nous avons donc $\text{FrameSize} = \text{FrameLengthInBytes} / \text{SampleRate}$. Ce qui nous permet de définir un chiffre réel (compris entre -1 et 1 selon le système de décompression). Le moyen le plus utilisé pour une décompression rapide est les transformations rapide de Fourier

(Fast Fourier Transform : FFT), qui permet de passer d'une représentation binaire en une sinusoïde discrète pour un coup en mémoire moindre.

La phase Analytique

Analyse Globale

Comme dit précédemment, la musique peut être caractérisée par une suite de notes répétées. Cependant ce n'est pas la seule donnée qui peut être extraite dans une analyse globale du son.

Le tempo : Le tempo définit le rythme de la musique ou autrement le nombre de battements par minutes. Il ne peut être évalué précisément que par une analyse globale de la musique. Nous devons réussir à déterminer les « pics » de son par rapport à une moyenne sur un temps donné. Prenons par exemple une grosse caisse de batterie simple frappant à intervalle régulier toutes les secondes.

La fréquence fondamentale : en analysant la source entièrement nous pouvons déterminer la fréquence la plus utilisée dans la chanson.

La gamme de note : il suffit d'exprimer les fréquences en fonction de la fréquence fondamentale.

Pour extraire ces informations, nous avons besoin d'analyser le spectre sonore entièrement. C'est donc une analyse qui ne peut pas se réaliser en temps réel (en même temps que la musique est jouée). Il nous permettrait de définir une tendance globale sur la musique. La gamme de note nous permettrait d'identifier les styles de musiques, par exemple les musiques de l'été ont très souvent les mêmes notes utilisées.

Analyse Spécifique d'une frame

Lors de l'analyse spécifique d'une frame nous ne possédons pas les informations de la frame actuelle et des frames précédentes s'il y en a eu. Les analyses ici sont surtout des analyses d'évolution dans le temps :

Evolution dans le temps (diminution, augmentation de l'amplitude)

Détection d'un battement (beats) si le son est X fois plus puissant que la moyenne des Y dernières frames, dans le même style détection d'un creux dans le son

Evolution globale de la somme des amplitudes (le son a-t-il tendance à être plus puissant ou moins puissant dans le temps ?)

Combinées à une analyse globale, ces fonctions vont nous permettre d'affiner la visualisation en temps réel, et permettre de coller à un instant T à ce que l'utilisateur entend.

Les outils d'analyses

Les filtres

Le son étant un ensemble de signaux, il est possible de définir des filtres au niveau des fréquences :

Les filtres passe haut et passe bas : qui permettent de ne laisser passer qu'une fréquence supérieure ou inférieure à un seuil donné. Ces filtres vont s'avérer très utiles, les fréquences basses et hautes n'ayant pas les mêmes puissances sonores si nous voulons déterminer l'impact sur l'utilisateur nous devrons souvent user de ces filtres pour mettre les deux puissances sur un pied d'égalité et pouvoir les comparer.

Les filtres passe bande : qui sont la combinaison d'un filtre passe haut et un filtre passe bas permettent d'extraire une fréquence dans un intervalle défini.

Outre les filtres, nous pouvons aussi avoir besoin de déclencheurs (triggers) qui émettront un signal positif si le son dépasse ou est inférieur à une certaine amplitude, dans le cas contraire ils émettront un signal négatif. Ces trigger pourront aussi servir pour analyser la tenue d'une fréquence dans le temps, par exemple, nous pourrions définir un trigger qui se déclencherait si la fréquence est tenue dans le temps.

Les Descripteurs

La famille des descripteurs sert à décrire l'évolution ou l'état actuel du signal. Prenons par exemple les descripteurs de variations, il peut être intéressant de savoir si sur une période donnée la musique a plutôt eu tendance à augmenter/diminuer ou varier constamment.

Ces outils-là se placent comme un inspecteur. Ils vont stocker à chaque frame la valeur de la fréquence ou des fréquences sélectionnées selon les modèles mathématiques classiques.

