

Son et Jeux vidéos

Comment faire du son un média interactif ?

L'axe de la qualité pure

« Tant que les capacités sonores des ordinateurs ressemblent plus à des téléphones, on ne pourra pas parler de bande son dans le jeu » entendait-on au début des années 90.

Mais après avoir résolu les problèmes de qualité purement audio en passant en une petite décennie de la qualité « téléphone » à une qualité proche du CD audio¹, force est de constater que le son dans la plupart des jeux reste souvent une superposition d'éléments sonores plus ou moins harmonieuse et aléatoire. Les coupures sonores intempestives, les boucles insupportables, les répétitions systématiques de son, les désynchronisations ou des décrochements importants entre le son et l'image, sont encore le quotidien des joueurs.

L'axe du rendu sonore en « temps réel »

Il a bien fallu se rendre à l'évidence, la qualité du son ne suffit pas pour créer une bande sonore dans un jeu comme on l'entend au cinéma. L'adéquation du son avec l'image est encore trop hasardeuse. Mais pour parvenir à suivre les choix des joueurs, les déplacements des personnages dans l'espace du jeu, le son ne peut être calculé à l'avance. Cette charge doit être du ressort de l'ordinateur lui-même. Cette modification du son en temps réel réclame une puissance de calcul importante au regard de ce que l'ordinateur doit déjà gérer par ailleurs. Qu'à cela ne tienne ! Comme il y a des cartes vidéo dédiées au calcul de l'image, les cartes sons équipées de processeurs dédiés (DSP) s'en chargeront !

¹ Sur PC et console de dernière génération, en fonction des format de codage des fichiers audio et à condition de brancher l'ordinateur sur des enceintes correctes !

Las, ces cartes ne se répandent pas si rapidement, et il n'existe pas de norme suffisamment installée pour encourager les développeurs à implémenter des technologies que par ailleurs ils ne maîtrisent pas². Par ailleurs le surcoût que cela pourrait occasionner n'est pas jugé rentable par la majorité des productions.

Cet axe est pourtant une évidence pour qui travaille sur le son des jeux vidéo mais l'usage de moteur audio 3D nécessite toujours des réglages complexes et des choix artistiques que l'ordinateur ne peut pas faire. On rencontre ce genre de problèmes par exemple lors des changements de position de caméra qui entraîne automatiquement des changements de rendu sonore des plus inaudibles (un champ/contre-champ par exemple qui fait passer un son de gauche à droite de façon très fatigante pour l'auditeur/joueur). Les règles qui rendent cohérent le *point de vue* et le *point d'écoute* dépendent d'un grand nombre de facteurs, pas toujours les mêmes en fonction de la scène, et qui sont très difficile à modéliser.

Dans un jeu vidéo, nous sommes souvent obligé de faire le constat suivant :

L'interactivité d'un jeu va à l'encontre de certaines lois et fonctions du son.

Ces lois quelles sont elles ?

Au cinéma, devenu parlant, le son a totalement modifié le montage des plans en introduisant la continuité. Des plans

² Par exemple la société Creative implémente sur ses cartes son la technologie EAX qui permet d'appliquer en temps réel des effets de déplacements et d'environnement sonore mais n'est disponible que sur PC équipés de cartes compatibles EAX. En outre elle est surtout efficace pour des jeux de type FPS (First Person Shoot) et demande, comme n'importe quel système audio 3D, beaucoup de réglages manuels pour être vraiment crédible.

discontinus se retrouvent, grâce au son, reliés entre eux dans le temps et aussi dans l'espace. Par exemple la même ambiance de foule donnera la cohérence d'ensemble au montage d'un concert filmé à des endroits et à des moments différents. Dans d'autres cas, le son relie deux plans en anticipant légèrement la scène à voir. Dans tous les cas le son « lisse », en quelque sorte, le montage image.

