IRCAM Centre Georges Pompidou

Cours de traitement des signaux audio

Notes d'un cours de 20h du mastère Multimédia et Hypermédia de Telecom Paris et l'École Nationale des beaux arts de Paris introduisant le B-A-BA des techniques de traitement du signal audio,

Sommaire

1. Introduction
    1. Introduction à l'acoustique
    2. Représentations temps-fréquence
    3. Numérisation
    4. Spatialisation
2. Perception des sons
    1. Introduction à la psychoacoustique
    2. Perception des niveaux sonores
    3. Perception de la hauteur tonale
    4. Perception de l'espace
    5. Perception du timbre
3. Analyse/Synthèse
    1. Modélisation
    2. Effets
    3. Synthèse
    4. Contrôle

Commentaires

J'ai écrit ce cours directement en HTML. Si cependant vous préférez lire ce document sur une version papier de 50 pages, vous pouvez consulter la transcription postscript (269KO) de toutes ces pages HTML (ou bien la version postscript compressée de 63KO).

Le cours est écrit en français. Si vous êtes intéressé pour m'aider à le traduire, n'hésitez pas à me contacter.

N'hésitez pas non plus à me faire part de vos remarques et commentaire. Je tacherais d'en tenir compte dans la prochaine version. Vous pouvez également contribuer à l'écriture d'un nouveau chapitre dans le document.

Le site officiel de cette page est :
http://www.ircam.fr/equipes/analyse-synthese/tassart/doc/beauxarts/index.fr.html.

Remerciements

Je tiens à remercier tout particulièrement C. Pottier, O. Cappé et D. Matignon qui m'ont fait confiance pour l'organisation de ce cours. Je remercie également M. Wanderley et P. Depalle pour les conseils et les idées qu'ils m'ont transmis tout au long de ce travail, et enfin S. Rossignol et R. Tassart pour leurs efforts de relecture.

Page remise à jour le
Jeu 2 Avr 1998 19:36:35

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

IRCAM Centre Georges Pompidou

Introduction à l'acoustique

Présentation

Qu'est-ce que l'*Acoustique* :

- c'est l'étude scientifique des sons,
- les champs d'application sont:
  - la production sonore,
  - la transmission des sons,
  - la réception et la perception des sons.

Dans l'arborescence des sciences, c'est une sous-branche de la *Mécanique*, puis de la *Mécanique des Vibrations* (production sonore) et de la *Mécanique Ondulatoire* (transmission des sons).

L'acoustique admet de nombreuses ramifications (voir transparent)

Page remise à jour le
Jeu 6 Nov 1997 16:53:32

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Introduction à l'acoustique

Survol historique

L'*Acoustique* a 2300 ans d'histoire :

- L'hypothèse que le son soit une onde émise par le mouvement d'un corps puis transmise par un mouvement de l'air remonte aux Grecs (**Chrysippe** 240BC., **Aristote** 384-322BC.).
- **Pythagore** aurait été le premier à étudier les sons musicaux (550BC.). Il remarque que deux cordes à l'octave ont leur longueur dans un rapport double :
- Toutes ces notions apparaissent sous des formes différentes chez :
  - **Vitruve**, architecte et ingénieur romain (25BC),
  - **Boethius**, philosophe romain (480-524).

  Après, il faut attendre le XVI^{ème} siècle (Renaissance).
- **Galilée** (1564-1624) : en 1638, étude de la vibration des corps, notions de résonance, et de vibration symphatique induite. Relation entre hauteur du son / longueur de la corde vibrante et nombre de vibrations par seconde.
- **Mersenne** (1588-1648) : moine au Mans, tenu pour le père de l'acoustique donne la loi des cordes vibrantes (*f* est inversement proportionnelle à la longueur de la corde). Première détermination absolue de la fréquence d'un son (1625).
- **Boyle** (1660) montre qu'il faut de l'air pour que le son se propage (expérience de la cloche).
- **Newton** (1642-1727) donne la première tentative de calcul de la vitesse du son. Il se trompe (mouvement non isotherme), mais donne le début de la formalisation mathématique des phénomènes sonores (Principia 1686).
- **C. Huygens** (vers 1690) fait une synthèse des connaissances de l'époque sur les phénomènes sonores.
- Le XVIII^{ème} siècle est très riche pour le développement de l'acoustique. **D'Alembert** (1717-1783), **Euler** (1707-1783) et **Lagrange** (1736-1813) établissent le formalisme définitif en développant la notion de dérivée partielle (d'Alembert, 1747) puis en jetant les bases de la mécanique analytique (Lagrange, 1759).
- À partir de cette époque, le formalisme est établi, le reste n'est que raffinement. **Helmholtz** (1821-1894) expérimente et développe la théorie de l'audition.
- **Fourier** (1768-1830) : décomposition des fonctions périodiques de la théorie de l'audition.
- **Rayleigh** (1824-1919) : oeuvre considérable en théorie de l'acoustique, publie en 1877 un ouvrage qui demeure un ouvrage de base de l'acoustique.

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Introduction à l'acoustique

Le son, c'est quoi ?

Caractéristiques mécaniques

C'est une modification des caractéristiques mécaniques du milieu de propagation.

En pratique, il s'agit des micro-variations de la pression de l'air ambiant.

Remarques:

Il faut un milieu de propagation pour que le son se propage :

- l'air pour ce que nous connaissons,
- l'eau peut également transmettre les sons, on parle d'*Acoustique Sous-Marine*,
- la terre peut également transmettre les sons, on parle alors de *Sismologie*,
- les matériaux solides, comme le métal, peuvent également transmettre des sons,
- dans le vide, les sons ne peuvent pas se propager.

La lumière (qui n'est pas une onde acoustique, mais une onde électromagnétique) est un rare exemple d'onde ne nécessitant pas la présence d'un support matériel pour se propager. On a cru jusqu'à la fin du XIX^{ème} siècle, que le support de propagation de la lumière était une substance inconnue baptisée l'*éther*.

Propagation

Les perturbations ont tendance à se transmettre de proche en proche. Le déplacement des perturbations donne lieu à une *onde acoustique*.

Exemples :

- la densité des voitures dans le trafic : des espaces vides de voitures peuvent se déplacer dans le sens ou dans le sens opposé du trafic,
- l'élévation locale du niveau de la mer donne lieu aux vagues et à la houle.

Il n'y a pas de lien entre la vitesse de propagation d'une onde, la *célérité*, et la vitesse des particules de matière.

- Les vagues ne se propagent pas forcément dans le sens du courant marin,
- le trou de voiture dans le trafic peut se déplacer dans le sens contraire des voitures.

La vitesse du son dans l'air est de l'ordre de 340 mètres par seconde, alors que le mouvement des particules est de l'ordre de quelques centimètres par seconde.

Intensité et énergie

Les ondes acoustiques propagent l'énergie, pas la matière :

- le bouchon sur l'eau n'avance pas,
- on peut recueillir l'énergie de la houle,

La vitesse maximale de propagation de l'énergie est limitée par la vitesse de la lumière.

L'énergie est proportionnelle au carré de l'amplitude de l'onde acoustique. On appelle *intensité*, l'énergie d'une onde acoustique.

$$I = P^2 \ / \ (\ \text{rho}\ c\ )$$

où P mesure l'amplitude de la pression acoustique, *rho* la masse volumique de l'air (1.2 kg/m^3), et c la célérité du son dans l'air (340 m/s).

Rayons sonores

Les ondes acoustiques suivent le plus court chemin pour se déplacer d'un point à un autre. Donc tout comme la lumière, on peut parler de rayons acoustiques.

Le *front d'onde* est la surface que dessinent tous les points dans le même état vibratoire (i.e. la même *phase*). Pour reprendre l'analogie de la vague, la crête de la vague dessine sur la mer un front d'onde. Les formes caractéristiques qu'adoptent les front d'onde sont :

- des sphères (ou des cercles concentriques pour les ronds d'eau)
- des plans (ou des lignes parallèles pour les vagues en bordure de rivage)

Dispersion

L'énergie ne se crée pas et ne disparaît pas. Elle se propage (et en seconde approximation, elle se dissipe sous forme de chaleur, c'est-à-dire qu'elle se transforme en énergie calorifique). L'énergie se répartit uniformément le long des fronts d'onde. Si le front d'onde s'élargit, alors l'énergie se disperse en proportion égale.

Sur une sphère, la surface est inversement proportionnelle au carré du rayon de la sphère. Donc, dans le cas d'ondes sphériques, si la distance à la source du bruit est doublée, l'intensité de l'onde est divisée par 4.

Par exemple, l'amplitude des vaguelettes qui se produisent sous la forme d'ondes concentriques quand un objet tombe dans un mare d'eau, diminue avec l'accroissement du rayon des cercles.

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Introduction à l'acoustique

Principe de fonctionnement

Haut-parleur

Un haut-parleur convertit un signal électrique en signal de pression. Le haut-parleur consiste en une membrane bafflée qui soumise au mouvement d'un moteur électrique oscille d'avant en arrière. Le mouvement rapide de la membrane entraîne avec lui des surpressions et des dépressions qui se propagent dans le milieu aérien.

Microphone

Un microphone est l'opposé d'un haut-parleur. C'est tellement vrai qu'il est possible d'utiliser un casque de baladeur comme microphone rudimentaire. Les surpressions et dépressions locales de l'air entraîne un mouvement infime de la membrane, qui par induction électromagnétique, génère un courant électrique.

Instruments de musique

Traditionnellement, distinction est faite entre :

- les instruments entretenus (voix, violon, orgue, clarinette ...),
- les instruments de type impulsionnel (piano, guitare, tambours, ...).

Les instruments que nous pouvons facilement décrire sont les suivants :

- guitare,
- guitare électrique,
- orgue Hammond,
- clarinette,
- violon,
- vibraphone,
- piano,
- flûte,
- voix...

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Introduction à l'acoustique

Quelques unités

Unité de pression

Le pascal (Pa) est l'unité de pression.

La pression d'une atmosphère est de l'ordre de 1020 hectopascals ($1{,}02.10^5$ Pa). Le seuil de sensibilité correspond à des variations de l'ordre de 20 micropascals (2.10^{-5} Pa). Le seuil de douleur correspond à peu près à des variations de l'ordre de 100 pascals (1.10^2 Pa).

Unité de temps et de fréquence

L'unité de temps (c'est une unité fondamentale) est la seconde (s). L'unité inverse est l'unité de fréquence : le hertz (Hz). Le hertz mesure la périodicité ou la cyclicité d'un phénomène.

- Au cinéma, les images défilent à 24 images par seconde. Cela correspond à 24Hz.
- L'électricité en Europe est caractérisée par 50 cycles par seconde, c'est le 50Hz.
- Le faisceau d'électron d'un téléviseur parcourt en théorie (codage SECAM) 625 lignes pour chaque image, à 25 images par secondes. En conséquence la fréquence de balayage des lignes d'un téléviseur est 25 * 625 = 15625 Hz. On parle de 15,625 kHz.

Le seuil de sensibilité de l'oreille varie en première approximation entre 30 Hz à 16 kHz.

Unité d'intensité

L'unité d'intensité est le watt par mètre carré (W/m^2).

Deux sources sonores (de même intensité *I*) font plus de bruit qu'une seule source prise séparément. Pour les *bruits*, ce sont les intensités qui s'additionnent. Donc dans le cas présent, l'intensité des deux sources réunies donne 2*I*. Donc les deux sources réunies sont deux fois plus bruyantes qu'une seule source prise séparément.

Bels et décibels

Le bel (B) donne une échelle logarithme pour les intensités. Le décibel (dB) est la dixième partie du bel (tout comme le décimètre est la dixième partie du mètre). Le principe en est le suivant :

I	->	I * 10
dB	->	dB + 10

Autrement dit, si un son a une intensité 10 fois plus grande, alors cela correspond à une intensité de 10dB supérieure.

L'intensité est proportionnelle au carré de la pression, donc on obtient le tableau suivant qui prend en compte les différences d'amplitude des pressions acoustiques :

P	->	P * 10
I	->	I * 100
dB	->	dB + 20

L'échelle des décibels est une échelle de comparaison : un son de 60dB est défini comme étant un million de fois plus fort qu'un son de référence à 0dB. La référence couramment utilisée est le seuil de sensibilité de l'oreille : P_r = 20 micropascals. Cette référence correspond à l'échelle des dB SPL (SPL comme Sound Pressure Level).