Implémentation et Optimisation

Implémentation

D'un point de vue purement fonctionnel nous allons utiliser la librairie FMOD. Cette dernière étant très utilisée en Jeu vidéo notamment dans les moteurs Unity et Unreal Engine. Ce choix se base sur 3 caractéristiques :

- Possibilité de son environnant

- Possibilité de son 3D

- Lecture de différent format (+ normalisation)

Le son environnant se démarque du son 3D par le fait que celui-ci ne sera pas spatialisé. La spatialisation est le fait de donner une certaine portée au son pour reproduire un effet de réalité. Si une balle tombe à un endroit, le son partira donc de cet endroit pour ensuite rebondir sur les parois du monde et peut être arriver à l'oreille du joueur. Contrairement au son environnant qui lui sera toujours présent dans les oreilles du joueur. Ici, l'idée est de faire de la musique courante une sorte de musique d'ambiance mais aussi d'ajouter des sons 3D en fonction de évènements sur le terrain.

Pour ce qui est de la lecture et de la normalisation du format, nous voulons laisser le plus de choix possible au joueur, et nous ne pouvons donc pas prédire le format de ses fichiers audio. Les formats de fichier audio sont souvent propriétaire et

l'extraction des données est spécifique à chacune. Ainsi avec FMOD nous bénéficions de toutes ces transformations pré faites.

Comme dit plus haut, il va falloir transformer les données binaires en spectre pour pouvoir être analysées. FMOD contient des outils pour réaliser la FFT rapidement.

Optimisation

Précédemment, nous avons vu que nous devions pouvoir réaliser deux types d'analyses :

Analyse en temps réel

Pré-Analyse

La première optimisation possible est sur la plage des fréquences. Plus la plage sera grande, plus la précision sera grande mais plus le temps de traitement sera long. Reprenons la définition du 1.b, ici le pas de fréquence est défini par le taux d'échantillonnage et la taille de la fenêtre. La taille de la fenêtre va définir la précision que nous voulons en bit pour l'ensemble des fréquences.

Soit :

$$\text{FrequencyStep} = (\text{WindowSize} \times 2) / \text{SampleRate}$$

$\text{SampleRate}/2$: définit le nyquistRate (en générale 22 000Hz)

On va donc exprimer le buffer de la fenêtre en fonction du nombre de fréquence et nous pourrons ainsi définir le pas de fréquence.

Les paramètres d'une musique étant représentés par un tableau a deux dimensions de taille :

$$\text{Durée de la musique} \times \text{windowSize}$$

Nous pouvons aussi optimiser le temps de calcul en utilisant du multithreading. Le principe du multithreading est de faire travailler en parallèle ou en pseudo-parallèles

la CPU (l'unité central de traitement, aussi plus communément appelé "processeur" ou, "Central Processing Unit", en anglais). Ici, n'ayant à traiter que des données nous pouvons facilement paralléliser le traitement de la FFT et ainsi gagner en temps d'exécution.

Pour ce qui est de l'analyse en temps réel, nous allons extraire les données à chaque frame, les traiter avec les mêmes fonctions que vu précédemment, et les passer ensuite à FMOD pour qu'il les joue.

Gestions des transitions musicales

Afin de simplifier la génération, nous ne proposerons pas de solution de génération basée sur les transitions musicales. Celles-ci se baseront directement entre les différentes zones mitoyennes établies. Nous estimons qu'il ne serait pas ou peu représentatif d'une émotion puisque la résultante de cette transition musicale serait établie de manière très aléatoire selon les musiques, la technique et le moment de la transition.

Cependant, la transition aussi bien visuelle que musicale est importante pour ne pas sortir l'utilisateur de son immersion et provoquer un changement d'humeur non prévu. C'est pourquoi nous avons cherché un minimum de technique de transition automatisée entre deux musiques.

Etude de l'harmonique

Dans un premier temps, nous nous devons de déterminer quelles musiques sont les plus à même d'être utilisées ? Pour cela nous appliquerons le principe de la roue de Camelot. La condition première d'utilisation de cette roue est de déterminer l'harmonique des morceaux d'un set de musique. Nous pouvons ensuite les trier par harmonique. Enfin, selon quelques principes, règles de mixage, nous pouvons tisser des liens entre les musiques et établir une RoadMap.

Il existe 6 principes de mixage :

Transition par même clé d'harmonique : Cette première technique utilisée aura pour effet de lasser après quelques minutes, car le mix n'aura pas d'évolution de hauteur et paraîtra plat.

Transition par quinte : Cette seconde technique permet de choisir un morceau de la clé harmonique supérieure de la même tonalité. Comme dans l'exemple, nous passons d'un La Bémol Majeur à un Mi Bémol Majeur. La transition aura pour effet d'une légère montée en puissance.