Mais dans un film, le réalisateur est totalement maître du rapport image/son et il peut créer l'effet qu'il souhaite en déclenchant les éléments sonores en parfait adéquation avec l'image. Dans un jeu en revanche le joueur se retrouve à la fois dans la position du **monteur**, déclenchant des événements sonores au moment qu'il décide et **mixeur**, par sa position dans l'espace du jeu. Les synchronisations ne sont déterminées que ponctuellement (une animation/un bruitage) pour le reste, les enchaînements sont imprévisibles.

Par conséquent comment le son peut-il encore assurer son rôle de « continuum spatio-temporel » dans un jeu où l'interactivité rompt constamment cette continuité ?

Comment faire du son un média interactif

Une approche serait de penser que le son n'est pas d'essence linéaire, nos oreilles doivent simplement s'habituer à une autre façon d'entendre le son. Le fait que nombre de joueurs ne satisfont d'une qualité très médiocre et d'un brouhaha sonore justifierait cette théorie. Pourtant quand certains jeux proposent un son de meilleure qualité, les joueurs le plébiscitent même si ça ne devient pas leur critère principal. Il paraît peu concevable que la pratique du son dans une narration linéaire devienne totalement caduque pour un jeu interactif. Le son reste un fil tendu dans le

temps quand l'image est un drap suspendu dans l'espace.

A partir de là, le champ est ouvert pour toutes les expérimentations qui respectent ces deux principes.

L'axe de la scénarisation et de l'intégration du son dans les jeux

Il existe une grande variété de jeux. Chaque typologie de jeu possède ses propres contraintes, ses propres codes d'interaction avec le joueur et requiert une nouvelle réflexion sur le design sonore. Pour autant tous ces jeux ont en commun d'associer du son et des images et ne dérogent pas à un certain nombre de règles qui ont depuis longtemps été énoncées par le cinéma. Et si nous prenons l'exemple d'un jeu vidéo d'aventure, sa forme s'inspire très fortement d'une narration linéaire classique.

Nous verrons à la fois ce qui les distingue mais comment nous pouvons tenter de les rapprocher en considérant que la qualité du son c'est d'abord la qualité de l'intégration.

L'intégration est par analogie, la phase de mixage, le moment où les ressources sonores sont placées et mises en valeur dans l'espace du jeu. Mais un mixage modulaire qui prendrait déjà en compte certains éléments de l'interactivité. C'est une étape cruciale qui n'est pas directement liée au moteur audio et aux capacités des ordinateurs mais avant tout aux questions posées par le design sonore : A quel moment la musique doit démarrer ? A quel moment l'arrêter ? Comment ne pas créer de ruptures sonores trop brutales ? Les bruitages doivent-ils jouer un rôle d'identifiant ? Comment les rendre moins répétitifs ? etc. A partir de là un moteur audio et une interface peuvent être conçus pour répondre au mieux à ces questions.

Musique

La musique au cinéma a essentiellement un rôle *empathique*. C'est à dire qu'elle a la propriété de faire ressentir des sentiments. C'est en quelque sorte un « catalyseur d'émotion » qui oriente la vision du spectateur vers une interprétation plus ou moins dramatique de la scène qu'il voit.

On peut aussi créer un décalage entre une musique "joyeuse" et des images "dures" pour produire un sentiment de malaise chez le spectateur/auditeur. Dans tous les cas la musique est présente à des moments précis de l'image pour renforcer l'émotion sans effets visuels supplémentaires.

Dans un jeu on utilise aussi cette fonction avec des jingles de victoire ou de game-over qui désignent les sentiments de réussite ou d'échec du joueur/personnage. Mais hormis ces cas particuliers, il est souvent difficile de synchroniser correctement la partie musicale avec l'action pour obtenir cet effet puisqu'un jeu n'a pas de durée fixe. En outre pour des raisons de simplicité technique, on a longtemps utilisé la musique dans les jeux de façon "géographique", comme *illustration sonore*. Tout effet d'empathie tombe alors à l'eau puisque la même musique peut être jouée dans le même décor à n'importe quel stade du scénario, selon que l'action est lente, rapide, mystérieuse ou angoissante.