La formule qui permet d'obtenir la valeur en dB à partir des valeurs d'intensité ou de pression est la suivante :

$$\text{dB SPL} = 10 \log_{10} (I/I_r) = 20 \log_{10} (P/P_r)$$

Effet de la dispersion

Au chapitre précédent on a vu que l'intensité décroissait avec le carré de la distance à la source sonore. Donc, si on double la distance qui nous sépare d'une source sonore, l'intensité du bruit décroît de 6dB (cela correspond à $10*\log_{10}(4)$).

Page remise à jour le
Ven 26 Déc 1997 15:37:12

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Représentations temps et fréquence

Représentation temporelle

Le microphone transforme un signal de pression acoustique en un signal électrique proportionnel à celui-ci. L'observation de ce signal peut se faire à l'aide d'un *oscilloscope*. On obtient l'évolution de la pression acoustique en fonction du temps.

L'observation des signaux acoustiques permet de mettre en valeur certaines caractéristiques temporelles du signal sonore, du moins pour les signaux sonores *stables* :

- la quasi-périodicité,
- la présence d'une *forme d'onde*.

Dans une certaine mesure, il est possible d'associer ces caractéristiques physiques à des phénomènes perceptifs :

- le carré de l'amplitude du signal est proportionnel (jusque dans certaines limites) à la sensation d'intensité sonore,
- la période du signal est caractéristique de la perception de *hauteur* du son. Plus la période est petite, plus le son est aigu ou haut. Réciproquement, plus la période est grande, plus le son paraît grave ou bas.
- la forme d'onde est caractéristique dans une certaine mesure du timbre du son (le timbre des instruments de musique). Dans le cadre du signal de parole, la forme d'onde est le seul critère qui différencie des *phonèmes* (par exemple [a] ou [e]) prononcés à la même hauteur.

Page remise à jour le
Ven 26 Déc 1997 15:37:12

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Représentations temps et fréquence

Représentation fréquencielle

La représentation fréquencielle n'est pas qu'un simple outil mathématique dénué de tout fondement perceptif. Mathématiquement, la représentation fréquencielle consiste à décomposer le signal sur une base de signaux élémentaires : des sons purs ou sinusoïdes.

Perception de la couleur

À chaque couleur de l'arc-en-ciel (ce sont des couleurs dites *simples*) correspond exactement une longueur d'onde (ou fréquence) et une amplitude (ou intensité). Pour chaque type de lumière la décomposition à l'aide d'un prisme indique l'amplitude respective de chacune des couleurs simples de l'arc-en-ciel. L'ensemble forme le *spectre* de la lumière. Le prisme ne fait que révéler des informations qui sont cachées dans la lumière. On appelle ce domaine, le *domaine spectral* ou bien encore, *domaine fréquenciel*.

La perception que nous avons de la lumière dépend de 3 types différents de cellules qui tapissent le fond de la *rétine* sensibles à trois longueurs d'onde différentes : le rouge, le vert et le bleu. C'est parce que nous avons à notre disposition 3 types de cellules différentes qu'il est suffisant de décomposer la lumière sur la base des 3 couleurs dites *primaires* afin de donner l'illusion des lumières et des couleurs non-primaires. En terme simplifié, l'oeil n'est sensible qu'à trois couleurs primaires (i.e. à trois fréquences différentes). Tout le reste n'est qu'interprétation par notre cerveau des stimuli électromagnétiques captés par les récepteurs visuels.

Perception des sons

Tout comme la lumière, le son cache également en son sein un *spectre* et notre oreille est spécialement équipée pour le révéler. Contrairement à la vision, l'oreille interne est équipée de plusieurs milliers de cellules, chacune spécialisée dans une gamme très sélective de fréquences (ce qui correspondait aux couleurs simples de l'arc-en-ciel dans l'exemple précédent). Symboliquement, un son peut donc se représenter par une courbe indiquant la degré d'excitation de chacune des cellules le long de la *membrane basilaire*, c'est-à-dire sur l'axe des fréquences : c'est une *représentation fréquencielle* ou *spectrale* du son.

Mise en garde :

Notre présentation semble indiquer que le domaine spectral ne correspond qu'à des phénomènes perceptifs. Il n'en est rien. La représentation spectrale a une existence en dehors de tout dispositif de perception. Elle est définie mathématiquement par la *transformée de Fourier*.

Décomposition des sons

Sons purs

On qualifie de son pur l'équivalent en terme sonore des couleurs simples de l'arc-en-ciel. Le son pur est donc caractérisé entièrement par son amplitude et par sa fréquence. La représentation fréquencielle d'un son pur à la fréquence f_0 est un pic situé à l'abscisse de sa fréquence. La représentation temporelle d'un son pur, est une *sinusoïde*. La représentation temporelle du son pur fait apparaître une périodicité dans le signal. Cette période est l'inverse de la fréquence.

Expérience harmonique

On prend un générateur de sinusoïdes, puis on ajoute successivement des sinusoïdes aux fréquences f_0, puis $2f_0$, $3f_0$, $4f_0$ ... La première sensation consiste à entendre chacun des *partiels harmoniques* entrer séparément dans le son. Mais rapidement, tous les partiels se fondent pour ne donner plus que la sensation d'un son complexe, de même hauteur que le son pur original. Il n'est plus possible de distinguer séparément chacun des partiels du son.

Une façon d'analyser le son original, i.e. de le décomposer en ces composants élémentaires, consiste à ajouter au fur et à mesure des sinusoïdes dans le son, jusqu'à ce que le résultat corresponde au son original. Cette procédure de décomposition/recomposition du son s'appelle *analyse par la synthèse*.

Nomenclature et caractéristiques

Tous les sons *stables* se décomposent en sons élémentaires. Chaque son élémentaire se nomme *partiel* du son. Quand le son original est périodique, les fréquences des partiels sont toutes en rapport *harmonique* les unes entre elles. Dans ce cas, les partiels prennent le nom d'*harmoniques* du son.

L'écartement fréquenciel entre chaque partiel est caractéristique de la période du signal temporel, et donc de sa hauteur. L'*enveloppe spectrale* que dessine les sommets des partiels est caractéristique de la forme d'onde et donc du timbre. Dans le cadre de la parole, les trois premiers maxima locaux de l'enveloppe spectrale s'appellent des *formants* et sont caractéristiques de la voyelle prononcée (et de la forme du conduit vocal).

L'énergie d'un signal peut être localisée dans une zone fréquencielle n'ayant rien à voir avec sa hauteur (la fréquence fondamentale). En particulier, la hauteur de la parole varie entre 100 et 200Hz tandis que l'énergie est transmise essentiellement dans la gamme de fréquences 800 - 3000Hz. Le téléphone d'ailleurs ne transmet que la bande de fréquence utile : 800Hz à 8kHz.

Sensibilité

Notre oreille est sensible en première approximation aux fréquences entre 30Hz et 16kHz. Le maximum de sensibilité se situe aux alentours de 3kHz, ce qui est en adéquation avec le mécanisme de production de la voix qui produit de l'énergie essentiellement autour de cette fréquence.

La réception des signaux se fait par des cellules cillées, qui sont la terminaison de cellules nerveuses, qui ne sont jamais remplacées. La destruction des cellules cillées est irréversible. Les cellules se détruisent avec l'âge, mais aussi avec des expositions trop violentes ou trop répétées à des stimuli de grande amplitude.

Théorie simplifiée de l'harmonie

Un accord musical sonne d'autant mieux que les sons fusionnent correctement. On a vu précédemment que des sons purs harmoniques avaient tendance à fusionner sans que l'on puisse les distinguer. L'analyse d'un accord se fait en superposant la représentation fréquencielle des sons constituant l'accord, et à observer comment se superpose les partiels harmoniques des sons.

Dans le cas d'un accord d'octave (1/2), un partiel sur deux fusionne. C'est l'accord le plus *consonant* (par opposition à *dissonant*). Dans le cas d'un accord de quinte (2/3), approximativement un partiel sur trois fusionne. C'est un des accords le plus consonant après l'accord d'octave.

Quand deux partiels se superposent mal, disons avec un écart de 10Hz, ils produisent des battements, c'est-à-dire à une modulation d'amplitude, dans le cas présent, de 10 battements par seconde. Ce type de battement est trop rapide pour être perçu comme un phénomène temporel, et trop lent pour être perçu comme un phénomène fréquenciel. C'est le phénomène de *rugosité*. La rugosité entraîne une ambigüité de perception qui induit un stress et une *dissonance* de l'accord.

Filtrage

Le filtrage consiste à atténuer ou amplifier *sélectivement* chacune des régions du spectre. Un filtre est caractérisé par sa *fonction de transfert* (ou *gain en fréquence* ou encore *réponse fréquencielle*) qui décrit le gain de chacune des régions du spectre.

Conclusion partielle

Tous les signaux, toutes les opérations de filtrage ou de modification des sons doivent être considérés à la fois dans le domaine temporel, et dans le domaine fréquenciel. Les deux domaines sont indissociables et complémentaires. On ne peut prétendre expliquer un phénomène sonore qu'en l'envisageant simultanément dans les deux domaines.

Page remise à jour le
Ven 26 Déc 1997 15:37:11

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Représentations temps et fréquence

Représentation temps-fréquence

Pourquoi

Les représentations temporelles ne montrent aucune caractéristique fréquencielle du signal et réciproquement, les représentations fréquencielles n'apportent aucune indication de nature temporelle sur le signal. Il est souhaitable d'obtenir une représentation hybride alliant les avantages des deux types de représentation.

Portée musicale

La portée musicale est une première tentative de représentation temps-fréquence :

- le temps est indiqué horizontalement,
- la fréquence est indiquée verticalement.

Toutefois, la portée musicale a ses limitations :

- *quantification* du temps,
- *quantification* des hauteurs,
- pas d'indication sur la répartition spectrale d'énergie ou sur le timbre.

Spectrogramme

Le spectrogramme est une représentation à court-terme adaptée pour figurer simultanément des informations fréquencielles et temporelles. Elle est réalisée à l'aide de l'outil mathématique appelé *transformée de Fourier à court-terme*. Pour comprendre cette représentation, il suffit de remarquer que :

- le temps est indiqué horizontalement,
- la fréquence est indiquée verticalement.
- une coupe verticale du spectrogramme donne exactement une représentation fréquencielle (à court-terme, c'est-à-dire localisée dans le temps),

Applications

Cette représentation consiste en une *analyse* du signal. Ce type de représentation est utilisé systématiquement dans la plupart des algorithmes sophistiqués de traitement des sons :

- repérage des clics, restauration des enregistrements anciens,
- segmentation des sons,
- filtrage variant dans le temps,
- séparation de sources (par exemple, séparation de la voix d'un chanteur perdue au milieu de l'orchestre),
- comprendre des sons (le son qui monte infiniment de J.-C. Risset)...

Qu’est-ce qu’on y voit

Les spectrogrammes permettent d’obtenir de nombreuses indications sur le son à partir d’indices visuels simples :

- les variations rapides du signal sont signalées par des composantes hautes-fréquences. En particuliers les clics dus à des discontinuités du signal, quasiment invisibles dans le domaine temporel, apparaissent clairement comme un afflux soudain et bref d’énergie à toutes les fréquences,
- les sons percussifs sont indiqués par des traces d’énergie assez brèves dans le domaine temporel, assez étendues dans le domaine fréquenciel, avec une décroissance plus rapide dans les aigus que dans les graves,
- les bruits (chuintements, souffles, sifflements...) sont indiqués par des zones grisées, visibles souvent à haute-fréquence,
- quand tous les partiels du son montent en même temps, cela indique que la hauteur du son monte continûment,
- quand tous les partiels suivent une ondulation, cela indique un vibrato de l’instrumentiste. Tous les partiels du même instrument oscillent en phase, ce qui permet d’isoler facilement un instrument dans un orchestre.

Page remise à jour le
Ven 26 Déc 1997 15:37:11

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Numérisation

L'opération de numérisation se réalise en théorie en deux étapes :

- échantillonnage,
- quantification.

Échantillonnage

L'échantillonnage consiste à passer d'un signal à temps *continu* (un signal électrique, un signal acoustique...), en une suite *discrète* de valeurs (valeurs mesurées à intervalles réguliers).

