



HAL
open science

La synthèse d'image sur les sentiers du sonore

Pauline Gravel

► **To cite this version:**

Pauline Gravel. La synthèse d'image sur les sentiers du sonore. *Resonance*, 1995, 9, pp.1-1. hal-01156748

HAL Id: hal-01156748

<https://hal.science/hal-01156748>

Submitted on 27 May 2015

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.

La synthèse d'image sur les sentiers du sonore

Pauline Gravel

Résonance n° 9, octobre 1995

Copyright © Ircam - Centre Georges-Pompidou 1995

A la fin des années 50 apparaissaient les premiers sons synthétisés par ordinateur. Une décennie plus tard émergeait la synthèse des images, qui évolua très rapidement, jusqu'à dépasser parfois son aînée. De la recherche industrielle à la médecine, l'architecture ou le cinéma, l'image de synthèse a depuis lors trouvé une multitude d'applications.

Outil de création, l'ordinateur offre désormais des possibilités d'une grande subtilité. Or, qu'elles visent à synthétiser du son ou de l'image, les stratégies retenues en la matière restent très proches. Cette similitude est due à la fois à l'outil informatique lui-même, à la structure des matériaux sonore et visuel, et aux processus de perception.

Tout comme les sons, les lumières visibles qui concourent à l'élaboration d'une image rétinienne sont des phénomènes vibratoires. De même que les fréquences vibratoires qui composent un son déterminent sa hauteur et son timbre, le contenu fréquentiel d'une lumière réfléchi par un objet définit une couleur. En matière visuelle, les longueurs d'onde des lumières visibles, correspondant à des fréquences de l'ordre de 10^{15} hertz, sont très supérieures à celles des sons audibles (10^3 hertz).

Le son et l'image se distinguent également au niveau de leurs manifestations temporelle et spatiale. Bien qu'ils émanent de sources réparties dans l'espace et que leur rayonnement comporte une dimension spatiale, les sons résultent fondamentalement de l'évolution au cours du temps des fréquences et des intensités des ondes qui les constituent. C'est donc la dimension temporelle qui prime ici. La perception visuelle est quant à elle immédiate, l'illumination et la couleur qui composent une image étant instantanément visibles et compréhensibles, même si l'animation des images introduit une notion de durée. Essentielle, cette distinction s'inscrira donc dans la représentation numérique du matériau. Ainsi, tandis que la synthèse sonore implique la génération d'une myriade d'échantillons, la synthèse d'image procède plutôt à l'élaboration de milliers de points minuscules (les « pixels », contraction de *picture element*), dont l'intensité lumineuse et la couleur sont codées numériquement. Lorsque les couleurs et les intensités de ces points obéissent à une certaine organisation, l'oeil ne les perçoit plus séparément, mais comme une image globale cohérente.

Dispositifs et modèles

Les systèmes de synthèse d'image comprennent généralement trois parties principales : un dispositif d'entrée et de mémorisation des données, une unité informatique qui traite ces données et une console de visualisation, dotée généralement de tubes à rayons cathodiques. Un canon émet soixante fois par seconde un faisceau d'électrons en direction d'un point précis de l'écran recouvert de luminophores. En bombardant l'écran, les électrons excitent les luminophores, qui émettent alors une lumière visible. La couleur de chaque pixel est quant à elle synthétisée « additivement », par un mélange approprié des trois couleurs primaires (rouge, vert et bleu), auxquelles les cellules réceptrices de la rétine sont particulièrement sensibles. Ces trois couleurs se combinent en chaque point de l'écran. Les systèmes de synthèse d'image actuellement disponibles offrent ainsi à l'infographiste une palette de seize millions de couleurs.

L'emploi de l'ordinateur comme outil de création exige l'élaboration d'un modèle qui, à l'aide d'équations mathématiques ou d'algorithmes, définit la structure d'un objet donné. Cette procédure permet de simplifier les manipulations. Les objets qui composent une scène sont d'abord modélisés au moyen de

figures géométriques élémentaires (point, ligne, rectangle, sphère, etc.), ce qui permet de les représenter sans qu'il soit nécessaire de définir chacun des points qui les composent.