Transition par quarte : Principe inverse à celle de la quinte : cette technique permet de choisir un morceau de la clé harmonique inférieure de la même tonalité. Comme dans l'exemple, nous passons d'un Si Bémol Majeur à un Mi Bémol Majeur. La transition aura pour effet semblable à une diminution de la pression.

Transition par tonalité : Ce quatrième pratique a pour principe de réaliser un changement de tonalité durant la transition. Comme dans l'exemple, nous passons d'un Ré Mineur à un Fa Majeur. La transition vers une tonalité Mineur permet d'apporter plus de "profondeur" tandis que la transition vers une tonalité Majeur donnera une impression "d'ouverture".

Transition par note supérieur : Cela consiste à monter l'harmonique d'une note supérieure sans changement de tonalité lors de la transition. Autrement dit, comme l'indique l'exemple, nous montons d'un Fa Majeur vers un Fa Dièse Majeur qui équivaut à la note supérieure la plus proche. Cela provoque un léger coup de boost dans le mix.

Transition par ton supérieur : Cette dernière technique de transition consiste à monter, lors de la transition, l'harmonique sur le ton supérieur. Ainsi comme le montre l'exemple, une clé harmonique Fa majeur deviendra un Sol Majeur. Cela est un second moyen (avec la transition par note supérieure) de provoquer un coup de boost dans le mix.

Ces règles établies nous permettent désormais d'avoir un choix cohérent de musique. Ces différentes transitions permettent également d'appuyer certains ressentis recherchés chez l'utilisateur. Mais cela ne fait pas tout, en effet les transitions peuvent sonner correctement à nos oreilles et pourtant nous paraîtront totalement à contre temps ou sur un faux rythme...

Analyse du rythme

Dans un second temps, nous nous devons de déterminer le tempo des musiques afin de pouvoir caler ces dernières, les unes sur les autres lors d'une transition. Le but est de réussir à déterminer une valeur pour le tempo des musiques sous une même unité de mesure. Nous avons choisi de déterminer le BPM (beats par minute). Une fois cette méthode établie et donc cette valeur obtenue, il sera relativement aisé de déterminer s'il faut ralentir ou accélérer leur vitesse de lecture pour les ajouter sur le même tempo.

Plusieurs méthodes existent cependant une semble plus pertinente tant elle est efficace et robuste : l'Inter-corrélation avec un peigne de Dirac. Cette méthode présente néanmoins un défaut incontournable, elle est très coûteuse et ne pourra donc pas être appliquée en temps réel. Nous proposons alors une autre méthode : La détection des crêtes.

Le principe est d'appliquer un filtre passe-bas pour écarter les bruits, impuretés et minimiser les variations.

Ensuite on sur-échantillonne le signal afin de diminuer le nombre de données à analyser, là encore nous diminuons le risque de bruit et ainsi nous rendons le temps de calcul plus efficace.

Par la suite on déterminera un signal "alpiniste" w tel que :

$$w[1]=x[1]$$

$$\{ \begin{array}{ll} w[k]=x[k] & \text{Si } x[k] \geq w[k-1]-\epsilon \\ @ \ w[k-1]-\epsilon & \text{Sinon (décroissance linéaire)} \end{array} \}$$

Avec ϵ une constante de décroissance.

Nous obtenons alors un signal w qui, pour chaque pente croissante, nous permet de déterminer que nous sommes sur un beat. Enfin il ne nous reste plus qu'à diviser le nombre total de beats (N) par la durée du morceau en seconde (T) pour obtenir le tempo tel que :

$$BMP = (60 \times N) / T$$

Application de la transition

Maintenant que nous avons un moyen de déterminer le BPM d'un morceau, nous pouvons appliquer une transition correcte basée sur trois principes : l'accord du rythme des musiques (également appelé time stretching), la superposition qui consiste à caler les BPM d'une musique sur l'autre et enfin le fondu qui revient simplement à basculer le volume sonore de la piste en cours de lecture vers la suivante.

Génération de l'environnement

Génération d'un environnement réel

Problématique : Comment générer un environnement cohérent via une donnée aléatoire.

Dans cette partie nous nous intéresserons aux méthodes possibles dans la génération procédurale. Notre outil doit être capable de construire son niveau en fonction de n'importe quelles analyses musicales et donc d'une donnée pseudo aléatoire.

La génération procédurale consiste à placer différentes ressources du monde de façon aléatoire. Cette dernière obéit à des règles prédéfinies ayant une probabilité de sortie différente.