Signal discret - signal continu

Signal à temps continu :

- la hauteur du bouchon qui flotte sur l'eau,
- le signal électrique qu'utilise un amplificateur audio,
- le signal hertzien de modulation d'amplitude, ou de fréquence,
- la vitesse d'une voiture...

Signal à temps discret :

- les mesures quotidiennes du taux de globules rouges dans le sang,
- la donnée de la température au bulletin météo tous les matins,
- le pourcentage de spectateurs regardant le journal de 20h de la Une,
- des mesures régulières de l'activité volcanique d'un volcan...

Interprétation temporelle

L'interprétation temporelle est très simple : on mesure périodiquement la valeur d'un signal à temps continu. Par exemple, on mesure la vitesse d'une voiture toutes les 10 secondes et on reporte les points sur un graphe. Chaque mesure s'appelle un *échantillon*. La *période d'échantillonnage* est la période de temps séparant deux échantillons successifs. La *fréquence d'échantillonnage* ou *taux d'échantillonnage* s'exprime en hertz, et correspond à l'inverse de la période d'échantillonnage (un période d'échantillonnage de 10s correspond à une fréquence d'échantillonnage de 0.1Hz).

Dans un premier temps, la reconstruction du signal n'est possible que si les variations de celui-ci sont assez lentes, ou réciproquement si la période d'échantillonnage est assez fine.

La reconstruction en pratique consiste à maintenir constante la valeur de l'échantillon jusqu'à l'arrivée de l'échantillon suivant. On appelle ce dispositif un *bloqueur d'ordre 0*.

Interprétation fréquencielle

D'un point de vue théorique, l'échantillonnage correspond à la *périodisation du spectre*. En conséquence, l'intégrité du signal est maintenue tant que les copies (les *alias* en anglais) du spectre ne se superposent pas l'une sur l'autre. Le phénomène de recouvrement des spectres est nuisible et s'appelle le *repli spectral* (ou *aliasing* en anglais). Une conséquence de cette interprétation est la suivante le théorème d'échantillonnage : pour éviter le repli spectral, il faut et il suffit que le signal original soit à bande limitée et que la fréquence d'échantillonnage soit supérieure à deux fois la bande utile du signal. En pratique, le signal audio utile est limité par notre perception, c'est-à-dire 16kHz, donc, la fréquence d'échantillonnage doit être supérieure à 32kHz. Pour que le signal audio respecte les conditions du théorème d'échantillonnage, il faut s'assurer d'avoir éliminé toutes les composantes hautes fréquences en filtrant par un filtre anti-repliement (anti-aliasing).

Effet du repli spectral

Le repli spectral (*aliasing* en anglais) est nuisible:

- en vidéo, la chemise à rayures fait un moirage à l'écran,
- au cinéma ou à la télévision les roues des voitures et des charrettes semblent tourner au ralenti dans un sens ou dans l'autre ,
- la décomposition stromboscopique du mouvement : le stromboscope permet de décomposer les mouvements rapides et périodiques, il agit selon le principe du repli spectral,
- avec un taux d'échantillonnage de 44.1kHz, une sinusoïde inaudible à 40kHz se replie en une sinusoïde audible et gênante à 4.1Hz,

Pratique de l'échantillonnage

Les signaux sonores ont en général peu d'énergie à haute fréquence.

La qualité de l'échantillonnage et de la restitution sonore dépend essentiellement de la qualité du filtre analogique anti-repliement. En particulier, le prix des cartes audio pour les ordinateurs personnels est essentiellement déterminé par la qualité des convertisseurs (et donc de la qualité des filtres anti-repliement). En particulier, de nombreuses cartes bon marché ne possèdent pas de filtres anti-repliement adaptées à toutes les fréquences d'échantillonnage proposées. Par exemple, de nombreux ordinateur Macintosh ont été vendus sans filtre anti-repliement à 32kHz, ce qui entraîne un très mauvais rendu sonore à cette fréquence d'échantillonnage.

La reconstruction avec des dispositifs bloqueurs induisent une génération de composantes haute-fréquences non-désirées. Il est nécessaire d'utiliser un filtre du même type que le filtre anti-repliement pour la conversion numérique-analogique.

Les techniques évoluées d'échantillonnages consistent à sur-échantillonner / sous-échantillonner. D'un point de vue théorique, cela consiste à déplacer le problème du filtrage anti-repliement du domaine analogique dans le domaine numérique, ce qui coûte beaucoup moins cher. C'est ce que l'on voit affiché sur les spécifications techniques des lecteurs de CD-audio.

Ce document a été formaté par htmlpp.

Numérisation

Quantification

Définition

En première approximation, la quantification consiste à remplacer un nombre réel par un nombre entier, par exemple à *arrondir* un nombre réel par le nombre entier le plus proche. De façon plus précise, la quantification associe un symbole logique à une quantité réelle. La terminologie associée à cette technique :

- pas de quantification q,
- quantification scalaire,
- quantification sur N bits, 8 bits, 16 bits, 24 bits,
- quantification vectorielle,
- quantification linéaire ou pas, A-law et mu-law,
- arithmétique en virgule fixe...

Le pas de quantification est en rapport avec le nombre de bits alloué pour la quantification scalaire linéaire (la plus couramment utilisée) :

$$q = 2^{N}$$

Effets sur le son

La quantification a pour effet de rajouter du bruit dans le signal : c'est le *bruit de quantification*. En première approximation, le bruit de quantification est un bruit blanc (c'est-à-dire réparti sur toutes les fréquences possibles), uniformément réparti (c'est-à-dire que les valeurs du bruit prennent de façon équiprobable toutes les valeurs comprises entre $-q/2$ et $q/2$).

La puissance du bruit généré est proportionnelle au carré du pas de quantification : $I = q^{2}/12$.

Le rapport signal à bruit correspond à la *dynamique* du support, c'est-à-dire le rapport entre la puissance du bruit de fond du support d'enregistrement ou de stockage et celle du signal le plus fort possible d'enregistrer sans distorsion sur ce support. Pour la quantification linéaire, le rapport signal à bruit est approximativement de (en décibel) $6*N$, où N est le nombre de bits sur lequel se fait la quantification.

Par exemple pour les CD-audio : 16 bits donnent une dynamique (théorique) de 96dB. Pour donner un ordre d'idée, la dynamique d'un orchestre symphonique peut s'élever à 100dB.

Dynamique (théorique) de différents supports

CD-audio (16 bits linéaire)	96dB
Cassette magnétique	50dB
Cassette magnétique + Dolby	60dB
Disque vynil	60dB Dynamique (théorique) de différents supports

Mise en forme spectrale du bruit

Pour minimiser les effets du bruit de quantification, il est possible de mettre en forme le bruit de quantification, de rejeter toute la puissance du bruit à haute-fréquence par sur-échantillonnage, puis de réduire le bruit par filtrage passe-bas. C'est tout l'intérêt de la technique de sur-échantillonnage associée à la technique sigma-delta de certains composants de conversion analogique-numérique.

Page remise à jour le
Ven 26 Déc 1997 15:37:11

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Numérisation

Technologie

Les dispositifs qui numérisent le signal s'appellent des convertisseurs analogiques-numériques ou numériques-analogiques (ADC ou DAC). Ils sont essentiellement caractérisés par la fréquence d'échantillonnage (44.1kHz, 48kHz...), le nombre de bits alloués pour faire la quantification (16 bits). On ne sait pas vraiment faire mieux que 17 à 18 bits. Pour obtenir mieux, il faut utiliser des convertisseurs sur-cadencés, par exemple avec la technique sigma-delta. En audio, les convertisseurs sigma-delta 1 bit fonctionnent en interne avec un seul bit de quantification, mais sur-cadencent au moins 256 fois le signal, ce qui correspond à une fréquence d'échantillonnage en interne d'au moins 11MHz. En externe, tout se passe comme si le convertisseur fonctionnait, par exemple en 24 bits à 44.1kHz.

En audio grand public, on parle essentiellement de quantification scalaire linéaire. Les technologies de compression et de transmission numérique de la parole (téléphone numérique) utilisent d'autres types de quantification, dites vectorielles.

Intérêts / inconvénients

Les intérêts sont multiples:

- stockage numérique (CD-audio, DAT, DAB...), et reproduction à l'identique possible,
- codage numérique, résistance sans faille à l'erreur,
- traitement numérique, donc pas de traitements destructeurs,
- on peut faire des choses complexes (en général) plus facilement avec des ordinateurs qu'avec de l'électronique analogique...

Il subsiste quelques inconvénients :

- gros volumes de données, difficulté des transmissions numériques,
- problèmes des formats de données.

Page remise à jour le
Ven 26 Déc 1997 15:37:13

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Spatialisation

Rayonnement acoustique

En pratique, les dispositifs acoustiques ne rayonnent pas régulièrement d'énergie acoustique dans toutes les directions. On a une perception intuitive de ce phénomène acoustique :

- une clarinette émet le son essentiellement dans l'axe de l'instrument,
- la guitare acoustique, dans la direction de la rosace,
- le violon acoustique, dans l'axe du manche,
- un haut-parleur, dans la direction de la membrane...

Chaque instrument ou source sonore est donc caractérisé par son *diagramme de directivité* qui indique quelles sont les directions privilégiées selon lesquelles le son se propage.

Chaque récepteur acoustique est également caractérisé par un diagramme de directivité. Par exemple les microphones omni-directionnels (micros omnis) sont sensibles aux sons provenant de toutes les directions, alors que les microphones directionnels (micros cardioïdes) ne sont sensibles qu'à une seule direction.

Nous sommes sensibles au phénomène de directivité essentiellement quand la source bouge par rapport au récepteur, ou quand le récepteur bouge par rapport à la source (mouvement du musicien, gestuel de l'interprète...).

Reproduction

Les caractéristiques de spatialisation d'un système de sonorisation sont en général :

- le nombre de pistes audio,
- le nombre d'enceintes à disposition,
- leur répartition spatiale.

Il n'existe pas à l'heure actuelle, de système reproduisant fidèlement et dynamiquement le champ acoustique d'une (ou plusieurs) source(s) sonore(s). En particulier, il est illusoire de croire pouvoir reproduire fidèlement le champ acoustique d'un instrument acoustique à l'aide d'une ou deux enceintes acoustiques. En fait, on ne sait pas combien d'enceintes sont nécessaire pour restituer « virtuellement » l'acoustique d'un instrument.

En pratique, les installations sonores dans les grandes salles de cinéma prennent en compte 4 pistes audio, réparties sur une petite dizaine d'enceintes pour donner l'illusion de la localisation et du mouvement. Dans les installations artistiques, le nombre de pistes audio et d'enceintes peut être bien plus grand.

Il est difficile de confondre le son d'un véritable instrument acoustique restitué à l'aide d'enceintes acoustiques, non plus à cause de la distorsion induite par le médium (cassette numérique, bande magnétique, CD-audio, chaîne d'amplification...), mais simplement parce que nous percevons clairement la directivité de l'enceinte, et pas celle de l'instrument acoustique.

Page remise à jour le
Ven 26 Déc 1997 15:37:13

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Introduction à la Psychoacoustique

Qu'est-ce qye c'est ?

C'est la relation entre le phénomène physique vibratoire acoustique, la perception que nous en avons, et l'organisation que nous en faisons.

Parallèle avec la vision

La perception visuelle fait état de :

- peu de couleurs (radiations) perçues indépendamment,
- 3 couleurs primaires...

L'organisation visuelle consiste :

- détections de formes simples,
- détection des directions...

Caractères perçus

Tous les stimuli acoustiques ne sont pas forcément perçus :

- sons trop faibles,
- sons trop aigus : ultrasons,
- sons trop graves : infrasons...

L'organisation est complexe. L'identification et l'organisation se fait sur différents critères plus ou moins simples :

- critères temporels,
- critères fréquenciels,
- critères énergétiques,
- critères timbraux...

Le phénomène de la perception audio est en général indissociable du contexte :

- critères visuels,
- passé,
- sémantique,
- autres phénomènes perceptifs...

L'organisation des événements sonores correspond finalement à un problème d'organisation de *percepts* qui tient de la *psychologie*.