Musique scénarisée

La musique *scénarisée*, quant à elle, essaie d'appliquer au jeu les mêmes recettes qu'au cinéma. Pendant une phase de recherche, on jouera une musique exprimant le mystère, à l'inverse dans une phase d'action violente, c'est une musique rythmée et tendue qui sera déclenchée. Cette façon de procéder en thèmes musicaux plutôt qu'en thèmes géographiques est beaucoup plus efficace car elle renforce la dramaturgie sans dialogues supplémentaires. Il est en outre

permis de ménager des phases de silence, sans musique, contrairement aux boucles musicales rapidement inaudibles des premiers jeux. Il n'est plus ainsi nécessaire de composer de très longs morceaux pour sonoriser tout le jeu sans qu'un ennui musical n'envahisse le joueur. Au lieu d'une heure de musique par exemple, vingt minutes suffiront avec pourtant un sentiment de lassitude bien moindre.

Cette amélioration « se paye » par contre en temps d'intégration puisque la musique est programmée en fonction de toutes les phases de jeu. Dans la plupart des jeux d'aventure où le scénario est « scripté » dans un fichier au format « texte » extérieur au programme, les commandes de son pour déclencher la musique doivent être intégrées à l'intérieur de ce script en suivant le déroulement de l'aventure. Mais cette tâche peut être confié au sound designer puisqu'il n'intervient pas dans le code du jeu. Le programmeur peut ainsi se consacrer uniquement sur le moteur du jeu en déléguant l'intégration du son au designer dont c'est le métier. Car l'intégration d'une musique scénarisée demande une certaine réflexion propre à un jeu interactif puisqu'il ne faut pas que la musique empêche le joueur d'agir en le prévenant de tous les dangers. Quand les violons stridents de Bernard Hermann dans *Psychose* annoncent la scène du meurtre, le spectateur n'a pas d'autre choix que d'assister à ce qui va suivre ; le joueur en revanche peut décider de continuer ou de rebrousser chemin, ce qui n'est pas forcément l'effet voulu par le *game-play*. Ce peut-être un écueil dans certains jeux. Dans les jeux d'aventure, dont la forme se rapproche du cinéma, la musique scénarisée est particulièrement adaptée.

Musique interactive

On restait précédemment dans une création musicale classique adaptée à la non-linéarité d'un jeu, un pas

supplémentaire est franchi avec **la musique interactive**. Cette fois-ci, c'est la musique elle-même qui évolue avec le jeu. On peut ainsi imaginer d'ajouter des instruments aigus, lors d'une phase de danger, de ralentir ou d'accélérer le tempo en fonction de l'urgence, changer la tonalité en fonction des personnages etc....

Paradoxalement les premiers jeux vidéo dans les années 80 intégraient déjà une sorte de musique interactive de par leurs limitations techniques. Comme la mémoire était chiche, il était plus économique de développer un mini synthétiseur pour générer les sons en temps réel plutôt que de jouer de la musique préenregistrée. On sépare ainsi le fichier musical (la partition en quelque sorte) des sons eux-mêmes. Il est ainsi très facile de modifier les propriétés du fichier « partition » en temps réel (tempo, notes...) comme ce fut le cas sur le jeu Space Invaders où le rapprochement des « envahisseurs » se traduit par l'accélération du tempo de la musique. Celle-ci avait beau être extrêmement pauvre musicalement, ce qu'on perdait en musicalité on le gagnait en plaisir de jeu. C'est d'ailleurs ce critère qui doit présider à tous choix de musiques et des techniques associées : son impact sur le *gameplay*.