Les modèles géométriques procédant par assemblage de formes simples (construction modulaire) s'organisent souvent selon une structure hiérarchique. Celle-ci permet de formaliser les relations existant entre les différents éléments constituant un objet ou une scène. L'élaboration d'un squelette humain s'effectuera par exemple selon l'organisation hiérarchique qui lui est propre, membre après membre : tête, corps, jambes et bras, ces derniers divisés chacun en plusieurs parties, dont une main comprenant elle-même des doigts formés de phalanges, etc. Les modèles géométriques s'avèrent cependant impuissants à construire certains objets complexes et irréguliers, tels que les montagnes ou les nuages. Des algorithmes de génération d'objets fractals intégrant les notions d'homothétie interne et de hasard énoncées par Benoît Mandelbrot constituent alors une alternative fort prisée des infographistes.

Les formes qu'adoptent certains objets (par exemple, les plis d'un tissu drapant un meuble) sont également délicats à modéliser géométriquement. La synthèse de tels objets nécessite le recours à des modèles physiques, qui définissent les principales propriétés physiques de l'objet et de l'environnement dans lequel il se situe. Un modèle physique de tissu tiendra compte ainsi de la friction avec la surface de l'objet qu'il recouvre, ainsi que des tensions s'exerçant entre les fils du tissage. Les modèles physiques permettent de générer des images d'un réalisme saisissant et constituent pour cette raison un champ de recherche en pleine expansion. Il en va de même en synthèse sonore, où l'on s'applique de plus en plus à développer des modèles physiques qui décrivent le fonctionnement d'un instrument à l'aide des lois de la mécanique et de l'acoustique.

En trois dimensions

Incluse dans plusieurs modèles, la définition du caractère tridimensionnel des objets offre de nombreuses possibilités. Elle permet notamment de réaliser des transformations géométriques (rotations, translations, changements d'échelle, etc.) et de faire évoluer un objet dans l'espace pour en révéler successivement toutes les facettes.

La représentation des objets en trois dimensions pose toutefois un problème crucial, du fait de l'inadéquation existant entre la troisième dimension et les deux dimensions de la surface de visualisation. La synthèse sonore est elle aussi confrontée à un problème similaire lors de l'enregistrement, qui ne permet pas la restitution du rayonnement tridimensionnel des sons dans l'espace, malgré l'effet de relief acoustique qu'apporte la stéréophonie. En infographie, ce problème se résout au moyen de la projection. Le passage d'un espace tridimensionnel à une surface plane s'effectue communément par une projection en perspective. Les règles de cette projection s'expriment notamment par des effets de lointain sur la coloration (due à l'absorption de la lumière), de mouvement (les objets éloignés se déplacent moins rapidement que les objets rapprochés), ainsi que par des réductions (ce qui est loin apparaît plus petit que ce qui est proche). En synthèse sonore, on parvient de même à simuler certains effets de profondeur en ajoutant des réverbérations au son synthétisé, ou par l'intermédiaire d'un spatialisateur, qui permet de contrôler la localisation de sources sonores ainsi que la projection des sons dans un espace réel ou virtuel.

Les modèles géométriques permettent d'élaborer essentiellement des images en « fil de fer », où n'apparaissent que les lignes de construction représentant les caractéristiques topologiques des objets. Ce type de représentation « squelettique », où les surfaces sont symbolisées par un ensemble de courbes, ne permet qu'une vision très schématique de la réalité, qui suffit toutefois aux besoins industriels de la conception assistée par ordinateur. De nombreuses applications, notamment artistiques, exigent cependant une simulation plus raffinée. Pour accroître le réalisme d'une image tridimensionnelle en mode filaire, il est nécessaire de lui adjoindre d'autres modèles destinés à simuler l'apparence visuelle des objets.

Effets d'optique

De même que les modèles géométriques, la plupart des modèles décrivant les attributs des objets (tels que leur éclairage et leur texture) s'inspirent des mécanismes de la perception visuelle. Pour améliorer le rendu de l'image 3D, une première démarche consiste à éliminer les parties « cachées », c'est-à-dire invisibles à l'observateur. En effet, nous ne percevons la profondeur de notre environnement que parce que l'information qui nous en parvient est partielle. Un objet apparaît par exemple derrière un autre, parce que celui-ci le cache en partie. Deux types de stratégies permettent de déterminer quelles portions des objets d'une scène sont effectivement visibles du point de vue de l'observateur. Une première méthode consiste à comparer les objets entre eux, afin de circonscrire les surfaces qui ne sont pas obstruées par d'autres, puis à éliminer finalement celles qui ne sont pas visibles. La seconde méthode vise plutôt à repérer les objets les plus proches de l'oeil de l'observateur. Une technique, qui découle de cette approche, consiste à lancer des rayons de lumière imaginaires à partir de l'oeil de l'observateur en direction des objets de la scène à synthétiser, un rayon étant émis pour chaque pixel de l'écran. Les objets ou les portions d'objet que ces rayons rencontrent en premier représentent les surfaces visibles ; ceux qui ne sont pas coupés par un rayon donné sont éliminés.