Champs d'application

De nombreux champs d'application :

- sociologie : musique d'ambiance, de publicité, d'annonce...
- urbanisme, ergonomie : protection contre la fatigue auditive au travail, dans la rue, chez soi, sur son ordinateur...
- psychologie, acoustique, marketing : perception et classification des bruits de voiture, de moteurs, de portières,
- militaire, ergonomie : aide à la navigation (pour les voitures, les avions, les chars...), message d'information ou d'alarme, reconnaissance des bruits-marins, système experts de reconnaissance, aide à la décision, aide à la spatialisation (pilote de chasse)...
- musical : aide à la composition,
- traitement du signal : codage psychoacoustique (MiniDisque de Sony, DCC -fini-, DAB, MPEG-audio, téléphone cellulaire, INMARSAT...),
- physiologie et neurologie : traitement des troubles auditifs,
- restitution sonore : spatialisation, matériel HIFI...
- multimédia : spatialisation (virtualisation de l'espace sonore) des sources sonores, richesse de l'environnement sonore des jeux, synthèse musicale...

Critère acoustique et attribut perceptif

Attribut perceptif	Phénomène acoustique	Unité psycho-acoustique
Niveau ou intensité sonore	dB (SPL)	Sonie (dB-A) et Phonie (en Sones)
Perception de la hauteur	Hz	Tonie
Perception des durées	s	Chronie
Autre	???	Timbre

Nomenclature et précautions d'usage

Son simple ou pur : sinusoïde,

son complexe : bruit blanc gaussien faible bande.

Remarques:

- Il n'est pas possible de comparer deux sinusoïdes directement, car la somme de deux sons purs de fréquences voisines produit des battements facilement discernables.
- Le problème de la représentation mentale des événements sonores est très important. En effet les processus de mémorisation font appel (à un état conscient ou pas) à une étape de représentation symbolique des stimuli perçus. Par exemple, dans le domaine visuel, on ne souvient pas de l'image d'une scène, mais de ce qui a été reconnu dans la scène. Dans le domaine sonore, on se souvient par exemple de la mélodie d'une chanson, parce que nous sommes capable d'obtenir une représentation mentale de la succession des notes, et chaque note est également représentée de façon plus ou moins consciente, par un symbole correspondant par exemple à sa notation dans la gamme occidentale.

- On remarquera dans les paragraphes qui suivent la parfaite adéquation entre les systèmes de production sonore que notre espèce utilise (la voix humaine produit essentiellement de l'énergie entre 1 et 3kHz), et les systèmes de réception (notre oreille est la plus sensible entre 1 et 3kHz). Il existe une autre similarité troublante entre la forme d'onde des impulsions *glottales* (i.e. la forme d'onde produite par chaque impulsion des cordes vocales), et la réponse impulsionnelle du filtre d'analyse développé par l'oreille (*gammatone filters*) pour discriminer les fréquences des sons. Nous nous contenterons de dire que la nature est bien faite.

Page remise à jour le
Ven 26 Déc 1997 15:37:13

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Perception des niveaux sonores

Courbes isosoniques

Principe expérimental

L'expérience consiste à régler un son pur de fréquence variable à la même intensité subjective d'intensité, qu'un son de référence à 1kHz. Par comparaison à l'échelle des décibels, on obtient l'échelle des *phones*, les courbes de même intensité sonore sont dites *isosoniques*.

Limites perceptives

La perception d'un son pur existe dans l'intervalle 20Hz-20kHz. Cet intervalle se réduit inexorablement avec l'âge. La *presbyacousie* correspond à la perte de cette acuité auditive. La destruction de cellules est irrémédiable et irréversible.

Limites de l'interprétation

Les limites perceptives correspondent à la perception des sons *stables*. On ne peut rien en déduire sur la résolution temporelle de l'oreille qui induit d'autres circuits de perception. En d'autres termes, les attaques brèves sont susceptibles d'être altérées par échantillonnage, même si la fréquence de coupure (moitié de la fréquence d'échantillonnage) se situe bien au delà du seuil de perception des sons stables.

Seuil d'audibilité

La courbe à 0 phone correspond au seuil d'audibilité. En deçà, un stimulus sonore ne produit pas de réaction sensible.

Effet Loudness

Les courbes isosoniques aux alentours de 50 phones permettent d'égaliser la répartition fréquencielle d'un son afin que le *rendu* sonore perceptif à faible niveau sonore soit identique que celui qu'on obtiendrait au niveau de jeu original. Cette égalisation s'obtient avec le bouton *loudness* que l'on trouve sur la plupart des équipements HIFI.

Dynamique

L'oreille n'est sensible qu'à 50dB de dynamique dans les graves, à comparer avec les 120dB de dynamique aux alentours de 3kHz (à comparer également avec la dynamique plus faible des instruments de mesure).

La chaîne des osselets (marteau, étrier et enclume) permet d'*adapter* l'impédance acoustique du milieu extérieur à celle de l'oreille interne. Il existe des mécanismes réflexes permettant de modifier dynamiquement le facteur d'adaptation acoustique de la chaîne des osselets afin d'augmenter ou de diminuer le ratio d'énergie transmis à l'oreille interne. Ce mécanisme s'apparente à celui de la *pupille* de l'oeil agissant comme un *diaphragme*, laissant entrer plus ou moins de lumière à l'intérieur de la

cornée.

Décibelmètre

Pour mesure l'intensité perceptive, il faut appliquer une correction sur les sons. Il existe plusieurs courbes normalisées qui inversent les courbes isosoniques. Elles sont référencées sur les instruments de mesure (les décibelmètres) sous le nom de dB(A) et dB(B). La législation française fait référence aux mesures de bruit, exprimées en dB(A) ou dB(B) pour signifier les normes et les maximales admissibles. Les normes européennes tendent à multiplier les échelles d'intensité sonore subjective pour la mesure de la nuisance des bruits appliquée à une multitude de situations différentes.

Page remise à jour le
Ven 26 Déc 1997 15:37:13

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Perception des niveaux sonores

Perception des rapports d'intensité

Principe expérimental

L'expérience consiste à demander à un utilisateur de régler le niveau d'un son 2 fois plus fort qu'une version de référence. L'expérience a été initiée par le professeur Steven entre 1955 et 1972. C'est l'échelle des sones.

Relation de Steven

La relation que Steven a mise en évidence est une loi logarithmique indiquant que la sonie est approximativement doublée tous les 10 phones. Autrement dit, notre perception des rapports d'intensité est logarithmique.

$$N = k\, P^{0.6}$$

Les échelles logarithmiques sont très générales dans la plupart des phénomènes de perception : nous ne percevons souvent que des rapports de sensation. Par exemple, le passage de 100 à 110 grammes nous procure la même sensation d'accroissement de poids que le passage de 10 à 11 kilogrammes, c'est-à-dire un accroissement de 10%.

Page remise à jour le
Ven 26 Déc 1997 15:37:14

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Perception des niveaux sonores

Perception des différences d'intensités sonores

Principe expérimental

Il existe deux expériences différentes mettant en évidence la perception différencielle d'intensité :

- réduction de l'index de modulation d'amplitude d'un son pur, jusqu'à ce que l'utilisateur ne perçoive plus la modulation,
- comparaison de deux stimuli A et B, dont les intensités peuvent être légèrement différentes ou identiques.

Seuil différentiel de perception

Le seuil différentiel de perception d'intensité varie avec la fréquence du son pur, ainsi qu'avec son intensité, mais globalement reste toujours aux environs de 1 phone. Il est donc inutile d'indiquer les mesures subjective d'intensité sonore en décibel avec une précision supérieure à l'unité puisque l'oreille ne fait pas la distinction entre le résultat d'une mesure de 60dB(A) et à 60.5dB(A).

Page remise à jour le
Ven 26 Déc 1997 15:37:14

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Perception des niveaux sonores

Phénomène de masquage

Principe expérimental

La mesure des courbes de masquage s'obtient en présentant simultanément à un auditeur un son pur et un bruit blanc faible bande situés dans des gammes de fréquences voisines. L'expérience montre que si l'intensité du son pur est suffisamment faible, et si les gammes de fréquences sont suffisamment voisines, l'auditeur *ne perçoit pas* le son pur qui est couvert par le bruit. On déduit de cette expérience une *courbe de masquage* qui quantifie les niveaux et les fréquences pour lesquels un bruit *masque* un son pur (et réciproquement).

Intuition du masquage

Le phénomène du masquage est à rapprocher de la situation d'un bruit important (marteau piqueur, circulation automobile, réacteur d'avion) couvrant la conversation. Dans un premier temps, le bruit est tellement important que la voix de l'interlocuteur est entièrement couverte par le bruit. La seule possibilité qu'a votre interlocuteur pour se faire entendre consiste à élever la voix. Auquel cas, seules quelques bribes de paroles parviennent à émerger du bruit ambiant. De plus, les bribes elles-mêmes restent peu compréhensibles, car souvent entachées de bruits parasites dans des zones fréquencielles cruciales pour la reconnaissance de la parole. Au total, vous ne disposez pas d'éléments suffisants (ni temporels, ni fréquenciels) pour décoder correctement le message de votre interlocuteur. L'écoute en milieu bruité est donc bien plus fatigante que celle en milieu calme puisque toute l'attention est nécessaire pour comprendre, déchiffrer la parole, en faisant bien souvent appel à des mécanismes très complexes de *contextes* (mouvements des lèvres, gesture et posture etc.) pour *intuiter* plus qu'entendre le message sonore.

Il est facile d'obtenir l'intuition du masquage sonore en utilisant la métaphore visuelle de l'ombre projetée. Le bruit projette sur l'axe des fréquences une ombre, tout comme l'arbre projette également une ombre sur le sol. Tous les événements se situant à l'intérieur de l'ombre sont invisibles, c'est-à-dire imperceptibles.

Mels et bande critique

L'étude précise de ce phénomène fait apparaître une largeur de *bande critique* à l'intérieur de laquelle un son peut être masqué et à l'extérieur de laquelle un son ne peut pas être masqué. Cette expérience prouve entre autre que notre oreille est équipée de récepteurs sélectifs en fréquence, traitant des zones fréquencielles dont la largeur est précisément la largeur de la bande critique. Donc deux sons séparés de plus d'une bande critique excitent des récepteurs complètement disjoints ; ils sont ainsi complètement discriminés.

Le concept de bande critique intervient dans de nombreux autres phénomènes de perception sonore, comme par exemple le phénomène de *rugosité*.

La largeur de la bande critique n'est pas une constante en fonction de la fréquence. L'échelle des mels (ou barks) est une échelle déduite de l'échelle des fréquences (en Hz) de telle façon que la bande critique soit de largeur constante. Expérimentalement, on constate que l'échelle des mels correspond à celle des hertz jusqu'à 1kHz, puis la relation entre les deux échelles devient encore une fois logarithmique.

Utilisation musicale

Les compositeurs et musiciens ont une connaissance assez intuitive de ce phénomène. En utilisant une palette sonore la plus large possible, ils permettent aux sons instrumentaux d'occuper au mieux l'espace sonore afin que nous puissions discriminer chacun des instruments. Les zones de l'espace des fréquences occupées par chaque instrument ne doivent pas se recouvrir pour que nous puissions les isoler facilement. Par exemple, une mélodie aigüe sur un celesta sera perçue très distinctement au milieu d'une orchestration à base de guitare basse, guitare électrique, batterie, choeurs... sauf si la guitare électrique, en utilisant un effet de distorsion, vient occuper dans les aigus l'espace du celesta.

En bref, notre capacité à discriminer un certain nombre d'instruments dans un orchestre tient aux caractéristiques de masquage fréquenciel (en l'occurrence de non-masquage).

Utilisation pour le codage et la transmission

Dans les chapitres précédents, nous avons vu que la numérisation des sons entraînait un flux de données numériques considérable, qui est souvent incompatible avec la technologie actuelle de stockage ou de transmission. Un choix cohérent de la fréquence d'échantillonnage et de la méthode de quantification permet de réduire la taille des données sonores, mais en général, ce n'est pas suffisant. Par exemple, un CD-audio de 76 minutes (16 bits linéaires, 44.1kHz, stéréo) correspond à plus de 750 mégaoctets de données (soit plus de 500 disquettes 3 pouces et demi formattées à 1.4MO)!