La génération par itération (exemple : no man's sky) :

No man's sky est un jeu encore en développement qui a pour ambition d'utiliser la génération procédural pour générer 18,446,744,073,709,551,616 mondes différents (source : <http://iq.intel.co.uk/no-mans-sky-procedural-generation/>) . Leur algorithme se divise en plusieurs parties :

Création de la planète ou ici carte des hauteurs. Ils vont ainsi définir le relief global de la planète de façon aléatoire. (Perlin Noise)

Choix des biomes présents sur la planète. En fonction de la carte des reliefs ils vont se baser sur des faits pseudo scientifique pour déterminer les types de paysage que la planète aura : par exemple une carte ayant peu d'eau, aura surement des biomes plutôt arides.

Placement des éléments de décors (arbres, eau, nuage, neige etc...). Ici, le choix est fait en fonction des biomes définis, et de différents paramètres comme la hauteur ou la distance à l'eau la plus proche. Mais aussi une partie d'aléatoire.

Pour terminer, il implémente les créatures en fonction de leur besoin vital (air ou eau, carnivore, herbivore).

La génération par automate cellulaire

Notamment utilisé dans le jeu Terraria pour la génération de cave, les automates cellulaires sont représentés par une grille de X cases par Y dans laquelle chaque case à un état (souvent « vivant » ou « mort »). A cette grille est associés un ensemble de règles qui définit l'état suivant. Prenons pour exemple le jeu de la vie de John Horton Conway défini par les règles suivantes :

Chaque cellule vivante avec moins de deux cellules vivantes autour d'elle meurt pour cause de sous population.

Chaque cellule avec 2 ou 3 voisines vie à la génération suivante

Chaque cellule avec plus de 3 voisines meurt pour cause de surpopulation

Chaque cellule morte avec exactement 3 voisines vie à la prochaine génération.

Avec ces règles et un pattern de départ, différents types d'automates arrivent à émerger :

Automate stable : après une génération X il n'évolue plus

Automate oscillatoire : leur état oscillent entre deux ou plus états en boucle.

Automate « vaisseaux » : leur états sont récurrents après un certain nombre de générations

Automate « mathusalems » : il est défini comme un petit objet qui explose en de nombreux objets stables.

Il existe d'autres structures pouvant apparaître mais étant plus rares nous ne nous attarderons pas dessus.

Pour en revenir à la génération procédurale une structure stable du jeu de la vie peut amener à des motifs représentant des caves, de vastes espaces plus ou moins irréguliers, comme dans l'exemple ci-dessous.

Assemblages de parties prédéfinies.

Cette technique a notamment été utilisée dans la série des jeux Diablo, et plus généralement dans la génération de donjons aléatoires. Il s'agit ici de disposer d'une bibliothèque de salles, avec un marqueur définissant les entrées et sorties de chaque salle, nous allons ensuite mettre bout à bout les salles pour générer un donjon. Cet algorithme permet de gérer plus facilement les choix possibles et les chemins possibles par l'utilisateur et il est moins coûteux en mémoire. Cependant il peut aussi être manipulé en forçant la chance.

Ces différentes techniques permettent de générer des environnements virtuels statiques.

Monde dynamique

Les mondes dynamiques se construisent au fur et à mesure du besoin du joueur, sur une « seed » commune. Une « seed » est un chiffre généré aléatoirement permettant d'influencer la partie aléatoire de la génération. Ainsi avec une même « seed » nous aurons des mondes semblables mais pas identiques. C'est le cas pour le jeu Minecraft qui génère un monde faisant 8x la taille de la terre. Ce monde n'est pas généré en une seule fois mais dynamiquement en fonction des déplacements du joueur. Pour que le monde reste cohérent, le système s'appuie sur une « seed » qui permettra de générer un monde similaire.

Génération via un flux audio

Définitions Lexical

Pour introduire notre concept de génération d'environnement via un flux audio nous allons devoir commencer par définir les concepts de bases :

Les biomes : Nous allons définir par biomes les éléments de plus haut niveau, définissant une zone de même type dans leurs caractéristiques globales. Les caractéristiques globales définissent le type de zone tel que désert, jungle, maritimes, urbain etc... Elles nous permettent de définir le type de décors, les types d'évènements, l'éclairage et la skybox. Elles donnent un univers global à une partie de l'environnement crée. Par exemple dans le cas d'un biome pour la jungle nous aurons :

Densité de la faune et flore

Illumination

La skybox (le paysage lointain)

Les régions : Elles sont un niveau en dessous des biomes, elles en prennent donc les caractéristiques globales. Mais subdivisent les biomes en plusieurs parties. Les régions servent à définir un ensemble de sous régions ayant les mêmes propriétés au niveau musical.