La conquête du réalisme passe également par le coloriage des surfaces. Le rendu de couleur que revêt chaque surface dépend en effet d'une multitude de facteurs : des propriétés intrinsèques de la surface en jeu (sa couleur, sa texture, sa réflectance), de l'éclairage ambiant, ainsi que de sa position et de son orientation par rapport aux sources lumineuses, à l'observateur et aux autres surfaces qui l'entourent. L'éclairage joue donc un rôle déterminant, ne serait-ce que par les indices qu'il procure sur le relief et la perspective de la scène (ombres de reflets et de transparences).

Les modèles d'éclairage existants reposent sur les règles de l'optique géométrique et les lois de propagation des ondes lumineuses dans les différents milieux (transparents, translucides ou opaques). Ces modèles prédisent essentiellement la quantité de lumière émise, transmise ou réfléchi dans la direction de l'oeil, en chaque point de l'image.

Ombre et lumière

En premier lieu, l'éclairage comporte une composante diffuse : l'éclairage ambiant. Celui-ci est constant en tout point de l'espace, et fait apparaître les surfaces uniformes, quelles que soient leur position et leur orientation. Dans la réalité, des sources ponctuelles et directionnelles interviennent également, dont les effets varient en fonction de l'orientation des surfaces qu'elles atteignent. L'éclairage d'une surface dépend en effet de l'angle d'incidence des rayons lumineux. Ainsi, l'éclairage est maximal pour une source située à l'aplomb de la surface, tandis qu'il est nul sous un éclairage rasant. En outre, la réflexion d'une lumière par une surface n'est pas uniforme. Elle est nettement plus intense dans la direction symétrique du rayon incident. Lorsque l'oeil de l'observateur se situe exactement sur le chemin de cette réflexion (dite « spéculaire »), la couleur qu'il perçoit est celle de la source lumineuse plutôt que celle de l'objet. Par exemple, à l'endroit où s'effectue la réflexion spéculaire sur une pomme, celle-ci n'apparaît pas rouge mais blanche, comme la couleur de la lumière incidente. D'autre part, l'image vue à travers un objet translucide est distordue en raison du phénomène de réfraction, et sa couleur est parfois même modifiée.

La modélisation d'un éclairage réaliste implique évidemment la reproduction des ombres, qui apportent des informations additionnelles sur la profondeur. Tandis que les sources lumineuses ponctuelles produisent des ombres bien définies, les sources plus étendues engendrent des zones plus floues d'ombres et de pénombres. De même, la modélisation de l'éclairage devient d'autant plus complexe que le nombre de sources lumineuses s'accroît. Les éclairages indirects et les multiples rebondissements des rayons lumineux entre les différents objets d'une scène présentent également des situations dont la modélisation est toujours en exploration.

La texture des surfaces représente un autre aspect primordial dans la synthèse de scènes réalistes. La texture désigne autant la microstructure d'une surface, que les motifs bigarrés qui la colorent (grain du bois, veinures du marbre, etc.). Pour résoudre cet épineux problème, les infographistes utilisent souvent

une image de texture réelle numérisée, que, dans un second temps, ils collent sur l'objet. Cette méthodologie d'échantillonnage est également utilisée en synthèse sonore, lorsque l'objet à modéliser est trop complexe. Des échantillons d'instruments naturels sont alors prélevés afin d'éviter leur synthèse intégrale.

Vers l'animation

Les applications de synthèse d'images animées (cinéma, réalité virtuelle) exigent naturellement une procédure supplémentaire, puisqu'il s'agit alors de faire varier un phénomène dans le temps. Tout changement ayant un effet visuel peut devenir objet d'animation : l'évolution de la position d'un objet, de l'observateur ou d'une source lumineuse, aussi bien que la modification de la forme, de la couleur, de la transparence ou de la texture.

Ce genre d'effets est obtenu par la technique de l'animation par « positions-clés » (*keyframing*), qui consiste à spécifier, à divers instants, certaines positions stratégiques et valeurs extrêmes d'attributs (couleur, texture et taille). A partir de ces images-clés, l'ordinateur calcule les positions et les valeurs intermédiaires, créant ainsi une animation régulière. L'interpolation peut reposer à la fois sur des modèles géométriques ou sur des modèles physiques, ou encore sur les deux à la fois. Des modèles s'inspirant des principes de la robotique permettent par exemple de simuler de façon très réaliste les mouvements articulatoires d'un personnage élaboré selon une structure hiérarchique. Grâce à ce type de modèle dit de « cinématique inverse », l'ordinateur interpole automatiquement les orientations et positions intermédiaires de toutes les articulations déplacées lors du repositionnement d'une extrémité du squelette du personnage.