L'utilisation du phénomène de masquage sonore permet de réduire considérablement le volume des données à stocker ou à transmettre. Le principe consiste à ne coder (ou transmettre) que ce que nous entendons. Il est en effet inutile de transmettre les sons que nous n'entendons pas. Le phénomène psychoacoustique du masquage est à présent suffisamment compris pour que les ingénieurs en télécommunications soient capables de préciser la part de ce que nous pouvons percevoir de celle que nous ne pouvons pas percevoir ; ils ne transmettent que ce que nous percevons. Ce type de codage s'appelle un *codage psychoacoustique*.

D'un point de vue technique, le codage psychoacoustique consiste à décrire très précisément à un instant donné la courbe de masquage du son à transmettre. Cette courbe de masquage indique le seuil à partir duquel un bruit n'est plus perçu. Nous avons vu dans le chapitre précédent que l'effet de quantification correspondait essentiellement en une génération de bruit blanc large bande. L'idée consiste à procéder à une *allocation dynamique* des bits pour que le bruit de quantification soit caché au mieux par le son transmis. Cette méthode permet donc de minimiser *en moyenne* le nombre de bits alloué pour la quantification en *adaptant* localement la répartition des bits en fonction des caractéristiques de l'oreille et du son à transmettre.

Ce type de codage psychoacoustique est décliné selon différentes variantes que l'on retrouve par la suite dans le MiniDisc de Sony, dans les (ex-)cassettes numériques de Philips, dans les téléphones cellulaires, pour la radio numérique (DAB), et dans les normes de MPEG-Audio (MPEG-Audio Layer 2 et Layer 3, Musicam...).

Page remise à jour le
Ven 26 Déc 1997 15:37:14

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Perception des hauteurs

Hauteur tonale

Périodicité

La périodicité (phénomène temporel) est le principal phénomène physique à mettre en rapport avec la perception de hauteur. Par exemple, tous les sons possédant une période de 10 ms, seront jugés comme des sons possédant la même hauteur (*hauteur tonale*), et en particulier la même hauteur qu'une sinusoïde à 100 Hz.

Harmonicité

D'un point de vue fréquenciel, la périodicité d'un son entraîne une répartition *harmonique* de ses partiels. Donc, si nous devions énoncer une règle pour mesurer la hauteur perçue d'un son périodique à partir de son spectre, nous dirions qu'il s'agit de déterminer le plus grand commun diviseur (PGCD) des fréquences de tous les partiels harmoniques. Cette fréquence s'appelle *fréquence fondamentale* (ou encore *la fondamentale*) d'un son.

Pièges

La fréquence fondamentale d'un son n'est pas :

- la fréquence du premier partiel harmonique, dit *le fondamental*. Nous sommes alors dans le cas connu du *fondamental absent*. Il s'agit par exemple de sons creux, tels que celui du basson. À part un timbre un peu pauvre, cette situation n'a rien d'extraordinaire ou étonnante ; au niveau de la forme d'onde, rien de particulier ne distingue ce cas du cas où le fondamental est présent,
- l'écartement entre les partiels. Nous sommes dans le cas où il manque de nombreux partiels dans le son. Dans le cas de la clarinette, il manque approximativement un partiel harmonique sur deux, caractéristique de cette sonorité un peu nasillarde,
- un maximum d'énergie du spectre. La perception du maximum d'énergie spectrale est à mettre en rapport avec un autre phénomène de perception de la hauteur, dit de *hauteur spectrale* par opposition à la hauteur tonale.

Ambigüité d'octave

La hauteur des sons est ambigüe à une octave près. Un son à 200Hz et un son à 400Hz produisent tous les deux une sensation de hauteur assez semblable. Cela tient au fait que si mathématiquement 2.5ms est une période du signal (400Hz) alors, 5ms est nécessairement une autre période du signal (200Hz).

L'importance du rapport d'octave est très largement utilisée en musique, en particulier pour définir des classes de hauteurs (Do, Ré, Mi... sont définis à une octave près, et définissent ainsi une classe de hauteur).

La position particulière du rapport d'octave conduit à représenter les hauteurs sur une hélice circulaire (en trois dimensions), ou sur une spirale (en deux dimensions), de telle façon que deux hauteurs séparées d'une octave se fassent face sur ce graphe. Il est possible de passer continûment de la sensation d'une hauteur à celle de la hauteur double, sans passer par la sensation des notes

intermédiaires. Il s'agit de l'*octaviation*.

Perception différentielle

La perception de la hauteur, est, comme la plupart des phénomènes perceptifs, régie essentiellement par une échelle logarithmique.

- la perception de la hauteur du son change en fonction de son intensité sonore du son et en fonction du niveau du bruit ambiant. Cette déviation de perception de hauteur dépend également de la hauteur du son ; en particulier, la direction de la déviation change à 1kHz,
- le seuil de discrimination est également logarithmique. Il est à peu près de 1%, c'est-à-dire, que présentés séparément, deux sons à 400 et 404Hz provoquent la même sensation de hauteur,

Oreille absolue

Normalement, nous ne sommes capable de percevoir que des rapports de hauteurs. En d'autres termes, nous nous souvenons sans difficulté de la mélodie de « Au clair de la Lune », mais nous reconnaissons toutes les mélodies *transposées* également comme « Au clair de la Lune ». Donc les mélodies de hauteurs reposent principalement sur l'enchaînement des rapports de hauteur, et non pas sur les hauteurs proprement dites.

Certains individus sont toutefois capables de percevoir la hauteur des sons, de la mémoriser, et de la comparer avec d'autres hauteurs. Cette caractéristique s'appelle *l'oreille absolue*. C'est une caractéristique génétique, et fait donc partie de l'inné. Si on la possède, elle se cultive, sinon elle ne s'apprend pas.

Hauteur spectrale

La hauteur spectrale est un phénomène concurrent de la perception de la hauteur.

Ambigüités entre hauteur tonale et hauteur spectrale

Page remise à jour le
Ven 26 Déc 1997 15:37:14

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

IRCAM Centre Georges Pompidou

Perception des hauteurs

Organisation des hauteurs tonales

Les rapports musicaux

Gamme de Pythagore

Gamme naturelle

Gamme tempérée

Références

Page remise à jour le
Lun 22 Déc 1997 15:56:03

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Perception de l'espace

Critères binauraux

Les critères binauraux regroupent tous les indices qui impliquent les deux oreilles pour nous donner des indices sur la position dans l'espace de la (des) source(s) sonore(s) :

- différence d'intensité entre les deux oreilles, c'est le critère utilisé par la stéréophonie en HIFI pour restituer une impression d'espace,
- dé-synchronisation ou déphasage des signaux parvenant aux deux oreilles : les distances que le son parcourt entre la source et les deux oreilles sont différentes. Une impulsion générée à ma droite arrive donc d'abord sur mon oreille droite puis sur mon oreille gauche. Pour les sons stables et périodiques, cela induit un déphasage entre la voix gauche et la voix droite,

Critères monauraux

Les circonvolutions du pavillon de l'oreille entraînent des atténuations différentes pour les ondes sonores en fonction de leur direction de provenance. Notre cerveau a une connaissance intuitive de cet effet de directivité, et est capable d'en extraire des indications sur la direction des sons.

En première approximation, l'intensité d'un son nous donne une indication sur sa proximité. En effet, plus la source sonore est éloignée, moins elle est forte. Toutefois, dans le cas des enregistrements audio, les niveaux d'écoute sont relatifs, et sont donc insuffisants pour nous donner une indication d'espace ; pourtant nous sommes capables de percevoir un *effet de présence*. L'effet de salle nous donne donc des indications simultanément sur la salle et la position de la source. En général, on distingue successivement dans une salle :

- le son direct,
- les premiers échos,
- les réflexions tardives.

Les durées et les amplitudes respectives de toutes ces phases sont des critères qui nous aident à juger de la proximité (ou de l'éloignement) de la source sonore.

Critères de mouvement

Le mouvement d'une source sonore (ou du récepteur) entraîne une signature acoustique très caractéristique : l'effet Doppler. Si la source et le récepteur se rapprochent l'un de l'autre, les sons se décalent vers les aigus. Si ils s'éloignent, les sons se décalent vers les graves. C'est le même effet, appliqué aux ondes lumineuses, qui nous permet de mesurer les vitesses d'éloignement des astres par rapport à la Terre.

Ce document a été formaté par htmlpp.

Perception du timbre

Par exclusion, on nomme *timbre* tout ce qui ne relève ni de l'intensité, ni de la durée, ni de la hauteur, ni de la perception de l'espace.

Espace des timbres

La mesure de l'espace des timbres consiste à demander à des auditeurs de juger la *dissemblance* entre des sons d'instruments calibrés (même intensité, même hauteur...), en la notant sur une échelle de 1 à 10, 10 représentant deux sons très différents (une trompette et un piano), 1 représentant deux sons très semblables (saxophone et clarinette). Le problème consiste à trouver une interprétation géométrique où chaque instrument est un point et la distance séparant deux points correspond au jugement de dissemblance entre les deux instruments. Le premières tentatives mettent en oeuvre une distance euclidienne (la distance commune) comme mesure de la distance entre deux points, et conduit à utiliser un espace à 3 ou 4 dimensions (selon les expériences) pour représenter les timbres. Les dernières études tendent à prouver que des effets de catégorisation s'ajoutent à ce jugement de distance, et qu'il est nécessaire de faire intervenir un autre type de distance, une distance *p*-adique (mesure de la distance séparant deux feuilles dans un arbre) pour prendre en compte cet effet de catégorisation.

Les trois premières dimensions perceptives de l'espace des timbres ont été expliquées de la façon suivante :

- temps d'attaque (la qualité de l'attaque est primordiale pour reconnaître un son d'un autre. Pierre Schaeffer dans les années 50 a mis en évidence que le son du piano, sans l'attaque percussive du marteau sur la corde, n'était pas reconnu comme un son de piano),
- le *centre de gravité spectral* (la hauteur spectrale),
- le *flux spectral* (mesure de l'évolution du spectre avec le temps).

Caractéristiques spectrales

Certaines caractéristiques spectrales sont associées à certains types de timbre :

- inharmonicité des partiels : plusieurs hauteurs peuvent être entendues ; caractéristique des cloches d'église, et des sons *métalliques*,
- disparition d'un partiel harmonique sur deux : les clarinettes (tube cylindrique et anche simple) ne sont pas capables de générer d'harmoniques paires dans leur spectre, tous les sons possédant cette caractéristique peuvent sans doute se rapprocher d'un son de clarinette,
- absence des premiers partiels : un son creux,
- décroissance de 6dB par octave des partiels : son agressif et nasillard,
- peu de partiels harmoniques, décroissance de plus de 18dB par octave : un son rond.

Parole et formants

Conclusion

Le timbre est une mesure très subjective, prenant en compte des caractéristiques fréquencielles, mais également temporelles, voire même d'autres natures. En particulier, un son, pour être vivant, doit être modulé, doit vibrer (vibrato, trémolo...). La nature de ces modulations (des micro-variations du son) doit être mise en relation avec le geste de l'instrumentiste qui joue également une grande part dans notre perception du timbre des instruments de musique, mais qui est difficilement quantifiable.

Page remise à jour le
Ven 26 Déc 1997 15:37:15

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Modélisation

Enjeux

Les enjeux de l'analyse/synthèse dépassent largement le (pas si) simple but de créer des synthétiseurs musicaux. De façon générale, l'étape de *modélisation* est préalable pour :

- comprendre la nature des signaux acoustiques (comment ils ont été générés, comment ils ont été propagés, comment ils ont été mesurés ou perçus),
- trouver des *représentations* efficaces pour la transmission, le stockage et la compression,
- pour prévoir le comportement des structures,
- pour contrôler des modifications (i.e. pour prévoir l'effet des modifications),
- pour sortir du champ des possibles.

Introduction

Un modèle est une *représentation conceptuelle* de la nature d'un système que nous ne pouvons qu'observer et mesurer. Un modèle par nature n'est qu'une vue de l'esprit, qui, à l'aide d'équations, de raisonnements mathématiques, et de quelques postulats de base, tente d'expliquer les observations que nous faisons du monde physique.