Disposition des zones de flores et faunes

Les sous régions / Les zones d'évènements : Elles sont un niveau en dessous des régions. Elles définissent des zones dans lesquelles un évènement particulier a lieu. Ces évènements sont liés au biome auquel il appartient, par exemple pour une jungle cela pourrait être :

Eléments mobile : Meutes d'animaux

Eléments fixe : temple maya

Eléments évolutif : plante qui pousse

Figure 4 Exemple de réparation possible d'un biome au moyen de la méthode Voronoï

Implémentation et optimisation

Implémentation

Dans un premier temps, nous allons implémenter un système de génération procédurale dynamique qui se créera en même temps que l'utilisateur se déplacera dans l'environnement. L'idée est d'utiliser un diagramme de Voronoï pour découper les zones dans une région donnée. Le diagramme de Voronoï est un algorithme de subdivision de plan permettant de définir des zones d'influences par rapport à des germes. Nous tirons deux avantages de cette technique :

L'ajout Delaunay : qui nous permettra d'ajouter des germes pour étendre notre environnement en fonction des déplacements du joueur, et donc de rendre l'environnement dynamique.

Le déplacement des germes : pour les sous régions qui permettent aux éléments mobiles de voir le jour.

L'avantage du diagramme de Voronoï est qu'il permet un déplacement dynamique des zones pour les régions mobiles, ce qui nous permet de « forcer » l'utilisateur à vivre certaines expériences en fonction de ce que nous analyserons. Etant divisé en zone nous pouvons aussi gérer facilement le Culling (technique de rendu permettant

de ne pas afficher les parties qui ne seront pas visibles par le joueur). Aidé de cette technique nous définirons une zone autour du joueur qui sera prédéterminée, et nous pourrons ainsi charger/décharger des zones en fonction de la musique courante. Cette technique a été imaginée pour limiter le possible décalage qu'il y aurait entre la musique courante et l'environnement dans le cas où l'utilisateur ne bougerait pas trop. Ainsi, même un utilisateur statique verra des événements arriver vers lui et le monde autour de lui évoluer.

Optimisation

Pour la réalisation de cette partie nous utiliserions OpenGL. Permettant de gérer le rendu de la scène dans son plus bas niveau nous pourrions donc réaliser plusieurs optimisations. Dans un premier temps, il faut rappeler qu'OpenGL s'utilise au sein d'un programme et permet d'appeler et d'exploiter la carte graphique via des shaders. Les shaders présents à différents niveaux vont permettre de réaliser l'affichage des points. Tout ce qui sera à l'intérieur du code du shader sera exécuté directement sur la carte graphique. Ainsi, une des optimisations possibles, est d'essayer d'exécuter le plus de code possible du côté carte graphique. Cependant, cela pose aussi une problématique supplémentaire, la transmission des données. Le fait de transférer les données en mémoire de la RAM vers la carte graphique peut prendre du temps. Le temps de transition des données entre les deux n'est pas le problème, mais la recherche de ces informations dans la RAM prend, elle, plus de temps. C'est pourquoi, de nombreux développeurs de Jeux Vidéo utilise le Data-Oriented Design. Ce design pattern, permet d'agencer les données dans la RAM pour que ces dernières soient contiguës. Ainsi lorsque que le CPU et le GPU veulent chercher une information, ils vont rapatrier une plage autour de la donnée dont ils ont besoin (de la taille du cache). Si la donnée suivante se trouve dans cette plage, alors la recherche n'a pas besoin d'être ré-effectué. En faisant cela sur plusieurs cycles CPU, nous économiserions une grande quantité de temps. Pour réaliser cette technique, il faut privilégier le passage par copie des éléments lors de l'appel d'une fonction, éviter les pointers et le « pointer chasing », le fait d'avoir un pointer sur un pointer qui va nous faire aller chercher en mémoire des adresses fixes et pas forcément contiguës.

