

Glossaire

musical

Quelques rappels de solfège pour mieux se repérer dans ce numéro.

→ Gamme

Une gamme est une succession de notes disposées dans l'ordre des fréquences croissantes (ou décroissantes). Pour former une gamme diatonique, on utilise sept notes (*do, ré, mi, fa, sol, la, si*) de noms différents, la huitième étant la répétition de la première à l'octave supérieure (*do* dans cet exemple). Les notes – ou degrés – de la gamme diatonique ne sont pas séparées d'un même intervalle.

→ Tons et demi-tons

Les intervalles qui séparent les notes de la gamme diatonique sont soit des tons, soit des demi-tons. Les notes *do* et *ré* sont séparées d'un ton ou de deux demi-tons (un demi-ton de *do* à la note intermédiaire et un demi-ton de cette note intermédiaire à *ré*). Cette note intermédiaire est nommée *do* dièse quand elle est obtenue en élevant la note inférieure ; obtenue en abaissant la note supérieure, elle se nomme *ré* bémol.

→ Degrés de la gamme

Chaque son peut être la première note d'une gamme. Chaque degré, quelle que soit la note qui le représente, a un nom qui caractérise la position qu'il occupe dans la gamme. À cette position correspond une fonction. Ainsi, le premier degré se nomme tonique, le deuxième sus-tonique, le troisième médiate, le quatrième sous-dominante, le cinquième dominante, le sixième sus-dominante, le septième sensible et le huitième octave ou tonique.

→ Accords et ensembles de notes

Un tétracorde est une succession de quatre sons (*do, ré, mi, fa*). Par extension, un hexacorde est une succession de six sons. Un accord est un ensemble de sons entendus simultanément. Un accord, par exemple *do-mi-sol*, peut subir ce que l'on nomme des renversements. Ainsi, dans le premier renversement, *mi* devient la note la plus basse (*mi-sol-do*), et dans le deuxième renversement, c'est le *sol* (*sol-do-mi*).

→ Mode

Le mode indique une disposition particulière des intervalles de la gamme. La musique tonale s'organise en deux modes, le majeur et le mineur. La gamme de *do* majeur est le modèle des gammes majeures, la gamme de *la* mineur, celui des gammes mineures. Dans la gamme de *do* majeur, les demi-tons sont situés

entre la troisième note et la quatrième (*mi* et *fa*) et entre la septième et la huitième (*si* et *do*). Dans la gamme de *la* mineur, les demi-tons sont situés entre la deuxième note et la troisième (*si* et *do*), entre la cinquième et la sixième (*mi* et *fa*) et entre la septième et la huitième (*sol* dièse et *la*).

→ Rythme

La mesure, le rythme et le mouvement fixent le déroulement temporel d'un morceau. La mesure est la division en parties égales d'un morceau de musique. Une mesure contient plusieurs notes et des silences ; elle est délimitée par des barres de mesure. Une mesure se subdivise généralement en deux, trois ou quatre temps. Tous les temps d'une mesure n'ont pas une égale importance. Dans une mesure à deux temps, le premier est un temps fort, le second un temps faible. Dans une mesure à trois temps, le premier est fort, le deuxième et le troisième sont faibles. Dans une mesure à quatre temps, le premier et le troisième sont forts, les deux autres faibles. Le rythme est caractérisé par les durées des notes et des silences dans une phrase musicale. La multitude des combinaisons possibles est une des principales richesses de la musique. Le mouvement est la vitesse à laquelle on doit exécuter un morceau : c'est le tempo.

→ Tessiture

La tessiture est l'étendue moyenne des hauteurs de notes que l'instrumentiste (ou le chanteur) couvre avec aisance. Ce terme suppose l'absence de proesse dans la formulation de la note. L'étendue complète d'un instrument ajoute à la tessiture des notes extrêmes dans le grave ou dans l'aigu.

→ Timbre

Un son a quatre qualités : la durée, la hauteur, l'intensité et le timbre. Le timbre traduit la perception que l'on a des sons. Il fait que l'on perçoit un son sourd, sec, moelleux, doux, métallique, cristallin, etc.

→ Fréquence

Dans une onde sonore, la fréquence correspond au nombre de périodes par unité de temps. Elle se mesure en hertz. Elle correspond à la perception auditive de la hauteur. Plus la fréquence est élevée, plus le son est aigu. L'oreille humaine perçoit entre 20 et 20 000 hertz. Le *la* du diapason a une fréquence de 440 hertz.

✓ Interprétation

Outre sa terminologie technique, l'écriture musicale utilise aussi des expressions spécifiques indiquant aux musiciens comment interpréter l'œuvre. Les termes, généralement indiqués en italien sur les partitions, sont expressifs : *affettuoso*, *agitato*, *con anima*, *con espressione*, *con spirito* (avec esprit), *grazioso*, *maestoso* (majestueux), *moderato*, *vivace*, voire *vivacissimo*. Ces termes évoquent les émotions que le musicien doit transmettre à l'auditeur.

RENTÉE 2008/2009

[L'EXAM DANS LA POCHE]

[MINI-MANUEL]

« Apprendre et comprendre facilement »

[A à Z]

« Découvrir toute une discipline de A à Z »



[MAXI-FICHES]

« Retenir l'essentiel et réviser facilement »

[EXPRESS]

« Comprendre et s'entraîner facilement »



[LES TOPOS] « Le savoir en deux mots »



Catalogue complet www.dunod.com

En vente en librairie





La musique en

Victor A. Stoichita et Bernard Lortat-Jacob

La musique ne peut pas être étudiée isolément, car elle est indissociable du contexte social où elle est pratiquée.

Ce qu'est la musique relève apparemment de l'évidence. On le sait par intuition et même les scientifiques les plus exigeants ne semblent pas s'inquiéter de sa définition. Mozart, c'est de la musique bien sûr. Le bruit du TGV, non. Qu'en est-il du « free jazz », de la techno ou du rap ? Dans les journaux et dans la rue, certains n'y voient que du bruit, et les avis commencent à diverger.

Le problème est ancien, par exemple, Berlioz comparait la musique chinoise à des hurlements de chats. Obsessionnellement attelée à la tâche généreuse de réconcilier tout le monde, la science mit un certain temps à lui trouver une réponse. La plus communément admise, celle proposée par l'anthropologue britannique John Blacking (1928-1990) en 1973, consiste à considérer la musique comme du « son humainement organisé ». Mais cette définition n'exclut pas la langue. Elle n'écarte pas non plus les cloches des vaches, les klaxons, les sonneries des téléphones portables, les indicateurs radiophoniques, ni même le TGV, dont les acousticiens travaillent à aménager les propriétés sonores pour qu'elles satisfassent l'oreille d'un son rassurant, confortable, plaisant enfin. Si le son musical se définit de cette façon, il nous faut revoir nos pratiques scientifiques, qui accordent une place dès lors disproportionnée à ses manifestations les plus prestigieuses !

À peine suffisante pour reconnaître notre propre musique, l'intuition l'est encore moins pour aborder et définir celle des autres. Dans une grande partie du monde arabe, peut-on parler de musique dans une situation où le chant est essentiellement un embellissement de la parole, et la musique instrumentale une imitation du chant ? Dans le Nord du Canada, les « jeux de gorge » sont, pour les Inuit, des jeux, au même titre que les jeux de balle, mais sans rapports avec les chants des chamanes, qui pour

eux sont d'une autre espèce. Les Pygmées Aka de la forêt centrafricaine utilisent des formes complexes de jodel (l'équivalent des yodels des Alpes) pour communiquer durant la chasse, mais n'y voient de la musique que lorsqu'une pulsation régulière sous-tend la polyphonie. En Roumanie, les lamentations funéraires ont beau être versifiées selon un rythme régulier et portées par des courbes mélodiques bien identifiables, elles ne sont pas classées parmi les chants (*cântece*), mais plutôt parmi les pleurs (*bocete*).

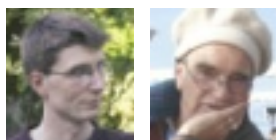
Les limites de l'intuition

Face à ce problème, certains spécialistes proposent d'abandonner purement et simplement le concept de musique ou d'en restreindre l'usage aux traditions où les acteurs eux-mêmes l'emploient (dans la tradition classique occidentale et dans certaines musiques populaires d'Europe, d'Amérique du Nord, de l'aire latino-américaine, pour l'essentiel). D'autres, au contraire, proposent de redéfinir la notion, et de l'aborder comme une forme d'action.

Dans la tradition classique occidentale, il est courant de distinguer les « œuvres » (matérialisées par des partitions écrites) de leurs « interprétations ». Les apprentis musiciens (et les musicologues) étudient les premières comme des objets finis, dûment référencés. Mais lorsqu'on la rapporte à l'ensemble des pratiques musicales observables dans le monde, cette manière de séparer les choses devient un cas particulier. Ce qui se donne à voir est le plus souvent une interaction, en partie sonore, dans un système de vastes échanges impliquant les sons bien sûr, mais aussi des objets, d'autres humains, des animaux, des divinités, etc.

Dans le Haut Atlas marocain par exemple, l'*ahwash* est non seulement une forme musicale exécutée collectivement, mais aussi la fête dans laquelle elle s'inscrit. Il est en fait bien difficile de séparer la musique, comme objet, de la fête, comme processus. Lorsque celle-ci « ne prend pas », on dit communément non pas que l'*ahwash* n'a pas réussi, mais, plus radicalement, qu'il n'a pas eu lieu. Les tambours

LES AUTEURS



Victor A. STOICHITA est chargé de mission à la Cité de la musique, à Paris. Bernard LORTAT-JACOB est directeur de recherche au CNRS.

action



résonnèrent pourtant, et l'on chanta. Ici, la musique n'apparaît pas comme un « objet », posé au milieu de la collectivité, mais comme une forme d'interaction qui l'infuse tout entière. D'autres exemples indiquent qu'à l'échelle du monde, cette façon de voir les choses n'est nullement anecdotique.

Dans la forêt tropicale de Nouvelle-Guinée, les Kaluli chantent leurs émotions (surtout les plus tristes) en imitant le chant de certains oiseaux qui les entourent. Aux yeux des Kaluli, ces derniers reflètent les âmes des morts, en chantant eux aussi leur nostalgie à proximité de certains lieux marquants de la vie collective (près du village, de certains cours d'eau, de certaines clairières, etc.). Dans un écosystème sonore et spirituel aussi étroitement imbriqué, aucun ensemble ne se prête à être isolé comme « de la musique », au sens que le terme recouvre habituellement. L'anthropologue Steven Feld, de l'Université du Nouveau-Mexique, propose donc d'abandonner cette notion, et de suivre plutôt la façon dont les Kaluli construisent et ressentent leurs émotions, en s'appuyant sur l'acoustique, la topographie et l'ornithologie vernaculaires.

D'autres anthropologues préfèrent conserver l'usage du concept de musique, en s'appuyant sur l'idée que l'homme dispose d'une capacité à détacher les sons de leurs causes physiques manifestes. Cette capacité serait à la fois commune et propre à l'espèce humaine. Comme le rappelle le philosophe britannique Roger Scruton : « Pour entendre la musique, il faut être capable d'entendre un ordre qui ne nous informe pas sur le monde physique, qui se tient à l'écart de l'enchaînement habituel des causes et des effets, et qui est irréductible aux réalités qui lui ont donné physiquement naissance. » Une telle aptitude est exploitée dans toutes les sociétés humaines, et il est improbable que les animaux la partagent (alors même qu'ils sont eux aussi sensibles aux sons et aux informations qu'ils véhiculent).

Toutefois, le fait que la musique soit une représentation mentale, requérant des aptitudes hautement spécialisées, n'implique pas pour autant qu'elle

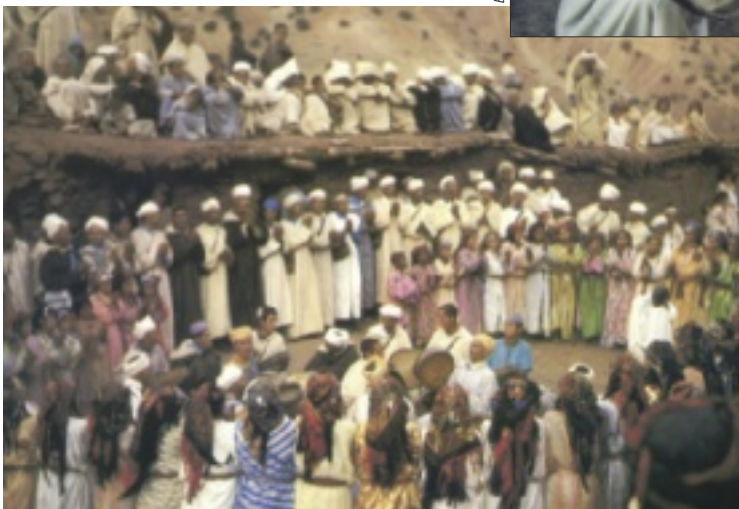
soit autonome. L'anthropologie a précisément pour vocation de relever les liens profonds qu'elle entretient avec les autres réalités de la vie sociale. S'y incarnent des normes et des valeurs collectives, mais aussi (et, surtout), elle est un moyen d'action. Le champ musical se dote ainsi d'une grande extension, qui recouvre toutes les organisations impliquant le sonore et auxquelles les humains attribuent des effets sur d'autres humains, des animaux, des divinités ou l'ordre cosmique.

D'une façon générale, l'anthropologie musicale s'attache à montrer les liens qui unissent les expressions sonores d'un groupe à des événements rituels, à des périodes de l'année, à des occasions festives, à des pas de danse, etc. Présentée ainsi, sa démarche semble fondamentalement distincte de celle des sciences exactes. L'expérimentation implique au contraire de mesurer et contrôler avec minutie les paramètres sonores et la manière dont les sujets interagissent avec eux. En d'autres termes, les sciences exactes paraissent étudier la musique « hors contexte », tandis que les sciences humaines et sociales la verraient « en contexte ». Mais cette opposition est fallacieuse, car – la sociologie des

1. L'AHWASH,
dans le Haut Atlas
marocain, est une forme
musicale collective,
mais aussi la fête
où elle s'inscrit.



Bernard Lorfaï-Jacob



Bernard Lorfaï-Jacob



2. CERTAINS RITUELS du Kerala, en Inde, associent des dessins au sol (à gauche), des offrandes et de la musique (à droite).

sciences l'a montré – une expérience de laboratoire est elle aussi saturée de déterminants. Elle met en œuvre des réseaux complexes d'humains et de machines – appareils de mesure, opérateurs, outils statistiques, locaux, revues, crédits, etc. – qui chacun participe à la construction du fait de science.

Ce que sait en tout cas l'anthropologue, c'est que la musique – sa production, son audition et toute opération qui vise à en appréhender la structure – mobilise toujours un ensemble d'humains et souvent aussi d'objets. Postuler une autonomie du son musical est une vue de l'esprit – selon nous étroite –, et ce d'autant plus que l'anthropologie contemporaine aspire à s'ouvrir à des pratiques nouvelles, centrées sur les techniques de reproduction elles-mêmes facilitées par l'utilisation du son numérique. Aisé-ment dupliqué, réutilisable de multiples façons, un même enregistrement peut alors se prêter à un jeu d'expériences esthétiques théoriquement illimité, chacune de ses écoutes offrant un cadre de compréhension renouvelé. La musique techno par exemple, selon Mark Butler, de l'Université de Pennsylvanie, à Philadelphie, s'écoute aussi bien sur une piste de danse qu'en voiture ou chez soi. La soumettre à une analyse musicologique n'est jamais qu'une façon parmi d'autres de l'écouter et de la comprendre. Cette diversification des contextes d'écoute fait écho aux pratiques de mixage des disc-jockeys – les DJ – qui donnent à écouter conjointement des disques différents, faisant émerger des propriétés musicales nouvelles.

Du fait des interactions complexes dans lesquelles elle s'inscrit, il est souvent difficile de déterminer les effets que la musique aurait en propre. Au Kerala par exemple, en Inde, les rituels propitiatoires destinés aux divinités serpents combinent des offrandes de nourriture et de musique (textes, chants et musique instrumentale) jouée par des musiciens spécialistes, qui réalisent également à cette occasion des

dessins de sol représentant les serpents grâce à du sable coloré. L'efficacité du rituel repose sur l'intime fusion de l'ensemble, dont la musique est un élément essentiel, mais qui ne déploie ses effets qu'en lien avec les autres. Elle ne se prête donc pas à être isolée comme un « stimulus » auquel correspondraient des effets spécifiques. Un tel découpage ne permettrait de comprendre ni pourquoi ni comment les formes sonores varient, et il ne rendrait pas non plus compte du pouvoir dévolu aux rituels où elle intervient. Pour comprendre ces aspects fondamentaux de la musique, il est nécessaire de la considérer comme un mode de relation et d'action.

Un mode de relations

En Transylvanie non plus, la musique ne peut être réduite à un simple stimulus : les musiciens professionnels tsiganes de la région savent qu'il est possible de faire pleurer l'assistance, dans un enterrement, avec des mélodies de danse destinées en principe aux occasions festives, comme les mariages. Au chevet d'un mort, l'allégresse et les souvenirs évoqués par une mélodie des jours heureux peuvent provoquer une nostalgie bien plus grande que les airs lents et tristes que l'on associe normalement à ce genre d'événement. Si la même forme musicale provoque des réactions à ce point contrastées, c'est que son efficacité dépend, là encore, autant du réseau où elle apparaît et qui lui donne un sens (ici le mort, ses proches, la communauté villageoise, le souvenir des autres morts...), que de ses caractéristiques structurelles intrinsèques.

Fondée sur des cas particuliers, l'hypothèse se prête à être généralisée. Elle stipule que la musique est une construction active, dotée d'une forme d'efficacité au sein d'un ensemble de relations sociales. Cette façon de voir les choses est commune à de nombreux travaux menés actuellement par des anthropologues. Ils prennent acte des difficultés à rechercher des universaux dans une musique arbitrairement isolée, et tendent à explorer les relations que cet objet, sous toutes ses formes et avec tous ses sens, entretient avec les pratiques sociales et l'imaginaire des Hommes. ■

✓ BIBLIOGRAPHIE

F. Bonini-Baraldi, *All the joy and pain of the world in a single melody. A Transylvanian case-study on musical meaning and emotion*, Music perception, à paraître, 2009.

C. Guillebaud, *Le chant des serpents. Musiciens itinérants du Kerala*, CNRS éditions, 2008.

V. A. Stoichita, *Fabricants d'émotion. Musique et malice dans un village tsigane de Roumanie*, Publications de la Société d'ethnologie, Nanterre, 2008.

R. Scruton, *The Aesthetics of Music*, Clarendon Press, Oxford, 1997.

B. Lortat-Jacob, *Musiques en fête*, Publications de la Société d'ethnologie, Nanterre, 1994.

B. Lortat-Jacob, *Musiques en fête*, Publications de la Société d'ethnologie, Nanterre, 1994.

musique & technique

revue professionnelle de la facture instrumentale

musique & technique

aborde le champ de la facture instrumentale dans ses différentes composantes professionnelles, technologiques, musicologiques et économiques. Cette revue francophone offre une tribune d'expression aux acteurs rassemblés autour de l'instrument de musique, du fabricant au distributeur, du musicien au preneur de son, du scientifique au gestionnaire culturel.

Rendez-vous annuel autour d'un dossier thématique, **musique & technique** propose des articles de fond afin d'aborder l'environnement de l'instrument : conception, fabrication, pratique, histoire, contexte et enjeux économiques...

musique & technique est édité par le Pôle d'innovation des métiers de la musique.

Le Pôle d'innovation des métiers de la musique

est un service d'appui, d'information et de conseil. Il s'adresse aux entreprises de la facture instrumentale : fabrication, réparation, restauration, accord des instruments de musique.

Mission spécifique de l'**Institut technologique européen des métiers de la musique (Itemm)**, son action s'inscrit au cœur des enjeux musicaux, environnementaux et économiques: rareté des bois et matériaux précieux, recherches de formes et de sonorités nouvelles, conception assistée par ordinateur, prototypages, ouverture des marchés à l'international, développement des services de conseil aux musiciens, analyse de la qualité sonore des instruments en piano, instruments à vent, guitare, lutherie, orgue, instruments anciens et traditionnels...

Le pôle d'innovation est membre du réseau national des pôles d'innovation coordonné par l'Institut supérieur des métiers.



Institut
technologique
européen
des métiers
de la musique

Le Mans - France

Pôle d'innovation des métiers de la musique
71 avenue Olivier Messiaen - 72000 LE MANS
tél. 02 43 39 39 00 - e mail : innovation@itemm.fr

Site Internet : www.itemm.fr



n°3 (2008)

déjà parus



n°0 (2005)

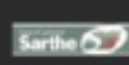


n°1 (2006)



n°2 (2007)

L'Institut technologique européen des métiers de la musique est soutenu par :



FEDER FSE

Faire vibrer l'air avec

François Gautier, Jean-Loïc Le Carrou et Vincent Doutaut

Du frottement d'une corde jusqu'au rayonnement acoustique de l'instrument entier, les physiciens élucident les propriétés de la harpe pour développer des outils d'aide à la facture instrumentale.

Le 9 mars 1967, un étrange instrument de « musique » est inventé... par Gaston Lagaffe dans le journal *Spirou*. Ce gaffophone tient plus de l'engin de démolition par les nombreux dégâts que ses vibrations ont entraînés que de la harpe dont il est pourtant une sorte d'avatar. À l'occasion d'un concours, quelques amateurs se risquèrent à construire de vrais gaffophones, mais aucun n'a intégré d'orchestre sympho-

objectifs, d'autres reposent sur la perception subjective de l'instrument par l'auditeur. Pourtant, les implications économiques font de la qualité un enjeu majeur, car elle justifie souvent des différences de prix établis.

Les instruments de musique résultent d'un processus d'optimisation, par approximations successives, sur lequel les contraintes économiques pèsent en permanence. Par exemple, aujourd'hui, les accords internationaux de protection de la faune et de la flore contraignent la production d'instruments nouveaux, en limitant par exemple les prélèvements de pernamboucs (un bois brésilien) pour les archets ou de palissandres pour les guitares et les percussions.

Ces contraintes entraînent un besoin de matériaux alternatifs et d'innovations qui ne peuvent venir que d'outils scientifiques. Ils sont aussi nécessaires pour améliorer la compétitivité des entreprises artisanales. Cela place l'étude de la physique des instruments de musique au cœur d'un processus conduisant à une facture performante. Nous prendrons comme exemple un instrument à cordes pincées, la harpe. Peu étudiée, elle a néanmoins fait l'objet d'analyses récentes dans notre laboratoire. Ainsi, après avoir rappelé comment fonctionne un instrument à cordes,

nous détaillerons les spécificités de la harpe. Enfin, nous verrons en quoi la compréhension de la physique de l'instrument est utile au luthier.

Un modèle à cordes

Les instruments à cordes sont classés selon la façon dont le son est produit : on distingue ainsi les instruments pour lesquels des sollicitations sont entretenues (la corde frottée du violon) et ceux à cordes libres qui peuvent être frappées (le piano) ou pincées (la harpe ou la guitare). Dans ce dernier cas, les oscillations résultent d'une perturbation de l'état initial et constituent un phénomène quasi périodique, s'amortissant avec le temps. Rappelons que les vibrations d'une corde sont la superposition de contributions élémentaires, nommées modes, dont les fréquences sont des multiples entiers d'une fréquence dite fondamentale. Chaque mode de vibration est doté de caractéristiques, tels sa fréquence propre et son coefficient d'amortissement qui traduit la rapidité avec laquelle la corde retrouve son état initial (avant pincement).

Le rôle du luthier est de concevoir un dispositif mécanique robuste qui transforme en son la vibration de la corde. En effet, la corde est, seule, quasi incapable de mettre en mouvement l'air, et donc de créer

TOUT GENTILHOMME DOIT AVOIR un coussin sur sa chaise, une femme vertueuse et une harpe bien accordée.

Proverbe irlandais du XII^e siècle

nique, les harpistes préférant sans doute des instruments dont la qualité a fait ses preuves, notamment ceux de facteurs renommés, tel *Camac* en France.

Ces instruments haut de gamme sont le plus souvent fabriqués de façon artisanale. Mais comment définit-on la qualité d'un instrument ? Il s'agit d'une notion globale qui recouvre le rendu sonore, l'expressivité ou la facilité de jeu, et associe des critères d'ergonomie, de solidité, de tenue dans le temps, notamment en ce qui concerne l'accord de l'instrument. Certains de ces attributs sont



un son audible. Elle est donc associée à un système vibrant plan, une plaque nommée table d'harmonie (voir la figure 1). Cet élément a un rôle essentiel, celui du rayonnement acoustique, mécanisme par lequel la structure vibrante (principalement la table) transmet son mouvement à l'air environnant, mouvement ensuite propagé de proche en proche jusqu'à nos oreilles.

Cependant, pour une plaque seule, les rayonnements des deux faces créent des contributions presque opposées qui ont tendance à s'annuler. L'introduction d'une cavité diminue ce phénomène de court-circuit acoustique en réduisant le rayonnement de la face arrière de la plaque. Enfin, l'aménagement de trous dans cette cavité, des ouïes ou des événements, produit une résonance des basses fréquences qui augmente l'importance du champ rayonné dans ce registre de fréquences. La plupart des instruments à cordes sont construits sur ce modèle, dont nombre de paramètres sont ajustables par le luthier. Le choix de l'artisan est un équilibre entre son intuition et des règles issues de l'analyse du fonctionnement acoustique des instruments.

Examinons maintenant une harpe. Connue depuis l'Antiquité sous sa forme actuelle (elle est représentée sur un sceau sumérien), elle a néanmoins beaucoup évolué et donné lieu à de multiples modèles selon la forme, la taille et le nombre de

cordes. La grande harpe, ou harpe de concert, utilisée dans l'orchestre symphonique, est constituée de 47 cordes, fixées entre la console, en haut, et la table d'harmonie, en bas. Cette dernière est constituée d'une plaque faite de plusieurs couches, en bois et en fibres de carbone, dont la composition exacte est à la discrétion du luthier.

Couplages sympathiques

Cette plaque, trapézoïdale, est renforcée par plusieurs raidisseurs et par une barre centrale métallique. L'arrière de la table est raccordé à une caisse de résonance tronconique, tandis qu'une colonne soutient la console à l'avant. La difficulté d'accorder l'instrument est connue, tant les tensions des cordes rendent l'équilibre mécanique précaire. Un jeu de sept pédales reliées à un système de tiges articulées près de 1 500 pièces mécaniques différentes permet de modifier la longueur vibrante de chaque corde, et donc sa fréquence fondamentale.

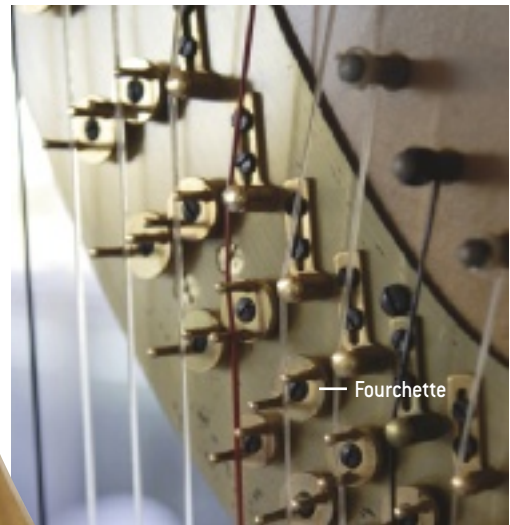
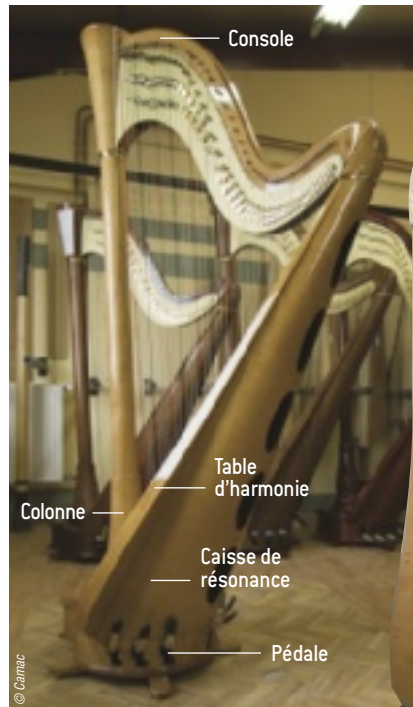
Autre particularité de l'instrument, les nombreuses cordes créent des résonances par sympathie: la vibration d'une corde donnée entraîne, essentiellement par couplage solide, c'est-à-dire *via* les points d'insertion des cordes et l'ensemble de l'instrument, la vibration de ses voisines selon des amplitudes variables. En effet, le cou-

L'ESSENTIEL

✓ Le nombre de cordes d'une harpe et les tensions qu'elles exercent influent sur le son de l'instrument.

✓ Les phénomènes vibratoires des cordes et des éléments de la caisse de résonance sont étudiés par les physiciens qui en identifient les caractéristiques principales.

✓ Les physiciens mettent au point des outils d'aide à la facture instrumentale qu'utilisent les luthiers pour fabriquer des instruments plus reproductibles.



1. LA HARPE DE CONCERT est constituée de 47 cordes tendues entre la console et la table d'harmonie, celle-ci étant adossée à une caisse de résonance ajourée d'évents. Une colonne maintient l'instrument. Les pédales actionnent des fourchettes qui, selon leur position, modifient la longueur vibrante des cordes.

plage entre des cordes est d'autant plus grand qu'il existe des relations simples entre les fréquences mises en jeu : une corde excite facilement les cordes situées à l'octave (le rapport des fréquences de vibrations est égal à 2) ou à la quinte de l'octave supérieure (le rapport est de 3/2). Ces couplages sympathiques engendrent un phénomène de halo sonore, qui est une signature acoustique de l'instrument. Trop important, ce halo est gênant et peut rendre l'instrument injouable. Le luthier doit donc le maintenir dans une limite raisonnable.

Du fait de ces couplages sympathiques, parmi d'autres, les sons produits ont un grand nombre de composantes spectrales, dont les fréquences sont parfois proches. Un autre type de couplage est dû à l'existence de deux polarisations pour les cordes, correspondant à deux mouvements vibratoires, l'un dans le plan des cordes de la harpe et l'autre dans un plan perpendiculaire passant par la corde jouée.

Ces différents couplages donnent lieu à des phénomènes de battements et de décroissances multiples. De quoi s'agit-il ? Les battements sont des fluctuations lentes de l'amplitude du son dues à des interférences destructives et constructives entre des composantes spectrales de fréquences proches. Le phénomène de décroissance multiple apparaît quand plusieurs composantes du son sont amorties différemment : certaines sont dominantes au début du son, mais s'éteignent plus rapidement que d'autres, qui perdurent et sont dominantes à la fin du son (le son est alors dit rémanent). Voyons maintenant le mouvement vibratoire de la table d'harmonie, élément essentiel du rayonnement acoustique.

L'analyse de la harpe a montré que, pour les premiers modes, ceux dont les fréquences propres sont les plus basses, le volume de la caisse de résonance reste constant. Ces modes, sans grande importance acoustique, correspondent à des mouvements vibratoires de la caisse autour de ses points de fixation.

Les modes de l'instrument

En revanche, les modes de fréquences plus élevées entraînent un mouvement de flexion important de la table d'harmonie et modifient le volume interne de la cavité ; ils engendrent un rayonnement acoustique notable. C'est le cas de deux modes nommés T_1 (le premier mode de la table, ou mode de table) et A_0 (le premier mode de l'air, aussi nommé mode d'air ou mode de Helmholtz). Les modes de table mettent en jeu principalement des vibrations de la table d'harmonie, tandis que les modes d'air correspondent à des mouvements importants des masses d'air situées au voisinage des événements ; on parle de pistons d'air.

Ces deux modes sont aisément identifiables sur une guitare : en bouchant l'évent avec un bloc de mousse, le mode A_0 disparaît ; en ajoutant un poids sur la table, la fréquence du mode T_1 est décalée vers les graves.

Le mode A_0 amplifie les basses fréquences, car il donne lieu à une résonance qui relève les niveaux acoustiques autour de sa fréquence. Cet effet, nommé *bass-reflex*, est utilisé dans certaines enceintes acoustiques. Son rôle est important dans le cas de la guitare classique, où les modes A_0 et T_1 ont des fréquences voisines respectivement de 100 et 200 hertz. Petite, la table d'harmonie de la guitare

rayonne mal, c'est-à-dire de façon inefficace, les fréquences inférieures à environ 100 hertz, cela à cause du phénomène de court-circuit acoustique. Aussi, accorder A_0 à cette fréquence renforce le rayonnement de la guitare dans une plage de fréquences où la table seule fonctionne mal.

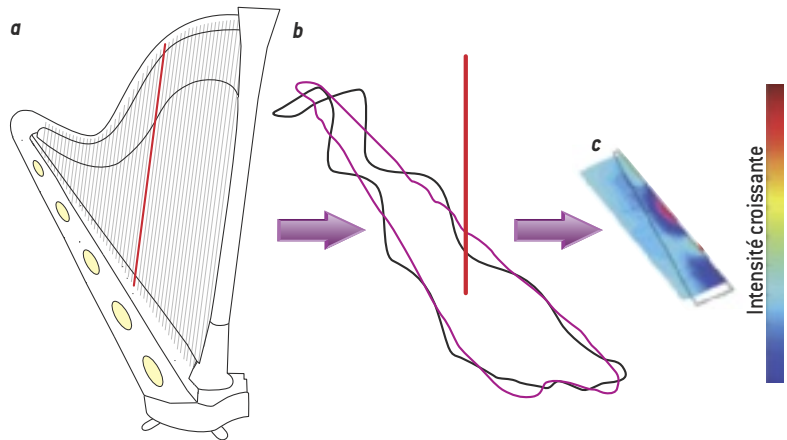
Un instrument en cours d'optimisation

Si la résonance du mode A_0 est une caractéristique acoustique importante pour une guitare, qu'en est-il pour une harpe ? Des expériences lors desquelles on a bouché tous les événements et ajouté des masses sur la table d'harmonie d'une harpe ont permis d'identifier A_0 et T_1 à des fréquences d'environ 170 et 150 hertz. Comment cela se traduit-il ? L'effet *bass-reflex* est moins marqué que dans le cas de la guitare, car les phénomènes d'amortissement sont plus importants et la présence d'autres modes que A_0 et T_1 rend la réaction de la table plus complexe et masque l'effet. Et surtout, la fréquence de A_0 est trop élevée ! Dans la harpe, le rôle attendu de l'effet *bass-reflex* serait de compléter le rayonnement en basses fréquences de la table, là où le court-circuit acoustique est important. Pour cela, la fréquence de A_0 devrait être faible, et au moins inférieure à celle de T_1 . Cette condition n'est pas remplie : la harpe n'utilise donc pas l'effet amplificateur que l'on pourrait attendre de A_0 .

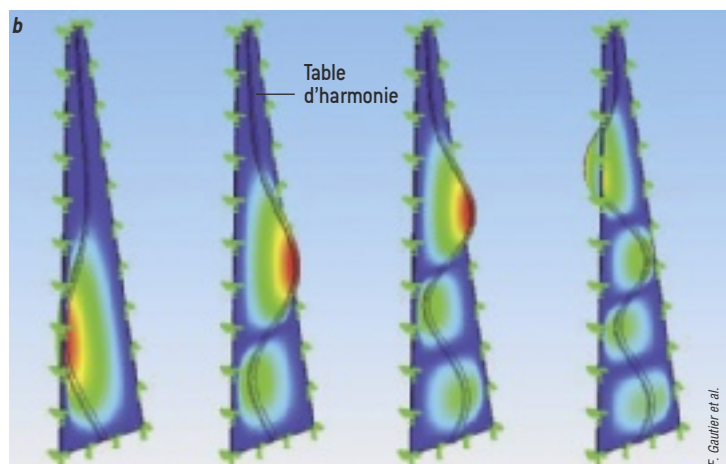
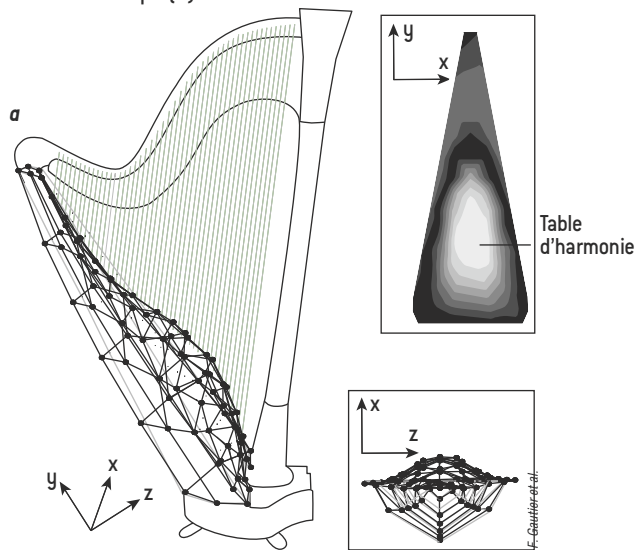
De fait, dans une harpe, les événements sont d'abord conçus pour faciliter... un accès aux cordes ! En outre, le voisinage des événements est modifié par la présence du corps de l'instrumentiste. Dans une harpe, l'effet *bass-reflex* n'est donc pas encore optimisé, ou plutôt le processus, opéré par l'histoire de la musique, est ici inachevé. Cela peut s'expliquer par le nombre de harpes fabriquées, qui est bien moins important que le nombre de guitares, pour lesquelles le processus d'optimisation est plus abouti. La harpe est donc un instrument en évolution.