Dans une première étape dite d'*analyse*, un modèle réduit les observations en un certain nombre de paramètres et de constantes. Nous appelons ces paramètres, dans notre domaine, des *contrôles*. Par la suite, la *synthèse* consiste, entre autre, à vérifier que les paramètres de contrôle ainsi que les lois d'évolution du modèle permettent effectivement de prendre en compte les observations originales. La différence entre la synthèse et l'original s'appelle l'*erreur de modélisation*. Dans beaucoup de cas, il s'agit de faire un compromis entre la concision du modèle (le nombre de paramètres nécessaires pour expliquer l'observation) et l'erreur de modélisation.

signal original - analyse - contrôle - synthèse - signal de synthèse

Il peut être intéressant d'insérer dans ce schéma une phase de modification / transformation au milieu.

Exemples

Quelques exemples d'application de ce schéma :

- numérisation : le signal original est le signal analogique, l'analyse devient l'étape de conversion analogique-numérique, la synthèse, l'étape inverse de conversion numérique-analogique, l'erreur de modélisation correspond finalement au repli spectral (aliasing), et au bruit de quantification.
- synthèse de la parole très bas-débit : le signal original est le signal de parole, les signaux de contrôle se résument à la transcription écrite du discours original. Dans ce cas, l'analyse prend la forme de la *reconnaissance automatique de la parole* ; la synthèse, de la *synthèse de la parole à partir du texte* (text-to-speech synthesis). Dans ce cas l'erreur entre l'original et la synthèse mesure seulement les distorsions sémantiques du discours (ambiguïtés de sens...) puisque les nuances d'expression, sont libres d'interprétation par le lecteur,
- le codage psychoacoustique met en oeuvre un modèle du récepteur (par opposition aux modèles

d’émission ou de propagation),

- la partition musicale : l’original consiste en un morceau de musique traditionnel, les signaux de contrôle deviennent tout simplement des notes et des indications sur la partition. Là encore, l’aspect de l’interprétation musicale n’est pas pris en compte par le modèle.
- synthèse par modèle physique : le signal original est par exemple le couple du trompettiste et de sa trompette, les contrôles deviennent alors des mesures du geste instrumental, ce qui inclut la partition jouée, mais également, les mouvements des lèvres, la façon de respirer...

À part dans le premier cas, tous les types de modélisation sont très complexes : elles font intervenir des connaissances de nature sémantiques qui sont très difficiles à modéliser. Dans certain cas, la partie synthèse existe sans que forcément la partie analyse existe.

Si la synthèse est utilisée sans la contrepartie de l’analyse, il faut trouver des dispositifs physiques de contrôle en adéquation avec la synthèse. Le plus connu, dans le domaine musical est le clavier.

Dans le cas des effets sonores (réverbération, distorsion, limiteur...) qui ne visent qu’à modifier le son, l’effort de modélisation est moindre, et le schéma d’analyse/synthèse ne s’applique pas forcément. Nous les incluons toutefois dans ce diagramme, quitte à considérer par signaux de contrôle les modifications apportées au signal.

Comme on vient de le voir, la nature des paramètres de contrôle peuvent tout à la fois être très abstraits (timbre, hauteur, partition...) ou très techniques (suite d’échantillons, codage psychoacoustique, mouvement des lèvres). On admet en général que la quantité des contrôles doit être moins grande que celle du signal original (application de compression), mais ce n’est pas une règle toujours respectée, notamment pour des opérations de transformations subtiles du signal.

Il faut distinguer, dans la partie de synthèse, le synthèse théorique, celle que décrit les lois mathématiques d’évolution, et celle faite en pratique avec des composants électroniques. Nous traiterons de la pratique de la synthèse dans un autre chapitre.

Page remise à jour le
Ven 26 Déc 1997 15:37:15

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Effets sonores

Introduction

Les effets sonores sont inclus dans la partie synthèse sonore, essentiellement parce que certaines utilisations musicales les utilisent tels quels pour sortir du *champ des possibles*, c'est-à-dire, soit comme un élément du discours musical, soit pour rajouter de l'expressivité à un instrument naturel - dans ce cas, nous parlons d'hyper-instruments - (la trompette de Jon Hassel, le saxophone de Jan Garbareck, mais on peut également parler de guitare électrique, de violon électrifié...).

Un effet consiste à modifier un son existant. En général, le schéma qui s'applique au cas des effets est le suivant, avec la condition que si aucun contrôle n'est appliqué sur le son, celui-ci n'est pas modifié :

(son original + contrôle) - effets - son modifié

Description d'un rack d'effets

Les termes adoptés pour qualifier les effets couramment utilisés sur un rack d'effets sont empruntés de l'anglais sans qu'aucun effort n'ait jamais été poursuivi pour les franciser.

Contrôle de la dynamique

expansion/compression:

Historiquement, le contrôle de la dynamique a été développé comme une solution à l'enregistrement de sources sonores dont la dynamique dépassait facilement les 80dB, sur un support magnétique qui n'en supportait pas plus de 50. Dans ces conditions, la solution ad-hoc adoptée fut la suivante : baisser le volume sonore quand le niveau sonore est trop fort, augmenter le niveau sonore quand la source est trop faible. Cette étape s'appelle la *compression*. L'opération inverse est connue sous le nom d'*expansion*.

Un mécanisme de compression ou d'expansion est entièrement décrit par la méthode choisie pour mesurer le niveau sonore et par la courbe mettant en relation le niveau d'amplification en fonction du niveau sonore observé.

Globalement, le *Dolby* (NR pour noise reduction) connu sur les cassettes magnétiques est une forme élaborée de compresseur/expanseur. Le principe est un peu différent : le bruit du support magnétique est uniformément réparti dans les graves et dans les aigus alors que l'on constate que les signaux enregistrés ont en moyenne une pente spectrale de -6dB/octave. Cela a pour conséquence d'exhiber le bruit nuisible du support (le souffle de la cassette par exemple) dans les aigus, là où aucun signal musical ne peut le cacher. Une solution au problème consiste à développer un filtre de compression/expansion qui amplifie (de façon adaptative avec le niveau sonore) les aigus à l'enregistrement, et les atténue à la restitution.

La technique de compression est couramment utilisée sur les stations de radios commerciales, et pendant les pauses publicitaires sur les chaînes de télévision. En effet, cette réduction de la dynamique permet d'augmenter artificiellement le niveau sonore : le niveau maximal ne change pas, tandis que le niveau minimum augmente! Or, actuellement, la loi française ne réglemente que les niveaux maxima

admissibles. Donc en réduisant la dynamique, on augmente le niveau moyen sans augmenter le niveau maximum.

Le *noise-gate* est un autre système agissant sur la dynamique. En fonction du niveau sonore constaté, le noise-gate décide soit de couper complètement le signal entrant quand le niveau sonore est trop faible, soit de le laisser inchangé. Cela a pour effet d'atténuer la nuisance des bruits (bruit de souffle, bruit d'enregistrement...) en présence de blancs sur la bande.

fade-in / fade-out / crossfade :

Cette technique consiste à éviter que le son n'apparaisse ou ne disparaisse trop brusquement. Un *fade-in* consiste par exemple à monter progressivement le volume sonore de 0 jusqu'au niveau nominal dans un intervalle d'une seconde pour éviter que la musique ne brise trop rapidement le silence. Le *fade-out* est l'opération inverse, et consiste donc à réduire progressivement le volume sonore. Enfin, un *cross-fade* (ou *fondu enchaîné*) consiste à mixer deux sources sonores afin de passer progressivement de la première à la seconde en gardant approximativement le même volume sonore.

Effets temporels

Chorusing/Harmonizer :

Le *chorusing* est un effet qui permet de multiplier virtuellement une source sonore. À partir d'un seule voix chantée, il est donc possible artificiellement de créer l'impression d'un choeur.

L'*harmoniseur* procède de façon complètement différente. À partir d'un signal original, il fabrique un signal à l'octave, à la quinte... permettant ainsi d'obtenir un accord harmonique à partir d'une seule source sonore.

Réverbérations :

Actuellement, en production audio, toutes les sources sonores sont enregistrées en milieu acoustiquement neutre (dans un studio) sur des pistes séparées. L'opération de *mixage* consiste alors à regrouper toutes les sources sonores ensemble, et à leur appliquer séparément, un effet de salle différent (concert hall, cathedral, ...). On maîtrise ainsi assez précisément les facteurs de présence pour chacune des sources sonores. En d'autres termes, les effets de salle ne sont quasiment plus jamais naturels (sauf pour les enregistrements de concerts de musique classique... encore que...). Les dispositifs qui simulent ces effets de salle sont les *réverbérateurs* (*réverbs*).

Parmi tous les effets de réverbération, l'*écho* est très largement utilisé pour induire une rythmique binaire. À chaque impulsion (par exemple de batterie ou de guitare basse), l'écho renvoie une autre impulsion moins forte quelques dixièmes de secondes plus tard et ainsi de suite périodiquement.

Modulation en anneau:

Le *modulateur en anneau* tire son nom du montage électronique utilisé pour multiplier deux signaux analogiques. Autant, l'opération de multiplication est simple dans le domaine des signaux numériques (si 23 est la valeur de l'échantillon du premier signal, 8 la valeur de l'échantillon du second signal, alors 8*23=184 est la valeur de l'échantillon du signal résultant de la multiplication des deux signaux), autant l'opération est délicate en électronique analogique. Le montage consiste en un pont de diodes que tous les électroniciens connaissent sous le nom de modulateur en anneau.

Quand les deux signaux multipliés sont quelconques, le signal résultant est du bruit, sans grand intérêt.

Si un des signaux est une sinusoïde (la *modulante* ou la *porteuse*), alors de la modulation résulte un signal dont toutes les fréquences se trouvent décalées à gauche et à droite (vers les graves et les aigus) d'un montant correspondant à la fréquence de la modulante. En conséquence, si le signal original était périodique (harmonique), alors le résultat à toutes les chances de ne plus l'être (et donc de devenir inharmonique). Ce peut être un moyen de rajouter de l'inharmonicité, ou de la rugosité dans un son.

Si la modulation concerne un signal original et un signal très basse fréquence (une *enveloppe temporelle*), alors le résultat aboutit à une modulation en amplitude, c'est-à-dire au rajout d'une enveloppe temporelle, qui peut induire des effets sonores intéressants, comme par exemple une rythmique...

Effets fréquenciels

Égaliseur :

Un *égaliseur* (ou *équaliseur*) consiste en un dispositif (un banc de filtres) qui permet d'amplifier ou d'atténuer sélectivement des gammes de fréquences. Sur les égaliseurs graphiques, chaque curseur correspond au niveau d'amplification pour un intervalle de fréquences précis. Un égaliseur n'est rien d'autre qu'un filtre dont le gain en fréquence est réglable graphiquement.

Ce dispositif équipe un certain nombre d'équipements HIFI de salon. Professionnellement, un égaliseur permet de corriger certaines résonances, certaines colorations que des salles de diffusion induisent sur le son. Ce dispositif permet donc d'obtenir le même rendu sonore dans des salles différentes par égalisation de la réponse sonore.

La résolution des égaliseurs est qualifiée en général d'octave, de demi-octave ou de tiers d'octave. Un égalisateur en tiers d'octave possède par exemple trois filtres différents pour le réglage du gain dans la bande de fréquence 400 - 800Hz. Des trois résolutions, l'égalisateur en tiers d'octave est celui qui a la plus fine.

Flaging/Phasing :

Il semble qu'historiquement, le *flanging* ait été mis en oeuvre par des DJs, en tentant de synchroniser deux mêmes disques vynils sur des platines différentes. Pour des raisons mécaniques, la synchronisation parfaite n'est pas possible. En plus de cette dé-synchronisation, il existe toujours un *jitter*, car les vitesses de rotation des deux platines ne sont pas stables. L'un dans l'autre, le son résultant du mixage des deux platines étaient qualifié de *phasy*, et intéressait énormément les DJs. Depuis lors, l'effet peut être reproduit artificiellement très simplement par un *flanger*.

Le flanging et le phasing consistent à colorer artificiellement un son en sommant deux versions déphasées du même signal. Un effet assez étrange de *rotation* résulte de la modification périodique du déphasage.

Il est plus facile de colorer les sources sonores qui ne possèdent pas de hauteur sonore clairement définie. L'utilisation la plus courante consiste donc à appliquer le flanging, exclusivement, aux pistes de batterie et de percussion. L'effet était très prisé en musique pop-rock au début des années soixante-dix.