Des animations particulièrement réalistes peuvent par ailleurs être réalisées en attribuant des propriétés physiques aux objets et en faisant intervenir certaines forces extérieures, telles que la force gravitationnelle, la poussée du vent ou la friction du sol. Dans ce type d'animation (appelée « dynamique inverse »), il est parfois extrêmement difficile d'identifier les forces qui sont en jeu, de sorte qu'il est souvent moins fastidieux de procéder par la cinématique inverse, en dessinant les positions-clés à l'aide de modèles géométriques.

Très récente, l'approche issue de la technologie du virtuel permet désormais de produire des animations en temps-réel d'un grand réalisme, basée sur des mouvements captés sur des personnes réels. Cette approche fait appel à des périphériques connectés à l'ordinateur, tels que des capteurs de position installés sur le corps d'une personne en mouvement. Les signaux émis par ces périphériques sont traduits en informations numériques, que l'ordinateur utilise pour générer l'animation. Des changements de position, d'orientation ou des modifications de rendu des objets ou des personnages synthétisés s'affichent alors à l'écran en temps-réel. De nombreux musiciens ont également recours à cette technologie du virtuel dans leurs compositions. Ils utilisent généralement des dispositifs mesurant différents aspects de l'interprétation de l'oeuvre musicale, afin de commander en temps-réel la génération d'événements sonores synthétisés. Cette approche se distingue par le fait qu'elle échantillonne non plus le signal lui-même, c'est-à-dire le son ou l'image, mais un paramètre de contrôle du modèle de synthèse.

De multiples applications

Les débuts de la synthèse d'image ont coïncidé avec la mise au point des premiers simulateurs de vol. Il s'agissait alors de créer un environnement virtuel indispensable à l'entraînement des pilotes. Depuis, ses applications se sont multipliées. Par le biais d'un attirail technologique sophistiqué, la réalité virtuelle permet une immersion dans l'image. Grâce au visiocasque, qui dispense une visualisation stéréoscopique des images de synthèse en cohérence avec la position de l'opérateur, il est désormais possible de pénétrer dans l'image virtuelle et de s'y déplacer. Un gant de données permet également d'interagir avec les objets virtuels de la scène.

De telles possibilités s'avèrent extrêmement utiles dans un domaine d'application telle que l'architecture, la

synthèse d'images permettant ici de représenter l'allure d'un bâtiment, de le situer dans le paysage environnant et de simuler une promenade à l'intérieur et à l'extérieur du bâtiment.

Certaines parties du corps humain dont la visibilité et l'accessibilité sont fortement réduites peuvent également être reconstituées sous forme d'images de synthèse tridimensionnelles. De telles images permettent alors au chirurgien de simuler une intervention avant de l'exécuter. De même, la synthèse d'image permet également aux chimistes de modéliser des structures moléculaires complexes. Elle a en outre donné lieu à des applications particulièrement remarquées au cinéma et en publicité, soit en introduisant un objet ou un personnage synthétique dans un décor réel, soit en incrustant à l'inverse un personnage réel dans un décor synthétisé.

Le morphing

Le procédé dit d'interpolation de formes (ou *morphing*) permet de transformer continûment une image en une autre. Il ne consiste pas en un «mixage» qui donnerait un fondu-enchaîné, mais en une déformation structurelle des objets constituant les images. Pour ce faire, on procède d'abord à la numérisation des images choisies comme point de départ et d'arrivée. L'infographiste réalise ensuite une analyse de ces images en délimitant, à l'aide d'une grille de contrôle, les parties qui seront déformées. Par exemple, lors de la métamorphose d'un visage en un autre, l'analyse consiste à caractériser la position, la forme, la texture et le contour des principaux éléments constitutifs de chacun des visages. Un modèle géométrique de l'image est ainsi défini et peut alors être transformé de multiples façons. On établit ensuite une correspondance entre certains objets de l'image de départ et des objets de l'image d'arrivée. L'ordinateur calcule ensuite les images intermédiaires, en interpolant les formes et les attributs des objets mis en correspondance. L'illusion de la métamorphose est enfin obtenue par l'enchaînement de toutes ces images, en utilisant un procédé d'animation classique. Une démarche similaire est adoptée en synthèse sonore pour transformer un timbre sonore en un autre.