Comment les cordes et la table sont-elles couplées ? Une image peut aider à répondre. L'énergie qu'une corde transmet à la table est comparable à l'eau d'un seau que l'on vide dans une mare. Quand le seau est vidé rapidement, la mare est agitée, ce qui traduit un couplage fort : la vibration et donc le son produit sont de grandes amplitudes, mais plutôt brefs. À l'inverse, quand le seau est vidé lentement, le couplage est faible : le son créé est modéré, mais dure longtemps.

Dans une harpe, ce degré de couplage est mesuré par l'admittance. Il s'agit du rapport, dans le domaine des fréquences, entre la vitesse vibratoire en un point (celle de son déplacement périodique selon un axe perpendiculaire au plan de la table) et la force appliquée pour entraîner cette vitesse. Ce paramètre dépend de la fréquence de la corde et de sa position sur l'axe où sont fixées les cordes : on peut donc représenter l'admittance sous forme d'un plan position/fréquence où les cordes sont notées par des



2. LA PRODUCTION D'UN SON avec une harpe se déroule selon plusieurs étapes : le doigt excite la corde [a, en rouge] qui communique son mouvement à la table d'harmonie [b, en noir] et crée un rayonnement sonore, visible sur la carte d'intensité acoustique [c].



3. L'ANALYSE MODALE d'une harpe consiste à déterminer ses modes de vibration, c'est-à-dire les fréquences et les champs vibratoires associés. On identifie ainsi le champ vibratoire de l'instrument, associé au premier mode d'air, nommé mode de Helmholtz et noté A_0 [a]. On identifie aussi les champs vibratoires correspondant aux premiers modes de la table d'harmonie [b].

LES AUTEURS



François GAUTIER est professeur au Laboratoire d'acoustique de l'Université du Maine (CNRS UMR 6613) et à l'École nationale supérieure d'ingénieurs du Mans.

Jean-Loïc LE CARROU est maître de conférences à l'Institut Jean Le Rond D'Alembert dans l'équipe Lutheries-acoustique-musique (CNRS UMR 7190) de l'Université Pierre et Marie Curie, à Paris.

Vincent DOUTAUT dirige le pôle d'innovation de l'Institut européen des métiers de la musique, au Mans.

points. Pour ces derniers, l'abscisse est la position du point d'attache et l'ordonnée, la fréquence fondamentale de la corde. L'ensemble de ces points constitue une trajectoire (voir la figure 4).

La carte d'admittance

Quelles informations en tire le luthier ? Les variations de l'admittance le long de cette trajectoire constituent un indicateur de l'homogénéité de l'instrument. Lorsque le long de la trajectoire l'admittance varie trop ou de façon chaotique, les conditions de couplage entre les cordes et la table sont alors très fluctuantes d'une corde à l'autre : certaines cordes sonnent trop, et d'autres trop peu. L'artisan en déduit alors les différences d'amplitude et de décroissance entre les cordes. Grâce à ces renseignements, il peut diagnostiquer les notes donnant lieu à un son de forte amplitude, mais s'atténuant rapidement (couplage fort) ou un son d'amplitude plus faible, mais durant plus longtemps (couplage faible). Cette caractéristique, importante pour le musicien, est ainsi mesurée objectivement par des paramètres physiques pertinents.

Le luthier influe sur la carte d'admittance, sans le savoir, en modifiant la géométrie de la table et par le choix des matériaux. La trajectoire est quant à elle fixée par les fréquences de jeu et les positions des points d'accrochage.

Le physicien Chris Waltham, de l'Université de Vancouver, au Canada, a développé une interprétation physique de cette carte d'admittance. La table d'harmonie est moins épaisse près de la console qu'à proximité du pied de la colonne. C'est pour-

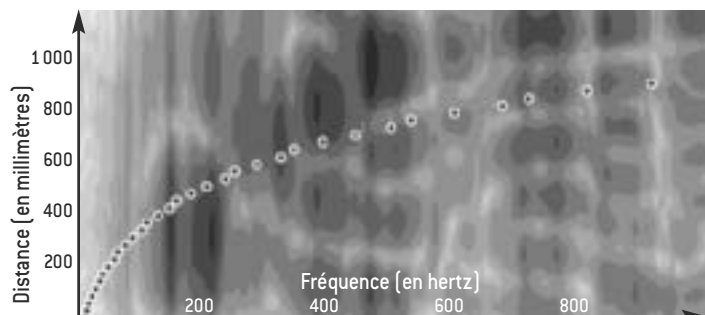
quoi, parmi les « ventres » des différents modes de flexion de la table (les régions pour lesquelles le champ vibratoire de ce mode est maximal), celui qui est dans la partie supérieure de l'instrument (la plus souple) a une amplitude supérieure à celle de ses voisins. Ce ventre, plus marqué que les autres, est d'autant plus haut sur l'instrument que l'ordre de ce mode est élevé.

La carte d'admittance met en évidence ce résultat : quand la fréquence augmente, le nombre de ventres visibles augmente (un ventre pour une fréquence de 200 hertz et six à 850 hertz), les régions pour lesquelles l'admittance est la plus élevée (les régions noires de la carte de la figure 4) correspondant aux ventres de grande amplitude. Du point de vue acoustique, ce phénomène a une conséquence importante en termes de directivité.

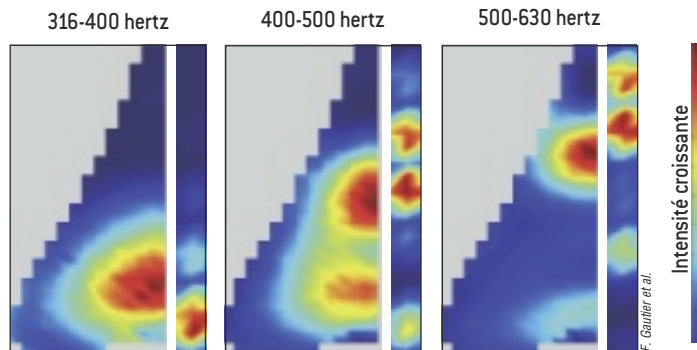
En effet, le rayonnement acoustique de l'instrument peut être décrit à l'aide de sources élémentaires équivalentes nommées monopoles acoustiques. La répartition dans l'espace de ces monopoles révèle les zones actives de l'instrument, c'est-à-dire les sources acoustiques du système. Les études menées dans notre laboratoire montrent que les sources acoustiques équivalentes à l'instrument se déplacent vers sa partie supérieure lorsque la fréquence augmente.

L'ingénierie mécanique a notablement évolué depuis 50 ans grâce aux modèles numériques qui mettent en évidence les effets de modifications structurelles et évitent la construction de prototypes longs et coûteux à réaliser. La lutherie utilise aujourd'hui ces outils pour guider ses évolutions.

Les artisans disposent ainsi d'un prototype virtuel dont la conception est assistée par ordinateur. Toutefois, la connaissance



4. LA CARTE D'ADMITTANCE d'une table d'harmonie de harpe décrit le degré de couplage entre cordes et table [plus il est fort, plus la zone est noire], qui dépend de la fréquence et de la position sur l'axe de l'instrument. Chaque corde est représentée par un point, dont l'abscisse est la fréquence fondamentale de la corde et l'ordonnée la position de son point de fixation.



5. CARTE D'INTENSITÉ ACOUSTIQUE rayonnée par une harpe de concert pour trois bandes de fréquences (pour chacune, à gauche, la table d'harmonie et à droite, les événements). En basses fréquences, le son est surtout émis par la partie inférieure, la plus large. Quand la fréquence augmente, les régions de forte intensité acoustique se déplacent vers le haut.

LE DOIGT ET LA CORDE

Pour émettre un son, le musicien pince la corde de l'instrument. Cette opération modifie l'état initial et met en oscillation des cordes, ce qui produit un son. Dans cette première étape, l'interaction du doigt et de la corde joue un rôle essentiel, car l'instrumentiste contrôle de nombreux paramètres rendant l'excitation dépendante de sa morphologie. Pour la harpe de concert, le pincement de la corde s'effectue par la pulpe du doigt et peut se décomposer en deux étapes avant le lâcher. D'abord, le musicien tire la corde jusqu'à exercer une tension maximale, puis laisse glisser la corde sur sa première phalange. Cette façon d'attaquer la corde est caractéristique du son de chaque harpiste, ce qui rend possible l'identification de l'instrumentiste dès les premières notes jouées. Pour approfondir notre connaissance sur ce pincement et en caractériser les paramètres physiques discriminants, nous avons mis au point un modèle de cette interaction tenant compte de l'élasticité du doigt et du frottement (similaire à celle d'un archet sur une corde de violon). Le geste instrumental étant rapide, nous avons obtenu les paramètres d'un tel modèle à l'aide d'une caméra rapide. Les déplacements du doigt et de la corde au cours du pincement ainsi mesurés, nous avons estimé les paramètres de l'interaction : l'élasticité du doigt et les caractéristiques de frottement, variables selon les instrumentistes, rendent compte de certaines différences acoustiques que l'on peut percevoir entre les musiciens. Là encore, le modèle physique permet d'éclairer le fonctionnement de l'instrument.



© C. Luzzati

des caractéristiques mécaniques d'un instrument, qui sont les données qu'utilise ce type de modèle, n'est pas toujours aisée. En effet, les paramètres mécaniques du bois varient souvent selon les échantillons et dépendent en outre des conditions hygrométriques et thermiques. Les caractéristiques des liaisons mécaniques, tels les collages, ainsi que les précontraintes introduites dans l'instrument lors de la fabrication sont également difficiles à quantifier. Néanmoins, les informations fournies par ces modèles sont riches et révèlent des quantités physiques difficilement mesurables, comme les transferts ou les couplages entre certaines parties du système. Le prix à payer pour faire « chanter les équations » est le temps de calcul.

L'aide à la facture

Les caractéristiques vibratoires et acoustiques d'un instrument sont utiles pour l'analyse de son fonctionnement. Ces caractéristiques sont accessibles *via* nombre d'outils, mais il s'agit le plus souvent d'instruments de laboratoire difficiles à mettre en œuvre chez les artisans luthiers.

Aujourd'hui, des outils métrologiques dédiés, à coût réduit, portables et robustes sont développés pour répondre aux besoins de la profession. À titre d'exemple, cette

démarche a été entamée au Laboratoire d'acoustique de l'Université du Maine (LAUM) depuis plusieurs années, en particulier grâce à des projets d'élèves de l'École nationale d'ingénieurs de l'Université du Maine (l'ENSIM) en lien avec le pôle d'innovation de l'Institut technologique européen des métiers de la musique (l'ITEMM) et les activités de l'Union nationale de la facture instrumentale (l'UNFI). Pour le luthier, quatre applications sont visées.

D'abord, l'établissement d'une sorte de carte d'identité technique, qui facilite la traçabilité des interventions tout au long de la vie de l'instrument. En cas de restauration ou de réparation, les comportements d'origine peuvent ainsi être retrouvés et l'on mesure la qualité d'une intervention, par exemple le revernissage. Ensuite, l'artisan s'assure de la reproductibilité d'un procédé de fabrication en archivant des fréquences modales à différents stades. Les outils métrologiques aident également au choix du matériau. Enfin, on peut confectonner des copies d'un instrument de référence, non pas uniquement en copiant sa géométrie, ses matériaux et son aspect, mais en reproduisant ses modes de vibration.

Si Gaston Lagaffe avait eu vent de ces outils, son gaffophone n'aurait sans doute pas été utilisé pour mettre en déroute des chasseurs de baleines! ■

✓ BIBLIOGRAPHIE

J.-L. Le Carrou, *Vibroacoustique de la harpe de concert*, Musique et Technique, Ed. ITEM, Le Mans, 2008.

C. Waltham *et al.*, *Vibrational characteristics of harp soundboards*, Journal of the Acoustical Society of America, vol. 124 [3], pp. 1774-1780, 2008.

J.-L. Le Carrou *et al.*, *Experimental study of A_0 and T_1 modes of the concert harp*, Journal of the Acoustical Society of America, vol. 121 [1], pp. 559-567, 2007.

A. Chaigne, *Comprendre la guitare acoustique*, Pour la Science, Dossier n° 52, pp. 74-75, juillet 2006.

I. Brémaud, *Diversité des bois utilisés ou utilisables en facture d'instruments de musique*, Thèse de doctorat, Université de Montpellier II, 2006.

Les cuivres :



Frederick P. Matzen/Shutterstock

L'ESSENTIEL

✓ La qualité sonore des trompettes et autres cuivres tient surtout à leur perce, la forme de la colonne d'air qui se trouve dans l'instrument. On pourrait obtenir des sons similaires avec des tubes... en bois.

✓ Les physiciens peuvent aider à concevoir la perce à l'aide d'algorithmes d'optimisation, de simulations numériques, et de descripteurs sensoriels établis avec des musiciens.

Miles Davis, Chet Baker, Wynton Marsalis... Tous les grands trompettistes vibrent avec leur musique. Sous l'effet de la puissance sonore, ils sentent sous leurs doigts les vibrations de leur instrument. À tel point, parfois, que le public « vibre » avec eux. De même, dans le registre classique, tous les auditeurs passionnés d'*Aïda*, l'opéra de Verdi, connaissent et ressentent le son « cuivré » des trompettes célébrant le triomphe des troupes égyptiennes. Verdi avait vu juste, car les Égyptiens, probables inventeurs de la trompette voilà 3 000 ans si l'on en croit la découverte de plusieurs exemplaires de cet instrument dans la tombe de Toutankhamon (1345-1327 avant notre ère), faisaient un usage guerrier de cet instrument, sans doute en raison de sa puissance sonore. Comme les Grecs, les Romains et les Celtes, qui aimaient défiler aux sons du salpinx, de la buccina et du carnyx.

Cet usage militaire s'est transformé au fil des siècles. Monteverdi et Bach ont

su introduire les cuivres avec brio dans leur musique. Aujourd'hui, la majorité des instruments à vent de type cuivre sont constitués de laiton, un alliage de cuivre et de zinc. Les plus courants dans les orchestres sont, outre la trompette, le trombone à coulisse, le cor, le tuba, l'euphonium et le saxhorn. Contrairement aux instruments à vent de type bois, ils présentent seulement deux ouvertures, l'embouchure, la pièce métallique sur laquelle le musicien pose ses lèvres, et le pavillon, qui émet le son vers l'extérieur. Certains cuivres plus anciens, comme le serpent, le cornet à bouquin et l'ophicléide, ont également des trous latéraux.

Au premier abord, on peut penser que le matériau est la cause principale de leurs qualités sonores. Or les recherches en acoustique ont montré, au cours du XX^e siècle, qu'il ne joue qu'un rôle secondaire sur le son produit. En adaptant une embouchure de trombone sur un tuyau de plastique de plusieurs mètres, un instrumentiste

le son et la forme

Joël Gilbert, Jean-François Petiot et Murray Campbell

Une meilleure connaissance du fonctionnement intime des instruments, mais aussi des préférences des musiciens, donne un souffle nouveau à l'art de concevoir trompettes et autres cuivres.

expérimenté obtient un son « cuivré » semblable à celui d'un instrument !

En réalité, un cuivre se définit, non par son matériau, mais par la façon dont le son est émis : la colonne d'air à l'intérieur de l'instrument – la perce – est mise en résonance par la vibration des lèvres du musicien. Aussi classe-t-on parmi les cuivres des instruments non métalliques, telle la conque marine, ou réalisés en bois, comme le cornet à bouquin et le serpent. Le laiton utilisé généralement dans les cuivres possède cependant des qualités intéressantes : il est notamment très ductile, donc relativement facile à mettre en forme. Lisse et non poreux, il permet d'obtenir une surface d'excellente qualité.

La perce se définit mathématiquement comme l'évolution du rayon intérieur de l'instrument en fonction de la ligne moyenne allant de l'embouchure au pavillon. Elle a une très grande influence sur le comportement de l'instrument. Indépendamment des qualités esthétiques de l'instrument, le fabricant de cuivres – le

facteur – doit donc surtout ajuster la perce en fonction du résultat musical désiré. Comment y parvient-il ? Les physiciens acousticiens peuvent-ils l'y aider en optimisant les paramètres qui influent sur les qualités sonores ? Intègrent-ils alors les caractères plus subjectifs qui influent sur les choix des musiciens ? Après avoir résumé les caractéristiques de la perce, nous examinerons les méthodes et les recherches grâce auxquelles les acousticiens contribuent à la déterminer, et son rôle essentiel dans la production des sons cuivrés – la cuivrabilité, qualité spécifique des cuivres. Enfin, nous reviendrons sur la contribution des parois de l'instrument au rayonnement sonore.

Reprenons notre trompette. Son registre courant – les notes couvertes du grave à l'aigu – s'étend sur deux octaves et demie, du *fa dièse* grave au *do* au-dessus de la portée. En faisant varier la conformation de ses lèvres, son « masque », et son souffle dans l'embouchure, le musicien expérimenté obtient, pour un même



UNE TROMPETTE CELTE, le carnyx.

LES AUTEURS



Joël GILBERT est directeur de recherche CNRS au Laboratoire d'acoustique de l'Université du Maine (LAUM UMR CNRS 6613), au Mans. Jean-François PETIOT est professeur à l'École centrale de Nantes et à l'Institut de recherche en communications et cybernétique de Nantes (IRCCyN - UMR CNRS 6597). Murray CAMPBELL est professeur d'acoustique musicale à l'École de physique de l'Université d'Édimbourg, en Écosse.

L'optimisation de la perce par algorithmes génétiques

Les algorithmes d'optimisation sont des procédures mathématiques ou numériques qui permettent de trouver le ou les minimums d'une fonction. Parmi eux, les algorithmes génétiques reposent sur deux postulats de la théorie darwinienne de l'évolution : premièrement, dans chaque environnement, seules les espèces les mieux adaptées perdurent ; deuxièmement, au sein de chaque espèce, le renouvellement des populations est essentiellement dû aux « meilleurs » individus de l'espèce, c'est-à-dire aux individus les mieux adaptés au milieu.

Chaque variable d'optimisation (le rayon de la perce à une position donnée du tube) est tout d'abord intégrée dans un vecteur x dit vecteur d'optimisation. Dans un premier temps, il faut coder ce vecteur pour pouvoir le manipuler numériquement. Ce codage permet de constituer le « chromosome », composé de

« gènes ». Un codage binaire des rayons de la perce est un exemple de chromosome possible.

Ensuite, on travaille avec des populations de solutions candidates à produire itérativement de nouvelles populations à l'aide de trois principes d'évolution : le croisement (les solutions sont croisées entre elles et donnent à leur « enfant » une partie de leur patrimoine génétique) ; la mutation (certains gènes peuvent être modifiés) ; et la sélection (on choisit par exemple de ne retenir que les « meilleures » solutions au sens du critère souhaité pour l'itération suivante). Des milliers d'itérations sont réalisées par ordinateur pour produire finalement un ensemble de « bonnes » solutions au problème. Une telle méthode donne des perces optimisées parmi lesquelles on retrouve les solutions déjà connues, mais aussi des perces inédites.

Cuivrabilité et classification

✓ Outre son importance musicale, la « cuivrabilité » est devenue un critère original de classification des cuivres. En collaboration avec le musée d'Édimbourg, nous avons proposé de l'utiliser pour répertorier les cuivres des collections. Pour caractériser ce paramètre, les organologues (les spécialistes de l'étude des instruments de musique) utilisent une grandeur calculée selon une formule mathématique intégrant les dimensions de la perce et la valeur de la fréquence fondamentale jouable par l'instrument. Cette grandeur, issue de considérations d'acoustique non linéaire, est facile à estimer dès lors que la perce de l'instrument est connue. Ainsi, la grandeur en question a été évaluée pour des centaines de cuivres. Par ailleurs, il est possible de suivre l'évolution de cette grandeur pour un instrument donné au cours des siècles, de comparer des cuivres d'une même famille ou un type d'instrument provenant de différents pays... Bref, de donner un nouvel outil d'analyse aux spécialistes de l'histoire des cuivres et de leur facture.

doigté sur les trois pistons de l'instrument, des notes nommées partiels. Elles sont proches d'une série harmonique, c'est-à-dire proches d'une suite de fréquences multiples d'une fréquence fondamentale.

La colonne d'air intérieure est excitée et mise en résonance par les vibrations des lèvres du musicien, une succession rapide d'ouvertures et de fermetures de la bouche. La maîtrise de cette technique, le *buzz* (le bourdonnement), nécessite des années de pratique. Lorsque le musicien y parvient, la pression variable que son *buzz* exerce dans l'embouchure excite la colonne d'air dans l'instrument ; celle-ci se met à vibrer, puis rétroagit sur les lèvres, qui excitent à nouveau la colonne d'air, jusqu'à l'obtention d'un son stable. En configurant son masque, le musicien produit l'un des partiels correspondant à un doigté particulier. Les physiciens associent d'ailleurs souvent les cuivres et les instruments à anche (clarinette, saxophone, hautbois, basson, etc.), car le *buzz* a le même rôle que la vibration de l'anche : moduler le souffle du musicien, et produire un son par « effet valve ».

Les cuivres fonctionnent donc grâce au couplage entre les lèvres du musicien, le système excitateur, et la colonne d'air contenue dans l'instrument, le résonateur, qui propage et amplifie l'onde sonore émise. La première caractéristique du résonateur est son impédance d'entrée acoustique : l'amplitude de sa réponse à une excitation donnée ; autrement dit, le rapport, à l'entrée de l'instrument, entre la pression acoustique, la variation de pression par rapport à la pression atmosphérique, et le débit acoustique, c'est-à-dire la variation de débit par rapport au débit

moyen. Lorsque l'on mesure l'impédance, on constate que son amplitude est maximale pour plusieurs fréquences particulières, les fréquences de résonance (voir la figure 1). Ces dernières correspondent aux partiels les plus faciles à émettre avec l'instrument.

Jouer sur la perce

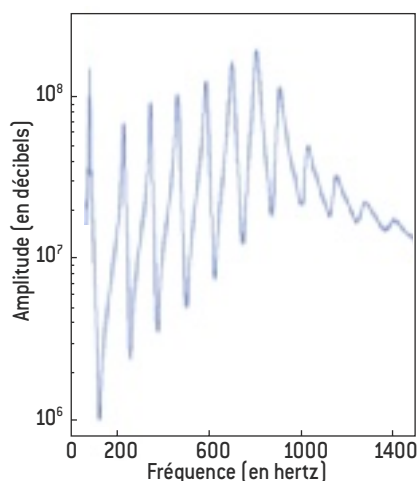
Puisque c'est la perce qui nous intéresse, comment faire pour la concevoir au mieux ? On peut d'abord considérer qu'elle doit être telle que les différentes fréquences de résonance soient très proches d'une série harmonique. Les facteurs de cuivres résolvent ce problème empiriquement grâce à des ajustements progressifs et à leur savoir-faire. Ils déterminent la forme intérieure qui minimise globalement l'« inharmonicité » des fréquences de résonance de l'instrument, mais aussi en tenant compte des qualités de timbre et d'émission. La perce est ainsi le résultat de compromis réalisés par le facteur entre plusieurs objectifs techniques, esthétiques et musicaux.

Réaliser une perce afin de fabriquer un cuivre demande bien sûr des compétences sur le travail des matériaux. La mise en forme d'un pavillon nécessite en effet des dizaines d'opérations, telles que découpage, martelage, recuit, repoussage, etc. Celle des tubes est effectuée généralement par étirage du tube sur un mandrin, puis cintrage. Les dimensions de la perce sont ajustées au moyen du ou des mandrins sur lesquels le métal est étiré ou repoussé.

Aujourd'hui, les acousticiens peuvent contribuer à la conception de la perce en proposant des modèles mathématiques

prédictifs et des simulations informatiques. Ces méthodes permettent d'optimiser divers paramètres physiques pour obtenir la réponse souhaitée. Ainsi, si les variables d'optimisation de la perce sont les rayons du résonateur à certaines positions, on peut modéliser ce dernier comme une suite de cônes et de cylindres ayant chacun un rayon particulier. Dès lors que l'on connaît les valeurs souhaitées des fréquences de résonance, la conception de la perce revient à rechercher les valeurs des rayons des différentes sections du tube qui minimisent l'écart des fréquences de résonance de l'instrument avec celles souhaitées. L'une des méthodes d'optimisation utilise ce que l'on appelle les algorithmes génétiques (voir l'encadré page ci-contre).

Bien sûr, l'approche par optimisation du critère « harmonicité » n'est pas sans défauts. D'abord, il existe huit doigtés possibles pour obtenir le registre de deux octaves et demie d'une trompette. Il est difficile d'obtenir une série de résonances harmoniques pour tous ces doigtés. En effet, le nombre de variables dimensionnelles sur lesquelles le facteur peut agir



1. LA CARACTÉRISTIQUE PHYSIQUE essentielle de la colonne d'air, ou résonateur, qui propage le son dans un cuivre est l'impédance d'entrée acoustique : le rapport entre la pression acoustique à l'embouchure et le débit acoustique. Mesurée sur une trompette, elle apparaît maximale pour certaines fréquences, les fréquences de résonance, qui correspondent aux « partiels » de l'instrument, c'est-à-dire aux notes les plus faciles à émettre.

pour corriger la justesse est limité par la forme de l'instrument ; par exemple, la majeure partie de la perce de certains cuivres est cylindrique. Le facteur parviendra difficilement à ajuster tous les doigtés avec le même jeu de variables. Il est contraint de faire des compromis, qui exigent généralement le « sacrifice » de l'un des doigtés ; sur la trompette, ce sont les notes *do dièse* et *ré grave* de la première octave qui restent toujours trop hautes par rapport à la fréquence « juste » ; le musicien peut compenser cette imperfection à l'aide d'une petite coulisse actionnable avec l'auriculaire. Le facteur peut aussi utiliser des formes évolutives, ni cylindriques ni coniques, pour « harmoniser » l'instrument. C'est le rôle principal du pavillon et de la branche d'embouchure.

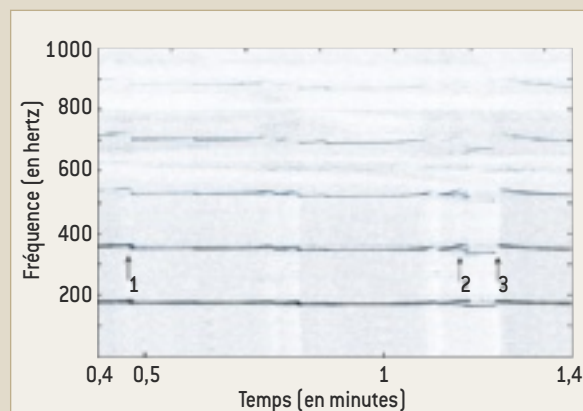
De plus, les modèles physiques traduisent le comportement acoustique de l'instrument, mais ils simplifient la réalité ; certains phénomènes secondaires ne sont pas encore pris en compte : effet des soudures et de l'état de la surface interne, influence de l'enroulement du résonateur, vibrations des parois, etc. Autre

QUAND LES SONS DEVIENNENT « PATHOLOGIQUES »

Dayton Miller, un acousticien américain, a publié en 1909 un article pionnier étudiant l'influence du matériau des instruments à vent sur la qualité sonore. Pour cela, il comparait des tuyaux d'orgue à embouchure de flûte d'épaisseurs et de matériaux différents (bois, zinc).

Dans l'une de ses expériences, Miller remplissait progressivement d'eau l'espace compris entre la double coque d'un tuyau d'orgue en zinc caréné. Il observait une série de comportements atypiques au cours du remplissage, notamment des sons qui « roulent », c'est-à-dire non tenus et instables au cours du temps. Ayant récemment reproduit l'expérience, nous avons mis en évidence des effets similaires aisément audibles, et reliés aux vibrations des parois : notes qui « roulent », mais aussi changements de hauteur de note, etc.

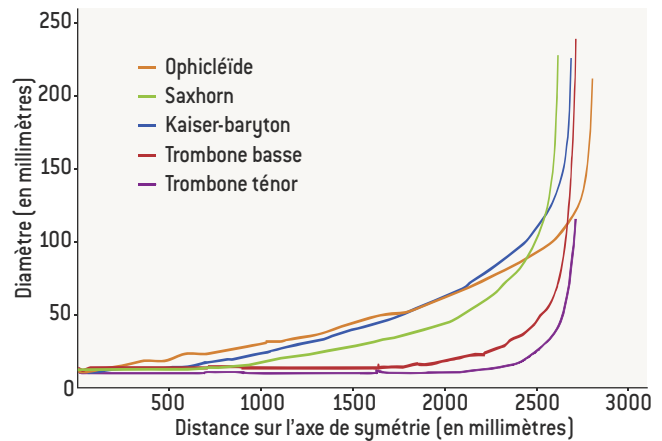
Ces effets sonores ont été analysés et nous les avons modélisés. Notre interprétation est la suivante : la colonne d'air de l'instrument peut être très perturbée, au point de modifier les sons émis, lorsqu'il y a coïncidence, c'est-à-dire égalité des fréquences de résonance, entre une résonance acoustique (celle de la colonne d'air) et une résonance mécanique (celle des parois). Dans le tuyau d'orgue, cette coïncidence ne se produit que pour une certaine hauteur d'eau, correspondant à des fréquences particulières de résonance mécanique. Or ce couplage est favorisé par une « brisure de symétrie » du résonateur (le changement de ses caractéristiques physiques), due par exemple à une ovalisation d'un tube originellement cylindrique. Dans ces situations « pathologiques », les perturbations du son ne sont pas la conséquence des ondes émises vers l'extérieur par les vibrations des parois, mais une conséquence indirecte de ces dernières : la colonne d'air de l'instrument est perturbée par les ondes émises vers l'intérieur par les parois, elles-mêmes mises en vibration par la colonne d'air.



CE SPECTROGRAMME montre les fréquences des sons émis par un tuyau d'orgue en zinc à double coque excité par un flux d'air à mesure que l'espace situé entre ses deux parois est rempli d'eau. À chaque instant, on enregistre la fréquence fondamentale (*en bas*) et quatre fréquences harmoniques, multiples de la fondamentale. Au début (quand l'espace est plein d'eau), le tube vibre sur toute sa hauteur. Quand l'espace est rempli à moitié (au bout d'une minute environ), seule la moitié supérieure du tube vibre. Les trois flèches pointées sur l'harmonique de rang 2 indiquent respectivement une variation de justesse importante (1), une note qui roule (2) et un silence (3). La note qui roule, dont le régime d'oscillation n'est plus périodique, est obtenue pour une certaine hauteur d'eau, donc pour une certaine fréquence de vibration des parois, qui coïncide avec la fréquence de résonance de la colonne d'air du tuyau d'orgue.



2. UN CUIVRE est composé d'une succession de sections à perce plus ou moins cylindrique ou conique (*ci-contre, la perce d'un clairon*), dont le facteur peut ajuster les dimensions pour parvenir au son recherché. Or la cuivrabilité des instruments dépend de la perce. Plus elle est cylindrique, plus elle propage le son en favorisant l'amplification des hautes fréquences, un phénomène caractéristique du son cuivré. En traçant le diamètre de la perce en fonction de la distance selon l'axe de symétrie de l'instrument déroulé, les auteurs ont classé ces cinq instruments par ordre de cuivrabilité : le kaiser-baryton, l'ophicléide, le saxhorn, le trombone basse et le trombone ténor (de gauche à droite). Le moins « cuivable » est l'ophicléide (*en orange*), puis viennent le saxhorn, le kaiser-baryton, le trombone basse et le trombone ténor (*en violet, le plus « cuivable »*).



inconvenient de l'approche par l'harmonicité, elle ne tient pas compte des préférences du musicien. Or la justesse peut dépendre du contexte musical dans lequel il s'exprime, selon ses goûts et sa sensibilité. C'est pourquoi une autre approche de définition de la perce, plus récente, s'appuie sur le musicien pour déterminer les valeurs « cibles » des fréquences de résonance : celles que l'instrumentiste recherche ; ensuite, on établit, à l'aide d'une méthode d'optimisation, la perce qui correspond aux fréquences de résonance proches de ces valeurs. La détermination des valeurs cibles passe par la réalisation d'une évaluation sensorielle d'un ensemble d'instruments, comparable aux tests réalisés par exemple en œnologie. Des travaux originaux ont été menés à cette fin par Émilie Poirson au sein de l'équipe de l'un d'entre nous (Jean-François Petiot), à l'Institut de recherche en communications et cybernétique de Nantes.

Ces mesures sensorielles nécessitent tout d'abord de définir un ensemble d'instruments de qualités différentes qui sont ensuite expertisés par un panel de musiciens essayeurs, préalablement entraînés à l'évaluation sensorielle. L'objectif est d'obtenir une mesure fiable des instruments selon des descripteurs sensoriels : la justesse, la facilité d'émission des sons et d'autres critères artistiques. Ces descripteurs sensoriels peuvent être ensuite reliés aux grandeurs physiques caractéristiques de l'instrument, principalement l'impédance d'entrée, à l'aide de techniques de modélisation et d'analyse de données.

Existe-t-il d'autres voies d'optimisation de la perce des cuivres ? Comme nous l'évoquions à propos de *Aïda*, la richesse sonore des cuivres dépend notamment de leur capacité à émettre un timbre cuivré au

« brillant » caractéristique. Ce timbre dépend lui aussi de la perce. En effet, la propagation de l'onde sonore nécessaire à l'effet cuivré est défavorisée lorsque la perce est conique ou évasée, si bien qu'un bugle, à perce majoritairement conique, sonnera avec beaucoup moins d'éclat qu'une trompette, dont la perce est plus cylindrique, alors qu'il a la même constitution et couvre la même échelle de sons – la même tessiture. Ainsi, les « cuivres doux » tels que les bugles et les saxhorns, instruments typiques des *brass bands*, orchestres composés uniquement de cuivres et de percussions, sont de perce à dominante conique, alors que les « cuivres clairs », comme la trompette et le trombone, ont une perce à dominante cylindrique.

L'origine de la cuivrabilité

Quelles sont les causes physiques de la cuivrabilité ? Jusqu'ici, la description du résonateur relevait de l'acoustique dite linéaire : l'onde de pression ne s'y déforme pas au cours de sa propagation. Or les sons cuivrés obtenus lorsque le musicien joue fort ont une intensité très élevée à l'intérieur de l'instrument (jusqu'à 170 décibels). Leur propagation est alors non linéaire : le maximum de l'onde acoustique se propage plus rapidement que le minimum, et l'onde distordue peut aboutir à une « onde de choc », une variation brutale de la pression acoustique. De la déformation de l'onde, il résulte un enrichissement spectral par les harmoniques de rang élevé, qui se retrouve également dans le son rayonné à l'extérieur de l'instrument – où l'approximation linéaire redevient valide. Ces timbres caractéristiques sont souvent qualifiés de « sons cuivrés » par les instrumentistes, car perçus comme des sons métal-

✓ BIBLIOGRAPHIE

E. Poirson et al., *Integration of user-perceptions in the design process: application to musical instrument optimisation*, J. Mechanical Design, vol. 129, pp. 1206-1214, 2007.

J. Gilbert et al., *A simulation tool for brassiness studies*, J. Acoust. Soc. Am., vol. 123, pp. 1854-1857, 2008.

G. Nief et al., *Influence of wall vibrations on the behaviour of a simplified wind instrument*, J. Acoust. Soc. Am., vol. 124, pp. 1320-1331, 2008.