Effets non-ordinaires

Au delà des effets ordinaires, banals, courants, que tout le monde utilise depuis une trentaine d'années se rajoutent à présent des techniques plus sophistiquées transformant la nature même du son. Souvent, ces techniques s'appuient sur des modèles de signaux dont nous parlerons au chapitre suivant.

Vocoders

En principe, un *vocodeur* est un dispositif générique permettant de coder puis de re-synthétiser la voix, dans le but d'une transmission efficace. Toutefois, les premiers dispositifs déformaient considérablement la voix, lui donnant une texture particulièrement *robotique*. Aussi, les premiers dispositifs furent-ils utilisés dans des applications musicales, profitant ainsi des effets apportés au timbre de la voix.

Vocodeurs en sous-bande :

Ces vocodeurs résultent directement du modèle source-filtre (ou soustractif) décrit au chapitre suivant. Ce dispositif permet de décorreller complètement la prosodie de la voix (la mélodie du langage) des phonèmes prononcés. L'effet consiste donc en une voix de robot, dont la hauteur peut-être contrôlée indépendamment du locuteur, par exemple par un clavier.

Vocodeurs de phase :

La technologie des vocodeurs de phase est beaucoup plus subtile. Elle permet un certain nombre d'effets, tels que la *synthèse croisée*, la *dilatation temporelle*, le *changement de hauteur*...

Dilatation temporelle

Cet effet, connu également en anglais sous le nom de *time stretching*, est délicat à mettre en oeuvre, tant il est difficile de clairement définir ce que chacun attend d'une dilatation temporelle.

En principe, sur les magnétophones à bande, ou sur les platines disques, ralentir le rythme de défilement de la bande devant la tête de lecture, permet certes de ralentir la cadence du signal original, mais au prix d'un décalage vers les graves quand on ralentit, ou vers les aigus quand on accélère. Autrement dit, la hauteur et la cadence de la bande sont intimement liées. La dilatation temporelle consiste alors à ralentir un signal original, sans pour autant porter atteinte à ces caractéristiques fréquencielles.

En fait, le problème est bien plus délicat, car nous n'attendons pas que tous les sons soient dilatés de la même façon. Prenons par exemple le cas de la voix. Les voyelles peuvent en effet être prononcées plus ou moins rapidement, mais ce n'est pas le cas des plosives ([p], [k], [t]...), dont la vitesse d'exécution ne peut pas être modifiée. Cela indique, dans un premier temps, que les sons ne doivent pas être rallongés uniformément, et que le résultat d'une dilatation temporelle uniforme n'a pas forcément d'équivalent naturel ou réaliste. En fait, le changement de cadence d'un son peut rapidement devenir une opération très complexe.

Changement de hauteur

Les mêmes réflexions que précédemment sont encore valables pour le changement de hauteur à savoir :

- a priori, cadence et hauteur sont indissociables,
- dans le cas de la voix, les consonnes n'ont pas (à proprement parler) de hauteur, donc, il n'y a guère de sens à changer « la hauteur » des consonnes,
- le résultat du changement uniforme de hauteur n'est pas forcément réaliste.

En pratique, les techniques de dilatation temporelle alliées à celles de changement de hauteur, permettent, à partir d'un signal original par exemple de parole, d'en changer complètement, et à volonté, la prosodie et l'évolution temporelle.

Changement de timbre

Les techniques simples de changement de timbre consistent par exemple à utiliser la sortie d'un harmonizer, associée à un égaliseur spécialement réglé pour renforcer certaines fréquences. On aboutit alors à une voix plus grave, ou plus caverneuse, ou plus nazillarde...

Changement de hauteur de la voix et consistance de la position des formants.

Toutefois, les arguments précédents concernant les transformations du signal de parole restent encore valables ici. En particulier on ne transforme pas de la même façon en parole les consonnes plosives, les voyelles, les consonnes fricatives... Pour faire des transformations de haute qualité, il faut *segmenter* le signal, c'est-à-dire à marquer la position des voyelles, des fricatives, des plosives... et appliquer différentes transformations à chacun des segments. Ce n'est pas encore (malheureusement) une tâche entièrement automatisable.

Page remise à jour le
Ven 26 Déc 1997 15:37:14

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Synthèse

Synthèses musicales

Suivant le point de vue selon lequel on se place, il est possible de définir trois types de modélisation :

- synthèse par modèles physiques : on ne modélise que le système producteur de son, c'est-à-dire l'instrument de musique lui-même (ou le système phonatoire humain dans le cadre de la synthèse de la parole). En général, il s'agit de simulations physiques pour vérifier la pertinence d'un modèle théorique. Le problème de la compression du signal n'est souvent ici qu'auxiliaire.
- modèles de signaux : on ne s'intéresse qu'à la mesure du signal acoustique, sans réellement se préoccuper de la nature de la production. C'est en grande partie les modèles utilisés en théorie du codage en télécommunications, et en synthèse musicale,
- modèles psychoacoustiques : on ne s'intéresse qu'aux caractéristiques du récepteur, c'est-à-dire de l'oreille. Cela devient de plus en plus une préoccupation majeure pour les applications de télécommunications.

Synthèses par modèle physique

Ce type de synthèse repose sur une description physique et acoustique relativement précise des mécanismes initiateurs de la vibration et du son. C'est en général une tâche très ardue que de développer un modèle physique d'un instrument, les disciplines impliquées étant nombreuses :

- l'acoustique et la mécanique pour comprendre et décrire les phénomènes physiques,
- l'automatique et le traitement du signal pour établir un schéma de discrétisation adéquat au problème, et fournir les lois de commandes adéquates au modèle afin d'assurer la stabilité numérique du système,
- l'informatique et l'électronique pour coder le synthétiseur temps-réel sur une architecture matérielle adéquate,
- la musique pour apprendre à jouer du modèle physique.

Il existe plusieurs méthodologies permettant d'obtenir des modèles physiques, le tout étant de préciser les couples de variables qui décrivent le système :

- pression entrante et sortante pour la théorie des *guides d'ondes* développée pour la première fois par J. O. Smith en 1982. C'est la description que l'on retrouve dans la plupart des synthétiseurs dits *virtuel* du marché. Dans cette description, le système excitateur est clairement dissocié du résonateur (mais toujours fortement couplé!).
- force et déplacement (ou débit) pour Modalys développé depuis 1985 à l'IRCAM. Les systèmes physiques linéaires sont dans ce cas entièrement décrits par leurs modes de résonance (description *modale*). Dans cette description, le système excitateur reste clairement dissocié du résonateur (et toujours fortement couplé!).
- force et position pour Cordis/Anima, développé depuis 1980 par Claude Cadoz au sein de l'ACROE à Grenoble. Le système est très général et permet de prendre compte beaucoup d'autres systèmes que les stricts systèmes acoustiques.

À l'exception de Modalys, tous les autres systèmes ont la possibilité de fonctionner en temps réel, et d'interagir directement avec l'instrumentiste. Nous verrons dans un chapitre ultérieur les implications que cela entraîne sur le contrôle de la synthèse.

La plupart des grands constructeurs de synthétiseur ont à l'heure actuelle dans leur gamme de produits au moins un synthétiseur fonctionnant sur le principe des modèles physiques. Pour l'instant, c'est la description sous la forme de guide d'ondes qui emporte l'adhésion des constructeurs. Le marché n'est pas encore réellement développé, mais tous les acteurs du petit monde de la musique et des constructeurs de synthétiseurs s'accordent à dire que le marché existe potentiellement, et qu'il ne manque pas grand-chose pour qu'il démarre véritablement.

Synthèses par modèle psychoacoustique

Ce sont des modèles qui sont essentiellement développés pour le stockage du signal sonore. D'un point de vue de la synthèse musicale, aucune application n'a été pour l'instant proposée. Le taux de réduction des données est considérable. Les principales normes faisant appel à ce type de codage :

- MPEG audio,
- Musicam (Digital Audio Broadcasting),
- SonyDisk,
- cassette numérique de Philips (disparu).

Ce type de représentation est terminale (puisque correspondant à ce que nous percevons). Il n'est pas question de faire autre chose que d'écouter des sons stockés sous un format psychoacoustique. En particulier, il est hors de question d'appliquer à de tels sons des algorithmes de traitement ou de modification. En effet, les artefacts du codage sont dissimulés dans le signal audio, mais après transformation (étirement temporel, filtrage, mixage, ...), ces artefacts n'ont aucune raison de rester cachés. Une transformation d'un son codé psychoacoustiquement risque de révéler des bruits de codage habituellement dissimulés!

Synthèses par modèle de signaux

Échantillonnage

Ce n'est pas à proprement parler une méthode de synthèse. Toutefois, il existe de nombreux détails qui font de l'*échantillonneur* plus qu'un simple magnétophone.

Décomposition temporelle d'un son « musical »

Traditionnellement, les sons « musicaux », du moins les sons issus d'instruments de musique électro-acoustiques, se décomposent en quatre phases distinctes, correspondant (en anglais) à :

- *attack* : c'est la phase qui correspond à la mise en action des phénomènes acoustiques générant le son. Cette phase dite *transitoire* se caractérise par une brusque montée en amplitude du signal sonore.
- *decay* : cette phase correspond à l'établissement du *régime permanent* quand il existe. Elle indique la fin des phénomènes transitoires et est caractérisée en général par une légère décroissance de l'amplitude du signal sonore qui tend à se stabiliser.
- *sustain* : cette phase n'existe que pour les instruments entretenus. C'est une phase où les caractéristiques du son restent globalement stables (on oublie pour faciliter la caractérisation de

cette phase tous les phénomènes expressifs du type vibrato, tremolo...)

- *release* : cette phase caractérise la fin des événements sonores quand la structure qui est à l'origine du son n'est plus soumise qu'à ses propres résonances. Autrement dit, c'est la décroissance terminale du son.

Les phases d'attack, decay et release sont définies par des temps caractéristiques de montée ou de descente. La phase d'attack et de decay sont également définies par leurs amplitudes relatives.

Édition d'un son échantillonné

Un échantillonneur est un appareil qui enregistre des échantillons, et qui peut les restituer à la demande (par exemple en pressant une touche d'un clavier qui lui est raccordé). En ce sens, un échantillonneur agit comme un magnétophone.

De la même façon que sur un magnétophone à bande, la variation de vitesse de lecture entraîne une variation de hauteur, sur un échantillonneur, la variation de vitesse de lecture des échantillons permet *artificiellement* de changer la hauteur (et la durée) d'un son. Ainsi un échantillonneur n'est-il autre chose qu'un magnétophone à vitesse variable. La corrélation entre les caractéristiques temporelles du son (sa durée par exemple), et ses caractéristiques fréquencielles (son timbre, sa hauteur) est la principale limitation de la technique d'échantillonnage qui ne peut prétendre à reproduire fidèlement le son d'un instrument acoustique dans toute sa variété.

Il reste toutefois un problème : la restitution des sons soutenus. Il s'agit de générer le son issu de l'enregistrement par exemple d'un saxophone tant que la touche du clavier est pressée, et d'arréter le son quand la touche du clavier se relève. Pour parvenir à cet effet, il suffit d'enchaîner convenablement les phases d'attack, decay, sustain, et release. L'appui de la touche déclenche successivement les phases d'attack, decay, sustain. On reste sur la phase de sustain tant que la touche reste appuyée. Le relâchement de la touche déclenche la phase de release.

L'édition d'un son échantillonné consiste à isoler les 4 phases précitées, pour que l'enclenchement d'une touche enchaîne convenablement des quatre phases du son.

Bouclage (looping)

Pour rester indéfiniment sur la phase de sustain, il est nécessaire d'user d'un certain nombre d'artifices. La technique traditionnelle consiste à isoler dans le son quelques périodes du signal, et de synthétiser la phase de sustain en répétant à l'infini ces périodes du signal. C'est la technique du *bouclage* (on boucle indéfiniment sur quelques périodes du son). Elle est relativement délicate à mettre en oeuvre. Les boucles de signal doivent se recoller exactement, sinon, les artefacts de synthèse sont très audibles. Cette édition se fait quasiment systématiquement à la main et à l'oreille sur les échantillonneur du commerce, bien que quelques techniques automatiques soient à présent à peu près au point dans différents laboratoires.