Il existe aussi des techniques d'optimisation propre à OpenGL. Comme évoquée ci-dessus, la technique du frustum culling, permet de n'afficher que les objets visibles

à l'écran. Pour ceci, on va diviser l'espace en zones équivalentes. Nous allons ensuite vérifier si ces zones sont visibles ou non par la caméra. Si elles le sont alors nous afficherons les objets qui appartiennent à ces zones. Mais la technique ne s'arrête pas là, il nous faut aussi savoir si les objets ne sont pas cachés par d'autres objets disposés devant eux. Nous pouvons appliquer une technique similaire pour la gestion des zones à charger en mémoire ou non. Les zones d'un Voronoï étant en triangle, nous pouvons savoir vers quelle arête se dirige le joueur. Ainsi en trouvant le centre du triangle, et en interpolant la position du joueur, nous sommes capables de définir des limites de chargements et déchargements des zones.

Les limites de chargements, déchargement sont des bandes dans lesquelles les deux zones sont chargées lorsque l'on sort

La zone de latence définit la zone où aucune map n'est chargée car pratiquement équidistantes de chaque zone.

Pour terminer, une des dernières grosses optimisations que nous pouvons apporter à OpenGL est instancing. Cette technique permet de réutiliser le « Mesh » (la structure d'un objet graphique). Ainsi avec une seule passe nous arrivons à dessiner à l'écran plusieurs objets pour le coup en mémoire d'un seul.

Calibrage / Adaptation au joueur

Gestion de la pression

D'après les recherches effectuées sur la médecine neurologique, avec un résumé explicatif dans la section "I, C" Cf. page 14. Nous proposons ci-dessous une manière d'exploiter une méthode pour contrôler au mieux les actions et réactions dues à un événement. Pour cela, nous devons considérer la réactivité de l'utilisateur et ainsi déclencher l'événement au moment le plus approprié.

Problématique : Comment pousser la limite des temps de réaction de l'utilisateur tout en lui assurant un contrôle ?

La difficulté ici est l'emploi d'une donnée (le temps de réaction) qui est très variable selon l'individu, l'âge, la concentration, ... sans oublier la variation des outils de captation.

Principe initial

Selon les études faites par Jean Bourgeois, nous pouvons déterminer le temps de réaction moyen en fonction du sexe et de l'âge d'un individu en passant par un simple calcul :

$$y = a \times \log_2 x + b \times \log_{10} [x+c]$$

En définissant x comme l'âge de l'utilisateur et des constantes comme définit dans le tableau suivant.

	a	b	c
Homme	0,640	-1,98	1,77
Femme	0,508	-1,78	1,64

Ainsi nous pouvons déterminer à partir de quand doit être effectuée une action. Cela laisse juste le temps de réaction nécessaire à l'utilisateur pour pouvoir réagir. Cependant les calculs ne nous permettent pas de personnaliser le temps de réaction. De plus, selon les machines utilisées et l'action du click nous estimons une variation de 25ms. C'est pourquoi nous proposons une boucle de correction d'erreur.

Correction d'erreur

Nous estimons le temps pris à l'utilisateur pour répondre à l'action et nous le stockons en mémoire pour en déterminer la valeur moyenne.

Avec l'ensemble des temps de réaction collectés, nous estimons un coefficient basé sur l'erreur type (ET) tel que :

$$ET = \sqrt{((\sum_{i=1}^n x_i)/n)/\sqrt{n}}$$

Avec x_i qui représente chacun des temps de réaction relevé et n est le nombre d'éléments de l'échantillon.

Soit TR= temps de réaction

$$\alpha = ([TR]_{\text{relevé}} - [TR]_{\text{attendu}}) / [TR]_{\text{attendu}}$$

Enfin nous pouvons fixer la prédiction de temps de réaction suivante (tmp) tel que :

$$tmp = [TR]_{\text{moy}} + ER \times \alpha$$

Etude de l'individu

Nous avons pu voir tout au long du chapitre deux comment nous envisagions un univers dynamique et évolutif en fonction des flux sonores. Cependant nous avons également pu voir dans le premier chapitre que les émotions ressenties au travers de

ce flux sonore, étaient très subjectives. Ainsi nous devons être en capacité de mesurer l'intensité des différentes émotions. Pour cela nous proposons plusieurs manières de procéder afin d'obtenir des informations rétroactives durant l'expérience : explicite et implicite.

Etude explicite

Parmi les études explicitées, nous avons pensé à mettre en place deux outils.