E. Poirson, *Thèse de doctorat*, 8 déc. 2005, disponible sur <https://hal.ccsd.cnrs.fr>

liques. Toutefois, deux cuivres de perces différentes n'ont pas la même aptitude à la cuivrabilité (voir la figure 2). Qui plus est, le lien direct entre la perce d'un cuivre et ses possibilités de créer des timbres cuivrés peut apparaître réducteur. En effet, un musicien chevronné peut toujours faire sonner son instrument plus ou moins « cuivré » en modifiant subtilement son masque. Et un instrument donné rendra plus ou moins facilement ce timbre selon l'instrumentiste.

Les vibrations de parois

Que les cuivres soient caractérisés essentiellement par la perce du résonateur ne doit pas faire négliger le choix du matériau. En effet, celui-ci détermine la méthode de fabrication, le poids et l'équilibre, l'aspect et le prix de l'instrument. De plus, la surface interne du cuivre doit être très lisse pour limiter les pertes sonores par frottement contre les parois : un matériau possédant une surface interne rugueuse ou poreuse rend l'instrument plus difficile à jouer, voire injouable. Cette problématique pourrait être liée au vieillissement de l'instrument : on ignore encore s'il correspond à une perte de matériau avec le temps, ou s'il traduit un changement des caractéristiques de l'état de surface interne.

En outre, la question du rôle des vibrations des parois des cuivres et du rayonnement acoustique

qui en découle est un sujet de recherche actuel. Des expériences spectaculaires appliquées à d'autres instruments à vent, des tuyaux d'orgue à embouchure de flûte, ont été décrites voilà un siècle. Dans certaines conditions, que nous qualifions de « pathologiques », le tuyau d'orgue produit des sons particuliers, notamment une « note qui roule », traduisant un régime d'oscillation quasi périodique, à la place d'une note tenue et stable, de régime périodique (voir l'encadré page 55). Or, à l'occasion de la thèse de Guillaume Nief, nous avons pu reproduire ces expériences à l'Université du Maine. Ce travail nous a permis de décrire, de modéliser et ainsi de comprendre la production de ces sons atypiques. Ils proviennent de la perturbation de la colonne d'air à l'intérieur de l'instrument par le rayonnement interne des parois en vibrations. Le cadre théorique que nous avons décrit convient donc aussi pour ces sons pathologiques.

Pendant, une question reste ouverte : le matériau – au niveau du pavillon – joue-t-il un rôle dans l'émission du rayonnement sonore vers l'extérieur ? C'est une observation fréquemment rapportée par les facteurs d'instruments et les musiciens. Peut-être mettra-t-on un jour en évidence des situations expérimentales comparables à celles du tuyau d'orgue, des « pathologies » de cuivres qui produiront sans ambiguïté des effets sonores inédits dus aux vibrations de leurs parois... ■

✓ SUR LE WEB

LAUM
<http://laum.univ-lemans.fr/>
IRCCyN
<http://www.irccyn.ec-nantes.fr/>
École de physique d'Édimbourg
<http://www.ph.ed.ac.uk>
Tout sur la trompette
<http://la.trompette.free.fr/>
Edinburgh University Collection of Historic Musical Instruments
<http://www.music.ed.ac.uk/euchmi/>

POUR LA **SCIENCE** et le Palais de la Découverte vous invitent

le mardi 18 novembre 2008 à 18 h 30

Les limites des prévisions de la population mondiale



Avec **Gilles Pison**

Directeur de recherches à l'Institut national d'études démographiques
et rédacteur en chef de la revue *Population et sociétés*.

En partenariat avec



Entrée libre • Inscription obligatoire à population@palais-decouverte.fr
Palais de la Découverte • Avenue Franklin-D.-Roosevelt • 75008 Paris • www.palais-decouverte.fr



Le rythme

Jean-Pierre Dalmont et Jean Kergomard

Dans les instruments à anche dite faible, l'anche ne vibre pas à son propre rythme, mais à celui du résonateur en bois ou en métal dont elle commande l'entrée. Le fonctionnement physique d'une clarinette est typique de la production sonore de ce type d'instruments à vent.

L'ESSENTIEL


- ✓ Dans un instrument à vent, l'anche, une languette vibrante, peut servir soit à émettre le son, soit à moduler le flux d'air entrant.
- ✓ Quand elle sert à moduler le flux d'air tandis que la fréquence d'oscillation est imposée par le tuyau, il s'agit d'une anche faible.
- ✓ La façon dont le son se développe une fois émis peut être modélisée de façon simple dans le cas d'une clarinette.

Aux temps de vos lointains ancêtres, la tibia était fort en usage, et les joueurs de tibia furent toujours tenus en grand honneur: on en jouait dans les temples, on en jouait lors des jeux, et on en jouait lors des tristes funérailles.
Ovides, *Fastes*, Livres VI

La *tibia curua* évoquée par le poète romain Ovide (-43 à 17) était le principal instrument de musique romain. Probablement originaire de Phrygie (*Tibia* en latin), mais hérité des Étrusques, cet instrument était une sorte de hautbois en roseaux à deux tuyaux et anche double. Elle était proche de l'aulos grecque, autre hautbois composé d'un

double tuyau percé de trois ou six trous et à double anche lui aussi. Ces instruments, qui se prolongent au Moyen Âge dans les nombreux chalumeaux (du latin *calamus* pour roseau), attestent de l'ancienneté dans la musique occidentale du principe de l'anche faible, une languette de roseau, dont une extrémité est fixe tandis que l'autre se meut pour commander l'arrivée d'air d'un résonateur en forme de tuyau. Les chalumeaux médiévaux sont les ancêtres des instruments à anche faible modernes tels le hautbois, le basson, le saxophone, et la clarinette...

Si bien des aspects du son produit par ces instruments sont compris depuis longtemps, curieusement, la compréhension de leur production sonore n'a vrai-



des anches

Shutterstock

ment commencé qu'il y a une quarantaine d'années. Elle est aujourd'hui aboutie et validée par l'expérience. Cet article vise à expliquer le fonctionnement de base des instruments à anche, c'est-à-dire la façon dont le son se développe après avoir été créé par l'écoulement que module l'anche.

Familles d'instruments

Un instrument à anche est un instrument à vent, c'est-à-dire fonctionnant à l'aide d'une source d'air comprimé. Le plus souvent, cette source est le poumon du musicien, mais il peut aussi s'agir aussi d'une poche, comme dans le cas de la cornemuse ou d'un soufflet comme dans le cas de l'accordéon ou des orgues anciens. L'air alimentant un instrument à vent n'est que légèrement comprimé : la surpression dépasse rarement un dixième de la pression atmosphérique, alors qu'elle est de plusieurs pressions atmosphériques dans le cas des avertisseurs sonores à air comprimé...

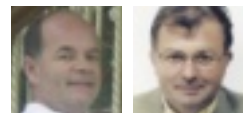
On distingue deux grandes familles d'instruments à vent. Dans la première, la famille des flûtes, le son résulte de l'oscillation d'un jet d'air de part et d'autre d'un

biseau. Aucune vibration mécanique n'entre en jeu, de sorte que le son provient des seules oscillations de la colonne d'air contenue dans le résonateur. Dans la seconde famille, celle des anches, le son résulte de la modulation d'un jet d'air par une lame - l'anche - jouant le rôle d'une valve. Sa vibration modifie la taille d'une ouverture éventuellement couplée à un résonateur acoustique, qui, le plus souvent, est un tube. Les notes obtenues avec l'instrument correspondent aux fréquences pour lesquelles les vibrations s'amplifient naturellement lorsqu'on excite le résonateur en continu : on les nomme fréquences de jeu. Sur le plan mécanique, on peut considérer l'anche comme une sorte de ressort, dont la masse et la raideur déterminent la fréquence propre d'oscillation.

Les anches sont exploitées dans trois situations types. Dans la première, l'anche oscille à une fréquence proche de sa fréquence propre (sa fréquence naturelle d'oscillation), et le résonateur, s'il existe, ne joue qu'un rôle d'amplificateur : on parle alors d'anche forte puisque c'est l'anche qui impose sa fréquence propre à l'instrument. Les anches fortes sont souvent métalliques. On les trouve dans les jeux d'anche des orgues, l'accordéon,

1. LES INSTRUMENTS À ANCHE comprennent les orgues, certains bois – clarinettes, hautbois, saxophones... –, les harmoniums, les harmonicas, et même les cornemuses. On en distingue deux groupes, les instruments dont l'anche est dite forte parce qu'elle impose sa fréquence propre à tout l'instrument, et les instruments à anche dite faible, objets du présent article, où le résonateur constituant le corps de l'instrument impose ses fréquences propres.

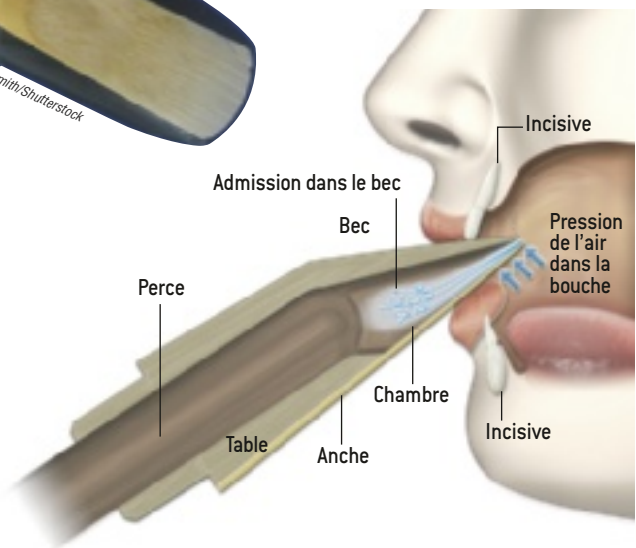
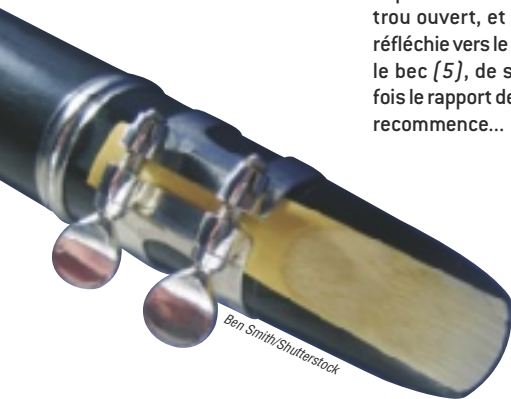
LES AUTEURS



Jean-Pierre DALMONT est chercheur au Laboratoire d'acoustique de l'Université du Maine [LAUM, UMR 6613]. Jean KERGOMARD est chercheur au Laboratoire de mécanique et d'acoustique de l'Université d'Aix-Marseille.



2. UNE SURPRESSION (en rouge) se propage depuis le bec jusqu'au premier trou ouvert (1) situé à une distance L . Elle passe devant le premier trou ouvert et une zone de dépression d'amplitude moindre est réfléchi vers le bec (2). Celle-ci se réfléchit dans le bec, puis se propage vers le premier trou ouvert (3). Elle passe devant le premier trou ouvert, et une surpression d'amplitude moindre est réfléchi vers le bec (4). Cette surpression est réfléchi dans le bec (5), de sorte qu'au bout d'un temps égal à quatre fois le rapport de la longueur L par la vitesse du son, un cycle recommence...



3. UN JOUEUR QUI SOUFFLE dans le bec y produit un jet d'air rapide dont l'élan se dissipe en turbulences dans le bec, ce qui met en mouvement la colonne d'air présente dans la perce. Une petite fluctuation de pression – surpression ou dépression – se propage dans la colonne d'air mobile, et, à chaque retour, est amplifiée grâce aux mouvements de l'anche.

l'harmonica ou l'harmonium. La voix appartient également à cette famille des anches fortes : dans ce cas, les anches sont les cordes vocales.

Dans le deuxième type de situation, l'anche oscille à une fréquence proche d'une des fréquences de résonance du résonateur. Ces anches, qui subissent en quelque sorte la fréquence imposée par le résonateur, sont qualifiées d'anches faibles. Les instruments à anche faible comprennent le hautbois, la clarinette, la corneuse... Le plus souvent, ils fonctionnent avec une anche en roseau, même si le roseau est parfois remplacé par du plastique ou par un matériau composite.

Dans le troisième type de situation, que nous ne ferons qu'évoquer ici, la fréquence de résonance de l'anche est accordée sur l'une des résonances du résonateur : dans cette situation, l'anche et le résonateur imposent tous deux leurs fréquences. C'est le cas des instruments de la famille des cuivres dans lesquels les lèvres du musicien jouent le rôle de l'anche, et c'est pourquoi on parle parfois d'anche lippale.

Le résonateur, un amplificateur

Ainsi, les instruments à anche proprement dits sont constitués d'une anche faible, simple ou double, couplée à un résonateur, en général un tuyau cylindrique ou conique percé de trous. En fermant ces trous, on change la longueur effective du tuyau, et, par là, la fréquence de jeu, soit, en d'autres termes, la note jouée. La famille des instruments à anche comprend principalement le hautbois, le saxophone, le basson (de perce conique) et la clarinette (de perce cylindrique). Nous allons nous concentrer sur cette dernière parce qu'elle est l'instrument à anche faible auquel les physiciens se sont le plus intéressés, sans doute parce qu'elle est aussi la plus simple à modéliser.

La clarinette est un instrument à anche simple qui a des propriétés inhabituelles. En effet, comme le hautbois ou la flûte, elle fonctionne sur deux registres. Toutefois, et c'est en cela qu'elle est particulière, son second registre est à la douzième (une octave, soit huit tons, et une quinte, soit cinq tons) du premier. Cela implique que les notes de ce second registre sont aux fréquences triples de celles des notes du registre fondamental, au lieu d'être à l'octave (c'est-à-dire au double des fréquences de base) comme c'est le cas sur un hautbois ou une flûte traversière. On dit que la clarinette quintoye. Pour saisir l'origine de cette propriété particulière de la clarinette, il faut s'intéresser au résonateur, puisque c'est lui qui impose à l'anche faible ses fréquences propres, c'est-à-dire les fréquences qu'il aura naturellement tendance à amplifier, que l'on nomme aussi fréquences de résonance.

Le résonateur de la clarinette est un cylindre que l'on peut considérer, pour simplifier, comme fermé à l'entrée et ouvert à la sortie. Comment obtient-on les fréquences de jeu d'un tel dispositif ? Rappelons que, dans l'air, le son est une variation de pression qui

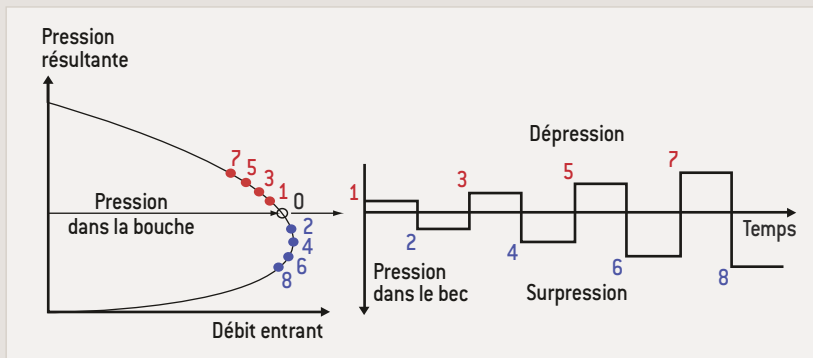
se propage à la vitesse du son, soit environ 340 mètres par seconde. Dans un tuyau plein d'air, si une telle onde parvient à une extrémité fermée, elle se réfléchit ; en revanche, si elle parvient à une extrémité ouverte – un trou par exemple –, elle est partiellement transmise vers l'extérieur et partiellement réfléchi, mais une surpression devient une dépression. Ainsi, on peut déterminer la période d'une note brève, et par là sa fréquence, en examinant les allers-retours que fait l'onde associée dans le résonateur, subissant quatre réflexions, dont deux avec inversions au cours de chaque cycle (voir la figure 2).

Pour expliquer complètement le fonctionnement d'un instrument à anche, par exemple d'une clarinette, il importe de considérer aussi l'autre partie de l'instrument, c'est-à-dire l'embouchure. Dans une clarinette,

celle-ci est constituée d'une anche simple fixée sur un bec, dispositif dont on pense que c'est Johann Christoph Denner qui l'inventa à la fin du XVII^e siècle. L'anche de clarinette est une lame de roseau plate et effilée, que l'on pose et fixe sur le bec – un bout de tube biseauté. Comme la forme du bec est légèrement courbe à son extrémité, l'anche posée sur le biseau ne l'obture pas complètement, et une ouverture d'environ un millimètre subsiste (voir la figure 3). En fléchissant, l'anche réduit cette ouverture, allant jusqu'à l'obturer.

On incurve l'anche soit en la pressant de la lèvre, soit en soufflant dans le bec, ce qui crée une pression sur la partie supérieure de l'anche et un jet d'air dans l'interstice entre l'anche et le bec. Quand on souffle dans le bec en augmentant progressivement la pression, on observe que le débit augmente, passe par un

L'ANCHE, UNE PORTE QUI BAT

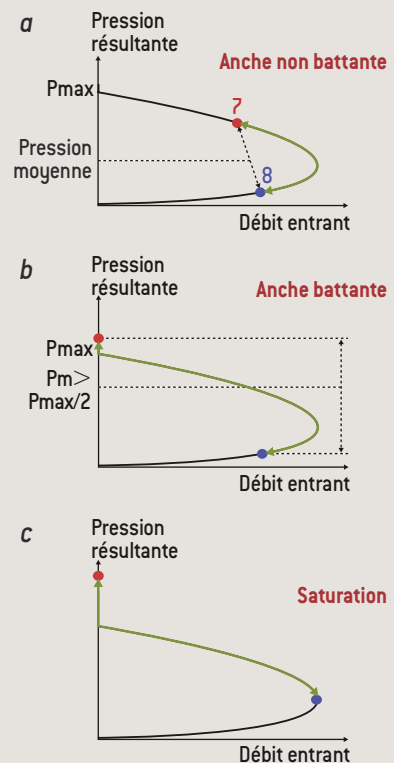


Dans une clarinette, l'évolution du débit en fonction de la pression exercée sur l'anche donne une courbe que l'on nomme la caractéristique non linéaire de l'instrument. Lorsqu'un joueur de clarinette attaque, il faut qu'il souffle assez fort pour mettre la colonne d'air en mouvement, et, par l'intermédiaire des oscillations de l'anche, communique de l'énergie aux ondes de pression ou de dépression qui vont et viennent entre le bec et la sortie. Ces ondes sont rapidement amplifiées. Voyons comment. Supposons une dépression dans le bec (point 1). Elle se propage jusqu'au premier trou ouvert, et s'inverse en une onde de pression. Quand elle atteint l'anche, elle la repousse, ce qui augmente l'ouverture. Sur la caractéristique non linéaire, on atteint le point 2, où le débit et la pression dans le bec ont augmenté. Une nouvelle bouffée d'air entre dans l'instrument, de sorte qu'une pression supplémentaire s'ajoute à l'onde de pression revenue de l'entrée de l'instrument. Le phénomène se reproduit ensuite, de sorte que les impulsions successives (points 3, 4, 5, 6, 7, 8, ...) amplifient l'onde sonore associée.

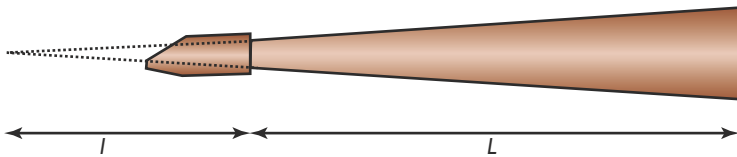
Cette résonance prend de l'ampleur jusqu'à ce que, sur la caractéristique non linéaire, les deux points de fonctionnement extrêmes (ici 6 et 7) aient des débits comparables (voir la courbe a). Toutefois, le point indiquant la phase de dépression (*en haut*) correspond à un débit légèrement inférieur (la différence dépend des pertes) ; les deux points sont à égale distance du point de départ.

Un équilibre instable s'est installé : l'instrument oscille entre deux états de pression et de débit, l'un à anche plus ouverte, et l'autre à anche plus fermée.

Au-delà d'une pression moyenne exercée sur l'anche (la pression de plaquage, qui vaut la moitié de la pression de plaquage, l'anche se ferme totalement durant la phase de dépression (voir la courbe b). On parle alors de régime en anche battante : l'anche est comme une porte qui bat dans des courants d'air. L'anche fonctionne alors presque comme une valve qui ne connaîtrait que deux états, l'un ouvert et l'autre fermé. Le rôle de cette valve est de transformer la pression continue dans la bouche en une pression alternative de quasiment la même amplitude, ce qui est le moyen le plus simple pour



produire des sons très intenses. Si la pression sur l'anche croît encore, l'amplitude de l'oscillation et donc la puissance sonore émise augmentent. Toutefois, pendant le cycle de pression, l'anche passe de plus en plus de temps plaquée sur la table, de sorte qu'à partir d'un certain seuil – le seuil de saturation –, la puissance sonore émise stagne, puis décroît. Elle finit par s'éteindre complètement quand le seuil d'extinction est atteint : la pression moyenne exercée par le joueur est alors égale ou supérieure à la pression de plaquage (voir la courbe c).



4. DANS UN INSTRUMENT À ANCHE CONIQUE, tout se passe comme si à la création d'une impulsion (une onde de dépression), la moitié allait se réfléchir au début (imaginaire) du cône à une distance l , tandis que l'autre moitié se propage jusqu'à la sortie du cône à une distance L . Lorsqu'elles se croisent au niveau du bec, les deux demi-impulsions échangent leurs rôles : c'est comme si elles se retrouvaient dans le bec après un aller-retour entre le début (imaginaire) et la fin du cône, après avoir parcouru une distance de $2(L+l)$ en un temps égal à $C/[2(L+l)]$, où C est la vitesse du son dans l'air.

maximum, puis diminue jusqu'à s'annuler lorsque l'anche, plaquée contre le bec, le ferme. Ce verrouillage du bec se produit à une pression dite de plaquage. L'évolution du débit en fonction de la pression exercée sur l'anche donne une courbe que l'on nomme la caractéristique non linéaire de l'embouchure (voir l'encadré page 61).

Cette courbe n'est pas facile à obtenir expérimentalement (il faut pour cela éviter que l'anche vibre pendant la mesure), mais elle est essentielle, car elle caractérise le fonctionnement de l'embouchure au même titre que l'impédance du résonateur – la grandeur qui mesure sa capacité à résonner – caractérise le fonctionnement du résonateur. En particulier, tous les états d'une clarinette en fonctionnement sont représentés par des points qui sont situés sur la caractéristique non linéaire. En outre, la plage de jeu est encadrée et caractérisée par trois seuils de pression : le seuil d'oscillation auquel l'instrument entre en vibration, le seuil de saturation à partir duquel la puissance sonore émise n'augmente plus, et le seuil d'extinction pour lequel le son s'éteint.

Tant la caractéristique non linéaire que ces trois seuils fondamentaux peuvent être calculés à l'aide d'un modèle physique simple, dans lequel on considère l'instrument comme un simple tube, dans lequel le débit d'entrée est contrôlé par l'ouverture d'une valve oscillante (l'anche). Les expressions algébriques issues de ce modèle révèlent que les paramètres essentiels sont la longueur du tube, la raideur de l'anche, l'ouverture au repos (en l'absence de pression) et le coefficient de perte dans l'instrument (l'amortissement de l'onde). Au plan expérimental, tant la caractéristique non linéaire que ces seuils sont difficiles à mesurer lorsqu'un musicien joue, mais ils peuvent l'être à l'aide d'une bouche artificielle, un banc de mesure conçu à cet effet. Exploitée pour mesurer les caractéristiques non linéaires et les seuils de plusieurs instruments, la bouche artificielle a montré la bonne adéquation du modèle avec la réalité physique. Dès lors, nous allons maintenant nous appuyer sur ce modèle (qui s'exprime dans la caractéristique non linéaire) pour expliquer le fonctionnement d'une clarinette.

Considérons ce qui se passe lorsque la pression exercée sur l'anche augmente jusqu'à atteindre le

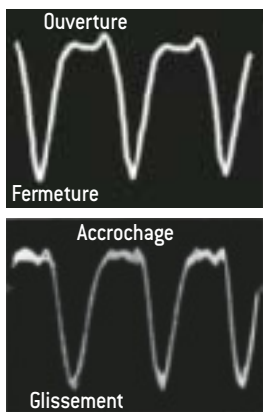
maximum de la caractéristique non linéaire. Au-delà du maximum, l'anche fléchit, jusqu'au plaquage sur le bec si la pression augmente assez. Toutefois, l'application d'une pression ne fait pas osciller l'anche. La contribution des ondes sonores qui se propagent au sein du résonateur est nécessaire.

L'auto-oscillation

Comment la production sonore se déclenche-t-elle ? Supposons que le clarinettiste fasse régner dans sa bouche une pression qui nous place au-delà du maximum de la caractéristique non linéaire. Le jet d'air qu'il injecte met la colonne d'air en mouvement. Pour autant, il n'est pas à l'origine de l'onde sonore. Celle-ci a pour origine de petites perturbations de pression – dépression ou surpression – issues de mouvements erratiques de l'air dans la bouche et le bec. Supposons par exemple qu'il s'agisse d'une petite dépression créée juste au-dessus de l'anche. Sur la caractéristique non linéaire, l'apparition de cette dépression dans le bec correspond à un léger déplacement dans le sens d'une plus grande pression résultante sur l'anche, et d'une pression à l'intérieur du bec et d'un débit moindres. Une onde de pression se propage alors jusqu'à la sortie de l'instrument, se réfléchit, et, revenant, pousse l'anche, ce qui correspond sur la caractéristique non linéaire à un déplacement dans le sens des pressions résultantes sur l'anche moindre et de pression et débits augmentés dans le bec. Une nouvelle onde de pression augmentée est alors émise qui se propage vers la sortie. À partir de là, un équilibre instable s'établit dans lequel les allers-retours de l'onde sonore s'accordent à ceux de l'anche tout en amplifiant les surpressions ou les dépressions qui reviennent dans le bec. Notons que l'énergie nécessaire à cette amplification est fournie seulement par le souffle de l'instrumentiste.

Ainsi, dans le bec, des oscillations de pression ont lieu autour d'une valeur moyenne qui correspond à la pression dans la bouche du joueur. Tout se passe, comme si, sur la caractéristique non linéaire, le système anche-colonne d'air oscillait entre un point à débit et pression dans le bec inférieurs situés au-delà du maximum, et un point à débit et pression dans le bec supérieurs situés en deçà du maximum. Pendant cette oscillation, à chaque période, une partie de l'énergie introduite dans la clarinette par le souffle du joueur est rayonnée sous forme de son à partir du pavillon, à l'extrémité de l'instrument.

Cette émission sonore ne s'établit qu'à partir d'une pression minimale dans la bouche qui définit le seuil d'oscillation. Dans le modèle physique qui nous guide ici, les calculs situent le seuil d'oscillation aux environs d'un tiers de la pression de plaquage, ce que les mesures ont validé. Si le clarinettiste augmente assez la pression dans sa bouche, la pression maximale résultante sur l'anche – qui, sur la caractéristique non linéaire, correspond au point à



5. LE SIGNAL DE PRESSION dans les instruments à anche (en haut) est similaire au signal de vitesse du contact corde-archer (en bas) des instruments à cordes frottées (violon, violoncelle, etc.). Étudié dès le XIX^e siècle par le physiologiste et physicien allemand von Helmholtz, ce type de signal conduisant à une production sonore porte pour cette raison le nom de mouvement de Helmholtz.

débit et pression minimales – atteint et dépasse la pression de plaquage. L’anche se comporte alors comme une porte qui bat dans les courants d’air : elle se plaque sur la table, y reste un moment, puis s’ouvre avant de recommencer. On parle alors de régime à anche battante. Si, alors qu’on est entré dans le régime à anche battante, la pression dans la bouche augmente encore, on observe qu’à partir d’un certain seuil de pression – le seuil de saturation –, la puissance sonore rayonnée n’augmente plus, puis décline. En effet, quand, enfin, la pression dans la bouche dépasse le seuil de saturation, l’anche bat de moins en moins, puis cesse de battre jusqu’à ce que plus aucun son ne soit émis : la clarinette a atteint son seuil d’extinction.

Peu de clarinettes ont conscience de l’existence du seuil d’extinction, mais ils le perçoivent sans doute. En effet, une fois le seuil de saturation atteint, tout effort supplémentaire est inefficace, tandis que la fermeture du canal anche-table provoque une sensation plutôt désagréable. Jouer au voisinage du seuil d’extinction est toutefois un exercice instructif, et certains clarinettes, tel Pierre-André Taillard, le font pour obtenir une nuance *pianissimo*.

En pratique, le jeu est compliqué par le fait que le musicien peut modifier l’embouchure (en pinçant l’anche de la lèvre), ce qui permet de varier considérablement la dynamique de jeu. Ce n’était pas le cas avec les ancêtres de la clarinette que sont le cromorne (un instrument à anche double de la Renaissance) ou les chalumeaux de certaines cornemuses. L’alimentation en air à l’aide d’une poche empêche les nuances. Lorsque le musicien pince l’anche pour modifier l’ouverture, cela diminue le débit maximal, et donc la valeur du seuil d’extinction. Notons que la lèvre a également pour rôle d’amortir la vibration propre de l’anche, ce qui évite les couacs. L’expérience montre que le seuil d’oscillation dépend peu de l’embouchure : quels que soient l’ouverture et donc le seuil d’extinction qu’il a choisi, le musicien peut attaquer la note avec à peu près la même pression, ce qui est confortable dans l’interprétation. Cette remarquable propriété résulte de la façon dont les bords sont construits, et tout l’art du facteur d’anche et de bec – les deux sont indissociables – est justement de l’obtenir.

Le clarinettiste, lorsqu’il joue, a donc deux façons de faire varier l’amplitude de l’oscillation : la pression dans la bouche et l’ouverture de l’anche. Mais quelles que soient la puissance sonore choisie et la façon de l’obtenir, toute sa plage de jeu s’étend entre le seuil d’oscillation et le seuil d’extinction, et passe par un seuil de saturation !

Le principe régissant le fonctionnement des clarinettes s’applique aussi aux instruments coniques,

tels le hautbois et le basson. Leur embouchure n’est pas fondamentalement différente de celle d’une clarinette, mais ils fonctionnent avec une anche double. Les travaux d’André Almeida et de ses collègues de l’IRCAM et du LMA ont montré que leurs caractéristiques non linéaires ne sont pas fondamentalement différentes de celles de la clarinette. Quant au saxophone, il est muni d’un bec semblable à celui d’une clarinette. Récemment, une entreprise américaine a commercialisé un bec pour basson à anche simple sur le modèle d’un petit bec de saxophone, mais donnant vraiment des sons de basson ! Ce n’est donc pas dans le principe de l’anche (simple ou double) que réside la différence entre bassons et clarinettes, mais dans la forme de leurs résonateurs, cylindrique pour la clarinette et conique pour le basson. Une analogie avec le cas des instruments à cordes (*voir la figure 5*) montre que la réponse impulsionnelle d’un cône peut être schématisée d’une façon voisine de celle d’une clarinette. La différence principale est que l’impulsion initiale est divisée en deux : tout se passe comme si la moitié de l’impulsion était réfléchi peu après la sortie de l’embouchure, à l’entrée du cône, et l’autre moitié au bout de l’instrument (*voir la figure 4*).

En régime permanent, un modèle simplifié du même type que celui de la clarinette donne le même régime à deux états, avec cette nuance que les durées des deux états sont différentes. Précisons que lorsqu’un saxophoniste ou un bassoniste monte une gamme, la durée de l’épisode de fermeture est invariable alors que la durée de l’épisode d’ouverture dépend de la note. Il semble que la constance de l’épisode de fermeture confère à l’instrument une caractéristique de timbre commune aux différentes notes, permettant ainsi de distinguer un instrument d’un autre. Comme dans le cas de la clarinette, il est possible de calculer l’amplitude du signal en fonction des différents paramètres, ainsi que les seuils d’oscillation, de saturation et d’extinction. Toutefois, un phénomène nouveau apparaît lorsque l’on passe le seuil de saturation : au lieu de s’éteindre, comme c’est le cas pour la clarinette, le signal change de forme et semble même parfois s’inverser. Signalons une autre différence importante : au voisinage du seuil d’oscillation, il ne peut y avoir de petites oscillations, ce qui explique la difficulté à attaquer *pianissimo* au saxophone, en particulier dans les graves.

Ces modèles simples de clarinette ou de saxophone suffisent pour comprendre les bases du fonctionnement des instruments à anche. Pour autant, la réalité physique de ce qui s’y produit est bien plus complexe. Aujourd’hui, les acousticiens travaillent notamment sur la mécanique de l’anche, ainsi que sur le rôle du conduit vocal dans le jeu ou encore sur l’optimisation de la forme des résonateurs... ■

PEU DE CLARINETTES ont conscience de l’existence du seuil d’extinction...

✓ SUR LE WEB

Des informations complémentaires se trouvent sur le site du Laboratoire d’acoustique de l’Université du Maine : <http://laum.univ-lemans.fr>

Ce site contient en particulier une page consacrée à la physique des instruments de musique : <http://laum.univ-lemans.fr/spip/spip.php?rubrique107&lang=fr>.

On pourra aussi consulter : <http://www.phys.unsw.edu.au/music/clarinet/>

✓ BIBLIOGRAPHIE

A. Chaigne et J. Kergomard, *Acoustique des instruments de musique*, collection Échelles, Belin, novembre 2008.

J.-P. Dalmont et C. Frappé, *Oscillation and extinction thresholds of the clarinet : Comparison of analytical results and experiments*, J. Acoust. Soc. Am., vol. 122, pp. 1173-1179, 2007.

S. Ollivier, J.-P. Dalmont et J. Kergomard, *Idealized models of reed woodwinds. Part I : analogy with the bowed string*, *Acustica-acta acustica*, vol. 90 (6), 1192-1203, 2004.

Jean-Pierre Dalmon et Joël Gilbert, *Des instruments à vent harmoniques*, Pour la Science, n° 238, août 1997.

Suzanne et Franck Laloë, *La clarinette*, Pour la Science, n° 91, mai 1985.

De la BOUCHE ARTIFICIELLE...

Christophe Vergez et Didier Ferrand

Une machine qui souffle dans une trompette permet d'étudier la production de notes. Si l'on équipe une telle bouche artificielle d'un système d'asservissement, l'analyse des phases dynamiques du jeu devient possible.

L'ESSENTIEL

✓ Les bouches artificielles servent à mesurer et à figer les paramètres qui contrôlent l'émission d'une note.

✓ Dans une bouche artificielle asservie, le débit d'air à l'entrée est régulé en fonction de la pression souhaitée. On étudie ainsi les transitions entre les notes.

Avez-vous déjà écouté un orgue de Barbarie dans les rues de Paris ? Les notes s'égrènent, mais quelle que soit la verve du joueur de manivelle, son interprétation reste à des années lumière du phrasé si unique d'un Miles Davis jouant de la trompette ! Pourquoi ? Parce que le génie du trompettiste, comme le talent de tout musicien, réside dans sa façon de réaliser les transitions entre les notes (dites transitoires). Après que les physiiciens ont construit une machine – une bouche artificielle – qui fait fonctionner des instruments à vent, ces transitoires ne pouvaient qu'être leur prochaine cible.

Nous expliquerons ici pourquoi et comment nous avons développé une bouche artificielle asservie, première étape vers l'analyse et la reproduction du jeu instrumental. Ce banc de mesure est aujourd'hui avant tout un moyen pour aborder enfin l'étude scientifique des phases dynamiques du jeu.

En musique, il existe depuis longtemps deux stratégies de pro-

duction sonore : l'une demandant une action directe du musicien, qui doit souffler dans un tube, pincer une corde, frapper une peau, etc., et une autre où il lui suffit de donner, en général sur un clavier, un signal à une machine conçue pour produire le son. Dans l'orgue de Barbarie, ou dans toute boîte à musique, cette dernière stratégie est poussée encore plus loin, puisqu'un programme – codé par des trous dans une bande cartonnée dans le cas de l'orgue de Barbarie – remplace le musicien qui n'a même plus à donner de signal. Une bouche artificielle est un dispositif de ce genre, mais construit pour reproduire et mesurer l'action physique de la bouche du musicien qui crée une note avec un cuivre (trompette, trombone, etc.) ou un bois (clarinette, hautbois, etc.).