Séparer la musique des sons

Le principe de la génération de la musique pilotée en temps réel par un fichier « partition » fut à l'origine de la norme MIDI (Musical Instrument Digital Interface). Malheureusement le MIDI n'a pas très bonne réputation dans le jeu car les sons dépendent du matériel du joueur. Ces 128 échantillons d'instruments de la norme de base (Général MIDI) limitent d'une part la composition et « sonnent » de façon très différente selon les ordinateurs (notamment ceux qui ne possèdent pas de carte son avec des tables d'échantillons musicaux préenregistrées mais uniquement un petit synthétiseur à modulation de fréquence à la sonorité catastrophique pour les

musiciens). Les fameux fichiers MOD qui eurent leurs heures de gloire sur PC et AMIGA comprenaient aussi les échantillons musicaux. On s'affranchissait d'un synthétiseur et on pouvait intégrer les sonorités choisies par le musicien mais cela réclamait plus de ressource de la part de l'ordinateur. Entre temps les disques durs et/ou les CD commençaient à offrir une place suffisante pour que les jeux utilisent de la musique « audionumérique » préenregistrée pour le plus grand bonheur des musiciens « traditionnels » et non plus seulement « informaticiens » !

Aujourd'hui, après avoir représenté une amélioration substantielle pour la musique de jeu, et un portage facilité entre PC et console, l'audionumérique est un frein pour la musique interactive car les fichiers se prêtent mal aux manipulations. Un nouveau format DLS (Download Sound) est apparu en reprenant le principe des fichiers MOD appliqué au MIDI. Ne fonctionnant que sur PC, il est encore très peu utilisé dans les jeux mais on peut imaginer par exemple que le joueur choisisse son style de musique parmi un certain nombre de fichiers DLS aux sonorités différentes.

Mais si la musique interactive représente un domaine riche certainement d'une nouvelle expérience sonore, on entre ici dans un domaine encore très peu défriché en raison de sa complexité technique mais aussi artistique. En effet l'interactivité, en multipliant les combinaisons musicales peut vite dégénérer en bouilli inaudible. Comment faire en sorte d'établir des règles pour que la musique ressemble encore à une musique telle qu'un musicien peut l'imaginer dans son studio. Il semble en outre que cette technologie soit surtout adaptée à un certain type de musique électronique et moins à un orchestre symphonique. Encore une fois, tout est question d'usage, de typologie de jeu et de gamedesign. Il faut toujours se poser les bonnes questions comme : « Quel rôle doit jouer la musique dans le gameplay ? Ne risque-t-elle pas de sur-accompagner l'image, renforçant chaque action, chaque élément visuel à l'instar de

certaines cartoons d'après guerre où chaque battement de paupière était

accompagné par une rangée de violons !?"

Voix

Au cinéma déjà, depuis le parlant, l'histoire passe avant tout par la voix ; soit par la « **voix off** », hors de l'action du film, soit par les dialogues des personnages. En général la synchronisation est respectée par le fait même de filmer des comédiens en train de parler et **la fusion audiovisuelle** est assurée par la notion de **point de vue** qui concorde plus ou moins avec le **point d'écoute**³. Ainsi un personnage vu de loin dans une église, aura une voix atténuée avec de la réverbération. Dans certains cas toutefois, on crée un effet particulier si le point d'écoute est beaucoup plus proche que le point de vue (rarement l'inverse) et on entend un personnage lointain comme s'il était à quelques mètres. Ce dispositif est en général voulu par la mise en scène et n'est pas systématique.

Dans un jeu vidéo, au contraire, on fait plus difficilement concorder point de vue et point d'écoute. Avec l'apparition des jeux en vue subjective et en 3D, le problème est devenu encore plus crucial d'autant que l'on peut passer d'un point de vue à un autre dans la même partie. Si on désire que le joueur s'identifie complètement au héros il peut être plus efficace de traiter la voix du héros/joueur comme une voix off, c'est à dire comme **une réflexion introspective**, avec un effet de proximité. Si au contraire, on souhaite que le joueur prenne le rôle du demiurge qui manipule un personnage (on privilégie la vue à la 3^{ème} personne), il vaut mieux considérer la voix comme faisant partie des dialogues "IN" avec un traitement sur le timbre et l'enveloppe adéquat à l'espace visuel. Dans ce dernier

cas, il n'est pas toujours facile d'assurer un rapport étroit entre le son et l'image dans le cas d'un jeu en 3D temps réel où le point de vue change constamment sauf à utiliser des technologies de modification de l'enveloppe du son en temps réel⁴. Beaucoup de jeux, éludent le problème en gérant le son de tous les dialogues comme des voix off, indépendamment de l'environnement visuel. Les joueurs s'y sont habitués par défaut et c'est dommage car il est beaucoup plus efficace pour l'immersion de traiter les voix en fonction du point de vue quitte à faire quelques compromis entre les contraintes techniques et la justesse du point d'écoute.