Piano numérique

Tous les pianos numériques du type Clavinova de Yamaha utilisent le principe de l'échantillonnage pour restituer des sonorités voisines de celle d'un piano acoustique. Il y a 5 ans, les notes d'un piano de concert était très proprement enregistrées en tiers d'octave (i.e. 3 notes toutes les octaves), pour trois vélocités différentes (piano, mezzo, forte). Le son du piano est restitué, par interpolation, et par une technique de bouclage astucieuse. Toutefois, les constructeurs restent très discrets sur les technologies employées et il est très difficile d'obtenir des renseignements utiles des documentations techniques.

Synthèse additive

Historiquement, il s'agit de la première méthode utilisée pour synthétiser des sons sur ordinateur (dès les années 60!). Toutefois, la méthode est très lourde à mettre en oeuvre, et pour l'instant, même s'il existe quelques synthétiseurs utilisant explicitement le principe de la synthèse additive, la synthèse additive n'est pas encore intégralement exploitée sur le marché des synthétiseur commerciaux.

Le modèle a déjà été décrit dans les chapitres précédents. Il s'agit de décomposer un son, en sons élémentaires, que nous avions qualifiés de *partiels* du son. Chaque partiel est représenté dans le domaine temporel par une onde sinusoïdale. La synthèse consiste donc à superposer des sinusoïdes les unes avec les autres.

L'intérêt de ce type de représentation : elle est entièrement temps-fréquence. À chaque instant, un son est caractérisée par les fréquences, les amplitudes et les déphasages respectifs de chacun de ses partiels. Autrement dit, il y a dissociation complète entre les caractéristiques temporelles et les caractéristiques fréquencielles. Toute la complexité réside dans l'analyse du son, c'est-à-dire trouver les bonnes fréquences, les bonnes amplitudes et les bons déphasages pour chaque partiel. La difficulté et la lourdeur d'analyse explique qu'il n'existe pas encore sur le marché de synthétiseurs additifs convaincants.

Ce type de synthèse reste toutefois parfaitement adapté (quand les techniques d'analyse sont convenablement maîtrisées) pour toutes les modifications subtiles du son. La voix du castrat Farinelli a été en partie synthétisée à l'aide de cette méthode. Les dilatations et contractions temporelles fonctionnent correctement avec cette technique, ainsi que les procédés de changement de hauteur ou de timbre. Le morphing de deux sons fonctionne également assez bien avec ce type de modèle.

Les stations de travail musicales des années 80, de type Fairlight, Synclavier, Korg Wavestation... utilisaient, parmi d'autres, ce type de synthèse. L'édition des paramètres de synthèse était toutefois particulièrement pénible : le seul contrôle disponible pour modifier un son, consistait à éditer à la main, à l'aide d'un crayon optique, l'évolution temporelle de chaque partiel du son.

Synthèse soustractive

C'est le prototype même de la synthèse populaire, que tout le monde utilise sans même le savoir. Elle tient sa popularité à différents facteurs :

- sa simplicité de mise en oeuvre,
- c'est une modélisation source-filtre très intuitive,
- c'est un modèle physique de l'appareil phonatoire,
- les contrôles sont également très intuitifs.

On la retrouve ici et là sous des noms différents :

- modèle auto-régressif, AR, ARMA,
- modèle source filtre,
- prédiction linéaire
- codeur CELP,
- synthèse soustractive,
- synthèse granulaire...

Le principe consiste simplement à partir d'une source sonore très riche en harmoniques (du bruit, un train d'impulsion, un signal carré ou triangulaire...), à filtrer sélectivement les fréquences, autrement dit, à sculpter l'enveloppe spectrale, d'où le nom de synthèse soustractive.

Synthèses par distorsion

Ce ne sont pas à proprement parler des modèles de synthèses, mais plus précisément des modèles de modification et d'enrichissement des sons qui doivent leur popularité à la simplicité de la technologie mise en oeuvre.

Synthèse par modulation de fréquence

La modulation de fréquence ou synthèse FM est sortie pour la première fois sous la forme d'un produit commercial en 1983, avec la série DX7 de Yamaha. Le principe consiste à *moduler* (à changer périodiquement) très rapidement la fréquence d'un oscillateur. Il suffit de connecter la sortie d'un oscillateur sur le contrôle en fréquence d'un autre oscillateur pour obtenir une modulation de fréquence.

Les contrôles possibles avec ce type de synthèse restent très génériques :

- contrôle dynamique de l'enveloppe temporelle,
- contrôle dynamique de la richesse spectrale,
- contrôle non-dynamique d'un indice d'enveloppe spectrale.

Il n'existe pas vraiment de méthode d'analyse. En conséquence, les bibliothèques de sons FM ne peuvent se faire qu'à la main, et uniquement grâce à l'expérience et à la manipulation de paramètres qui n'ont rien d'intuitifs.

Synthèse par distorsion d'amplitude

C'est une autre méthode qui permet de générer des sons relativement riches en harmonique à peu de frais. Ce n'est pas à proprement parler une méthode de synthèse, puisqu'il s'agit simplement de distordre le signal électrique. L'effet est bien connu des joueurs de guitares électriques qui utilisent des modules de distorsion pour changer la sonorité de leur instrument.

Page remise à jour le
Ven 26 Déc 1997 15:37:14

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Dispositifs de contrôle musical

Interaction homme-machine

La relation instrumentale ne peut être prise en compte que si l'instrument, naturel ou synthétique, en réaction à l'action de l'instrumentiste, réagit quasi immédiatement, de façon sonore, mais également de façon mécanique. Cela implique de faire de la synthèse en temps-réel, ou plus précisement que le retard entre l'action (celle par exemple d'appuyer sur une touche) et la réaction (le son qui en résulte) soit suffisament insignifiant pour nous ne le percevions pas.

La raison de l'importance de la loi de l'action et de la réaction dans le geste instrumental est simple : l'instrumentiste doit continuellement *adapter* sa loi de contrôle (i.e. le mouvement, le geste) en fonction de ce qu'il entend et ressent. Par exemple, un violoniste modifie en permanence le mouvement de son geste en fonction du son du violon, en fonction de la vibration de la corde qui se transmet à l'archet et qui est ressentie dans la main de l'instrumentiste, en fonction de la vibration du corps du violon qui remonte dans l'épaule de l'instrumentiste.

La première génération de contrôleurs musicaux consiste uniquement à capter le geste, et à fournir des informations aux synthétiseurs qui prennent en chargent la génération en temps-réel des sons. Une seconde génération de contrôleurs commence à apparaître sur le marché, qui permet de restituer des sensations *tactiles* (tactilo-kinesthésiques) dans les mains de l'instrumentiste. Historiquement, MIDI est devenu depuis 1981 le protocole officiel pour échanger des informations de contrôle entre les dispositifs de capture du geste (par exemple un clavier) et les dispositifs de synthèse (le synthétiseur ou l'expandeur). A priori, un contrôleur est incapable de fournir aucun son.

Dispositifs unidirectionels

Clavier

Historiquement, le premier contrôleur artificiel d'un dispositif de synthèse au succès commercial incontestable fut le clavier, qui dans une première approche, restitue une interface assez similaire à celle disponible pour les orgues, les clavecins, les pianos...

Un clavier-maître est un clavier MIDI dont la seule charge est d'envoyer des contrôles à une chaîne de synthétiseurs placés en aval, mais qui est incapable de fournir aucune sortie sonore par lui-même.

C'est encore à l'heure actuelle le dispositif le plus couramment utilisé pour contrôler des dispositifs de synthèse.

La plupart des claviers sont sensibles aux éléments suivants :

- vitesse de frappe de la touche : *vélocité* (NOTE-IN),
- durée de la frappe : *duration* en anglais,
- vitesse de relachement de la touche (NOTE-OFF).

En plus de ces caractéristiques, certains claviers réagissent à d'autres éléments, tels que que :

- l'*after-touch*

Guitare

Depuis 1978, il est possible de contrôler un synthétiseur à partir des paramètres extraits du jeu d'une guitare. Les possibilités de jeu de telles guitares étaient à l'origine très différentes de celles des guitares acoustiques. Quelques guitaristes utilisent cette interface pour jouer des synthétiseurs : Pat Metheny, Robert Fripp, le guitariste de Uzeb... Les premières véritables guitares dites MIDI sensibles à la plupart des techniques de jeu habituellement utilisées par les guitaristes sont sorties en 1993. Les premières guitares basses MIDI sortent actuellement en 1997.

Mentionnant également les interfaces de types « pédales d'effets » disponibles depuis longtemps pour les guitares électriques et qui traditionnellement ont pour rôle de piloter un module d'effets à l'aide des pieds de l'instrumentiste.

Violon

Il n'existe pas encore de produits commerciaux correspondant à un système de contrôle de type violon permettant de piloter des modules de synthétiseurs. Toutefois, quelques réalisations universitaires aboutissent à des prototypes intéressants.

Contrôleur de souffle

On compte depuis une dizaine d'années de nombreux contrôleurs MIDI de type clarinette, saxophone... On les désigne tous sous le nom générique de contrôleurs de souffle ou *breath controlers*. Ce sont tous des capteurs de pression ou de débit mesurant la pression statique à l'intérieur de la cavité buccale. L'habillage des différents dispositifs diffèrent selon que l'on recherche le doigté du saxophone, ou celui de la clarinette...

Autres

Bien d'autres dispositifs de mesure et de capture du geste ont été mis au point, pas forcément en relation avec l'interface d'un instrument naturel :

- les capteurs de mouvements (i.e. des capteurs d'accélération : des accéléromètres) et de positions permettent d'enregistrer différents types de positions. Par exemple, il existe des dispositifs permettant de capter les positions des articulations d'un danseur afin d'assujétir la musique à la danse, et non pas la danse à la musique comme c'est le cas traditionnellement. Les gants de données (*datagloves*) enregistrent la configuration des doigts, ainsi qu'éventuellement les pressions qui s'exercent sur eux.
- des capteurs vidéo (tels que Big Eyes), optiques ou radios (radio baton, theremin) permettent d'enregistrer la position d'un élément visuel ou émetteur.

Il existe beaucoup d'autres dispositifs divers et variés pour enregistrement le mouvement, pour détecter le geste. Chacun donne lieu a une pratique instrumentale nouvelle, mais difficile à maîtriser : il manque un canal de communication, celui du retour mécanique.

Dispositifs bidirectionnels

Les communautés scientifiques et musicales commencent à comprendre la nécessité d'une boucle de retour mécanique dans le corps de l'instrumentiste pour parvenir à contrôler finement les réactions de son instrument. On cite souvent l'exemple de l'individu sourd qui ne peut pas parler car il ne peut pas entendre le son de sa voix, et donc ne peut rétro-agir par rapport à ce qu'il entend et ressent. Les dispositifs mécaniques qui réagissent par contact avec l'instrumentistes sont dits *haptiques*.

D'un point commercial, il n'existe encore aucun dispositif haptique appliqué à la musique. Par contre dans d'autres domaines (le marché du jeu, de la santé, de l'interaction homme-machine, de la simulation...) il existe de nombreux dispositifs qui renvoient un effort soit dans les mains de l'utilisateur, soit par d'autres moyens (vibro-mécanique, inertiel...) :

- retour d'effort dans le volant des simulateurs de voitures d'arcade,
- simulation des vibrations mécaniques toujours dans certains simulateurs d'arcade,
- simulation cinétiques des forces centrifuges et d'accélération toujours dans certains simulateurs d'arcades et dans certaines salles de cinéma spéciales (Cinaxe),
- retour des forces de frottement et des textures pour certaines applications de type scalpel virtuel pour la chirurgie assistée par ordinateur...

Page remise à jour le
Ven 26 Déc 1997 15:37:15

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

IRCAM Centre Georges Pompidou

Dispositifs de synthèse

Oscillateurs et patches

Dispositifs matériels

Dispositifs logiciels

Page remise à jour le
Ven 26 Déc 1997 15:37:15

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.

Glossaire

Termes scientifiques

- Mécanique (1),
- Vibration (1),
- Sismologie (3),
- Acoustique sous-marine (3),
- Acoustique sous-marine (3),
- Propagation (3),
- Onde (3),
- Énergie (3),
- Intensité (3),
- Pression atmosphérique (3),
- Pression acoustique (3).
- Célérité (3).

Termes techniques

- Microphone (3),
- Haut-parleur (3),
- (3),

Page remise à jour le
Mar 4 Nov 1997 16:14:43

Tassart Stéphan
IRCAM

Ce document a été formaté par htmlpp.