Les deux vertus d'une bouche artificielle

La caractéristique première d'une bouche artificielle est de reproduire de façon simplifiée les fonctions essentielles de l'instrumentiste : son souffle bien sûr, mais aussi l'action de sa bouche et plus particulièrement de ses lèvres. Cela exige une interface mécanique adaptée. Ainsi, pour les cuivres, une bouche artificielle doit disposer de lèvres artificielles qui vont vibrer sous l'effet du couplage avec l'instrument. Pour les instruments à anche, la fonction de « pince », assurée par les lèvres, doit être reproduite.



au MUSICIEN ARTIFICIEL ?

La deuxième caractéristique d'une bouche artificielle est de faire jouer l'instrument en figeant tous les paramètres de contrôle, qui sont normalement modifiés par le musicien en cours de jeu – par exemple la pression avec laquelle il souffle, la raideur des lèvres, la configuration du conduit vocal... Plusieurs grandeurs physiques sont mesurées simultanément: la pression d'alimentation en air comprimé, le débit d'air résultant, la pression dans la bouche et dans l'instrument, l'oscillation des lèvres ou de l'anche...

Pour un chercheur qui étudie la production sonore des instruments à vent, les mesures avec une bouche artificielle sont préférables à toutes celles qu'il pourrait mettre en place directement, mais avec difficulté, sur un musicien. Contrairement à un musicien, un tel banc de mesure est toujours disponible, rend les mesures reproductibles à volonté, accepte des capteurs dans sa bouche, ses lèvres ou son conduit vocal...

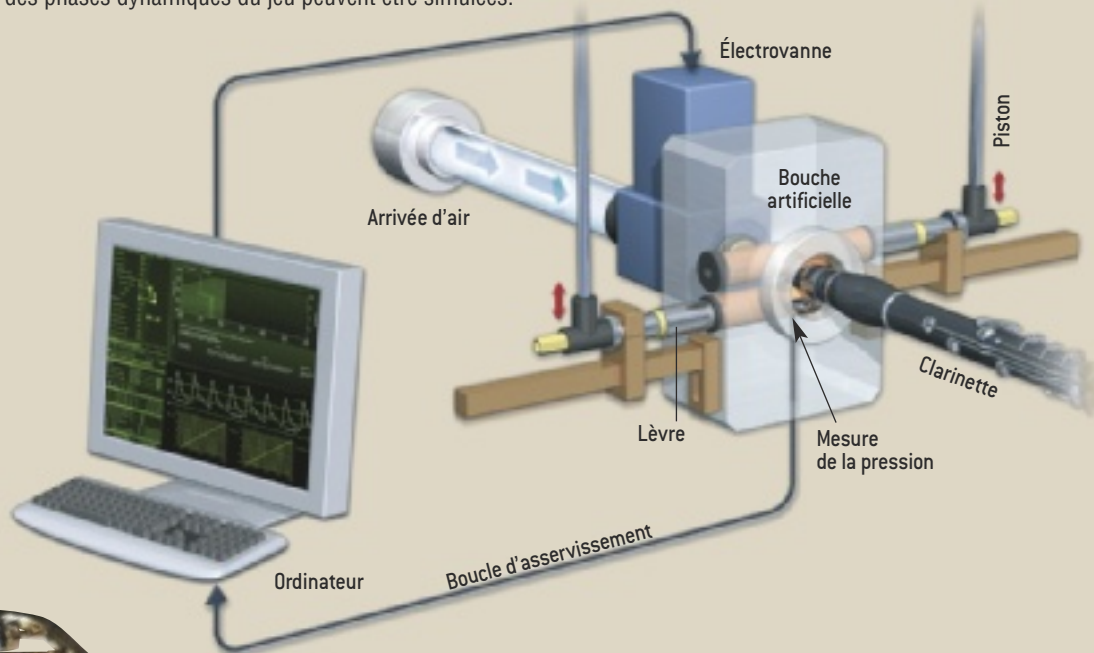
Par ailleurs, concevoir une bouche artificielle est aussi une façon de se concentrer sur l'essentiel du phénomène de production des sons. Un musicien en train de jouer module simultanément une multitude de paramètres. Avec une bouche artificielle, au



1. LE MUSICIEN ARTIFICIEL est encore loin, mais la mise au point d'une bouche artificielle capable de reproduire les transitoires entre notes constitue une étape importante pour s'en approcher.

UNE BOUCHE AU SOUFFLE ASSERVI

Dans la bouche artificielle asservie, un flux d'air sous une pression comparable à celle régnant dans la bouche d'un musicien (moins de 0,5 bar en plus de la pression atmosphérique) est contrôlé par une électrovanne (*en bleu*). On envoie de l'air vers la bouche et l'on mesure la pression en avant des lèvres. Cette dernière est transmise à l'ordinateur qui la compare à celle que l'on veut appliquer à l'instrument (la consigne). Toutes les 40 microsecondes, l'ordinateur envoie à l'électrovanne une correction destinée à ramener la pression dans la bouche à une valeur proche de cette consigne. De cette façon, les évolutions de la pression au cours des phases dynamiques du jeu peuvent être simulées.



Bruno Bourgeois



contraire, l'acousticien ne mesure que les quelques paramètres disponibles sur son dispositif, ce qui simplifie l'étude – une démarche habituelle en physique.

Les bouches artificielles construites jusqu'à présent ont eu pour principale vocation de servir à valider et à améliorer des modèles physiques utilisés pour décrire la production sonore des instruments de musique. Toutefois, elles sont aussi utiles pour inventer des instruments, tel ce saxophone cylindrique créé au Laboratoire d'acoustique de l'Université du Maine, au Mans, ou encore un nouveau type de clarinette développé dans notre laboratoire...

Pour le moment, les bouches artificielles n'ont servi qu'à étudier le « régime établi » de la production sonore, en d'autres termes le moment où la note sonne. Une note commence par une attaque, et s'éteint de façon plus ou moins longue ou abrupte. Pendant qu'elle sonne, la note peut être enrichie d'effets divers, par exemple un *vibrato* (on fait varier légèrement et rapidement la fréquence et l'am-

plitude de la note). Ainsi, si le son stationnaire (le régime établi) des cuivres et des instruments à anche est bien connu aujourd'hui, leurs transitoires ne le sont pas.

Les transitoires jouent un rôle majeur dans l'expression musicale, mais ils ne peuvent être étudiés à l'aide des bouches artificielles que nous avons décrites : celles-ci, en figeant les paramètres de contrôle, privent le scientifique de la possibilité d'étudier le pilotage fin exercé par le musicien. Il fallait donc développer un appareil dédié à l'étude des transitoires : la bouche artificielle asservie.

Asservir le dispositif à une consigne

De quoi s'agit-il ? Comme son nom l'indique, c'est une bouche artificielle classique augmentée par un système d'asservissement du son à émettre. Rappelons qu'un asservissement est un dispositif de correction conçu pour que le système commandé se comporte conformément à une certaine consigne. L'une de ses formes les plus simples, pour laquelle nous avons

LES AUTEURS



Christophe VERGEZ et Didier FERRAND sont chercheurs au CNRS et membres du Laboratoire de mécanique et d'acoustique de l'Université d'Aix-Marseille.

opté, est la régulation en boucle fermée. Dans ce principe, l'un des plus connus et des mieux éprouvés de l'automatique, la grandeur à réguler agit sur la grandeur régulante au sein d'une boucle fermée (dans une régulation en boucle ouverte, une grandeur perturbatrice supplémentaire agit en plus sur la grandeur régulante).

Conduire une voiture, par exemple, consiste pour l'essentiel à corriger en permanence sa direction pour la maintenir parallèle à la trajectoire définie par la médiane de la route. La consigne à laquelle le véhicule doit se conformer est donc de maintenir « une direction parallèle à la médiane de la route ». Pour effectuer les réglages nécessaires, le chauffeur mesure en permanence l'angle entre la direction de la voiture et la médiane à l'aide de ses capteurs (les yeux), puis déduit de la valeur trouvée la correction angulaire requise, qu'il impose ensuite à l'aide d'actionneurs (ses mains) agissant sur la direction de la voiture (la grandeur à contrôler).

Dans le cas qui nous intéresse, la grandeur à contrôler est la pression de l'air à l'entrée de l'instrument. Pour ce faire, le moyen le plus simple consiste à régler le débit d'air entrant dans la bouche. Nous avons utilisé pour cela une électrovanne, un actionneur commandé électriquement qui modifie le débit par une action mécanique. On intercale cette électrovanne entre la bouche artificielle et le dispositif d'alimentation en air (voir l'encadré page ci-contre), un détendeur qui produit, à partir de la source d'air comprimé (sept bars), une pression plus faible en amont de l'électrovanne (deux bars). C'est enfin par l'action de l'électrovanne qu'on obtient une pression comparable à celle qui règne dans la bouche d'un musicien. Pour sa part, le signal à contrôler, c'est-à-dire la pression régnant dans la bouche, est mesuré en permanence à l'aide d'un capteur de pression placé en retrait des lèvres.

Comment agit-on sur la vanne en fonction de la pression d'alimentation mesurée ? Il faut pour cela un régulateur, c'est-à-dire un dispositif capable de déterminer les ordres à fournir à l'électrovanne en fonction de l'écart entre la pression mesurée et la pression désirée (la consigne). Une étape essentielle consiste à en choisir le principe. Un conducteur qui braque trop vite dans un virage risque de produire un écart si dangereux qu'il lui faudra le corriger par un nouvel écart en sens contraire, etc. Ainsi, un mauvais mode de régulation

peut produire de dangereuses instabilités dans le contrôle, ce qui obligera à multiplier les corrections sans parvenir à coup sûr à réduire l'écart à la consigne. L'instrumentiste évite ce genre d'inconvénient grâce à un asservissement ultraperformant, son oreille et sa sensibilité tactile lui permettant de percevoir finement le retour d'efforts sur sa bouche et ses doigts.

Nous ne pouvons prétendre imiter une telle finesse de contrôle, aussi avons-nous opté pour un algorithme très utilisé dans l'industrie, qui a fait ses preuves dans de très nombreuses situations: le régulateur proportionnel intégral dérivé.

Connu sous son acronyme PID, ce régulateur combine trois actions de corrections à partir de la mesure de l'erreur par rapport à la consigne (ici un écart de pression): une action dite proportionnelle (l'erreur est multipliée par un gain), une action intégrale (l'erreur est intégrée sur

LA RÉGULATION EN BOUCLE FERMÉE, principe de régulation parmi les plus simples, suffit à asservir le débit d'entrée à la pression d'alimentation.

un certain intervalle de temps) et une action dérivée (la dérivée de l'erreur par rapport au temps est calculée à l'instant de mesure). Même si ce principe de régulation a des limites, il a suffi pour reproduire une grande variété de transitoires.

Ainsi équipés d'un système de mesure asservi à la pression dans la bouche, nous avons commencé à explorer les productions sonores associées à diverses consignes correspondant à des situations rencontrées par les musiciens. Par exemple, pour la flûte à bec, nous avons mis en évidence deux caractéristiques importantes de la production sonore. Nous avons d'abord constaté que l'augmentation progressive de la pression dans la bouche est associée à une modification continue du timbre, c'est-à-dire de la contribution des différentes fréquences à chaque note.

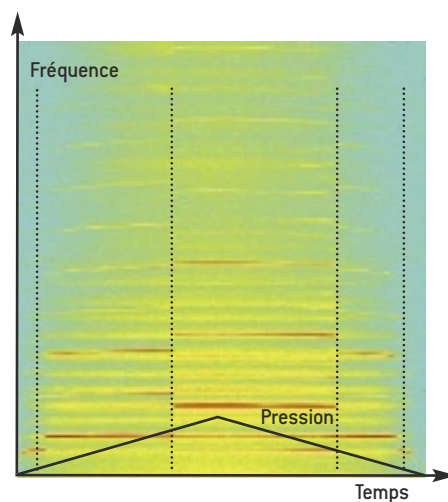
Ensuite, on peut vérifier que, pour certaines valeurs de la pression, des changements brutaux dans le son se produisent. Ces « bifurcations » correspondent à la naissance ou à l'extinction de notes ou au changement de la fréquence d'un son. Elles jouent un rôle essentiel dans l'analyse d'un

✓ BIBLIOGRAPHIE

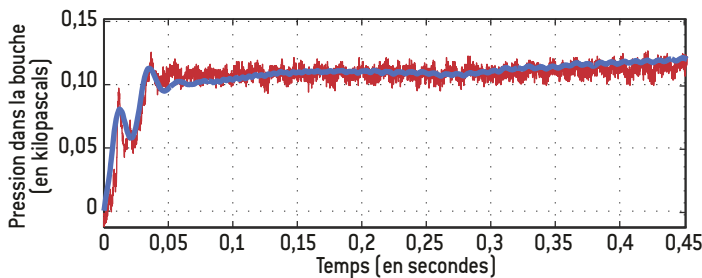
D. Ferrand et Ch. Vergez, *Blowing machine for wind musical instrument : toward a real-time control of the blowing pressure*, 16th Mediterranean Conference on Control and Automation, juin 2008.

M.J. Newton, M. Campbell et J. Gilbert, *Mechanical response measurements of real and artificial brass players lips*, J. Acoust. Soc. Amer., vol. 123, n° 1, 2007.

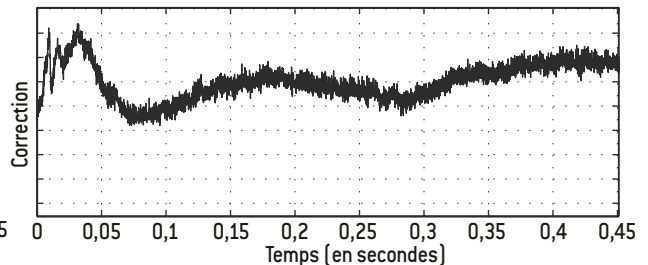
Ph. Guillemain, *A digital synthesis model of double-reed wind instruments*, EURASIP Journal on Applied Signal Processing, numéro spécial Model-based Sound Synthesis, vol. 4, n° 7, pp. 990-1000, juin 2004.



2. LE SIGNAL DE PRESSION d'une flûte à bec Bressan est ici tracé en fausses couleurs allant du vert (*faible pression*) au rouge (*forte pression*) en fonction de la fréquence et du temps, lorsque la bouche artificielle asservie impose à l'instrument une pression qui croît, puis décroît linéairement (*trait noir*). On constate des changements de régime, nommés bifurcations (*pointillés*), où la répartition en fréquences (le timbre) se modifie brutalement.



3. UN TRANSITOIRE DE PRESSION (la transition entre deux notes) d'une flûte à bec a été mesuré (à gauche en rouge) et comparé à la consigne (en bleu). La structure dentée du signal de pression correspond au bruit parasite qui règne à l'intérieur de la bouche artificielle, nettement plus fort que celui qui règne dans la bouche d'un musicien. La pression



dans la bouche artificielle est sans cesse ajustée pour se rapprocher de la consigne au moyen d'une correction (en noir à droite). Cette correction correspond à la commande envoyée à l'électrovanne. Sa forme diffrère du profil de pression durant la transition entre notes, ce qui illustre la non-linéarité de la bouche artificielle.

instrument de musique, puisqu'elles en constituent les caractéristiques essentielles. Ainsi, la confrontation des bifurcations mesurées à celles que prédit le modèle physique d'un instrument est un indicateur de la qualité de la modélisation effectuée pour en décrire la production sonore.

Des retombées pédagogiques

Toutefois, la grande application de la bouche artificielle asservie est l'étude des régimes transitoires. Pour cela, on établit comme consigne le profil de pression nécessaire pour imiter la façon dont un musicien attaque une note, puis passe d'une note à l'autre. Là encore, l'étude des transitoires sera une occasion d'améliorer les modèles physiques utilisés pour décrire la production sonore des instruments, les modélisations existantes ayant presque toutes été développées dans un contexte quasi statique.

Une troisième application de la bouche artificielle asservie est l'analyse des stratégies développées par le musicien expert pour piloter son instrument. On peut en effet tenter de reproduire avec la bouche artificielle l'évolution de la pression accompagnant les sons émis par les musiciens.

Des retombées pédagogiques sont même envisageables: on peut, par exemple, l'utiliser pour démontrer à des clarinettes qu'il est possible, sous certaines conditions, de produire un son *pianissimo* en soufflant de plus en plus fort dans l'instrument, ou encore pour leur révéler que s'ils n'appliquent pas certaines techniques, c'est par habitude plus que parce qu'elles sont à proscrire. La bouche artificielle asservie est autant un banc de mesure qu'un banc d'essai pour musicien.

La bouche artificielle peut aussi être considérée comme une étape préliminaire

vers la mise au point d'un robot musicien. Avec nos collègues de l'IRCAM, nous avons lancé le projet *Consonnes*, qui vise à produire un robot capable de contrôler la pression dans sa bouche et de se servir de ses lèvres. Tandis que nous sommes chargés de la réalisation des « poumons artificiels », les chercheurs de l'IRCAM travaillent à la réalisation de lèvres asservies dont on peut contrôler indépendamment la masse, la raideur et l'amortissement. En fin de projet, il est prévu de construire deux prototypes. Puisqu'ils seront dépourvus de moyens mécaniques d'action sur l'instrument lui-même, ces robots ne sont pas destinés à produire de la musique, mais seulement à étudier en détail le contrôle complexe exercé par le musicien par l'intermédiaire de son souffle et de ses lèvres.

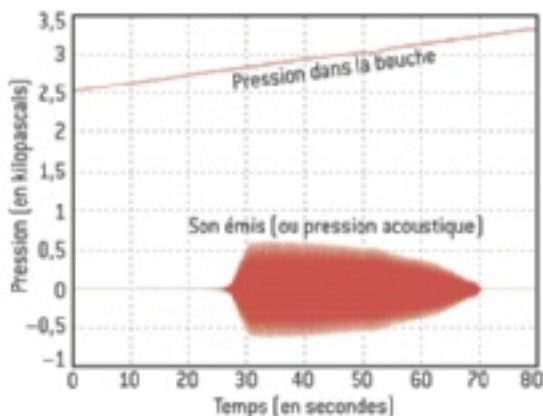
Après des années de pratique, un musicien expert acquiert une maîtrise si complexe de son instrument qu'elle en devient extrêmement difficile à analyser, tant pour les autres musiciens (ses élèves) que pour les physiciens. La bouche de joueur d'instrument à vent robotisée que nous développons autorisera une telle analyse.

En effet, l'étude scientifique des instruments à vent ne progressera vraiment que le jour où l'on pourra spécifier en quoi consiste la maîtrise instrumentale particulière d'un musicien. En effet, dans les mains d'un piètre instrumentiste, même un instrument de très bonne facture sonne mal, tandis que même un piètre instrument sonne bien dans les mains d'un excellent instrumentiste!

Ainsi, pour que l'analyse scientifique du jeu sur cuivres, bois ou flûtes entraîne des progrès tant en facture instrumentale qu'en musique, il faut parvenir à séparer la part de l'homme de celle de l'instrument. Un robot capable d'imiter finement les actions d'un musicien, de façon mesurable et reproductible, en est le moyen. ■

✓ SUR LE WEB

Robots musiciens ou parlants : <http://www.takanishi.mech.waseda.ac.jp/research/index.htm>



4. LE SON ÉMIS PAR UNE CLARINETTE en fonction du temps est ici représenté lorsque la pression dans la bouche croît linéairement alors que l'anche est fortement pincée. On constate que souffler toujours plus fort conduit à l'extinction progressive du son, ce qui fournit une technique inattendue pour produire un son *pianissimo*.

C'est quand on croit en soi qu'on se dépasse. Tous les chercheurs vous le diront.

L'innovation ce n'est pas un concept, c'est une stratégie. En tant que chercheur, vous le savez mieux que personne. Avant de publier, il est essentiel de déposer un brevet pour protéger et valoriser vos travaux... et pour qu'ils vous soient attribués. De plus, une veille technologique effectuée dans les brevets existants optimise votre potentiel de recherche et oriente au mieux vos projets à venir. **Connectez-vous sur www.inpi.fr**

On gagne tous à innover.

inpi
INSTITUT NATIONAL
DE LA PROPRIÉTÉ
INDUSTRIELLE



Des instruments de musique...

En modélisant les phénomènes physiques et en utilisant des outils mathématiques et informatiques, on sait aujourd'hui créer des instruments de musique virtuels aux sonorités réalistes.

Thomas Hélie et Christophe Vergez

Pourquoi, lorsqu'on souffle plus fort dans une flûte, le son produit est non seulement plus fort, mais la note ou le timbre changent aussi ? Quelle est l'influence de la géométrie et du matériau sur l'instrument ? Pourquoi un musicien a-t-il besoin d'un apprentissage long et difficile pour donner de l'expressivité à son art ? De telles questions laissent entrevoir la complexité des phénomènes mis en jeu dans le fonctionnement des instruments de musique.

Cette complexité est à l'origine de ce qui rend le son de l'instrument naturel, unique et vivant. Mais elle n'a pas empêché les scientifiques de chercher à comprendre les instruments de musique et à reproduire électroniquement leurs sonorités. Les progrès réalisés sont tels qu'on sait aujourd'hui simuler sur ordinateur et piloter en temps réel des instruments de musique virtuels avec un rendu réaliste.

Dans l'histoire de la synthèse sonore, qui vise à créer des sons de façon artificielle (par voie électronique ou informatique), on peut distinguer deux types de procédés. Dans la « synthèse par modèles de signaux », on s'attache à construire directement le signal électrique à envoyer dans les haut-parleurs, pour obtenir des sonorités insolites qu'on trouve par exemple dans la musique dite électronique. La « synthèse par modèles physiques » s'intéresse, au contraire, au générateur de l'onde sonore, c'est-à-dire à l'instrument et à l'instrumentiste, plutôt qu'au

son lui-même. Dans ce cas, qui fait l'objet du présent article, on veut mimer la réponse de l'instrument à tous les gestes possibles du musicien : attaques, transitoires entre notes, évolution du timbre avec la nuance.

Quel est l'intérêt de la synthèse par modèles physiques ? Il est, bien sûr, de disposer de versions virtuelles et jouables d'instruments réels, anciens ou même disparus. Mais il est aussi de tester virtuellement des modifications de la géométrie et des matériaux de l'instrument, voire d'optimiser la puissance de celui-ci, sa brillance sonore, etc. Il est par ailleurs intéressant de concevoir des instruments imaginaires ou farfelus, mais qui restent physiquement sensés : par exemple des instruments de taille immense, un conduit vocal de dragon ou de dinosaure, un hautbois évoluant en saxophone, etc.

Les travaux pionniers du savant allemand Hermann von Helmholtz, à la fin du XIX^e siècle, et de nombreux autres, en partie rassemblés par les Américains Neville Fletcher et Thomas Rossing à la fin du XX^e siècle, ont permis de comprendre le fonctionnement élémentaire des instruments de musique. En n'en retenant que les éléments clefs, plusieurs équipes dans le monde ont proposé dès la fin des années 1980 des premières synthèses sonores en temps réel, c'est-à-dire des synthétiseurs qui calculent suffisamment vite pour que l'opérateur humain ne perçoive aucun délai entre son action et le résultat sonore.

L'ESSENTIEL

- ✓ Scientifiques et ingénieurs conçoivent des instruments virtuels, dont on synthétise les sons.
- ✓ Cette synthèse sonore requiert une modélisation physique de l'instrument qui soit à la fois soigneuse et assez simple pour procéder en temps réel.
- ✓ La prise en compte du jeu de l'instrumentiste, en plus de la physique de l'instrument, reste un défi.

Instrument

Trompette
Saxophone
Trombone

Préréglages

Trompette Big Band 1
Trompette Big Band 2
Trompette
Trompette Big Band
Trompette Big Band 3

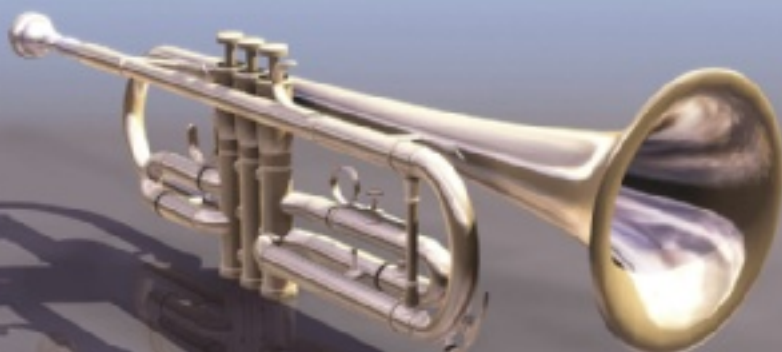
Fichier

Nouveau Enregistrer Enr. sous
Supprimer Importer Exporter

Timbre Vibrato

virtuels

Synthétiseur trompette



Attaque Pression Hauteur Timbre Souffle Vibrato Position Sourdine



Configuration

Trompette 1



Numéro 2

Spatialisation



Réglages MIDI

Contrôleur
de souffle
+ clavier

Canal MIDI



D'après l'interface du logiciel BRASS commercialisé par la Société grenobloise Arturia

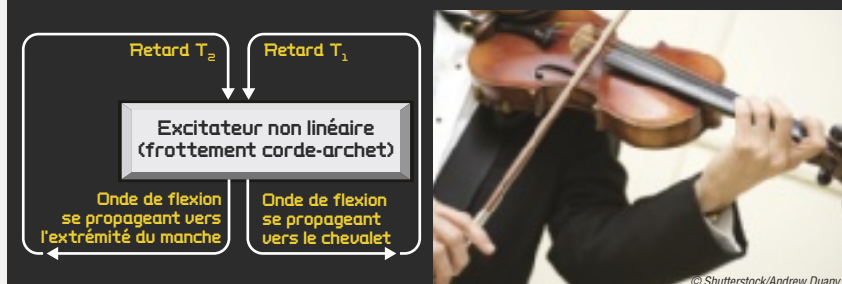
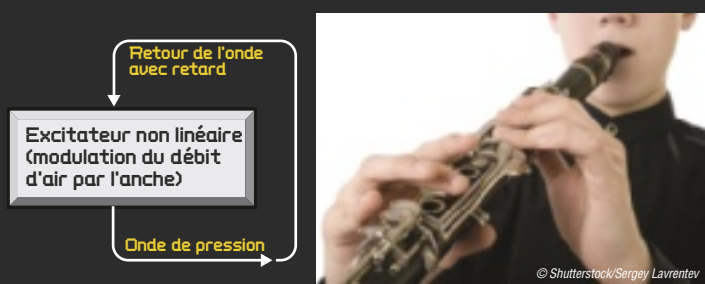
1. UN EXEMPLE D'INTERFACE graphique d'un logiciel permettant de piloter des modèles de trompette, de saxophone ou de trombone. La partie grise du clavier virtuel ici figuré correspond aux notes physiquement réalisables avec une trompette réelle.

Les simulations les plus simples du fonctionnement d'un instrument de musique intègrent deux ingrédients clés. D'une part, il faut représenter le mécanisme non linéaire d'excitation (modulation du débit d'air par l'anche dans le cas d'une clarinette, frottement entre l'archet et la corde dans le cas d'un violon, etc.) par un système simple d'équations non linéaires, bâti à l'aide de fonctions élémentaires. D'autre part, on doit prendre en compte les retards des ondes, c'est-à-dire les temps mis par les ondes pour effectuer un aller et retour au sein de l'instrument. Ainsi, une onde acoustique qui se propage d'une distance L dans un tube droit, puis se réfléchit, subit un retard $T = 2L/c$, imposé par la vitesse du son c .

Considérons l'exemple d'une clarinette. Chaque déplacement de l'anche crée une fluctuation de pression qui se propage à la vitesse c le long de l'instrument, jusqu'au premier

trou ouvert (situé à une distance L), où elle est réfléchiée avec un changement de signe. Cette onde revient donc à l'anche après un retard $T = 2L/c$ et y modifie les conditions de pression. En conséquence, l'anche change de forme, ce qui modifie le débit d'air entrant et crée une nouvelle perturbation de pression.

Autre exemple : un violon et son archet. Le frottement entre le crin colophané et la corde engendre, par frottement, des ondes de flexion sur chaque partie de la corde. Ces ondes se propagent à une certaine vitesse, l'une vers le chevalet, l'autre vers l'extrémité du manche. Parvenues à l'extrémité, elles y sont réfléchies. Après des temps de parcours différents T_1 et T_2 (sauf si l'archet frotte la corde en son milieu, auquel cas les temps de parcours sont égaux), les deux ondes reviennent à l'archet et interagissent à nouveau avec lui.



En particulier, dans l'équipe de Xavier Rodet, à l'IRCAM, des modèles physiques simplifiés au maximum ont permis dans les années 1990 d'expliquer et de reproduire des comportements caractéristiques de chaque famille d'instruments. Depuis, nos travaux s'attachent à améliorer les modèles pour augmenter le réalisme tout en préservant le traitement en temps réel. Un autre défi pour le réalisme reste le contrôle de l'instrument virtuel : même avec une excellente modélisation de l'instrument, le premier son que l'on produit en jouant sur un instrument virtuel est souvent... un canard ! Il s'agit de construire les « réflexes coordonnés » acquis par le musicien expert et qui lui permettent de bien jouer. Parmi les pistes explorées, nos recherches portent sur les méthodes d'inversion : elles consistent à retrouver les gestes de pilotage du modèle nécessaires à l'obtention d'un son voulu donné. Passons en revue ces différents aspects successivement.

Avant toute tentative de synthèse par modèle physique, il faut étudier et comprendre le fonctionnement d'un instrument de musique. Depuis Helmholtz, on sait que ce fonctionnement repose sur une boucle de rétroaction où un excitateur (ancho, archet, lèvres, plectre, jet d'air...) se

couple à un résonateur (tuyau, corde, membrane...). Ce faisant, le résonateur cède un peu d'énergie sous la forme d'un rayonnement acoustique, perçu par nos oreilles.

Dans le cas des instruments entretenus – c'est-à-dire capables d'auto-osciller spontanément à partir d'une excitation statique telle qu'un souffle constant, par opposition aux instruments pincés ou frappés –, l'émission d'une note résulte d'un couplage compliqué entre l'excitateur et le résonateur, et requiert en permanence des ajustements très fins de la part du musicien.

Pour la synthèse sonore en temps réel, on doit réduire au maximum la complexité du modèle de l'instrument afin d'avoir des temps de calcul courts. Mais le modèle simplifié doit aussi pouvoir expliquer une grande partie du fonctionnement d'un instrument. Une telle « formulation minimale » peut intégrer des descriptions sommaires de l'excitateur et du résonateur, mais, pour expliquer et reproduire le phénomène d'auto-oscillation, elle ne peut faire l'impasse sur la nature non linéaire de l'excitateur et de son couplage avec le résonateur (voir l'encadré ci-dessus). L'expression « non linéaire » signifie que la réponse (ou « sortie ») du système n'est pas proportionnelle à l'action (ou « entrée ») qu'on lui applique.

Il a été montré au milieu des années 1990 que ces modèles simples et peu coûteux à simuler incorporent assez la physique de l'instrument pour produire des comportements comparables à ceux de l'instrument réel : le son synthétisé est déjà caractéristique de telle ou telle classe d'instruments. Cependant, il faut raffiner les modèles pour réussir à leurrer l'oreille quant à la nature synthétique du son. Comment le faire de façon compatible avec une synthèse en temps réel, c'est-à-dire en n'allongeant pas trop les temps de calcul ?

Assimiler des tubes à des cylindres ?

L'une des approches est de modéliser de façon simplifiée la forme géométrique du résonateur. Les résonateurs d'instruments, qu'ils soient mécaniques (cordes, plaques, etc.) ou acoustiques (tubes) sont le lieu où se propagent des ondes. Simuler précisément cette propagation dans une géométrie tridimensionnelle se révèle souvent trop coûteux en calculs. Prenons le cas d'un résonateur d'un instrument tel qu'une trompette, qui a la forme d'un tube à section variable – on parle de la « perce » de l'instrument. Comment le représenter

de façon simplifiée? La première idée pour la synthèse sonore était d'approcher cette perce par un profil en marches d'escalier, comme si l'on raccordait des tubes droits (voir la figure 2).

Dans ce cadre, quand une onde acoustique plane, c'est-à-dire dont les fronts sont plans et perpendiculaires à l'axe du tube, se propage le long du tube et arrive à une discontinuité de section, elle subit une réflexion partielle instantanée (comme si une partie de l'onde rebondissait sur cette jonction) et se transmet aussi partiellement au tronçon voisin.

En fin de compte, on aboutit à une « structure de guide d'ondes » où, du point de vue mathématique et informatique, n'intervient qu'un petit nombre d'opérations élémentaires : retards de l'onde proportionnels à la longueur du tube, additions, multiplications des amplitudes ondulatoires par des coefficients de réflexion qui dépendent des sections du tube. Ce modèle a été utilisé à l'origine dans les années 1960 pour simuler le conduit vocal humain ; une fois optimisée, la structure informatique correspondante a été nommée, d'après les deux chercheurs qui l'ont introduite, structure de Kelly-Lochbaum.

Autour de 1995, diverses équipes ont modélisé un peu plus finement la forme du tube. L'idée était de considérer des tronçons de tubes coniques dans lesquels voyagent des ondes sphériques (c'est-à-dire dont les fronts sont sphériques). La structure de Kelly-Lochbaum finale est identique à celle du modèle à tubes droits, sauf que les discontinuités de pente aux jonctions entre tubes rendent les réflexions d'ondes non plus instantanées, mais à mémoire avec réponse exponentielle – en d'autres termes, la fraction d'onde réfléchie laisse une trace qui décroît exponentiellement avec le temps. Par rapport à la structure à tubes droits, ce modèle est à peine plus coûteux à simuler numériquement.

Ces modèles à tronçons de tube droits ou coniques sont-ils suffisants pour donner un rendu réaliste? Pas vraiment. Les discontinuités de la section ou de la pente du profil, aux jonctions des tronçons de tube, engendrent des sauts de pression audibles dans la réponse de l'instrument. Ainsi, pour un instrument à perce régulière, la simulation crée des artefacts que l'oreille perçoit.

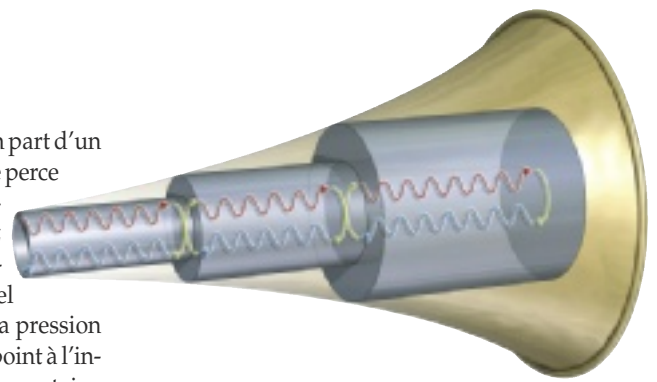
Pour renforcer le réalisme tout en restant compatible avec un traitement en temps réel, nous avons conçu vers 2002 un modèle à une dimension qui respecte

mieux la géométrie du tube. On part d'un tube à symétrie axiale, avec une perce dont le rayon dépend régulièrement de la coordonnée z mesurée le long de l'axe. Le calcul des sons produits par un tel tube implique de déterminer la pression acoustique régnant en chaque point à l'intérieur du tube, en réponse à une certaine pression appliquée à l'entrée. Ainsi, pour chaque position z le long de l'axe, on s'intéresse à la pression acoustique régnant à la distance r de l'axe. Cette pression obéit à une équation, qui décrit la propagation des ondes acoustiques, et dépend à la fois de la position et du temps. Dans le plan (z, r) et à un instant donné, les points associés à une même valeur de la pression constituent des courbes, nommées isobares comme sur une carte météorologique (voir la figure 3).

Réduire la dimension du modèle

Les isobares dessinent en général des lignes courbes, ce qui traduit le fait que la pression dépend à la fois de la coordonnée radiale r et de la coordonnée axiale z . L'idée de notre modélisation unidimensionnelle revient en quelque sorte à construire un miroir déformant qui redresse les isobares, pour les transformer en lignes droites perpendiculaires à l'axe. Dans cette nouvelle image, la pression devient indépendante de la coordonnée radiale, ce qui réduit la dimension effective du problème à un et simplifie beaucoup les calculs.