Dans un premier temps, l'idée serait de mettre en place un test sur la psychologie de l'utilisateur avant même que l'expérience ne commence. Avec un test tel que le "Myers Briggs type indicator" (ou MBTI), qui est le plus utilisé au monde, cela permettrait de catégoriser les utilisateurs. Ainsi en recoupant les informations recueillies sur une personne appartenant à une "cast X", nous serions en mesure d'avoir des indicateurs déjà prédéfinie, conceptualisant ainsi une source d'information de base pour un nouvel utilisateur appartenant à cette même "cast X". Bien entendu, cette base ne serait qu'un moyen de prédéfinir les options de configuration par défaut avant de pouvoir les adapter à l'utilisateur.

Dans un second temps, en partant d'une config par défaut, nous pourrions questionner l'utilisateur, au moyen d'une pop-up par exemple, sur son ressenti pendant l'expérience. Ainsi nous aurions à coup sûr une correction de l'estimation correcte. Cependant, cette seconde méthode est à utiliser avec délicatesse et parcimonie car elle a pour effet de couper l'utilisateur de l'expérience qu'il vit. Ainsi cela peut empêcher la pleine immersion de l'utilisateur.

Etude implicite

Enfin nous pouvons envisager une étude implicite, basé sur la compréhension des réactions comportementales de l'utilisateur vis-à-vis du monde dans lequel il se trouve. Cette méthode permettrait d'obtenir des informations plus floues que celles que nous pourrions obtenir par une méthode dite explicite. Cependant elle a le net avantage d'être totalement invisible pour l'utilisateur. Nous n'avons donc aucun risque de perte d'immersion.

Comme exemple d'implémentation nous pourrions imaginer l'utilisateur écoutant une musique avec autour de lui un certain nombre d'environnements différents se générant tout autour de lui. Nous pouvons enregistrer que selon le thème de la

musique courante, l'utilisateur s'est dirigé plus vers tel ou tel environnement, plutôt qu'un autre.

Exemple : faire demi-tour et retourné dans le désert aride ou bien avancé dans une forêt humide.

Conclusion

L'intérêt de ce mémoire était, pour nous, de faire un premier pas dans ce qui s'apparente à importer l'émotion d'un utilisateur au sein d'un univers virtuel. Pour cela nous avons décidé de capter les émotions à véhiculer en fonction d'un flux sonore.

Comme nous avons pu le voir dans le chapitre I, la principale difficulté de création d'un environnement réel en fonction d'un flux audio, est dans l'interprétation d'un signal non uniforme et additif. Les fréquences hautes et les fréquences basses n'ont pas la même échelle d'intensité, qui plus est nous ne pouvons pas séparer les instruments de musique et leur instrumental. C'est pourquoi notre solution est de prévoir des outils venant se poser bout à bout et fonctionnant un peu comme des circuits électroniques.

Pour ce qui est de la génération, nous pensons avoir proposé suffisamment d'outils pour en extraire des caractéristiques permettant de transcrire l'émotion perçue par l'utilisateur final en un élément graphique. Avec une bibliothèque d'éléments nous pouvons donc construire un environnement réaliste, et avec les outils d'analyse en temps réel, nous pouvons changer son état en fonction du temps. Ce point de départ nous a également amené à trouver des moyens de rétroaction afin de faire évoluer l'univers dans le bon sens selon l'utilisateur.

Ce mémoire a été l'occasion de comprendre les limites de la retranscription des émotions dans un univers virtuel : graphique et sonore. Un projet comme celui-ci est relativement conséquent du fait de ces possibles voies d'évolution et demande

beaucoup de précisions et peaufinage pour trouver la bonne méthode. Il mettrait du temps à mûrir. Effectivement, non seulement ce dernier demande un grand nombre d'objets graphiques, mais aussi beaucoup de techniques de réglage pour coller avec l'esprit que s'en fait l'utilisateur. Cela demande un grand nombre de phases de test qui pourront dégager de nouvelles problématiques mais également de nouvelles possibilités.

En effet, depuis quelques années, l'évolution des technologies émergentes dans le domaine de la réalité virtuelle et dans le domaine de la réalité augmentée pourrait apporter de nouvelles méthodes d'intégration complémentaires à notre solution. En outre, nous pourrions imaginer les possibilités de rétroaction qu'offrirait un oculus rift ou bien un myo...

Ainsi nous pourrions finir par nous demander : Quelle limite rencontrée ici, pourrait être repoussée par l'emploi de nouvelles technologies telles que l'oculus rift ?