Pour cela il faut déterminer précisément l'emplacement visuel de chaque dialogue et appliquer des filtres et de des effets de réverbération. C'est un travail délicat mais pas plus fastidieux qu'un mixage de dialogues au cinéma et cela contribue à la crédibilité du jeu des acteurs.

³ Même s'il n'y a pas de son direct, ce principe est aussi respecté dans les films de synthèse ou le dessin animé.

⁴ Technologie du Type EAX de la société Creative Labs qui fonctionne uniquement sur PC avec une carte son compatible.

Bruitages

On peut distinguer 2 types de bruitages : **Les bruitages d'ambiance** qui "habillent les décors" et **les bruitages d'entités ou d'évènements** qui sonorisent l'action. A cela, dans un jeu, on peut rajouter une troisième catégorie qui est **les bruitages d'interface/menu**.

Les frontières entre ces 3 catégories ne sont pas forcément très étanches mais dans un jeu elles aident à la conception du moteur audio.

Ambiances décors

Au cinéma, les ambiances sont constituées de ce qu'on appelle le « *champ* » qui distingue les sons dont la source est présente à l'écran et le « *hors champ* » qui prend en compte les sons présents dans l'action (sons *diégétiques*) mais qu'on ne voit pas à l'écran. Avec le *multicanal* (Dolby ou DTS), le hors champ a pris une place importante sans pour autant apporter beaucoup plus d'information au spectateur ; l'immersion est simplement meilleure. Dans un jeu en revanche, la notion de hors champ est beaucoup plus floue puisque la plupart des sons peuvent se trouver tour à tour, hors et à l'intérieur du champ, en fonction des déplacements du joueur. En fait, on pousse la notion de ***hors-champ*** à l'extrême puisque que ce n'est plus seulement ce qui est hors du cadre (le cadre n'a plus beaucoup de sens dans un jeu en 3D) mais ce qu'on l'on serait sensé voir si on avait pu le réaliser. Cela sert beaucoup dans un jeu pour simplifier l'image de synthèse. Par exemple, il arrive que l'on entende la pluie sans la voir réellement à l'image ou plus souvent des bruits de foule dans un endroit plutôt désert (les animations de personnages coûtent cher !). Ce principe de faire dire au son ce que l'image n'est pas en mesure de faire, permet une belle économie de moyen mais pose quand même un certain nombre de problèmes

de perception. A force de faire prendre au joueur des vessies pour des lanternes, il en arrive à ne plus faire fusionner ce qu'il voit avec ce qu'il entend et qu'il "décroche". Il n'est donc pas forcément évident qu'il faille systématiquement avoir des ambiances sonores beaucoup plus riches que les décors. Par contre il est important que ces ambiances soient variées, que l'on n'entende pas la répétition d'une boucle (risque encore de décrochage).

Un principe technique efficace, car ne nécessitant pas de ressources énormes de la part de l'ordinateur, est de gérer une ambiance comme une superposition d'événements sonores courts (du vent, un animal, une voiture qui passe...) joués aléatoirement à des intervalles irréguliers. Encore une fois, on déporte le travail sonore du mixage pré calculé, comme c'est le cas au cinéma, vers l'intégration dans le programme, pour un résultat bien supérieur.