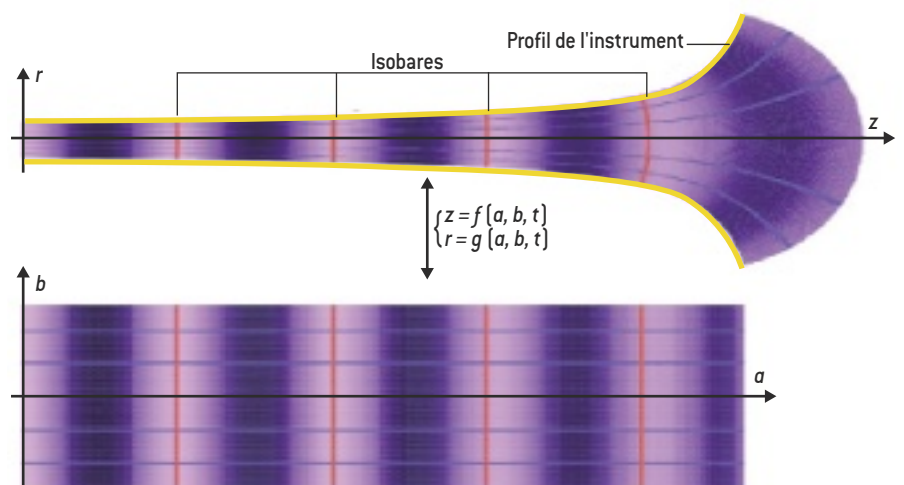
L'outil mathématique qui effectue cette transformation est un changement de coordonnées dynamique : on passe des coor-

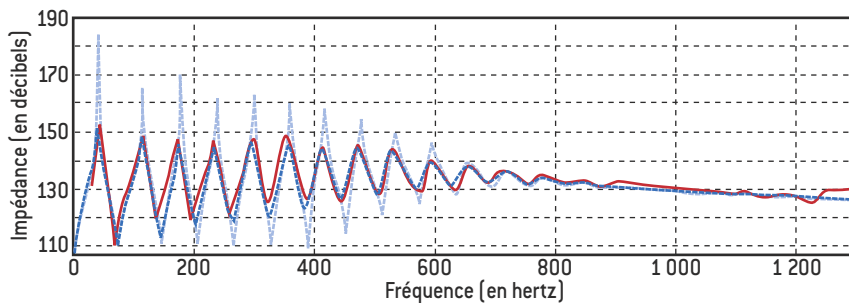


2. POUR MODÉLISER UN INSTRUMENT à vent en forme de tube de perce (ou profil) variable, on peut en première approximation l'assimiler à un ensemble de trois tubes droits. À chaque jonction entre tubes, une partie des ondes acoustiques est réfléchie (en bleu) dans le même tube et une autre transmise (en rouge) au tronçon voisin. Le modèle correspondant, après optimisation, est nommé structure de Kelly-Lochbaum.

Bruno Bourgeois

3. LES ONDES ACOUSTIQUES qui se propagent dans un tube symétrique autour de son axe Oz représentent un problème à deux dimensions, où les coordonnées pertinentes sont z et r (distance à l'axe). Les isobares sont ici des courbes. En effectuant un changement de coordonnées dynamique de (z, r) en (a, b) , à l'aide de fonctions appropriées f et g qui dépendent du temps et du profil du tube, on peut redresser les isobares. Autrement dit, les isobares deviennent des droites perpendiculaires à l'axe de la coordonnée a , ce qui signifie que la pression est indépendante de la coordonnée b . La dimension du problème se réduit à un, ce qui simplifie les calculs.





4. LES PERTES D'ÉNERGIE PAR FROTTEMENTS et échanges thermiques entre l'air en vibration et le tube d'un instrument à vent doivent être prises en compte pour qu'une simulation d'un instrument soit réaliste. Ces courbes l'illustrent : elles comparent l'impédance d'entrée d'un trombone telle qu'elle est mesurée (en rouge) avec l'impédance calculée avec les modèles des auteurs (sans pertes viscothermiques, en pointillés bleu clair ; avec pertes, en bleu foncé).

données z et r à deux autres coordonnées a et b grâce à des fonctions appropriées f et g , qui dépendent du temps t (et de la perce de l'instrument). On réécrit ensuite l'équation des ondes en termes des coordonnées a et b , de façon que la pression y apparaisse comme une fonction de la nouvelle coordonnée a uniquement, et non de b .

Autrement dit, en posant $z = f(a, b, t)$ et $r = g(a, b, t)$, on peut se ramener, connaissant le profil du tube et moyennant quelques approximations (justifiées lorsque les fréquences considérées sont assez basses), à la géométrie d'un problème à une dimension, comme pour des ondes planes. Nous avons montré que le modèle ainsi obtenu suffit pour simuler la contribution géométrique de tubes à symétrie axiale dont le profil et sa pente sont continus.

Pour les cordes, coques et plaques, des modèles dits de Reissner permettent de prendre en compte les principales caractéristiques géométriques du résonateur. Des méthodes de résolution de ces modèles, efficaces en temps réel, sont en cours d'étude avec David Roze et Joël Bensoam à l'IRCAM, et en partenariat avec le LIST (Laboratoire d'intégration des systèmes et des technologies) du CEA, dans la région parisienne.

La géométrie du résonateur n'est pas le seul aspect à prendre en compte dans la synthèse sonore. Un autre est l'amortissement. Si l'archet ou le souffle du musicien s'arrête, le son finit par s'éteindre : un instrument de musique est un système amorti, où l'énergie perdue par l'instrument est essentiellement le son émis lui-même. Or les émissions acoustiques des tubes, barres, plaques sont connues, et on sait les prendre en compte avec une bonne approximation dans les simulations du fonctionnement d'un instrument.

Cependant, d'autres types d'amortissements intrinsèques au résonateur ont

un effet important sur le son. Reprenons le cas des tubes. La colonne d'air en vibration dans le tube crée un frottement sur ses parois et échange de la chaleur avec elles. Ces phénomènes viscothermiques dépendent de la texture de la surface du tube, du type de matériau et de son aptitude à conduire la chaleur. De plus, leur effet sur les ondes sonores dépend de la fréquence de celles-ci. Si la modélisation n'intègre pas bien cet amortissement, le son produit par l'instrument virtuel semble, à l'oreille, synthétique. Or la prise en compte du bon amortissement se révèle délicate sur le plan mathématique. En particulier, elle fait apparaître une opération étonnante et, en théorie, impossible à simuler exactement : une dérivée temporelle fractionnaire.

Pertes viscothermiques

Pour comprendre de quoi il s'agit, prenons l'exemple d'une particule en mouvement. La dérivée temporelle d'ordre un (la dérivée par rapport au temps) de sa position est sa vitesse. En dérivant une fois de plus, on obtient la dérivée temporelle d'ordre deux de la position, qui est l'accélération de la particule. Une dérivée d'ordre $1/2$ est un opérateur qui, appliqué deux fois à la position, donne la vitesse. Il est difficile de donner une interprétation simple de cet objet mathématique, sauf peut-être dans le domaine fréquentiel. Une dérivée temporelle d'ordre un appliquée à un signal correspond à un filtre qui augmente de six décibels par octave les amplitudes des ondes sinusoïdales qui composent le signal (autrement dit, une onde de fréquence double verra son amplitude doublée) et déphase ces ondes de 90 degrés. Pour une dérivée d'ordre $1/2$, on aura alors une augmentation de trois décibels par octave et 45 degrés de déphasage... L'influence sur le timbre du son est manifeste : les dérivées d'ordre fractionnaire n'agissent pas de la même façon sur les composantes graves et aiguës du son.

Avec Denis Matignon et Rémi Mignot, nous avons récemment développé de nouvelles méthodes d'approximation (compatibles avec le temps réel) pour résoudre, en incluant les pertes viscothermiques, le modèle unidimensionnel de tube à section variable décrit précédemment. Les résultats sont très satisfaisants. Par

LES AUTEURS



Thomas HÉLIE est chargé de recherche au CNRS et travaille à l'IRCAM, à Paris (UMR 9912).

Christophe VERGEZ est chargé de recherche au CNRS et travaille au Laboratoire de mécanique et d'acoustique (UPR 7051), à Marseille.

✓ BIBLIOGRAPHIE

A. Chaigne et J. Kergomard, *Acoustique des instruments de musique*, Belin, à paraître, 2008.

Th. Hélié et D. Matignon, *Representation with poles and cuts for the time-domain simulation of fractional systems and irrational transfer functions*, Journal of Signal Processing, vol. 86, pp. 2516-2528, 2006.

Th. Hélié, *Mono-dimensional models of the acoustic propagation in axisymmetric waveguides*, J. Acoust. Soc. Amer., vol. 114, pp. 2633-2647, 2003.

Ch. Vergez et X. Rodet, *Trumpet and trumpet player: a highly nonlinear interaction [...]*, International Journal of Bifurcation and Chaos, vol. 11(7), pp. 1801-1814, 2001.

exemple, si l'on compare l'impédance d'entrée (rapport entre la pression acoustique et le débit d'air à l'entrée de l'instrument ; cette impédance mesure la « résistance » acoustique de l'instrument) d'un trombone réel à celle obtenue avec notre instrument virtuel, on constate un comportement presque identique en fonction de la fréquence, ce qui n'était pas du tout le cas lorsqu'on négligeait les pertes viscothermiques (voir la figure 4).

Pour les résonateurs à cordes, barres ou plaques, les amortissements peuvent avoir des complexités similaires, selon les matériaux. On devrait pouvoir leur adapter les méthodes mathématiques utilisées pour le cas des instruments à vent, ce à quoi nous nous attelons.

Comme on l'a vu, la géométrie du résonateur et les phénomènes d'amortissement viscothermiques sont deux points clefs pour élaborer un instrument virtuel à

rendu réaliste. Un troisième est que, dans un instrument de musique, les ondes d'amplitude élevée sont souvent distordues par des effets non linéaires.

Des distorsions non linéaires

Voyons comment sur l'exemple de la guitare. Pour accorder votre guitare, vous réglez la tension des cordes grâce aux mécaniques : plus la tension est élevée, plus la célérité de propagation des ondes (et donc la hauteur du son !) l'est aussi. Or pendant que la corde vibre, sa longueur varie par élasticité, ce qui modifie la tension et, partant, la vitesse de propagation des ondes. Il s'ensuit une distorsion des ondes, effet d'autant plus marqué que l'on pince fortement la corde.

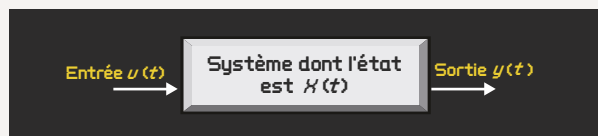
Autre exemple : pour accorder votre trompette, vous réglez la longueur du tube

à l'aide d'une petite coulisse... mais seulement après avoir joué un peu pour chauffer l'instrument. Car plus l'air est chaud, plus la célérité du son y est grande et la note est haute. Or une surpression d'une onde sonore chauffe l'air où elle se propage, ce qui augmente aussi la célérité. Ainsi, comme pour la corde de guitare que l'on pince plus fortement, souffler plus fort dans la trompette distord l'onde sonore produite.

En résumé, modestes à faible amplitude, ces effets non linéaires deviennent décelables aux nuances *fortissimo* sur certains instruments. Par exemple, le « cuivrage », ou « cuivrabilité » (la brillance, ou renforcement des hautes fréquences par rapport aux basses), des sons forts des trombones et trompettes provient de telles distorsions. Pour la synthèse sonore, les techniques mathématiques évoquées précédemment ne conviennent plus. Pour simuler des systèmes entrée/sortie (un

CONVOLUTION ET SÉRIES DE VOLTERRA

En traitement du signal et en automatique, on considère souvent deux grandeurs privilégiées relatives au système étudié : l'entrée $u(t)$ et la sortie $y(t)$, qui sont des fonctions du temps t . L'entrée peut correspondre à un signal électrique à émettre, à traiter ou à transformer, ou encore à la commande d'un système physique (une force appliquée à un système mécanique, le flux d'un produit chimique à transformer, etc). La sortie correspond alors au signal reçu, traité ou transformé, ou encore au déplacement du système mécanique, à la quantité de médicament synthétisé, etc. Les liens entre entrée et sortie font souvent intervenir des grandeurs intermédiaires (les signaux électriques de l'ensemble d'un circuit, les concentrations de composés intermédiaires pendant la réaction chimique...): l'ensemble de ces grandeurs constitue l'état $X(t)$ du système.



Il existe de nombreux cas où l'on ne sait pas écrire directement la relation entre $u(t)$ et $y(t)$, mais où les relations qui décrivent la dynamique du système sont de la forme :

$$X'(t) = f(X(t), u(t))$$

$$y(t) = g(X(t), u(t)).$$

Ici, f et g sont des fonctions de l'état interne X du système et de l'entrée u , et $X'(t)$ désigne la dérivée temporelle de $X(t)$. La première équation régit la dynamique de l'état du système, la seconde fournit la sortie.

Quand sait-on trouver une relation explicite entre l'entrée u et la sortie y ? Une telle relation existe pour un système linéaire où f et g sont de la forme $f(X, u) = aX + bu$ et $g(X, u) = cX + du$.

Plus précisément, pour un système causal (dont la sortie ne dépend pas du futur de l'entrée), excité par l'entrée u supposée nulle avant l'instant $t = 0$ et initialement au repos ($X(0) = 0$), la relation entre u et y est de la forme :

$$y(t) = \int_0^t h(s) u(t-s) ds$$

où $h(t)$ est la réponse temporelle du système à une excitation infiniment brève, une impulsion idéale. On dit que $y(t)$ est la convolution de l'entrée $u(t)$ avec la réponse impulsionnelle $h(t)$ du système. L'évolution de h au cours du temps décrit comment se comporte la mémoire du système. Ainsi, la vitesse à laquelle la réponse h se rapproche de zéro indique à quelle vitesse la sortie « oublie » l'influence d'une entrée passée.

Lorsque les fonctions f et g ne sont pas linéaires, on n'a généralement pas de relation explicite entre l'entrée et la sortie. Toutefois, si $f(0, 0) = 0$ et si f et g admettent un développement en séries de puissances entières autour de $(X, u) = (0, 0)$, on peut encore trouver une relation explicite. Cette relation met en œuvre une « série de Volterra » et est de la forme :

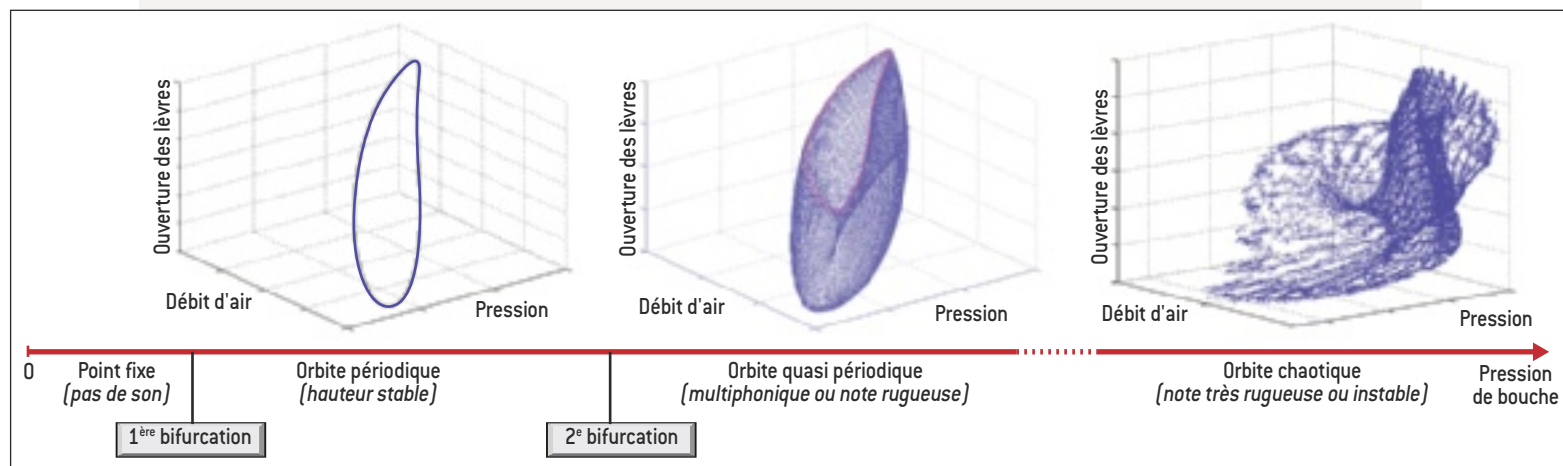
$$y(t) = \int h_1(s_1) u(t-s_1) ds_1 + \int \int h_2(s_1, s_2) u(t-s_1) u(t-s_2) ds_1 ds_2 + \int \int \int h_3(s_1, s_2, s_3) u(t-s_1) u(t-s_2) u(t-s_3) ds_1 ds_2 ds_3 + \dots$$

Il n'y a plus une seule réponse impulsionnelle $h(t)$, mais une infinité de fonctions h_1, h_2, h_3, \dots à une, deux, trois, ... variables respectivement. Le premier terme, en h_1 , correspond à la contribution linéaire du système. Une telle représentation est suffisamment souple pour mener des calculs explicites.

Mathématiquement, la question de la convergence de ces séries est délicate. Cependant, des travaux récents effectués avec Béatrice Laroche, à l'IRCAM, ont permis de préciser le domaine de convergence pour des systèmes dont l'état X est caractérisé par un nombre fini de variables.

Si un système physique voit son comportement qualitativement modifié quand on fait varier continûment l'un de ses paramètres, on dit qu'il est le siège d'une bifurcation. Pour les instruments de musique, l'exemple le plus commun est la bifurcation dite de Hopf, qui signe l'apparition d'un son correspondant à une onde périodique : c'est le cas en soufflant de plus en plus fort dans un instrument à vent, ou en augmentant la vitesse de l'archet sur la corde du violon. Mais les régimes d'oscillation observables sur un instrument de musique (ou sur les modèles physiques correspondants) ne sont pas toujours périodiques, loin s'en faut. Lors d'une bifurcation, un son peut apparaître, disparaître, changer de fréquence, devenir apériodique (quasi périodique ou

même chaotique)... On peut représenter les différents régimes d'oscillation dans l'espace des configurations, où les axes mesurent les différentes variables du système. Ainsi, sur la figure ci-dessous, le système considéré est un modèle de trompette, les variables étant la pression, le débit d'air à l'entrée et l'ouverture des lèvres. L'état du système à un instant donné correspond à un point dans l'espace des configurations. Au cours du temps, ce point dessine une trajectoire dont la nature caractérise le régime d'oscillation. En augmentant peu à peu la pression fournie par la bouche, on passe progressivement du silence au chaos. Cette richesse de comportements est l'une des difficultés auxquelles on se heurte pour contrôler un instrument virtuel.



signal d'entrée, qui représente l'action de l'instrumentiste, donnant un signal de sortie, c'est-à-dire les sons musicaux produits), nous avons utilisé une généralisation au cas non linéaire d'un outil mathématique bien connu applicable aux systèmes linéaires : la convolution. Cette généralisation de la convolution prend la forme de séries dites de Volterra, des sommes d'une infinité de termes où le premier est une intégrale à une dimension, le deuxième une intégrale à deux dimensions, etc. (voir l'encadré page 75).

Avec Martin Hasler, nous avons mis au point en 2004 une méthode pour résoudre les équations de propagation d'ondes faiblement non linéaires à l'aide de cet outil, en ne prenant que les premiers termes de la série de Volterra. Sur le plan musical, son application aux tubes droits (avec Vanessa Smet) et aux cordes (avec D. Roze) a abouti tout récemment à des simulateurs en temps réel qui sont réalistes pour l'oreille. Outre la synthèse, grâce aux algorithmes obtenus, nous avons aussi pu modifier de façon naturelle le cuivrage dans l'enregistrement d'un solo de trompettiste, comme on pourrait le souhaiter en studio de postproduction.

Maintenant, rêvez un peu et supposez que, grâce à la modélisation physique et aux techniques mathématiques, votre trompette en or massif au profil opti-

misé et votre stradivarius virtuels sont là. À vous de jouer ? Pas tout à fait... sauf si vous êtes trompettiste, violoniste et que vous disposez des capteurs électroniques de pression de bouche, d'état des muscles et de position des lèvres, de position et vitesse d'archet, de sa force d'appui, etc. Il faut en effet envoyer les bons gestes de pilotage à l'instrument virtuel pour ne pas faire de canard (au son certes réaliste) ou pour que l'oreille ne s'aperçoive pas que le son entendu est synthétique...

Calculer les bons gestes

Il existe ainsi pour certains instruments des dispositifs électroniques, allant des plus simples aux plus élaborés, qui enregistrent convenablement les gestes de l'instrumentiste. Et si celui-ci est chevronné, on arrive à restituer l'expressivité artistique sur l'instrument virtuel. Avec Philippe Guillemain, nous avons ainsi proposé une méthode non intrusive pour mesurer les variations du conduit vocal des saxophonistes. On peut ensuite utiliser ces informations pour contrôler des modèles physiques d'instrument.

Mais pour produire les bons gestes qui sélectionnent une infime partie des régimes acoustiques possibles de son instrument, le musicien a

✓ SUR LE WEB

Logiciel *Modalys* de l'IRCAM, développé par R. Caussé, J. Bensoam, N. Ellis : <http://forumnet.ircam.fr/701.html>

Projet CONSONNES [Contrôle des sons naturels et synthétiques] <http://www.consonnes.cnrs-mrs.fr/>

investi des années d'apprentissage... Aussi, nous explorons une autre démarche que celle consistant à mesurer les gestes d'un musicien expert: l'inversion. L'idée est la suivante. Enregistrez, en fonction du temps t , le son $s(t)$ produit par un instrument réel. Prenez un modèle physique de cet instrument. Comment piloter votre modèle physique pour obtenir le son voulu $s(t)$? C'est la question que pose l'« inversion de système ». Autrement dit, il s'agit, étant donné le son $s(t)$, de calculer les gestes $G(t)$ qui permettent de le produire (et non de partir des gestes pour arriver au son).

L'intérêt de l'inversion est multiple: élaboration de morceaux expressifs « à la manière de » à partir de bibliothèques de gestes analysés sur des musiciens professionnels, aide pédagogique par comparaison des gestes de l'élève et de son professeur, etc. Malheureusement, comme souvent pour les problèmes d'inversion, on se heurte à d'importantes difficultés. Une première raison est qu'il existe une infinité de gestes possibles conduisant à un son cible donné. Nous l'avons montré mathématiquement pour des modèles de cuivres (et cela même si le son cible est un son de violon... ce qui ne signifie pas qu'un trompettiste pourrait jouer un son de violon, car on doit se restreindre aux gestes humainement possibles!).

Une deuxième difficulté de l'inversion est qu'un bruit externe (page tournée par le musicien, réverbération de la salle) ou une partie non modélisée sur l'instrument (bruit de souffle, de piston, de clapet, de crin de l'archet) peut perturber notablement le système inverse s'il est sensible à de petites erreurs de mesure. Enfin, il est difficile de retrouver les gestes de pilotage sans reconstruire l'état vibratoire et acoustique interne (*a priori* inconnu) de l'instrument, et, pour les instruments à sons entretenus, on doit faire face à la grande diversité de régimes acoustiques possibles (voir l'encadré page ci-contre).

Nos premiers travaux en 1999 ont montré que, connaissant l'état vibro-acoustique du système, les gestes de pilotage d'un modèle de cuivre peuvent être retrouvés, et ce pour tout type de régime acoustique (transitoires, notes rugueuses, etc.). Avec Brigitte d'Andréa-Novel et Jean-Michel Coron, nous étudions aujourd'hui des techniques d'observation de l'état interne de l'instrument, qui permettraient de réaliser une inversion complète. Mais pour l'instant, on peut dire qu'obtenir une inversion complète et robuste est aussi délicat que d'effectuer une prévision météorologique fiable...

Ainsi, prendre en compte le jeu de l'instrumentiste en plus de la physique de l'instrument pour la synthèse sonore reste un défi: des techniques doivent être améliorées, explorées ou même inventées. Terminons en soulignant que résoudre ce problème d'interaction entre l'instrument et l'instrumentiste est encore plus audacieux lorsque les deux parties sont le plus intimement liées, à savoir dans le cas de la voix... ■



MUSÉUM NATIONAL
D'HISTOIRE NATURELLE

Clipperton

Environnement & biodiversité
d'un microcosme océanique

Loïc Charpy (coordinateur)

L'île de Clipperton appartient à une chaîne sous-marine jeune de la plus grande plaque tectonique du globe. Située entre la région Indo-ouest pacifique et le Pacifique est, Clipperton est très éloignée de l'épicentre de la biodiversité situé dans l'arc Indo-malais. C'est un atoll complètement fermé et son lagon de type méromictique est le siège de processus biogéochimiques originaux dont une phosphatogenèse actuelle. Aujourd'hui inhabitée, Clipperton a connu plusieurs périodes d'occupation humaine pendant la fin du XIX^e siècle et le début du XX^e siècle, puis ultérieurement des occupations de quelques jours à quelques mois. La cinquantaine de scientifiques qui participent à cet ouvrage rendent compte de leurs résultats qui portent sur le déplacement de la plus grande plaque tectonique du globe, la géomorphologie, la géochimie du lagon et la biodiversité terrestre, marine et lagunaire. Ils proposent aussi leurs réflexions sur l'avenir de Clipperton. Les données à l'origine de cet ouvrage proviennent pour la plupart de l'expédition organisée par Jean-Louis Étienne en 2005.



ISBN : 978-2-85653-612-4 | r. 500 p. | 59 € TTC

PUBLICATIONS SCIENTIFIQUES DU MUSÉUM

Commandes et renseignements :
Muséum national d'histoire naturelle
Publications scientifiques
Case postale 39 • 57 rue Cuvier • 75231 Paris cedex 05
Tél. : 01 40 79 48 05 • Fax : 01 40 79 38 40
diff.pub@mnhn.fr • www.mnhn.fr



Laurent Daudet

Les scientifiques cherchent des représentations numériques des sons qui soient non seulement compactes, mais aussi porteuses d'information sur le contenu musical.

Représenter

Dans leur version électronique, les sons musicaux peuvent être stockés sur disque, transformés par les ingénieurs du son, écoutés sur de petits appareils portables, échangés sur le réseau Internet, etc. C'est pourquoi la musique est aujourd'hui davantage présente sous la forme de signaux électriques numérisés, circulant dans des câbles et traités par des ordinateurs, que sous sa forme acoustique originelle.

De tels traitements nécessitent de choisir une « représentation » numérique appropriée des sons musicaux. Dans un sens littéral, représenter signifie transformer un objet de manière à ce que certaines de ses caractéristiques apparaissent le plus clairement possible. L'action de représenter désigne ici tout traitement préalable d'un son qui n'en change pas sa nature. On peut toujours, en théorie, passer d'une représentation à une autre. Mais le choix de la « bonne » représentation est essentiel pour l'ensemble des tâches du traitement du signal musical –

analyse et extraction d'information, transformations (effets sonores, filtrage, etc.), compression, débruitage, réduction d'écho... Nous verrons ici que certains développements récents des mathématiques appliquées annoncent des représentations numériques dotées de deux qualités cruciales : des représentations à la fois compactes (mettant en œuvre peu de paramètres, donc propices à la compression) et intuitives (les paramètres portent de façon explicite de l'information sur le contenu sonore).

Numérisation et code PCM

Commençons par rappeler les rudiments de la numérisation d'un signal. Par rapport à un signal analogique, par exemple la pression acoustique arrivant au tympan d'un auditeur, ou le signal électrique sortant d'un microphone, un signal numérique a subi deux opérations essentielles. Tout d'abord, il a été « échantillonné » :

l'amplitude X du signal n'est pas spécifiée à tout instant, mais seulement à intervalles de temps réguliers. Plus précisément, le signal est échantillonné par les valeurs X_n que prend X aux instants nT , où n décrit les entiers successifs et où T est la période d'échantillonnage choisie. Par ailleurs, les échantillons X_n sont « quantifiés », c'est-à-dire arrondis avec une précision finie pour être représentés numériquement (généralement de façon binaire, c'est-à-dire par des séquences de 0 ou 1). En traitement du signal, cette représentation est nommée code PCM (pour *Pulse Code Modulation*, ou code à impulsions modulées) ; elle a été utilisée pour la première fois durant la Seconde Guerre mondiale par les Alliés.

Le code PCM est la représentation numérique des sons la plus simple et la plus générique, celle employée par exemple dans les disques audionumériques (CD). Dans ce cas, la fréquence d'échantillonnage (nombre d'échantillons par seconde, soit l'inverse de la période T) est de 44,1 kilohertz (soit 44 100 échantillons par seconde



les sons musicaux

de son, pour chacun des canaux droit et gauche), et la précision est de 16 bits par échantillon, ce qui autorise le codage de $2^{16} = 65\,536$ niveaux différents.

Le choix de ces paramètres d'échantillonnage et de quantification est dicté par les caractéristiques de la perception humaine : d'après la théorie classique de l'échantillonnage (théorème de Whittaker-Kotel'nikov-Shannon, 1949), la fréquence d'échantillonnage doit être au moins deux fois supérieure à la fréquence maximale utile dans le signal – qui est d'environ 20 kilohertz, la fréquence maximale audible par l'homme. De même, d'après des résultats classiques, chaque bit de quantification permet d'augmenter la dynamique de six décibels, la dynamique étant le rapport entre l'intensité du son le plus fort et celle du son le plus faible. Une quantification sur 16 bits offre donc une dynamique théorique d'environ $6 \times 16 = 96$ décibels, ce qui couvre bien la perception humaine (par convention, 0 décibel correspond au seuil absolu d'audition pour un son pur de fré-

quence égale à un kilohertz ; les sons dépassant 100 décibels peuvent provoquer des lésions auditives).

Ainsi, la représentation de type PCM est particulièrement simple et générique, ce qui fait sa popularité. Elle est cependant très gourmande en espace mémoire : il suffit d'environ six secondes de son stéréo pour occuper un mégaoctet. De plus, elle est peu informative sur le contenu du son codé et n'autorise facilement qu'un nombre limité de transformations, telles que filtrage et édition de type « copier-coller ».

Analyse de Fourier à fenêtre fixe

Or on sait depuis les débuts de l'acoustique musicale que les sons entretenus (par exemple une note tenue créée en soufflant en continu dans une flûte) peuvent être vus comme une somme de composantes sinusoïdales (voir l'encadré page 86). L'opération mathématique permettant de réaliser cette décomposition est nommée

L'ESSENTIEL

- ✓ La représentation des sons par un ensemble de signaux purement sinusoïdaux n'est pas bien adaptée à la musique.
- ✓ Des ondes élémentaires d'extension temporelle variable permettent de représenter un son musical à l'aide de relativement peu de paramètres.
- ✓ Une telle « représentation parcimonieuse » porte de l'information sur le contenu musical du son examiné.

L'AUTEUR



Laurent DAUDET, maître de conférences à l'Université Pierre et Marie Curie, à Paris, est membre de l'équipe LAM (Lutheries, acoustique, musique) de l'Institut Jean le Rond d'Alembert. Il est coordinateur adjoint du double cursus de licence « Sciences et musicologie », en partenariat avec l'Université Paris 4.

analyse ou transformation de Fourier, d'après le mathématicien français qui l'a introduite au début du XIX^e siècle.

Une représentation du son comme une simple somme de composantes sinusoïdales de différentes fréquences ne serait-elle alors pas mieux adaptée que le codage PCM ? Le problème est qu'un morceau de musique n'est pas stationnaire : les fréquences contenues dans le signal se modifient au cours du temps. Cette évolution temporelle est essentielle et, pour en rendre compte, on doit recourir à des transformations de Fourier locales, c'est-à-dire en analysant le signal intervalle de temps par intervalle de temps, la « fenêtre » temporelle d'analyse étant suffisamment petite pour que, durant ce laps de temps, le signal soit bien approximé par une somme de sinusoïdes. En examinant le signal dans sa globalité, on peut ainsi tracer des « trajectoires » des paramètres fréquence et amplitude correspondant à chaque composante sinusoïdale.

Ce type d'analyse conduit à une représentation graphique très informative sur le signal, le « spectrogramme », qui visualise la répartition de l'énergie dans le plan

temps-fréquence (voir l'encadré ci-dessous). C'est une transformation temps-fréquence de ce type (dite « transformée en cosinus discrète modifiée ») qui est à la base du succès des codeurs audio numériques à haut débit, tel le codage MP3. Un des choix critiques pour une telle représentation est celui de la durée de la fenêtre d'analyse : trop longue et le signal ne peut plus être considéré comme stationnaire, trop brève et l'on en perd le caractère sinusoïdal (voir la figure 1). Le compromis à faire dépend du type de son analysé et de la tâche visée.

Analyse de Fourier à fenêtre variable

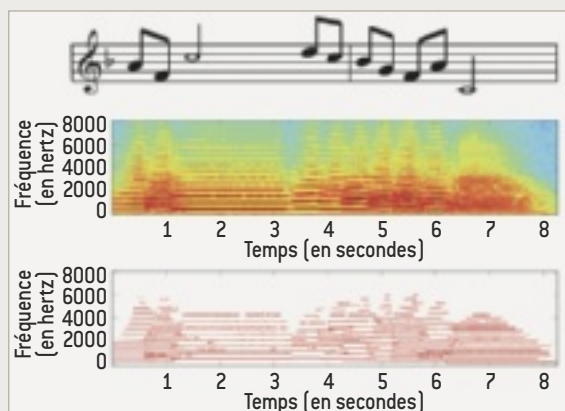
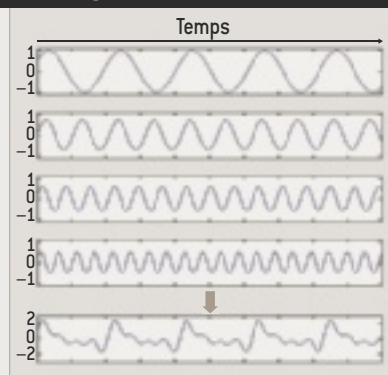
Depuis le milieu des années 1990, les scientifiques ont développé des algorithmes susceptibles de résoudre le dilemme d'une taille de fenêtre d'analyse fixée : il s'agit des méthodes de décomposition parcimonieuse des signaux. Comme l'analyse de Fourier, elles sont fondées sur l'idée intuitive « atomistique » d'une décomposition des signaux en combinaison linéaire (une somme pondérée par des coefficients) de formes d'ondes élémentaires,

LES REPRÉSENTATIONS SINUSOÏDALES

La plupart des instruments de musique sont des structures vibrantes dont la réponse à une excitation peut se représenter par une somme de sinusoïdes de fréquence et d'amplitude lentement variables (c'est-à-dire qui varient sur des échelles de temps bien plus grandes que les périodes caractéristiques des vibrations). Pour une excitation ponctuelle, la réponse de tels systèmes (instruments à percussion par exemple) est un ensemble de sons sinusoïdaux dont l'amplitude

décroit exponentiellement avec le temps. Si l'excitation est entretenue (souffle du flûtiste, archer se déplaçant le long de la corde, etc.), la réponse est stationnaire en amplitude et harmonique : les fréquences des sinusoïdes sont des multiples entiers d'une fréquence unique nommée fréquence fondamentale.

Ainsi, dans le schéma ci-contre, l'onde parfaitement périodique tracée en bas est obtenue par addition des quatre sinusoïdes tracées au-dessus.



On peut généraliser ce principe pour les sons non strictement périodiques, en considérant le signal comme une somme de sinusoïdes dont la fréquence et l'amplitude varient lentement dans le temps. C'est le principe de la « synthèse additive », qui a des applications importantes pour l'analyse et la synthèse des sons. Dans ce cadre, le signal $X(t)$ s'écrit sous la forme : $X(t) = A_1 \sin(\omega_1 t + \varphi_1) + \dots + A_N \sin(\omega_N t + \varphi_N)$, où N est le nombre de composantes sinusoïdales du signal et où les paramètres A_i , ω_i et φ_i varient sur des échelles de temps beaucoup plus longues que les périodes d'oscillation des sinusoïdes. On peut représenter ces variations par des trajectoires tracées dans le plan temps-fréquence. Un exemple est donné ci-contre, avec un solo de trompette. Le spectrogramme du morceau joué indique la composition en fréquences en fonction du temps, ainsi que les intensités correspondantes (croissantes du bleu au rouge). Le graphique du bas montre les trajectoires décrites par les fréquences.

dites « atomes ». La collection d'atomes de base est nommée « dictionnaire ».

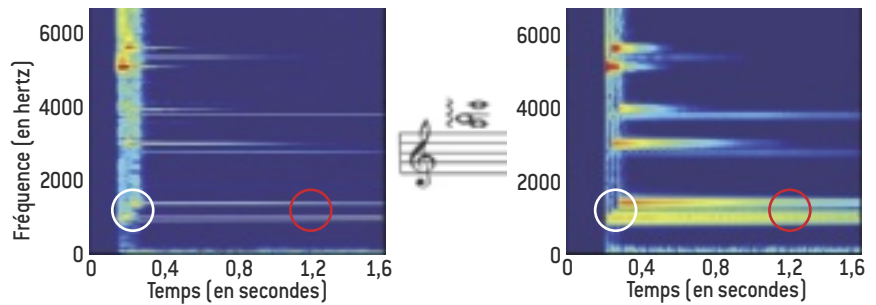
Dans le cas des sons, il peut s'agir par exemple de signaux oscillants dont l'enveloppe est une fenêtre localisée en temps. Cela signifie que l'amplitude du signal oscillant s'annule rapidement de part et d'autre d'un instant central (voir la figure 4). De tels signaux relèvent d'une représentation de type Fourier (analyse fréquentielle) tout en gardant une notion de localité temporelle. La fenêtre d'analyse est maintenant de taille (« d'échelle ») variable, et non fixée comme dans le cas de l'analyse de Fourier à court terme. On est donc passé d'une analyse temps-fréquence à une analyse temps-échelle-fréquence.

Dans de telles décompositions, on peut espérer représenter efficacement les parties sinusoïdales des sons avec des atomes à grandes fenêtres temporelles (et donc avec une bonne discrimination des fréquences), et les attaques des notes avec des atomes à petites fenêtres temporelles, bien adaptés aux signaux impulsionnels tels que les sons de percussions (voir la figure 2). Ces approches généralisent aussi des décompositions temps-échelle de type ondelettes (où l'échelle est là aussi variable, mais dépend de façon rigide de la fréquence).

La parcimonie : dire beaucoup avec peu

Une difficulté surgit alors : comment s'assurer que l'algorithme fera ce qui est attendu ? Il est en effet possible de représenter des parties tonales par des atomes à petite fenêtre, et des transitoires à grande fenêtre... mais avec un très grand nombre d'atomes ! On se donne alors une contrainte supplémentaire de parcimonie : on favorise les décompositions qui non seulement constituent une bonne approximation du signal, mais contiennent aussi le moins d'atomes possible – en d'autres termes, il s'agit de décrire précisément en employant peu de mots. Comme dans le langage usuel, plus le dictionnaire disponible est volumineux, moins nous aurons à utiliser de périphrases pour décrire un objet complexe...

Pour un son donné et un dictionnaire fixé de taille M , le problème consiste donc à trouver la décomposition la plus parcimonieuse, celle qui s'approche le mieux du signal en utilisant au plus N atomes. Ce problème n'est en général pas résoluble d'une manière exacte, car cela reviendrait à tester tous les groupes

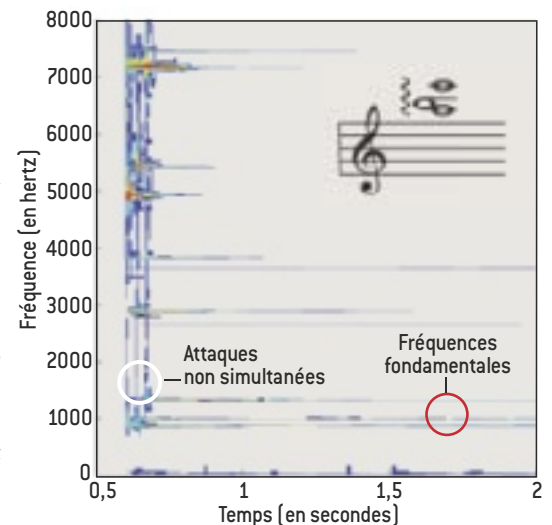


1. CES DEUX SPECTROGRAMMES d'un accord arpégé de trois notes [c'est-à-dire que les notes sont jouées avec un léger décalage dans le temps] illustrent la difficulté du choix de taille de la fenêtre temporelle d'analyse. Le spectrogramme de gauche a été réalisé avec une grande fenêtre d'analyse ; on distingue bien les fréquences fondamentales des trois notes (cercle rouge), mais, lors de l'attaque, on ne voit pas que l'accord est arpégé (cercle blanc). Le spectrogramme de droite a été réalisé avec une petite fenêtre d'analyse ; on distingue bien les événements du début (cercle blanc), mais les fréquences fondamentales des deux premières notes ne sont plus résolues (cercle rouge).

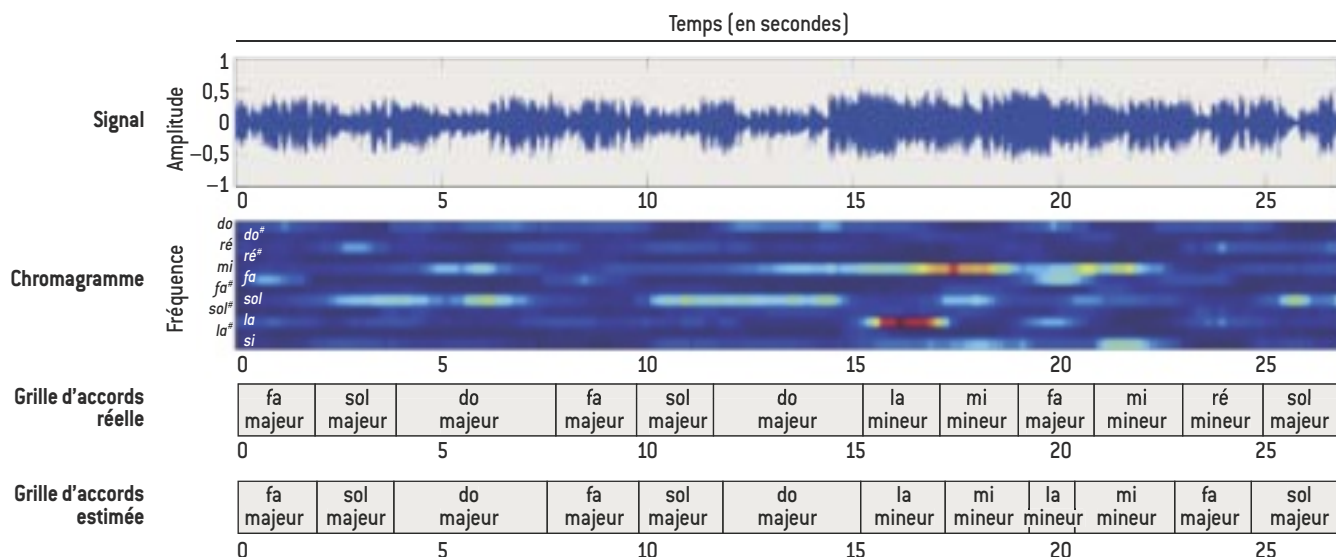
de N atomes piochés parmi M ... Pour les valeurs utiles de N et M , le nombre de combinaisons à tester serait très au-delà des capacités de n'importe quel ordinateur. Cependant, il existe depuis quelques années une grande diversité de méthodes utilisables en pratique pour résoudre ce problème de façon approchée. Pour sa simplicité, nous avons choisi d'utiliser l'algorithme dit de *Matching Pursuit*, qui sélectionne les atomes un à un : on sélectionne l'atome du dictionnaire qui présente la plus grande corrélation avec le signal, puis on soustrait sa contribution au signal et on répète la procédure jusqu'à ce que l'on ait sélectionné N atomes.

Une application naturelle du principe de parcimonie est la compression audio-numérique : si le dictionnaire est fixé à l'avance, on peut décrire un son d'une façon compacte juste en spécifiant les paramètres temps-fréquence-échelle des atomes de la décomposition parcimonieuse du son, ainsi que leurs poids respectifs. Comme nous l'avons vu, plus le dictionnaire est important, moins il faut d'atomes pour représenter les sons. Mais la transmission des paramètres temps-fréquence-échelle des atomes appartenant à un très grand dictionnaire requiert une grande quantité d'information. Le codage du paramètre variable « échelle » peut se révéler très coûteux par rapport aux approches à échelle fixée, moins parcimonieuses. De plus, utiliser un énorme dictionnaire est d'un coût calculatoire qui devient rapidement rédhibitoire. Ici encore, tout va être une question de compromis !

Dans l'expérience menée avec Emmanuel Ravelli, en collaboration avec Gaël Richard de *Telecom Paris Tech*, nous avons considéré un dictionnaire formé d'atomes

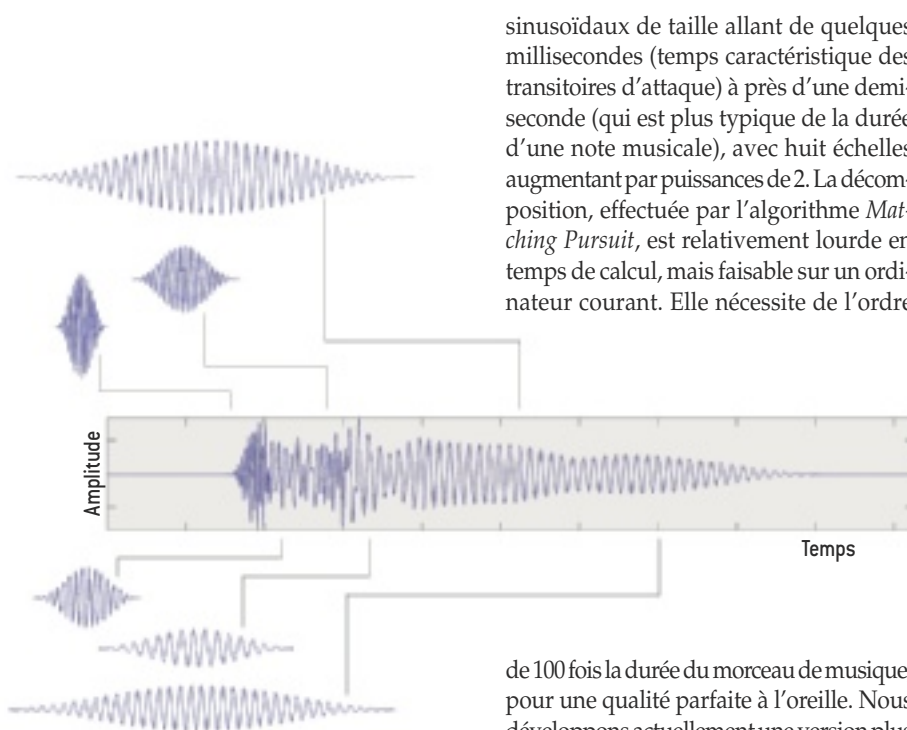


2. LE MÊME ACCORD de carillon que sur la figure 1 est représenté ici par une décomposition parcimonieuse. Chaque « atome » de la décomposition correspond sur le graphique à un rectangle de largeur proportionnelle à sa durée [son échelle], et de hauteur proportionnelle à sa largeur fréquentielle. Avec cette décomposition, on distingue à la fois la fréquence fondamentale des trois notes et les trois attaques non simultanées.



3. UNE REPRÉSENTATION PARCIMONIEUSE des sons musicaux permet de reconnaître assez bien la grille d'accords présente dans un morceau de musique, ici un extrait d'une chanson des *Beatles*. Le « chromagramme » construit à partir d'une représentation parcimonieuse temps-fréquence-échelle représente en fonction du temps le contenu en

fréquences correspondant aux demi-tons de la gamme chromatique (*intensités croissantes en allant du bleu au rouge*). On constate que la grille d'accords déduite du chromagramme correspond assez bien à la grille d'accords réelle (deux erreurs seulement) : sur ce type de musique, l'algorithme identifie correctement l'accord plus de 70 pour cent du temps.



4. DANS UNE DÉCOMPOSITION parcimonieuse, un signal complexe et fortement non stationnaire est approché par une somme pondérée d'un nombre assez faible d'« atomes » de différentes durées, fréquences et positions temporelles. En général, les atomes de grande taille (ou durée) correspondent aux parties les plus stationnaires du signal, tandis que les atomes de petite taille correspondent aux transitoires d'attaque. Si l'on représentait un tel signal complexe par de pures sinusoïdes, il en faudrait un très grand nombre.

sinusoïdaux de taille allant de quelques millisecondes (temps caractéristique des transitoires d'attaque) à près d'une demi-seconde (qui est plus typique de la durée d'une note musicale), avec huit échelles augmentant par puissances de 2. La décomposition, effectuée par l'algorithme *Matching Pursuit*, est relativement lourde en temps de calcul, mais faisable sur un ordinateur courant. Elle nécessite de l'ordre

de 100 fois la durée du morceau de musique, pour une qualité parfaite à l'oreille. Nous développons actuellement une version plus efficace capable d'obtenir des décompositions presque en temps réel.

Un avantage de cette classe de décompositions des sons est que les atomes obtenus sont classés par ordre d'importance décroissante. On obtient ainsi une qualité dite progressive : le début du fichier compressé contient les informations grossières, et la qualité s'améliore à mesure que l'on parcourt le fichier informatique – un peu comme la consultation de certaines images sur *Internet*, où l'image apparaît d'abord

avec une définition très grossière, puis avec des détails de plus en plus fins.

Les résultats des tests auditifs démontrent la pertinence de notre approche à très bas débit, au-dessous de 50 kilobits par seconde : pour de nombreux sons, la qualité auditive est supérieure à celle obtenue avec l'approche à simple résolution des codeurs actuels (ou, de manière équivalente, nous obtenons une qualité similaire avec une taille de mémoire notablement inférieure, le gain de taille pouvant aller jusqu'à plus de 50 pour cent). À ces débits, où l'on ne peut pas en général espérer de restitution perceptivement parfaite (c'est-à-dire identique, à l'oreille, à l'original non compressé), un faible nombre d'éléments à grande échelle suffisent à donner une fidélité raisonnable.

En revanche, dans la gamme des hauts débits nécessaires à un son de très bonne qualité (typiquement 128 kilobits par seconde), le surcoût dû au codage du paramètre d'échelle fait que ces représentations du son n'apportent aucun avantage par rapport aux méthodes classiques. Notre codeur prototype est toutefois loin d'avoir épuisé tout son potentiel, et nous espérons des progrès notables soit en optimisant le choix du dictionnaire, soit en améliorant les algorithmes de décomposition parcimonieuse.

Un domaine de recherche en plein essor est celui de l'indexation des bases de données musicales. Pour naviguer dans une énorme base de données de ce type, on peut en effet utiliser d'autres paramètres

que les informations textuelles classiques sur un fichier sonore (nom de l'album, de l'artiste, de la piste, etc.): navigation par sonorité (type d'instruments), par rythme (souvent très lié au genre musical), génération automatique de listes de lecture, etc.

Indexer les sons compressés

De nombreux algorithmes sont apparus récemment pour extraire automatiquement ce type d'informations avec une précision souvent étonnante: on arrive à des services du genre: « Vous avez écouté ce morceau, vous aimerez ces suggestions [dont la sonorité est semblable]. » Ces algorithmes effectuent des analyses parfois d'une grande complexité, qui combinent des critères temporels pour l'analyse du rythme et des critères fréquentiels pour l'analyse du timbre et de l'harmonie.

L'extraction de ces informations pour une très grande base de données est souvent très coûteuse en puissance de calcul. D'où notre idée: si l'on dispose d'une base de données musicales sous forme

compressée à l'aide de notre algorithme parcimonieux, pourquoi décompresser les sons pour effectuer ces analyses? Ne pourrait-on pas exploiter le fait que les représentations parcimonieuses fournissent directement un jeu compact de paramètres pour extraire cette information? Nous avons récemment montré que c'est en effet possible. Ainsi, à très faible coût calculatoire, nous pouvons obtenir les informations de rythme (reconnaissance de tempo) et d'harmonie (reconnaissance d'accords, voir la figure 3) avec une précision comparable à celle des algorithmes dédiés. Ce type d'études ouvre l'ère de la « compression intelligente » des sons.

Les représentations numériques des sons musicaux devront dans l'avenir être capables non seulement de stocker et transmettre les sons de manière efficace, mais aussi de rendre possibles l'analyse et la manipulation de leurs structures constitutives. À cet égard, les représentations parcimonieuses offrent une approche crédible, car leur complexité calculatoire, bien qu'élevée, est maintenant à la portée de la plupart des ordinateurs. ■

✓ BIBLIOGRAPHIE

E. Ravelli, G. Richard et L. Daudet, *Union of MDCT bases for audio coding*, IEEE Transactions on Audio, Speech and Language Processing, à paraître, 2008.

S. Mallat, *Une exploration des signaux en ondelettes*, Presses de l'École Polytechnique, 2000.

P. Flandrin, *Temps-fréquence*, Hermès [2^e édition], 1998.

✓ SUR LE WEB

Site du réseau « Music Information Retrieval » : <http://www.ismir.net/>

Articles récents de L. Daudet : <http://old.lam.jussieu.fr/src/Membres/Daudet/>

Une offre de formation en

ACOUSTIQUE

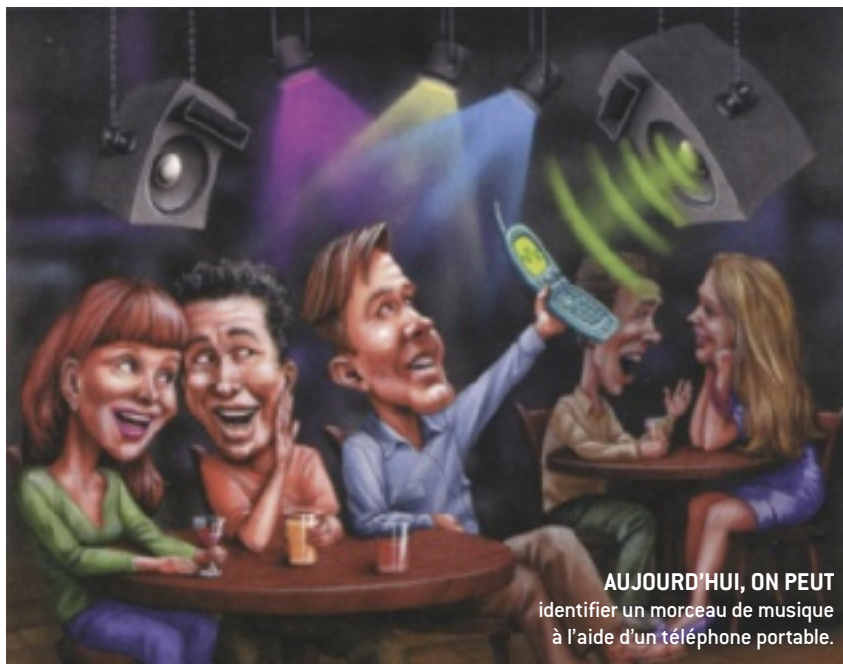
unique en Europe

Université du Maine

<http://ensim.univ-lemans.fr/>
<http://laum.univ-lemans.fr/spip/>
<http://www.univ-lemans.fr/sciences/>

Retrouver un texte dans une base de données en saisissant quelques-uns de ses mots est devenu une tâche banale. La recherche automatique de morceaux musicaux prend un chemin analogue.

Geoffroy Peeters



John Fraser

Indexer la musique

Le stockage de la musique, sa diffusion et sa consultation sont à l'ère numérique – une évolution rendue possible par le développement de techniques de compression adaptées (MP3, WMA, AAC ou OGG), qui ont réduit la taille d'un morceau à seulement quelques mégaoctets, et de réseaux rapides permettant de consulter en quelques secondes des catalogues musicaux immenses. En 2008, chacun peut ainsi se déplacer avec plus de 40 000 titres musicaux dans le creux de sa main et accéder en ligne à plusieurs millions d'autres. Mais comment trouver, dans ces musicothèques de plus en plus gigantesques, les morceaux désirés ? Tel est le défi auquel l'indexation automatique de la musique doit répondre.

Grâce à la numérisation massive des textes et aux moteurs de recherche, on peut depuis plusieurs années retrouver rapidement un texte, en utilisant le nom de son auteur et son titre, mais aussi en utilisant des éléments – mots ou phrases – de son contenu. De même, la prolifération actuelle des documents musicaux nécessite des moteurs de recherche adaptés à la musique,

donc des techniques d'indexation musicale. On peut bien sûr chercher un morceau par le nom de l'artiste, son titre ou le nom de l'album. Mais on souhaite aussi retrouver un morceau à l'aide d'éléments de son contenu, un exemple de requête étant : « Je recherche un morceau rock de tempo rapide avec un solo de saxophone. » Ces caractéristiques étant contenues dans le morceau de musique, on parle d'indexation par le contenu.

Depuis le milieu des années 1990, de nombreux centres de recherche – dont, en France, l'IRCAM, *Telecom Paris Tech*, l'INRIA et *Sony CSL* – se sont intéressés à l'indexation automatique de la musique. Les travaux ont d'abord porté sur l'indexation d'échantillons audio numériques, courts extraits renfermant un élément musical unique (un son de trompette, un tintement de verre, etc.) et utilisés pour la composition musicale ou la création de bandes-son. Dès 1998, l'application *Studio OnLine* de l'IRCAM permettait d'accéder en ligne à un catalogue de 130 000 sons. Le contenu sonore de chaque échantillon ayant été préalablement analysé, on pouvait effec-

tuer une recherche par similarité acoustique : « J'aime ce son, donnez-moi des sons qui sont similaires acoustiquement. »

Cependant, l'indexation automatique de la musique pose de nouveaux problèmes. Les descriptions recherchées sont différentes de celles des échantillons : alors qu'on décrit un échantillon par sa hauteur, son timbre ou sa cause (un son de flûte, de porte de voiture, etc.), un morceau de musique se caractérise par sa mélodie principale, ses accords, son tempo, les qualités vocales du chanteur. Les techniques d'analyse et de traitement du signal utilisées sont également plus complexes, car, dans un morceau de musique, de multiples sons se superposent. La première application publique d'indexation automatique a vu le jour en 2003, avec le *MusicBrowser* de *Sony CSL* : on pouvait désormais effectuer une recherche de morceaux musicaux par similarité acoustique.

Comment réaliser une telle indexation automatique de contenu ? Dans une première étape, on extrait du fichier musical des caractéristiques nommées descripteurs audio. On les choisit afin de mettre

en évidence certaines propriétés du signal liées à la perception du son (le son est-il brillant, terne, aigu, grave, rugueux, lisse?), à des concepts musicaux (hauteur des notes jouées, accords, tempo, métrique) ou aux sources présentes dans le signal (timbre du chanteur ou des instruments joués). Cette extraction repose sur les techniques de traitement du signal – en particulier la transformation de Fourier, opération mathématique qui décompose un signal en éléments purement sinusoidaux. On met ainsi en évidence les fréquences représentées dans le signal à chaque instant, ce qui permet d'extraire par exemple la hauteur des notes ou le timbre des instruments. Et avec la variation du contenu fréquentiel au cours du temps, on peut étudier la vitesse d'apparition des différentes notes ou percussions et en déduire le tempo du morceau.

Des descripteurs audio et des modèles statistiques

Après la détermination des descripteurs audio, la seconde phase de l'indexation automatique consiste à déterminer les valeurs prises par les descripteurs audio pour chaque entrée de l'index. Pour ce faire, on utilise des modèles statistiques. Par exemple, si l'on cherche à reconnaître automatiquement le genre musical dont les entrées d'index sont « rock », « jazz », « classique », on s'intéresse aux hauteurs, rythmes et instruments présents dans chacun de ces genres. La modélisation statistique indiquera alors par exemple que si le timbre d'une guitare électrique est observé dans un signal, ce morceau a une probabilité de 80 pour cent d'être du rock, 20 pour cent du jazz et zéro pour cent de la musique classique. Le même modèle statistique pourra ensuite indexer automatiquement des morceaux nouveaux.

Les portails de musique destinés au grand public intègrent peu à peu l'indexation automatique. Leurs gestionnaires cherchent aussi à présenter aux usagers la partie rarement consultée de leur catalogue, afin de diversifier et accroître les ventes. Des systèmes de recommandation sont utilisés pour cela. Initialement, ils reposaient sur les statistiques de vente (cas de *Amazon*, notamment) : « Si vous avez acheté ceci, sachant que les clients qui ont acheté ceci ont aussi acheté cela, nous

vous recommandons cela. » Les nouveaux systèmes de recommandation reposent sur les statistiques de métadonnées associées aux morceaux de musique (*iTunes Genius*, *Pandora*, *LastFM*) : « Si vous écoutez ce morceau-ci, sachant qu'il a tel type de chant, de tempo et d'instrumentation, nous vous recommandons ces morceaux-là, qui ont les mêmes caractéristiques. » Les métadonnées peuvent être saisies manuellement par du personnel spécialisé (*iTunes Genius*, *Pandora*), collectées auprès des utilisateurs (*LastFM*) ou extraites automatiquement du signal audionumérique (*FM4 SoundPark*). La taille des catalogues musicaux étant en croissance rapide, on privilégie de plus en plus la dernière solution.

Lorsque la description à laquelle on s'intéresse est absente du contenu du morceau – par exemple lorsqu'on cherche le nom de l'artiste, le titre du morceau ou celui de l'album –, on utilise une autre technique : l'identification audio. Elle consiste à prendre une empreinte numérique du morceau, fondée sur ses caractéristiques acoustiques. On recherche ensuite cette empreinte dans une base de données contenant des fiches d'identité pour chaque morceau. Une fois l'empreinte trouvée, la base de données fournit les descriptions manquantes correspondant à la fiche du morceau trouvé. Ainsi, avec la plupart des téléphones portables de dernière génération, il est possible d'enregistrer de la musique diffusée dans la rue ou dans un café et d'obtenir immédiatement les informations relatives au morceau entendu.

D'autres innovations voient le jour. L'une d'elles est l'identification de morceaux dont on chantonne la mélodie. Par ailleurs, dans le domaine de la consultation de musicothèques, on développe des algorithmes qui découpent automatiquement un morceau en chapitres (introduction, couplets et refrains), pour fournir à l'utilisateur une visualisation du contenu temporel du morceau, à la façon du chapitre d'un disque vidéonumérique. D'autres algorithmes sont capables de créer automatiquement des « résumés audio » ou « bandes-annonces musicales » grâce auxquels l'utilisateur peut écouter rapidement le contenu d'une musicothèque. Les performances de l'indexation automatique de la musique et les possibilités qu'elle offre ne cessent de croître, même s'il reste bien difficile d'extraire automatiquement la partition musicale correspondant à un morceau... ■

L'AUTEUR



Geoffroy PEETERS est chargé de recherche à l'IRCAM, à Paris. Il est responsable de l'indexation musicale au sein du programme français *Quaero* de recherche et développement sur les documents multimédias.

✓ BIBLIOGRAPHIE

EURASIP Journal on Advances in Signal Processing, numéro spécial *Music information retrieval based on signal processing* (sous la dir. de Ichiro Fujinaga et al.), vol. 2007, 2007.

G. Peeters, *A generic system for audio indexing : application to speech/music segmentation and music genre recognition*, Proceedings of the 10th Int. Conference on Digital Audio Effects, Bordeaux, 10-15 septembre 2007 [article téléchargeable sur le Web].

✓ SUR LE WEB

Communauté *Music Information Retrieval* : <http://www.music-ir.org/>

Publications de G. Peeters : <http://recherche.ircam.fr/equipes/analyse-synthese/peeters/ARTICLES/index.html>

La musique mise en algèbre

Moreno Andreatta et Carlos Agon

La musique du XX^e siècle se comprend mieux avec les mathématiques et notamment avec la théorie des groupes. Ce sont des concepts abstraits, Pourtant, ces structures seraient inhérentes à notre cerveau.

L'ESSENTIEL

✓ La musique du XX^e siècle, notamment la musique sérielle, obéit à des règles mathématiques dont la connaissance facilite notablement l'analyse des compositions.

✓ Au cœur de ces règles mathématiques, on trouve le concept de groupes et diverses transformations, notamment des symétries.

✓ Selon certains, cette structure mathématique serait naturelle et préexisterait dans notre cerveau, ce qui ouvrirait de nouvelles perspectives pour étudier la perception musicale.

En 1739, le mathématicien suisse Leonhard Euler (1707-1783) publie son *Tentamen novae theoriae musicae ex certissimis harmoniae principiis dilucide expositae* (Essai d'une nouvelle théorie de la musique, exposée en toute clarté selon les principes de l'harmonie les mieux fondés) dans lequel il souhaite expliquer pourquoi la musique apporte du plaisir. Selon lui, l'élément clef est la perfection, qu'il recherche dans les rapports de nombres représentant les accords.

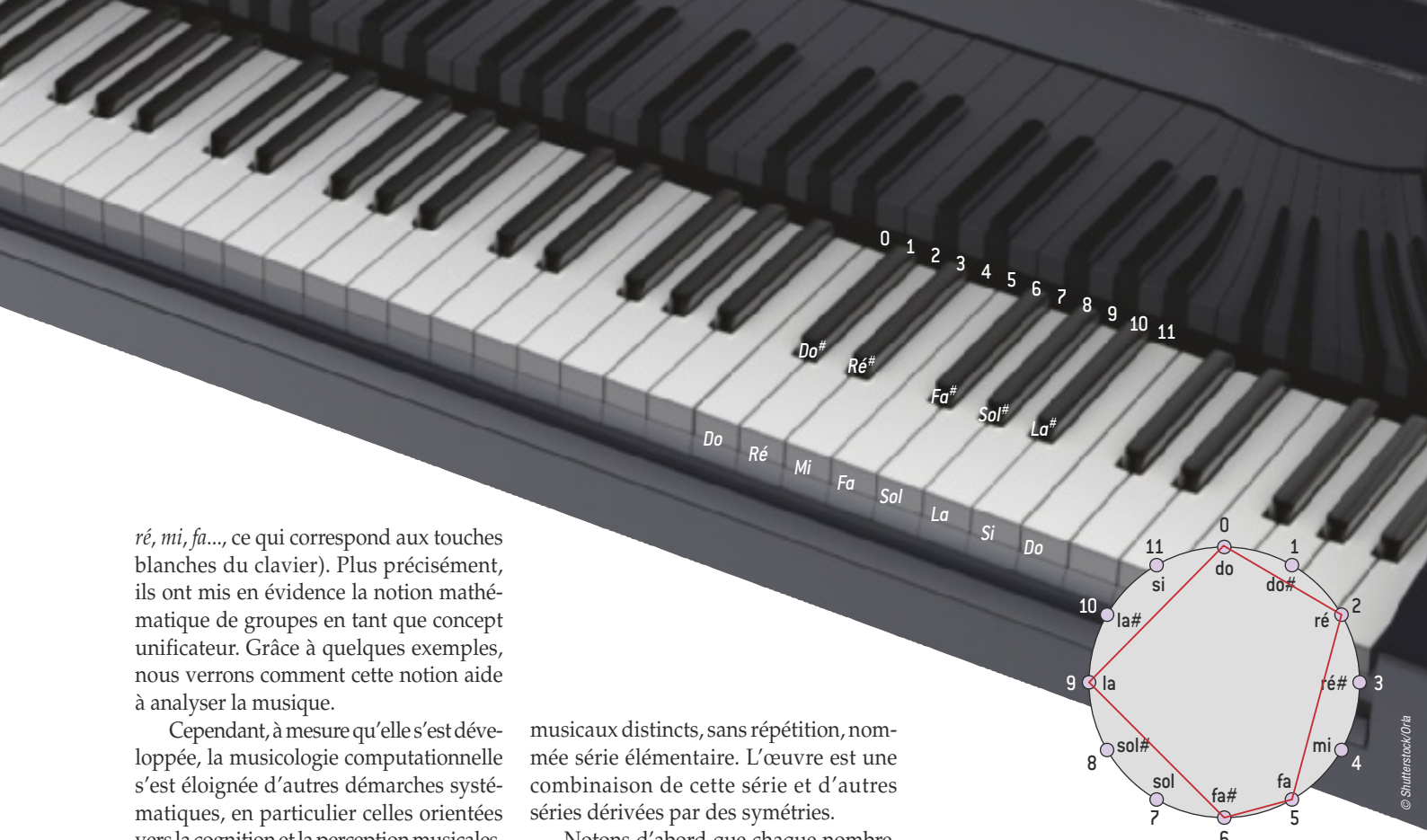
Euler est l'un des nombreux mathématiciens qui se sont intéressés à la musique depuis Pythagore jusqu'à nos jours. Depuis une dizaine d'années, l'étude des relations entre mathématiques et musique a connu de nombreux développements et la communauté des mathématiciens et informaticiens et celle des musicologues et musiciens y portent un intérêt croissant.

Ce nouveau champ de recherche, dont l'informatique a été l'un des catalyseurs, a accompagné et parfois accéléré la transformation de la musicologie en une discipline systématique, jusqu'à donner naissance à un nouveau champ d'études, la « musicologie computationnelle ». Il s'agit d'analyser les œuvres musicales de façon

à mettre en évidence les structures mathématiques sous-jacentes. Nous nous intéresserons ici plus particulièrement à la formalisation algébrique de la musique du XX^e siècle. L'utilisation des méthodes algébriques en musique met en œuvre trois aspects souvent liés, à savoir les aspects théoriques et analytiques, qui nous intéresseront ici, ainsi que ceux d'aide à la composition, que nous ne développerons pas.

La « trinité » algébrique

Trois compositeurs et théoriciens sont emblématiques de cette réflexion théorique sur la musique : l'Américain Milton Babbitt aux États-Unis, le Grec Iannis Xenakis (1922-2001) en Europe occidentale et le Roumain Anatol Vieru (1926-1998) en Europe de l'Est. Tous les trois sont arrivés, presque au même moment et d'une façon indépendante, à la découverte du caractère algébrique du tempérament égal, c'est-à-dire que dans une gamme, chaque note est séparée de sa voisine par un demi-ton (*do, do#, ré, ré#, mi, fa...*, soit une gamme à 12 demi-tons, ce qui diffère notablement de la gamme dite diatonique à 7 tons *do,*



ré, mi, fa..., ce qui correspond aux touches blanches du clavier). Plus précisément, ils ont mis en évidence la notion mathématique de groupes en tant que concept unificateur. Grâce à quelques exemples, nous verrons comment cette notion aide à analyser la musique.

Cependant, à mesure qu'elle s'est développée, la musicologie computationnelle s'est éloignée d'autres démarches systématiques, en particulier celles orientées vers la cognition et la perception musicales. Les deux visions sont-elles compatibles? Oui et, plus encore, elles s'enrichissent mutuellement. Nous montrerons qu'un dialogue est possible entre mathématiques et cognition, et que les premières peuvent éclairer la seconde.

De groupe en groupe

La notion de groupe est née au début du XIX^e siècle des travaux sur les racines de polynômes, notamment ceux d'Évariste Galois et de Joseph-Louis Lagrange. Cependant, cette structure ne fut utilisée en musique qu'à partir de la seconde moitié du XX^e siècle grâce au compositeur américain M. Babbitt. On lui doit l'observation fondamentale selon laquelle le système dodécaphonique est « un groupe de permutations qui est façonné par la structure de ce modèle mathématique ». En quoi consiste la musique dodécaphonique?

Elle se distingue de la musique tonale où une des sept notes de la gamme diatonique prédomine sur les autres et leur impose une hiérarchie. En 1923, Arnold Schoenberg (1874-1951) veut échapper à ce diktat et établit la méthode de composition avec 12 sons (d'où le nom dodécaphonique), qui donne le même mérite à chaque note de la gamme tempérée ou à tempérament égal.

Une composition dodécaphonique est fondée sur une séquence de ces 12 sons

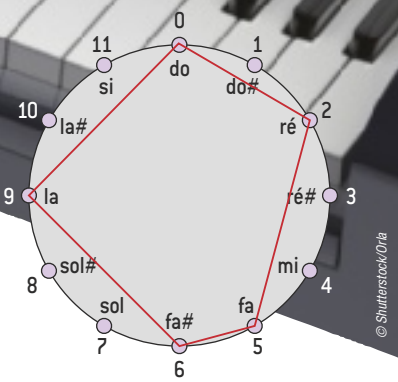
musicaux distincts, sans répétition, nommée série élémentaire. L'œuvre est une combinaison de cette série et d'autres séries dérivées par des symétries.

Notons d'abord que chaque nombre, représentant une note, est une classe d'équivalence modulo 12, c'est-à-dire que chaque nombre représente, d'une part, ce nombre, mais aussi ce nombre additionné d'un multiple de 12. Par exemple, 1 est équivalent à 49 ou à -11. L'addition de deux nombres devient une addition modulo 12 : par exemple, $(3 + 8)_{12} = 11$, $(5 + 9)_{12} = 2$.