Bruitages d'entités

Depuis longtemps, au cinéma on utilise le bruitage de façon créative pour renforcer l'impact visuel. Les coups de poing, les portes qui grincent (toujours !), les moteurs des vaisseaux dans le vide de l'espace, par nature silencieux, en sont quelques-uns des exemples courants. Le spectateur a ainsi acquis une culture du son au cinéma qui ne correspond en rien à la réalité mais qui « sonne » plus vraisemblable. Mis à part quelques sons très caractéristiques, la plupart ne sont pas reconnaissables sans l'image qui leur est associée d'où la difficulté d'utiliser des bruitages *acousmatiques* dont la seule écoute permettrait de visualiser parfaitement la source qui le produit. Si la source est facilement reconnaissable le son peut informer par ailleurs sur sa taille (si l'objet est gros ou petit, par la fréquence

du son), de sa matière (si c'est du métal, du bois, de la pierre, par son timbre...).

Dans les jeux vidéos cette difficulté d'une **écoute purement causale**, qui permet d'utiliser le son pour s'informer sur sa cause se retrouve accentuée par la plus faible qualité des sons et du système de diffusion ainsi que par l'impossibilité de modéliser en image tous ce que l'on veut entendre. D'où une utilisation (surtout dans les premiers jeux) intensive des sons « expressifs » qui privilégient une **écoute sémantique**, par référence à un langage ou un code d'interprétation comme les sons de bonus ou les onomatopées (cris) par exemple. Ainsi dans les premiers jeux, les bruitages étaient cantonnés à un rôle « signalétique » ou de « rétroaction » : « J'ai touché une brique », « J'ai reçu une balle », ou tout simplement « j'ai bien appuyé sur le bouton »...

Petit à petit la culture du jeu a, à son tour, créé un ensemble de stéréotypes sonores communément admis dont il n'est pas toujours facile par habitude ou par facilité.

Pourtant l'amélioration de la qualité sonore des ordinateurs aujourd'hui permet d'envisager une utilisation des sons « comme au cinéma ». Pourquoi dès lors ne pas utiliser des sons beaucoup plus riches d'information et se rapprocher d'une écoute causale. Pour certains jeux à l'interface déjà très chargée visuellement comme les jeux de sport ou de stratégie on a tout intérêt à privilégier l'information sonore dans l'interface. Il est plus judicieux d'utiliser des sons très riches

qui renseignent d'avantage sur sa cause plutôt que des « bips ».

Mais le problème reste encore celui de la redondance. En améliorant la qualité du son et donc la fidélité à la source, on passe quasi-obligatoirement par des échantillons enregistrés dont le nombre ne peut pas être infini et les possibilités de modification en temps réel sont réduites. Les sons qui privilégient une écoute sémantique supportent mieux la redondance car ils agissent plus comme de simples signaux. Il faut donc bien déterminer si le bruitage sera fréquemment répété avant de choisir son type de sonorité.

Au final de la même façon que pour les ambiances, la qualité des bruitages se fait par la qualité de l'intégration. Plus le sound designer peut confronter sa création sonore à la création du jeu, meilleur sera le résultat. Il n'est pas rare qu'un son très moyen seul, prenne toute sa texture et son grain mis en relation avec l'image. D'où l'importance particulière pour le bruitage de faire concorder étroitement point d'écoute et point de vue à l'aide d'une gestion par un moteur audio 3D réglé par le bruiteur lui-même.

En outre, dans un jeu, la gestion des bruitages **hors champ** peut revêtir un intérêt pour le *game-play* en signalant la présence d'ennemis avant d'être vu ou plus simplement pour se repérer dans un décor 3D complexe.

En conclusion

Il est tout à fait possible d'établir des parallèles entre le rapport image/son au cinéma et dans un jeu vidéo, à fortiori dans un jeu d'aventure. Le jeu doit même essayer de s'inspirer de certains principes dans la relation entre l'image et le son qui font l'efficacité narrative d'un montage traditionnel, pour les adapter à un **montage non linéaire**.

Pour cela, il est nécessaire que le designer sonore soit aussi l'artisan de la conception technique du moteur audio et de l'intégration des sons dans le jeu.