Examinons maintenant les symétries, à partir d'une série élémentaire P. La série rétrograde R est P jouée à l'envers. Dans la série renversée I, les nombres de la série élémentaire P sont remplacés par leurs nombres opposés modulo 12. La série renversée rétrograde RI est obtenue de P en appliquant les deux opérations précédentes. Enfin, la série transposée de P par k demi-tons est obtenue par l'addition modulo 12 de k à tous les nombres de la série P. De même, nous obtenons les transposées par k demi-tons d'une série rétrograde $T_k R$, d'une série renversée $T_k I$ et d'une série renversée rétrograde $T_k RI$.

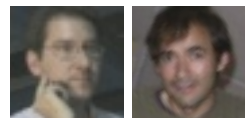
La musique sérielle est une extension du dodécaphonisme où l'idée de série est appliquée aux notes, mais aussi aux rythmes, aux intensités et à tous les paramètres du son.

L'ensemble des entiers modulo 12 offre un premier exemple musical de structure de groupe, celui noté $\mathbb{Z}/12\mathbb{Z}$. Il peut être interprété musicalement de plusieurs façons : soit comme le groupe des intervalles musicaux (avec l'addition modulo 12),



1. LA GAMME TEMPÉRÉE, c'est-à-dire celle où toutes les notes sont séparées par un demi-ton, correspond à l'octave d'un piano. Elle est représentée par un cercle où chaque note est numérotée (le do est noté 0, le do dièse (do#) ou le ré bémol (ré^b) est noté 1, et ainsi de suite). Par convention et afin de faciliter la lecture des exemples, on adopte dans cet article la notation en dièses (#). Dans cette représentation, un ensemble de n notes, par exemple un accord, correspond à un polygone à n côtés (en rouge). Dans tout l'article, on représente une série de notes dans ce cercle (leur position est notée entre parenthèses) et on s'intéresse aux polygones correspondants, ainsi qu'aux intervalles (en demi-tons) séparant les notes (dans l'exemple ci-dessus, 2 demi-tons entre do et ré, 3 demi-tons entre ré et fa, puis 1, 3 et 3)

LES AUTEURS

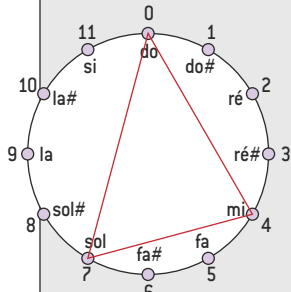


Moreno ANDREATTA est chercheur CNRS en musicologie computationnelle et Carlos AGON est chercheur en informatique dans l'équipe Représentations Musicales de l'IRCAM, à Paris.

LA STRUCTURE INTERVALLIQUE

La structure intervallique est constituée de la suite d'intervalles (en demi-tons) séparant les notes consécutives d'un ensemble. Dans la représentation circulaire, elle correspond aux intervalles entre les sommets du polygone à m côtés correspondant à l'ensemble de notes. Par exemple, la structure intervallique de l'accord majeur (constitué d'une fondamentale, d'une tierce majeure

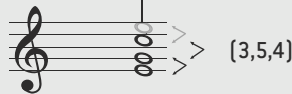
et d'une quinte, tels *do*, *mi* et *sol*) est égale à (4, 3, 5). En effet, quatre demi-tons séparent *do* et *mi*, trois séparent *mi* et *sol* et cinq séparent *sol* et le *do* supérieur. Les permutations circulaires de cette structure intervallique – (3, 5, 4) et (5, 4, 3) – correspondent aux renversements de l'accord. La représentation circulaire est identique dans les trois cas.



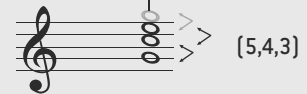
Do à l'octave supérieure



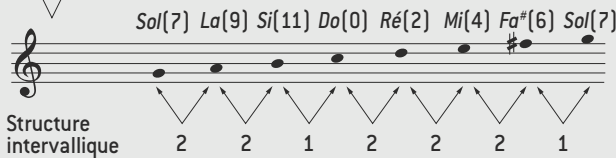
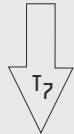
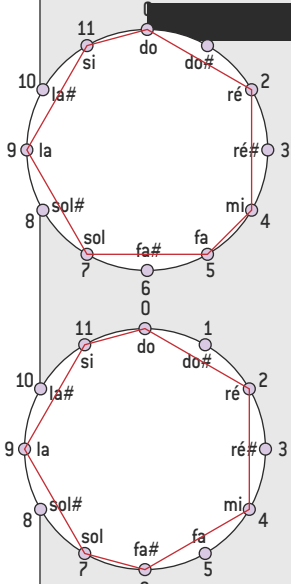
Mi à l'octave supérieure



Sol à l'octave supérieure

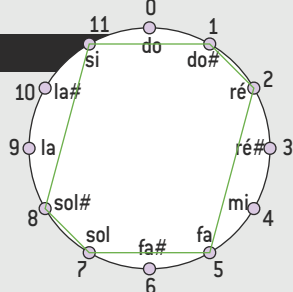
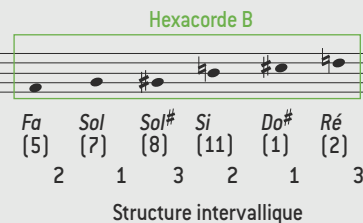
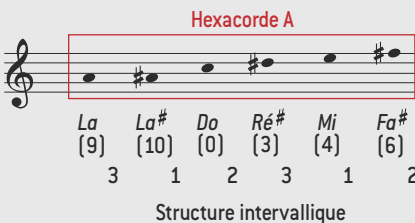
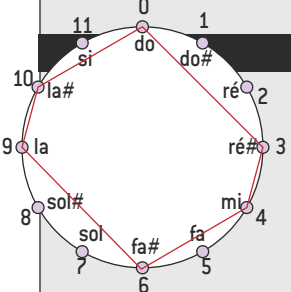


LA TRANSPOSITION



La transposition est une addition *modulo 12*. Elle laisse inchangée la structure intervallique d'une gamme ou d'un accord. Par exemple, imaginons une transposition de sept demi-tons (à chaque note, on ajoute 7 *modulo 12*) de la gamme diatonique : ainsi, de *do* (0) on passe à *sol* (7) et de *la* (9) on passe à *mi* (4), car $(9+7)_{12}=4$. Dans la représentation circulaire, cette transposition correspond à une rotation du polygone correspondant à la gamme diatonique de $(360/12) \times 7 = 210^\circ$. En d'autres termes, la structure intervallique est un invariant qui permet d'identifier de façon unique un accord et ses transpositions d'un nombre donné de demi-tons, celles-ci étant des rotations du polygone inscrit dans le cercle : la structure intervallique est préservée.

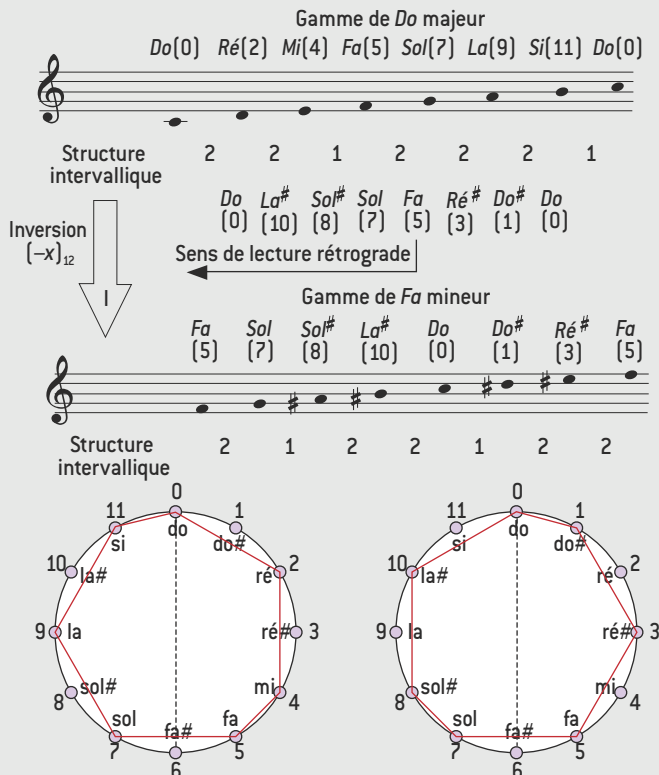
L'INVARIANCE TRANSPOSITIONNELLE



La structure intervallique met en évidence la symétrie interne de certaines séries de notes, tel un accord ou une gamme. Par exemple, des accords coïncident avec une, voire plusieurs de leurs transpositions. Une telle structure se nomme, d'après le compositeur français Olivier Messiaen (1908-1992), un « mode à transpositions limitées ». Tout mode de ce type est donné par une structure intervallique ayant des périodicités internes, c'est-à-dire des sous-structures intervalliques qui se répètent. On trouve de telles structures par exemple lorsqu'on partage une série dodécaphonique de Schoenberg dans le cinquième mou-

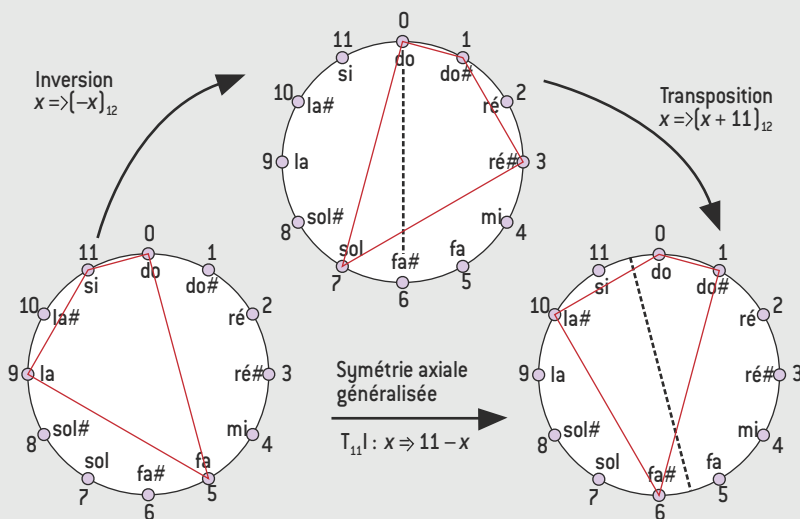
vement de la *Serenade op. 24* en deux hexacordes (des séries de six notes), notés A et B. Ces derniers sont dotés d'une propriété d'invariance transpositionnelle : dans les deux cas, la structure intervallique est « redondante », c'est-à-dire qu'elle se décompose en deux structures intervalliques plus petites identiques, respectivement (3, 1, 2) et (2, 1, 3). Notons que les deux hexacordes ne sont pas équivalents à une transposition près, car les structures intervalliques correspondantes ne sont pas les rotations circulaires l'une de l'autre. En fait, elles sont chacune la lecture rétrograde de l'autre.

L'INVERSION



L'opération d'inversion est une opération sur une gamme (ou un accord ou une série dodécaphonique) qui associe à chaque nombre x l'opposé $(-x)$ modulo 12, soit $(-x)_{12}$. Géométriquement, cela se traduit par une symétrie de la structure intervallique associée par rapport à un axe. L'exemple le plus simple d'inversion est celui d'une symétrie par rapport au diamètre principal (l'axe qui passe par les points $do = 0$ et $fa\# = 6$). Ainsi, la gamme de *fa* mineur mélodique (*en bas*) est une inversion de la gamme de *do* majeur (*en haut*).

LA SYMÉTRIE AXIALE GÉNÉRALISÉE



La symétrie axiale généralisée est la combinaison d'une transposition et d'une inversion, c'est-à-dire d'une transposition et d'une symétrie autour du diamètre principal ($do-fa\#$). On obtient alors des symétries génériques qui correspondent à des réflexions par rapport à des axes qui ne passent par aucun point du cercle de la gamme tempérée.

soit comme un groupe de transformations, celui engendré par les transpositions, l'un des types des transformations utilisées dans la technique dodécaphonique. On peut représenter ce groupe par un cercle, celui-ci correspondant à une octave, divisé en 12 parties.

L'hypothèse sous-jacente de cette représentation circulaire est qu'elle permet de formaliser tout accord musical : tout accord de m notes distinctes correspond, d'un point de vue géométrique, à un polygone à m côtés inscrit dans le cercle (voir la figure 1).

Dans la tradition américaine, la note *do* est notée 0, tandis que les compositeurs sériels européens de la seconde moitié du XX^e siècle, tels le Français Pierre Boulez et l'Allemand Karlheinz Stockhausen (1928-2007), ont privilégié une autre représentation faisant correspondre au *do* l'entier 1. La différence entre école américaine et européenne est moins une différence de notation qu'une distance conceptuelle qui sépare les deux traditions théoriques. Dans cet article, nous adopterons la notation américaine.

M. Babbitt propose au début des années 1950 d'exprimer les opérations sérielles comme des transformations de la structure du groupe cyclique $\mathbb{Z}/12\mathbb{Z}$. En effet, il remarque que l'on peut interpréter les quatre formes d'une série dodécaphonique (la série elle-même, la série rétrograde, la série inversée et la série renversée rétrograde) comme les quatre transformations suivantes :

- P : $(a, b) \Rightarrow (a, b)$
- I : $(a, b) \Rightarrow (a, (12-b)_{12})$
- R : $(a, b) \Rightarrow (11-a, b)$
- IR : $(a, b) \Rightarrow (11-a, (12-b)_{12})$

Ici, la série dodécaphonique P est représentée par une suite de couples (a, b) , a indiquant la position de la note dans la série et b , la note (relative à une origine 0).

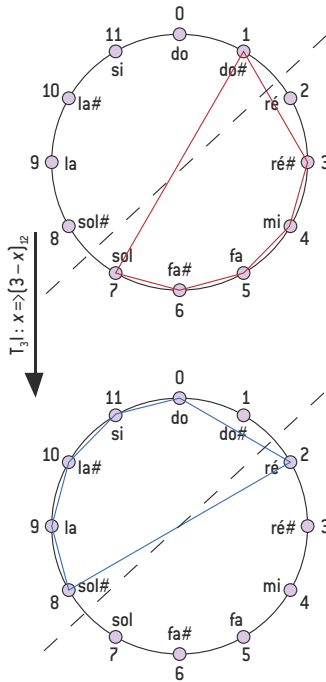
Ces quatre transformations d'une série dodécaphonique et, plus généralement, d'un profil mélodique constituent les éléments d'une structure algébrique nommée groupe de Klein de quatre éléments. Il tient compte de toutes les transformations de la musique dodécaphonique, alors que le groupe cyclique $\mathbb{Z}/12\mathbb{Z}$ ne rend compte que des transpositions.

Le groupe cyclique $\mathbb{Z}/12\mathbb{Z}$ et le groupe de Klein ne sont pas les seules structures algébriques intéressantes en musique. Deux autres groupes sont importants, mais avant de les décrire, nous devons définir deux

2. LA SYMÉTRIE AXIALE GÉNÉRALISÉE

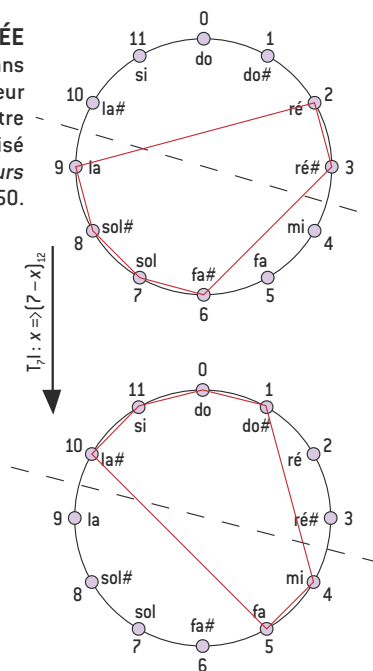
permet d'analyser cet extrait d'une polyphonie sérielle (*ci-dessus*) à deux voix. Ces deux voix correspondent à deux hexacordes (*l'un en bleu, l'autre en rouge*) liés par une symétrie axiale généralisée, comme le montre la représentation circulaire (*ci-contre*).

On peut aussi vérifier que si on segmente l'extrait en deux parties correspondant aux barres de mesure, on obtient également deux hexacordes en rapport de symétrie axiale généralisée.



3. LA SYMÉTRIE AXIALE GÉNÉRALISÉE

n'est pas utilisée uniquement dans la musique sérielle. Ainsi, le compositeur Olivier Messiaen en fait usage entre les deux hexacordes du mode utilisé dans la pièce *Mode des valeurs et d'intensités*, composée en 1950.



concepts importants, celui de la structure intervallique – l'énumération des intervalles, en demi-tons, qui séparent les notes successives d'un accord – et de la symétrie axiale généralisée, soit la composition d'une transposition et d'une inversion (*voir l'encadré page 94*). La représentation circulaire de ces deux idées, ajoutées à celles de la transposition et de l'inversion, nous fournira un ensemble d'outils d'analyse qui nous aideront à comprendre la structure de diverses œuvres.

Une boîte à outils

Ainsi, la notion de symétrie axiale généralisée nous éclaire sur la composition d'une polyphonie sérielle constituée de deux hexacordes (séries de six notes) liés par une telle symétrie (*voir la figure 2*). Le recours à une symétrie axiale généralisée dans le sérialisme n'est pas étonnant, car cette technique de composition est fondée sur une structure mathématique de groupe. En revanche, et cela soulève des questions quant au caractère universel de certaines constructions algébriques, on trouve de telles symétries chez des compositeurs utilisant d'autres techniques que le sérialisme, par exemple dans la musique modale du compositeur français Olivier Messiaen. De quoi s'agit-il ?

En musique, le mode d'une gamme est constitué des mêmes notes que la gamme dont il est issu, mais a une sonorité qui lui est propre, caractérisée par une tonique et par les intervalles entre cette tonique et les autres notes. Par exemple, à partir de la gamme de *do* majeur, dont la tonique est la note *do* (*do, ré, mi, fa, sol, la* et *si*), avec pour structure intervallique (en demi-tons : 2, 2, 1, 2, 2, 2, 1), on peut déplacer l'axe tonal sur la deuxième note afin d'obtenir un nouveau mode (en *ré*) ayant une nouvelle structure intervallique (2, 1, 2, 2, 2, 1, 2) en conservant les mêmes notes (*ré, mi, fa, sol, la, si, do*). Dans ce type de composition, on privilégie des notes ou des intervalles au détriment des autres. Dans la pièce *Mode de valeurs et d'intensités* de Messiaen, le mode qui engendre tout le matériau est constitué de deux parties liées par une symétrie axiale généralisée (*voir la figure 3*).

Nous pouvons maintenant aborder les deux autres groupes importants en musique. Commençons par le groupe diédral. Il s'agit de l'ensemble de toutes les compositions (au sens mathématique) des transpositions et des inversions. En d'autres termes, c'est le groupe des symétries axiales généralisées. Le nom diédral (à deux faces) indique que d'un point de vue géométrique, ce groupe correspond au groupe des symétries d'un polygone régulier de *n* côtés dans le plan. Ces symétries sont de deux types : rotations et réflexions (ou miroirs par rapport à un axe). Musicalement, les rotations correspondent aux transpositions et les réflexions sont des inversions par rapport soit à une note choisie comme pôle, soit à une note « imaginaire » qui se trouve entre deux notes à distance d'un demi-ton, quand l'axe de symétrie ne passe pas par une note du cercle chromatique.

Un premier exemple d'application du groupe diédral pour l'analyse musicale concerne la *Pièce pour piano op. 33a* écrite par Schoenberg en 1929. Les accords, obtenus par la segmentation de l'œuvre, se disposent d'une façon symétrique dans la partition (voir la figure 4).

D'une façon plus générale, les parties résultant de la segmentation de l'œuvre peuvent avoir des intersections; on parle alors d'une segmentation par imbrication. Un exemple d'une telle démarche est l'analyse que le théoricien américain David Lewin propose du *Klavierstück III* de Stockhausen (voir la figure 5).

Dans son analyse, D. Lewin distingue deux stratégies. Selon la première, les transformations sont organisées dans un ordre qui reflète le déroulement temporel de la pièce. Cette vision « chronologique » de l'organisation des transformations est une progression transformationnelle. Ici, le processus de segmentation par imbrication met en évidence une structure de pentacorde (une série de cinq notes), où l'on passe de l'un à l'autre (les deux ayant des notes en commun) grâce à une symétrie axiale généralisée: tous les pentacordes sont reliés par des transpositions et des inversions.

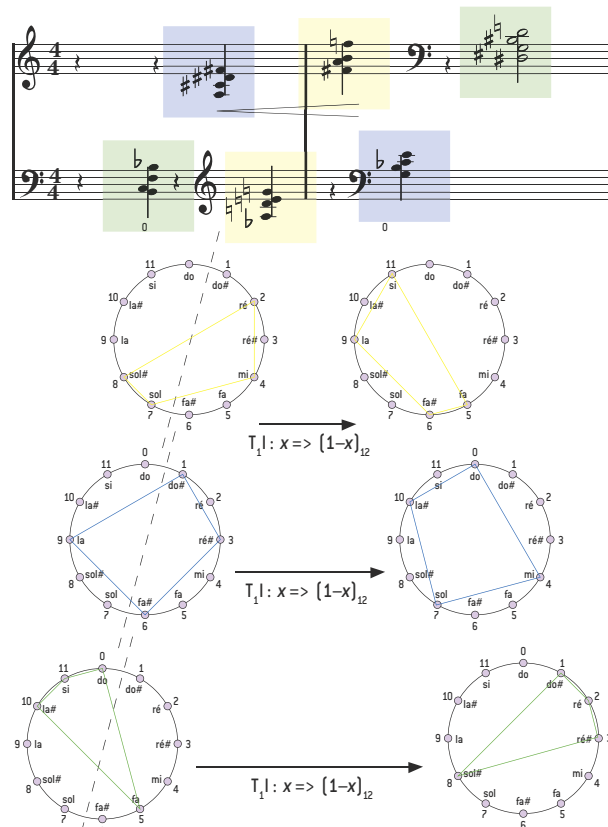
Progressions et réseaux

L'autre stratégie consiste à voir les transformations comme une structuration possible d'un espace abstrait, un réseau transformationnel, des formes du pentacorde dans lequel on analyse le déroulement de la pièce. Dans le réseau transformationnel du *Klavierstück III*, tous les pentacordes sont liés par des relations de transposition et d'inversion. Ici, à l'inverse de la progression transformationnelle, l'organisation des formes du pentacorde dans un réseau n'a aucun lien direct avec leur apparition chronologique.

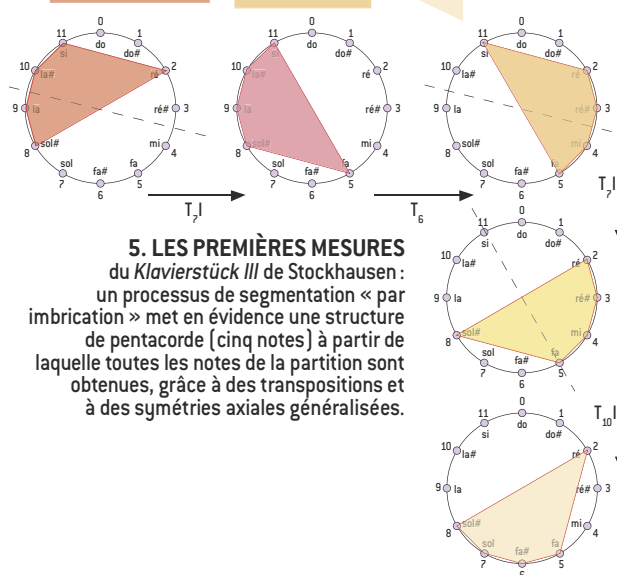
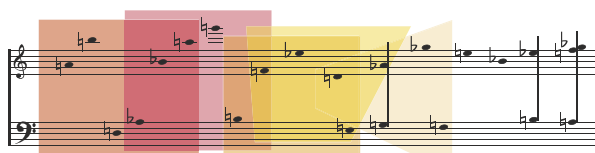
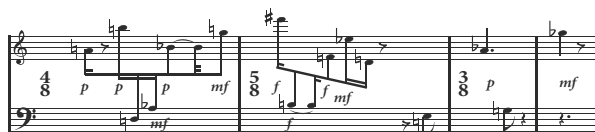
Dans un réseau, on retrouve parfois les mêmes configurations de correspondances entre pentacordes de la pièce dans des régions différentes: ce sont des isographies à partir desquelles on peut imaginer un lien étroit entre les réseaux transformationnels et la perception musicale. L'écoute de la pièce deviendrait ainsi l'un des parcours à l'intérieur de ce réseau avec la possibilité de repérer les isographies. Cependant, cette hypothèse doit encore être testée d'un point de vue de la psychologie expérimentale.

Passons maintenant au groupe affine d'ordre 48. Il est l'ensemble des fonctions f qui transforment un élément x de $\mathbb{Z}/12\mathbb{Z}$ en $(ax+b)_{12}$, où a est premier avec 12 et b appartient à $\mathbb{Z}/12\mathbb{Z}$. Le facteur multiplicatif a appartient donc à l'ensemble $U = \{1, 5, 7, 11\}$. Dans ce cadre, une transformation affine se réduit à une transposition quand $a = 1$ et à une inversion lorsque $a = 11$. Le théoricien américain Robert Morris a montré que les transformations affines sont compatibles avec des techniques utilisées par les musiciens de jazz, comme, par exemple, la substitution d'accords (voir l'encadré page 97).

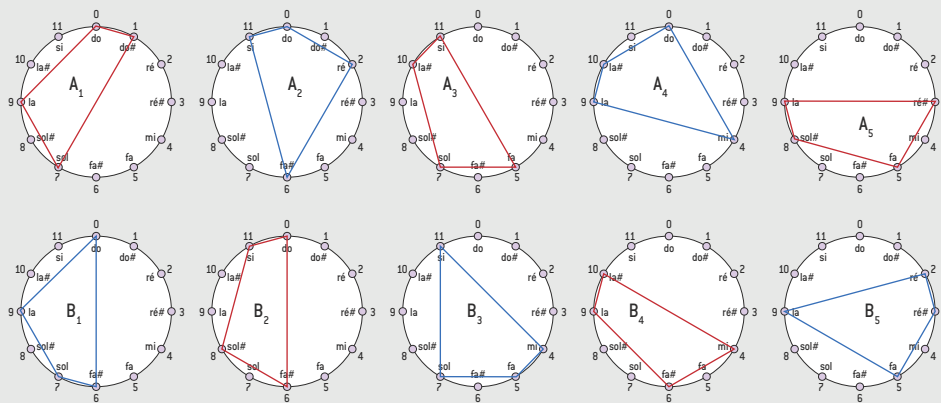
La description de ces différents groupes nous conduit à l'approche dite « paradigmatique » du problème de



4. LES ACCORDS DE LA PIÈCE POUR PIANO OP 33a de Schoenberg sont disposés de façon symétrique dans la partition. Il s'agit d'un exemple du paradigme du groupe diédral, car la symétrie axiale généralisée relie chaque couple d'accords.



5. LES PREMIÈRES MESURES du *Klavierstück III* de Stockhausen: un processus de segmentation « par imbrication » met en évidence une structure de pentacorde (cinq notes) à partir de laquelle toutes les notes de la partition sont obtenues, grâce à des transpositions et à des symétries axiales généralisées.



LES DEUX PROGRESSIONS (A et B) sont « préservées », dans leur caractère cadentiel, par les transformations affines : si on les analyse à l'aide des représentations circulaires, on se rend compte qu'elles sont constituées de seulement deux types d'accords (en rouge et en bleu) qui sont échangés à travers l'application affine.

la classification des structures musicales. Nous avons intégré ce concept d'analyse musicale, qui s'inspire des travaux du linguiste belge Nicolas Ruwet (1932-2001), dans notre démarche d'analyse musicale computationnelle. Grâce à *OpenMusic*, un langage de programmation visuelle pour la théorie, l'analyse et la composition assistées par ordinateur conçu et développé à l'IRCAM, un catalogue d'accords, ainsi qu'une analyse qui utilise ce catalogue, est valable à l'intérieur d'un paradigme qui sera plus ou moins pertinent en fonction du type de contexte qu'il essaie de décrire. Par exemple, nous avons vu que le paradigme du groupe diédral est valable dans le contexte de la musique atonale, tandis que le paradigme affine rendrait compte des techniques qu'on retrouve dans le jazz.

L'application de la théorie des groupes à la classification des structures musicales soulève des questions qui dépassent l'étude combinatoire du système tempéré. En effet, le philosophe allemand Ernst Cassirer (1874-1945) a étudié les relations entre le concept de groupe et les théories de la perception. Il a montré que la ressemblance perceptive entre différentes transpositions d'un même profil mélodique est liée « à un problème beaucoup plus général, un problème qui concerne les mathématiques abstraites ». On peut compléter : le groupe des transpositions engendre une relation d'équivalence entre structures musicales. En d'autres termes, une mélodie est reconnaissable à une transposition près : la structure du groupe cyclique et son action sur l'espace tempéré égal sont « transparentes » du point de vue de la perception musicale.

La perception algébrique

Peut-on généraliser cette propriété aux autres paradigmes ? Malheureusement, dans le cas d'autres structures algébriques, comme les trois autres groupes que nous avons détaillés, le problème n'a pas fait l'objet d'études approfondies. Pourtant, les compositeurs qui ont, consciemment ou inconsciemment, utilisé ce type de transformations en musique n'étaient sans doute pas insensibles à des considérations d'ordre perceptif. On a vu plusieurs exemples d'utilisation musicale de la symétrie axiale généralisée et d'inva-

riance par rapport à l'action d'un groupe sur un ensemble. Certains compositeurs, tel Elliot Carter, ont utilisé intuitivement le paradigme diédral, c'est-à-dire que ses techniques sont fondées sur les propriétés de ce groupe sans que l'auteur n'en ait eu conscience.

L'hypothèse d'une articulation entre progressions et réseaux transformationnels que nous avons décrite dans le cas de l'analyse du *Klavierstück III* de Stockhausen par D. Lewin conduit à proposer une alternative aux approches traditionnelles de l'analyse des formes temporelles.

L'analyse transformationnelle implique, d'une part, la « construction » d'un réseau, mais aussi, d'autre part, l'« utilisation » de cette architecture formelle pour dégager des critères de pertinence pour la réception de l'œuvre et pour son interprétation. Autrement dit, l'intérêt de construire un réseau transformationnel réside dans la possibilité de l'utiliser, à la fois pour « structurer » l'écoute et pour établir des critères formels utiles à son interprétation. En effet, la construction d'un réseau transformationnel s'appuie sur une volonté implicite de rendre « intelligible » une logique musicale à l'œuvre dans la pièce analysée.

Cette démarche analytique a probablement des implications théoriques inédites pour les sciences cognitives. Ainsi, l'approche transformationnelle non seulement représenterait un tournant en théorie et analyse musicales, mais elle déterminerait aussi une position singulière dans les rapports entre mathématiques, musique et cognition.

En somme, le concept de groupe est loin d'être simplement un outil technique de calcul. Comme l'a souligné le philosophe et épistémologue Gilles-Gaston Granger, « [C'est la notion de groupe qui] donne un sens précis à l'idée de structure d'un ensemble. [...] L'objet véritable de la science est le système des relations et non pas les termes supposés qu'il relie. » D'autre part, souvenons-nous de ce que disait le mathématicien Henri Poincaré : « le concept général de groupe préexiste dans notre esprit. Il s'est imposé à nous, non pas comme une forme de notre sensibilité, mais comme une forme de notre entendement. » C'est à nous, un siècle après, d'en tirer toutes les conséquences en musique. ■

✓ BIBLIOGRAPHIE

M. Andreatta et al., *Autour de la Set Theory. Rencontre Musicologique Franco-Américaine*, Collection « Musique /Sciences », IRCAM-Delatour, 2008.

M. Andreatta, *Méthodes algébriques en musique et musicologie du XX^e siècle : aspects théoriques, analytiques et compositionnels*, thèse, EHESS, Paris, 2003.

G. Assayag, *Computer assisted composition at IRCAM : PatchWork & OpenMusic*, Computer Music Journal, vol. 23(3), 1999.

M. Leman, *Music, Gestalt and Computing. Studies in Cognitive and Systematic Musicology*, Springer, 1997.

I. Xenakis, *Formalized Music (Rev. Edition)*, Pendragon Press, Stuyvesant NY, 1992.



La spatialisation du son

Olivier Warusfel

L'ESSENTIEL

✓ La perception auditive spatiale dépend de caractéristiques physiologiques et d'un traitement cognitif.

✓ La technique binaurale restitue l'illusion d'un son spatialisé dans un casque à l'aide de filtres encodant l'effet des directions de provenance.

✓ L'holophonie consiste à reconstruire le champ acoustique voulu dans toute une région de l'espace.

✓ La spatialisation du son autorise des expériences de réalité augmentée.

Confortablement installé dans votre canapé, vous écoutez le dernier *opus* de votre artiste favori sur votre tout nouveau système de *home cinema*. Les cuivres se lèvent à droite, les bois suivent à gauche, avant de laisser place aux ensembles de cordes droit devant vous. Vous avez l'impression d'être au beau milieu d'une salle de concert, devant un orchestre symphonique...

C'est en tout cas ce que vantent les publicités pour les systèmes de « son 3D ». En réalité, la spatialisation du son – les techniques de contrôle de la direction apparente de provenance des sons et la simulation ou la modification des effets acoustiques d'une salle de concert – proposée par les équipements audio grand public est loin d'être parfaite, et ce domaine de recherche loin d'être épuisé. Nous allons en exposer les principes et les développements récents.

L'introduction des techniques électroacoustiques en musique, au cours du XX^e siècle, a considérablement enrichi le répertoire des timbres, mais aussi la mise en scène des sons dans l'espace. La spatialisation du son est devenue un aspect incontournable de la création musicale. Grâce à l'outil informatique, il est désormais possible de placer et d'animer des

sources sonores de façon arbitraire, dans n'importe quelle direction, et ainsi de prolonger la scène traditionnelle où se trouvent les instrumentistes.

En parallèle, la spatialisation du son a connu ces dernières années une percée fulgurante chez le grand public par le biais des dispositifs de son 3D pour le *home cinema* ou les jeux vidéo. Dans ces derniers, l'ajout d'une composante sonore spatialisée est essentiel pour accroître le réalisme et la sensation d'immersion du joueur dans l'environnement de synthèse.

Quel que soit leur domaine d'application, les procédés de spatialisation font appel à plusieurs disciplines : informatique, acoustique, traitement du signal, psychoacoustique ou encore cognition. La synthèse de la direction et de la distance des sources sonores, de la propagation du son, ainsi que sa reproduction s'appuient à la fois sur la modélisation physique des phénomènes en jeu, mais aussi sur des modèles de la perception auditive spatiale.

Intéressons-nous en premier lieu à la perception spatiale du son. Elle est à la base de nombreuses techniques de spatialisation du son.

Contrairement à la vue, qui souffre d'un champ limité, l'audition couvre l'en-



© Shutterstock/Luisa Fernanda Gonzalez/Dash

Les techniques de spatialisation du son consistent à créer l'illusion que des sons proviennent de diverses directions de l'espace et à organiser ainsi des scènes sonores en trois dimensions. Elles connaissent aujourd'hui un essor important.

semble de l'espace. Notre capacité à localiser des sources sonores repose sur la comparaison entre les signaux parvenant à nos deux oreilles, qui dépendent bien sûr de la position de celles-ci, mais aussi de la diffraction des ondes acoustiques sur notre tête. Ces phénomènes dépendent de la direction d'incidence des ondes et permettent de localiser la source.

Une localisation *via* le décalage temporel et la différence d'intensité

Les performances de localisation du système auditif humain en azimut (dans le plan horizontal) sont de l'ordre de trois degrés pour des sources voisines du plan de symétrie de la tête. Elles tombent à une dizaine de degrés pour des sources situées sur le côté.

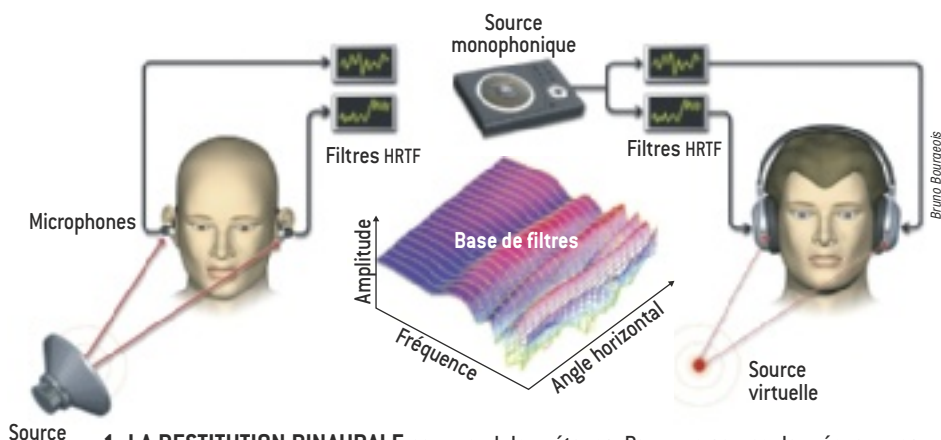
Comment fonctionne la localisation en azimut ? Les deux indices fondamentaux sont la différence entre les temps d'arrivée aux deux oreilles (ITD, pour *interaural time difference*), liée à leur écartement, ainsi que la différence d'intensité (ILD, pour *interaural level difference*), liée à l'obstacle que représente la tête. Ces indices dits

binauraux décrivent les relations entre les signaux captés par les deux oreilles. Toutefois, des considérations géométriques simples révèlent qu'à un couple de valeurs données des indices interauraux de retard et d'intensité correspond un ensemble de directions possibles. Ainsi, pour chaque direction dans le plan médian, les différences interaurales sont nulles. Des mécanismes plus fins sont nécessaires pour différencier les directions dans cette zone de confusion et plus généralement permettre la localisation en élévation.

En raison de l'interaction complexe des ondes sonores avec notre corps (diffraction par le crâne, le pavillon de l'oreille, les épaules, etc.), les variations de différences d'intensité dépendent de la fréquence du son. Dans les basses fréquences, les longueurs d'onde – plusieurs mètres – sont grandes par rapport aux dimensions de la tête. Celle-ci n'est donc pas un obstacle majeur et la différence de niveau sonore entre l'oreille du côté de la source (ipsilatérale) et l'oreille opposée (contralatérale) est faible. Plus la fréquence du son est élevée – plus sa longueur d'onde diminue –, plus les caractéristiques morphologiques contribuent à différencier les signaux parvenant aux deux oreilles. Ainsi,

chaque direction d'incidence s'accompagne d'une « coloration spectrale » particulière du son, que notre cerveau apprend à identifier. En d'autres termes, la perception auditive spatiale consiste à décoder l'effet de notre propre perturbation sur le champ d'ondes acoustiques. Les indices spectraux sont notamment sollicités pour discriminer les directions avant, arrière et, de façon plus générale, dans le plan vertical. La nature plus subtile de ces indices explique les performances assez médiocres de la localisation en élévation (angle vertical) – de l'ordre de 10 degrés de face, jusqu'à 30 degrés vers le haut ou l'arrière.

La perception de la distance dépend de plusieurs facteurs. L'éloignement de la source entraîne une baisse de l'intensité sonore et une atténuation relative des hautes fréquences liée à l'absorption par l'atmosphère. Ces indices ne sont cependant pas absolus ; ils supposent eux aussi un traitement cognitif fondé sur une connaissance préalable de la puissance et du spectre émis par la source. Lorsque nous identifions une voix, un cri d'animal, un véhicule, etc., nous en comparons le niveau et le timbre par rapport à d'autres événements du même type, dans diverses situations spatiales (proche, loin, derrière, etc.),



1. LA RESTITUTION BINAURALE comprend deux étapes. Pour une source donnée, on mesure d'abord les différences de temps d'arrivée et d'amplitudes aux deux oreilles d'un auditeur (à gauche). Ces différences, qui permettent de localiser la source sonore, sont caractérisées par une paire de filtres dits HRTF. On construit de même des filtres pour toutes les directions. Pour spatialiser un son monophonique, il suffit ensuite d'appliquer au signal émis la paire de filtres correspondant à la direction souhaitée : tout se passe comme si le son reconstruit venait d'une source virtuelle située au même endroit que la source réelle d'origine (à droite).

emmagasinés dans notre mémoire. D'autres facteurs s'ajoutent dans le cas d'environnements réverbérants comme une salle, pour lesquels notre perception s'appuie sur le rapport entre l'intensité directe de la source, qui varie avec la distance, et l'intensité réverbérée, qui est uniformément répartie dans la salle.

Trois familles de systèmes de restitution sonore

Les systèmes de restitution sonore spatiale visent, de façon générale, à reproduire aux oreilles d'un auditeur les valeurs appropriées de ces indices de localisation en fonction de la position souhaitée d'une source sonore virtuelle. Ces systèmes se rangent en trois familles : la stéréophonie et ses extensions, les techniques binaurales, réservées à la restitution sur casque, et les techniques de synthèse du champ acoustique. Au-delà de la fidélité avec laquelle elles reproduisent les indices de localisation, les mérites respectifs de ces différentes techniques se mesurent en termes de nombres de canaux à gérer et de contraintes sur la position et la mobilité de l'auditeur par rapport au dispositif de restitution.

Le modèle de restitution spatiale du son le plus utilisé encore aujourd'hui reste la stéréophonie. Deux haut-parleurs sont placés devant l'auditeur. En jouant sur le retard ou le niveau relatif des signaux envoyés à chaque haut-parleur, on contrôle de façon simplifiée les indices de localisation interauraux (ITD et ILD), ce qui donne

naissance à une source sonore fictive. La direction de cette « source fantôme » est limitée à la zone encadrée par les deux haut-parleurs. L'utilisation d'un simple retard entre les deux canaux aboutit à une zone de restitution très limitée. La plage de variation du retard pour laquelle les signaux fusionnent pour créer une image fantôme est en effet très restreinte. Dès que le retard dépasse une milliseconde (soit un décalage de 30 centimètres environ), le signal en retard est inhibé : il n'intervient plus dans le traitement perceptif de la localisation (effet Hass). En conséquence, un auditeur décalé, même légèrement, de l'axe médian des deux haut-parleurs aura l'illusion que l'ensemble de la scène stéréophonique se déporte de façon abrupte vers le haut-parleur le plus proche. L'utilisation conjointe d'une différence de gain contribue à limiter cet effet, sans toutefois le compenser totalement.

Par extension, il est possible de reproduire un son spatialisé sur des ensembles de haut-parleurs, plans ou tridimensionnels, disposés autour de l'auditeur. L'exemple le plus courant est la configuration « 5.1 » (cinq haut-parleurs répartis dans le plan horizontal) des formats sonores cinématographiques, également utilisés dans les jeux vidéo. Les installations sonores plus ambitieuses nécessitent la mise en place de dômes constitués de quelques dizaines de haut-parleurs. Pour synthétiser la direction d'une source sonore, on sélectionne la paire ou le triplet de haut-parleurs entourant la direction désirée et l'on règle les gains et éventuellement les retards relatifs en fonction de la direction souhaitée.

Ces systèmes partagent toutefois les mêmes limitations : l'illusion d'une source fantôme n'est valide qu'en un point particulier de l'espace (au centre du dispositif) et, à l'instar des trompe-l'œil, l'auditeur ne peut s'écarter de cette position sous peine de déformations, voire de dissipation totale de l'illusion spatiale.

La technique binaurale, pour sa part, court-circuite toute connaissance et modélisation physique ou perceptive des phénomènes en jeu et se contente de reproduire la réalité : son principe est de reconstruire directement l'ensemble des indices de localisation à partir de mesures acoustiques effectuées sur une tête humaine.

Longtemps confinée au laboratoire, elle a aujourd'hui atteint un degré de maturité qui rend possible son utilisation

L'AUTEUR



Olivier WARUSFEL est chargé de recherche à l'IRCAM, où il dirige l'équipe Acoustique des salles.

✓ BIBLIOGRAPHIE

N. Tsingos et O. Warusfel, *Modèles pour le rendu sonore. Dispositifs et interfaces de restitution sonore spatiale*, in *Traité de la réalité virtuelle*, P. Fuchs et G. Moreau, Les presses de l'École des Mines de Paris, 2005.

V. Larcher et A. Laborie, *Techniques de Spatialisation des sons*, in *Informatique Musicale : du signal au signe musical*, F. Pachet et J.-P. Briot, Hermès, 2004.

J. Blauert, *Spatial Hearing : the psychophysics of human sound localization*, MIT Press, 1997.

D. R. Begault, *3D sound for virtual reality and multimedia*, Academic Press Professional, 1994.

dans des contextes grand public, tels le jeu ou les télécommunications. La technique binaurale repose sur une procédure d'analyse et de synthèse du son perçu dans plusieurs directions. On peut l'assimiler aux techniques d'échantillonnage très utilisées dans le domaine de la synthèse sonore, où la simulation d'un instrument (piano, trompette, orgue,...) repose sur deux étapes : la constitution préalable d'une banque de sons enregistrés, parfois note par note, sur de vrais instruments, puis la restitution simple ou faiblement modulée de ces sons lors du jeu sur l'instrument virtuel.

Dans le cas de la synthèse binaurale, la première étape ne consiste pas à enregistrer des sons spatialisés, mais à mesurer la transformation – ou filtrage – subie par une onde sonore au cours de sa propagation, de sa source jusqu'aux tympans d'un auditeur. Pour cela, on insère des microphones dans les conduits auditifs et on enregistre la réponse à un signal test – en pratique, une impulsion – émis depuis une direction donnée. Ces mesures doivent se dérouler en chambre anéchoïque (sans écho) pour éviter les perturbations entraînées par la réflexion des parois. La réponse à cette impulsion élémentaire suffit à caractériser une fois pour toutes le couple de filtres (correspondant aux oreilles droite et gauche) associé à la direction d'incidence de l'onde sonore. Ces filtres, connus sous le nom de « fonctions de transfert de la tête » ou HRTF, (pour *head related transfer functions*), portent la trace de l'ensemble des phénomènes acoustiques survenus durant la propagation du son. L'opération est répétée pour un grand nombre de directions, de façon à baliser toute la sphère auditive et constituer une base de données de filtres directionnels.

Par la suite, lors de la synthèse spatiale, il suffit d'appliquer à tout signal monophonique – par exemple une voix captée par un microphone – le couple de filtres correspondant à la direction de la source virtuelle souhaitée. Les deux signaux obtenus, diffusés dans un casque, restituent de façon exhaustive les indices acoustiques responsables de la localisation auditive spatiale, et ce pour n'importe quelle direction présente dans la base de données.

Cette technique est réservée à la restitution sur un casque, faute de quoi, lors de la diffusion, le filtrage propre au trajet entre les enceintes et les oreilles de l'auditeur

se superpose au filtre choisi et perturbe l'illusion sonore. En adaptant « en temps réel » le couple de filtres appliqué en fonction de la direction de la source ou des mouvements de l'auditeur, on obtient sans peine une source fictive en mouvement.

La synthèse binaurale partage les mêmes limites que l'échantillonnage : ne reposant sur aucune modélisation, elle se prête mal aux manipulations. Elle nécessite par exemple la constitution d'une lourde banque de filtres. En effet, bien qu'il soit possible de reconstruire une direction manquante par interpolation entre deux directions voisines, celles-ci ne doivent pas être éloignées de plus de 15 degrés sous peine de voir apparaître des effets indésirables. La reproduction de sources en mouvement nécessite même des mesures tous les cinq degrés environ, ce qui conduit à des bases de données de l'ordre de 200 à 600 couples de filtres pour couvrir l'ensemble des directions.

En outre, les filtres HRTF dépendent par nature de la morphologie de l'auditeur. Ainsi, l'utilisation de filtres « génériques », mesurés sur une tête artificielle, entraîne en général des erreurs de localisation, en particulier des confusions avant-arrière.

Pendant, le couplage d'un dispositif de synthèse binaurale avec un capteur

d'orientation de la tête permet d'asservir de façon continue la scène sonore diffusée aux mouvements de l'auditeur. Cette rétroaction est déterminante pour l'auditeur, puisqu'elle lui permet de rétablir la corrélation entre ses propres mouvements et les variations des indices acoustiques transmis et réduit de façon considérable les confusions avant-arrière.

La technologie binaurale reste à ce jour le mode de restitution spatiale le plus rigoureux, et son utilisation se justifie non seulement dans le cadre d'études sur la perception et la cognition auditive spatiale, mais également sur le terrain du jeu ou des installations sonores interactives.

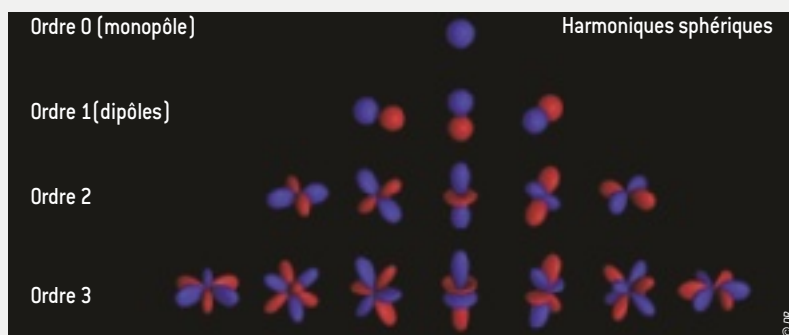
La synthèse du champ physique

Enfin, des approches plus récentes de la reproduction sonore s'attachent à modéliser et à reconstruire de façon précise les propriétés physiques du champ acoustique dans une zone plus ou moins large autour des auditeurs, en s'appuyant sur la résolution de l'équation d'onde. Parmi ces méthodes, on distingue les techniques holophoniques comme la *wave field synthesis* (WFS) et les approches par « décomposition harmonique », comme les formats *Ambisonics*.

LA DÉCOMPOSITION HARMONIQUE

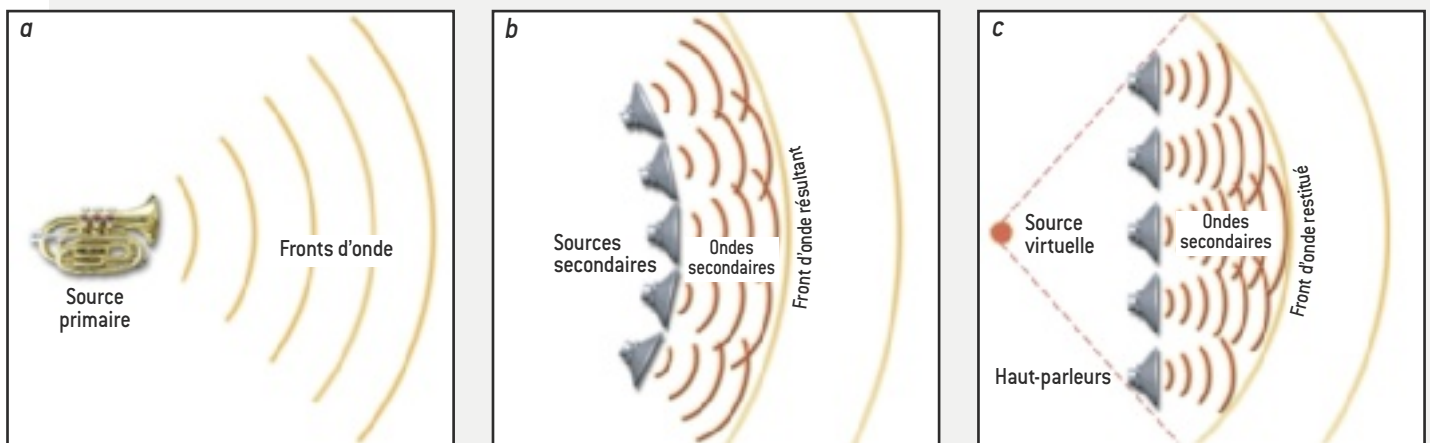
La spatialisation par décomposition harmonique consiste à représenter l'organisation spatiale du champ acoustique d'un point de vue angulaire (sans notion de distance) sous forme d'une superposition de fonctions périodiques de l'espace, les harmoniques sphériques, chacune associée à un canal audio.

L'harmonique d'ordre 0 (monopôle) correspond à une variation de pression homogène dans toutes les directions. Les harmoniques d'ordre 1 (dipôles) caractérisent chacune un gradient de pression selon une des trois directions de l'espace. Les cinq harmoniques d'ordre 2, puis celles des ordres supérieurs représentent des modes de vibration plus complexes ; elles permettent de préciser de façon progressive la dépendance angulaire du champ acoustique. La description exacte de celle-ci requiert en théorie un nombre infini d'harmoniques, et donc de canaux.



L'holophonie ou synthèse du champ acoustique (WFS) s'appuie sur le formalisme ondulatoire de Huyghens. À tout instant, chacun des points du front d'onde d'une onde originaire d'une source primaire (a) peut être vu comme une source secondaire qui réémet une ondelette sphérique de même phase et de même amplitude. Un instant plus tard, l'enveloppe de ces ondelettes secondaires correspond au front d'onde de l'onde primaire (b).

Pour reconstruire une onde acoustique, il suffit ainsi de disposer d'un ensemble de sources secondaires. Le champ acoustique s'exprime comme la somme des ondes émises par un réseau de sources secondaires situées sur une surface fermée entourant la région de restitution, ou sur un plan infini séparant les sources de cette région. Pour que la reproduction du champ sonore soit exacte, il faut une distribution infinie et continue de sources secondaires.



Développée dans les années 1980 à l'Université de Delft, aux Pays-Bas, la WFS suscite aujourd'hui un large intérêt, car le grand nombre de canaux (jusqu'à plusieurs centaines) que nécessite cette technique de synthèse spatiale n'est plus un obstacle technique.

Le signal sonore d'une source primaire que l'on désire spatialiser est diffusé sur un réseau de haut-parleurs très proches les uns des autres. Sous certaines conditions, la juxtaposition des fronts d'ondes émis par chacune de ces sources secondaires est alors équivalente au front d'onde d'une unique onde acoustique, provenant d'une source primaire virtuelle. Un contrôle adéquat du retard temporel et de l'intensité du signal envoyé sur chaque haut-parleur permet d'obtenir la création d'un front d'onde commun qui a pour origine n'importe quelle position souhaitée (voir l'encadré ci-dessus). Ce traitement doit être répété de façon indépendante pour chaque source composant la scène sonore.

En toute rigueur, la reproduction exacte du champ acoustique requiert une infinité de sources secondaires. En pratique, on se limite à créer des sources virtuelles dans le plan horizontal des auditeurs. Les sources secondaires sont alors elles-mêmes

placées dans ce plan, et moyennant quelques termes correctifs, il suffit d'installer des haut-parleurs le long d'une couronne autour des auditeurs, ou d'une rangée devant eux.

Le champ acoustique est reconstitué

De même, la distribution en théorie continue de sources secondaires est remplacée par un nombre fini de haut-parleurs régulièrement espacés. Cette discrétisation se traduit par une dégradation de la qualité de restitution spatiale dans les hautes fréquences. En pratique, pour assurer que l'indice interaural de retard soit reproduit de façon fiable dans la gamme de fréquences où il est primordial et sur l'ensemble de la région d'écoute, la fréquence limite est fixée au-dessus d'un kilohertz, ce qui conduit à un écartement maximal des haut-parleurs de l'ordre de 15 centimètres.

Si la stéréophonie s'appuie sur un modèle centré sur l'auditeur, la WFS adopte au contraire un point de vue exocentrique : le champ sonore est décrit et reproduit quel que soit le placement des auditeurs. La technique est donc particulièrement adaptée aux situations où les auditeurs sont mul-

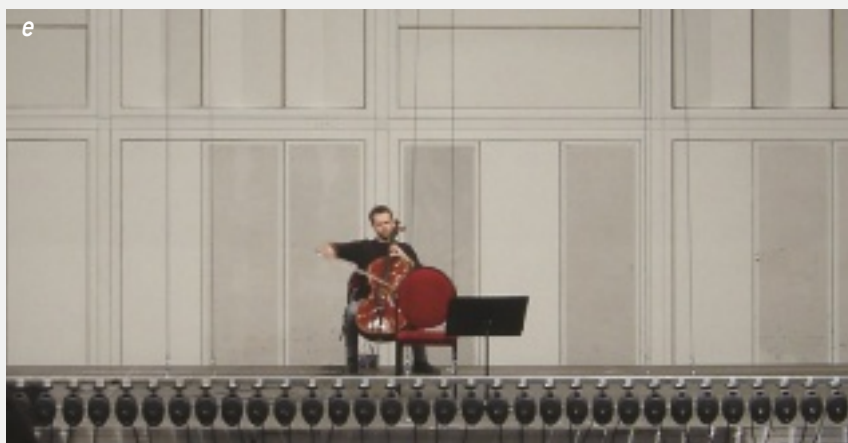
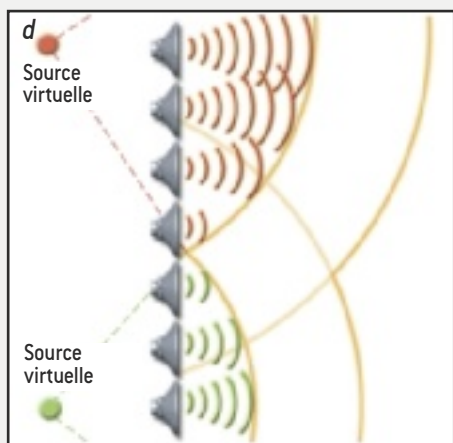
tiples et en mouvement, comme les concerts, les installations sonores, les expositions, etc. Cette technique a d'ailleurs été utilisée pour les expositions Dada, en janvier 2006, et Samuel Beckett, en 2007, au Centre Pompidou, à Paris.

L'autre approche pour reconstruire le champ acoustique est la décomposition en harmoniques – famille des formats *Ambisonics*. Introduite dans les années 1970 par Michael Gerzon, alors à l'Université d'Oxford, elle fait depuis une quinzaine d'années l'objet de nombreuses recherches. Elle repose également sur un formalisme physique, mais adopte un point de vue centré sur l'auditeur.

Par analogie avec l'analyse de Fourier, qui consiste à représenter un signal sous la forme d'une superposition de fonctions sinusoïdales, le champ acoustique entourant l'auditeur est représenté par une superposition de fonctions périodiques de l'espace (des harmoniques sphériques ou cylindriques), chacune associée à un canal audio (voir l'encadré page 103). Pour représenter un signal ayant une forte variabilité temporelle, il faut utiliser des composantes sinusoïdales de fréquences élevées ; de même, plus on désire décrire de façon précise l'organisation spatiale d'une scène sonore, plus il faut

En pratique, un réseau de haut-parleurs espacés de quelques dizaines de centimètres et disposés dans le plan des auditeurs joue le rôle des sources secondaires. En imprimant un décalage temporel (retard) et une variation d'amplitude (gain) au signal issu de la source primaire envoyé sur chaque haut-parleur, on crée un front d'onde cohérent qui semble provenir d'une source virtuelle (c), dont la position est identique pour tous les auditeurs. On peut de

même créer de multiples sources virtuelles (d). Le dispositif de synthèse du champ acoustique installé en front de scène à l'espace de projection de l'IRCAM (e) permet de spatialiser le son en temps réel, et de jouer sur la localisation du son de l'instrument et du son électroacoustique pour créer un instrument « augmenté » en les fusionnant, ou au contraire, les dédoubler en créant un ou plusieurs instruments virtuels éloignés.



tenir compte des harmoniques spatiales d'ordre élevé et donc utiliser un nombre important de canaux. Pour encoder l'intégralité de l'information spatiale, il faudrait en théorie une infinité de canaux.

Le principal intérêt du format *Ambisonics* est que l'encodage est indépendant du dispositif de restitution et hiérarchique. Le signal envoyé à chaque haut-parleur est une combinaison linéaire des canaux correspondant à chaque harmonique, dont les coefficients ne dépendent que du nombre et de la position des enceintes autour des auditeurs. Pour reproduire une scène sonore avec une précision spatiale – un ordre d'harmonique – donnée, il suffit de disposer d'un nombre de haut-parleurs supérieur ou égal aux nombres de canaux. Si l'on ne dispose pas d'assez de haut-parleurs, la scène peut être néanmoins décodée sans déformation, mais avec une précision spatiale inférieure.

Cette technique est bien adaptée aux applications de réalité virtuelle. On peut en effet enregistrer une scène sonore complexe comportant de nombreuses sources (ambiance urbaine, conversations dans un lieu public, etc.) en encodant son organisation tridimensionnelle, puis l'importer dans un environnement virtuel, à l'instar

des textures graphiques – des photos d'objets réels – appliquées sur un maillage. La facilité de manipulation permet de réorienter la scène en 3D pour compenser le mouvement de la tête de l'utilisateur.

Les systèmes de spatialisation ne peuvent se contenter de simuler la direction des sources composant une scène virtuelle. Ils doivent aussi rendre compte de l'interaction des ondes sonores avec l'environnement, en particulier la réflexion et la diffraction sur les obstacles et les parois de la salle. Reproduire ces effets est primordial pour que l'auditeur puisse estimer correctement la distance de la source et construire une représentation auditive cohérente du lieu virtuel où il évolue.

Reproduire l'effet de la salle

Parmi les différentes approches, on retrouve la distinction entre la synthèse par échantillonnage, par modèle physique ou par analyse et synthèse des propriétés acoustiques de la salle virtuelle. Toutes aboutissent cependant à un filtre, par lequel le son brut émis par la source doit être transformé pour reproduire les effets de l'environnement.

Une première possibilité est d'acquies des réponses acoustiques d'environnements réels. C'est l'extension de la technique binaurale, à ceci près que l'enregistrement est réalisé non plus dans une chambre anéchoïque, mais dans la salle que l'on veut simuler. Le filtre obtenu consigne l'ensemble des réflexions et absorptions subies par les ondes sonores. On constitue ainsi une bibliothèque de réponses de salles typiques (salon, amphithéâtre, salles de concert, église, ...). À l'instar de la technique binaurale, cette méthode garantit l'authenticité auditive de la simulation, mais elle souffre du caractère figé de la situation enregistrée. En particulier, elle interdit toute interactivité puisque la réponse est établie pour une position et une orientation données de la source et du récepteur, dans un lieu donné.

L'approche physique repose sur la simulation du comportement acoustique des parois du lieu virtuel. Les méthodes les plus courantes se fondent sur le modèle des « rayons sonores » qui se propagent en ligne droite dans l'espace, similaires aux méthodes employées dans le domaine graphique pour le rendu de la lumière. Les algorithmes construisent de façon itérative l'ensemble des chemins le long



B. Zöbeler, R. Gonzales-Arroyo, G. Eckel, O. Egger / photo : F. Schulz - Kunstmuseum Bonn.

2. LE PROJET *Listen* marie la spatialisation binaurale du son et des dispositifs de suivi de mouvement et d'orientation pour proposer des installations de réalité sonore augmentée. Les visiteurs d'une exposition sont dotés d'un casque sans fil qui diffuse un contenu audio adapté à leur comportement (ci-dessus, installation *Raumfaltung*, au Kunstmuseum de Bonn, en 2003).

desquels le son se propage de la source vers le récepteur, en s'appuyant sur des modèles de diffusion atmosphérique, de réflectance des surfaces, de diffraction des arêtes, etc. Les propriétés géométriques de l'ensemble des chemins permettent la construction d'un filtre encodant toutes les propriétés de la salle. Ces approches font l'objet de nombreuses recherches, notamment pour accélérer les calculs en vue d'une utilisation en temps réel. Le déplacement de la source ou l'auditeur dans l'environnement virtuel implique en effet de recalculer tout ou partie des chemins acoustiques.

La réalité sonore augmentée

Enfin, pour les situations de réalité virtuelle ne nécessitant pas un trop fort degré de cohérence entre le rendu sonore et visuel, on peut recourir à un modèle perceptif de l'effet de la salle, contrôlable par un ensemble de paramètres pertinents pour la perception (effet Doppler pour les sources en mouvement, effet d'occlusion par les parois, effet de réverbération). Sans passer par une description exhaustive des propriétés du lieu, ce modèle permet par exemple de déduire la durée de réverbération en fonction de données architecturales globales telles que le volume et l'absorption totale de la salle, et le rapport entre le niveau du son direct et le niveau de réverbération en fonction de la distance entre la source et le récepteur.

La prise en compte de ces propriétés perceptives permet aussi d'optimiser les algorithmes de simulation physique, ce qui est indispensable en cas de multiplicité des sources, faute de quoi les capacités de calcul des processeurs sont très vite dépassées.

L'équipe de recherche REVES de l'INRIA, par analogie avec des techniques de rendu graphique, exploite les propriétés de masquage auditif pour simplifier le traitement de la spatialisation sans dégradation du réalisme.

L'intérêt grandissant pour les environnements immersifs et interactifs, y compris dans le domaine artistique, incite à replacer les questions de spatialisation du son dans le cadre plus large de l'interaction entre les sens visuel, auditif et proprioceptif (perception de la position relative des parties du corps). À ce titre, la reproduction holophonique et la reproduction binaurale couplées à un dispositif de suivi du mouvement sont privilégiées, car elles assurent la correspondance en continu entre les indices auditifs spatiaux et les mouvements d'un utilisateur. Seuls ces systèmes permettent par exemple de reproduire le phénomène de « parallaxe auditive », qui se manifeste par le glissement relatif des plans sonores : lorsque l'auditeur se déplace, les sources d'arrière-plan semblent se décaler moins vite que les sources proches et les alignements entre sources se modifient.

Parmi les nouvelles applications de réalité augmentée, citons le projet européen *Listen*, dirigé par Gerhard Eckel, de l'Institut Fraunhofer de Bonn, en collaboration avec le KunstMuseum de Bonn et l'IRCAM. Le projet consiste ainsi à étendre le concept classique d'« audio-guide » employé dans les musées, en ajustant la progression du contenu sonore délivré en fonction du comportement du visiteur. En pratique, les visiteurs sont équipés d'un casque sans fil, et des dispositifs de mesure de la position et de l'orientation permettent d'asservir le contenu audio à leur parcours dans l'exposition, et ce de façon individualisée (voir la figure 2).

Les différentes technologies de spatialisation sonore que nous venons d'exposer témoignent de la vitalité des recherches dans ce domaine. Par ailleurs, il est intéressant de constater que les modèles ne précèdent pas toujours les technologies : si le formalisme décrivant la reproduction du champ acoustique précède de plus d'un siècle les premiers essais des systèmes WFS, *a contrario*, les premières réalisations de la stéréophonie précèdent les travaux qui en décrivent les fondements perceptifs. Le « théâtrophone » de l'inventeur Clément Ader, qui permettait à Marcel Proust d'écouter par téléphone et en stéréo des retransmissions de théâtres parisiens, précède ainsi de 20 ans les travaux de lord Rayleigh sur la localisation auditive!

De nos jours, la réalité virtuelle est à la fois un miroir de nos propres connaissances – on y expérimente le savoir qu'on y a injecté –, mais aussi un terrain d'approfondissement des connaissances, car elle permet de manipuler nos sens. Au-delà des applications artistiques, ludiques ou industrielles, la réalité virtuelle est un outil essentiel à la recherche sur la perception, en particulier auditive. ■

LA VILLE EUROPÉENNE DES SCIENCES

14 - 15 - 16 NOVEMBRE 2008
NEF DU GRAND PALAIS - PARIS

**ENTRÉE
GRATUITE**



vendredi 9 heures - 18 heures - samedi et dimanche 10 heures - 19 heures
www.villeeuropennedessciences.fr



MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE

SAISON
CULTURELLE
EUROPÉENNE
ue2008.fr



le nouvel **Observateur** Paris **Obs**



SCIENCE



Entendre dans un monde virtuel

Les environnements sonores réalistes dans des scènes de réalité virtuelle offrent de nouveaux moyens thérapeutiques pour certains troubles psychiatriques et perceptifs.

Isabelle Viaud-Delmon

Pour tout passionné de jeu vidéo, piloter un avion de chasse, voler dans un château hanté, combattre à mains nues une escouade de monstres devient vite une seconde nature. Aujourd'hui, vous pouvez même jouer au tennis ou au golf sans frapper aucune balle réelle avec votre raquette ou votre club. Ces mondes virtuels sont numériques, reconstitués par ordinateur, mais peuvent être d'un réalisme surprenant. Dans leur forme la plus aboutie, la réalité virtuelle, le joueur est plongé dans son monde grâce à une reconstitution en trois dimensions. Lorsqu'il se déplace, son environnement est « mis à jour » en temps réel : cet arbre sur la droite qui s'efface quand il avance, le voilà de nouveau si le joueur se retourne ; ce chien qui vient vers lui grossit comme il le ferait dans la réalité. Le joueur peut manipuler et modifier les objets qui se présentent à lui, ou faire un geste pour éloigner le chien.

Cette réalité virtuelle, un oxymore introduit par l'Américain Jaron Lanier au début des années 1980, se définit comme l'ensemble des techniques et des interfaces qui immergent un utilisateur dans un environnement artificiel et lui permettent d'interagir avec lui.

L'idée de réalité virtuelle repose sur le postulat qu'il existe une réalité non virtuelle, renvoyant au monde physique. Cependant, la réalité du monde physique se réduit à la perception que nous en avons ; sa représentation peut varier d'un individu à un autre, en fonction d'un nombre infini de facteurs tels que l'attention et l'acuité des sens. La différence essentielle entre les deux types de monde, le réel et le virtuel, est liée au manque de réalisme physique et fonctionnel de ce dernier, encore bien pauvre par rapport à ce que nos organes sensoriels savent décoder, et à l'équipement technologique dont un sujet a besoin pour le percevoir.

La réalité virtuelle a un autre inconvénient : ses applications, qu'elles relèvent du jeu, de l'apprentissage ou de l'art, reposent surtout sur notre sens de la vision, même si un certain nombre de sons sont restitués pour créer une ambiance sonore. Aujourd'hui cependant, les stimulations virtuelles qui associent vision et audition commencent à se révéler utiles en médecine, pour traiter des pathologies psychiatriques, mais aussi des troubles de la perception. Après avoir précisé les modes de fonctionnement de la réalité virtuelle

et ses applications en mode purement visuel, nous examinerons comment l'audition interagit avec la vision, et comment, dès lors, ces interactions permettent d'envisager de nouveaux traitements par réalité virtuelle.

Immersion et interaction

Si vous vous asseyez devant un ordinateur pour vous plonger dans un environnement artificiel, votre système perceptif n'est en général stimulé que par des informations visuelles. Qui plus est, celles-ci sont noyées parmi celles qui proviennent de l'environnement physique : la plante à côté de l'écran, le store derrière lui, etc. Un moyen d'améliorer l'immersion est de visualiser les informations sur un écran le plus grand possible, capable de couvrir l'ensemble du champ visuel. Autre solution : le visiocasque, composé de deux mini-écrans montés sur un casque ; vous ne visualisez alors que les informations artificielles, sans être perturbé par l'environnement.

Regardez maintenant, sur un écran géant, un jeu utilisant une manette, un



L. Vaud-Delmon/CHUS

joystick. Est-ce de la réalité virtuelle ? Oui et non. Oui, parce que votre champ visuel est entièrement stimulé par les informations provenant de l'environnement artificiel. Non, puisque celui-ci ne stimule que votre vision et que vous ne pouvez pas interagir naturellement avec lui : il vous faut utiliser une interface – la manette de jeu – pour relayer vos actions sur le monde artificiel.

En dehors des systèmes développés dans les laboratoires de recherche, les mondes synthétiques d'aujourd'hui ne reproduisent donc que faiblement la richesse du monde physique telle que nous la percevons normalement. Toutefois, l'industrie du jeu propose désormais des outils qui nous permettent d'utiliser naturellement notre corps pour interagir avec les mondes virtuels. Une telle implication du corps garantit que les diverses situations sont à la fois intéressantes et ludiques. L'utilisateur se laisse alors facilement convaincre qu'il agit dans l'environnement virtuel et non plus dans l'environnement physique : il a un sentiment de présence dans l'environnement virtuel.

Hormis ses multiples applications (voir page 110), la réalité virtuelle est de plus

en plus utilisée en médecine, notamment en psychiatrie. En 1994, l'équipe de Barbara Rothbaum, à l'Université Emory d'Atlanta aux États-Unis, a publié la première étude suggérant l'intérêt d'exposer des malades phobiques à des environnements virtuels. Deux groupes d'étudiants sensibles à l'acrophobie, la peur des lieux situés en hauteur, étaient comparés : les uns commençaient un traitement, tandis que les autres restaient en attente. Le traitement proposait, après une session de familiarisation avec le matériel, sept séances d'expositions progressives à différents environnements virtuels représentant des scènes de vide : un pont de hauteur variable au-dessus de l'eau, un balcon au rez-de-chaussée ou au 20^e étage, ou encore un ascenseur panoramique (voir page 113). On évaluait le degré d'anxiété, le comportement d'évitement et l'attitude des sujets avant et après le traitement. Ces mesures ont montré une nette amélioration pour le groupe traité.

Depuis cette étude pionnière, de nombreuses recherches ont mis en évidence l'efficacité de la réalité virtuelle pour traiter différents troubles anxieux. L'acrophobie, la phobie de l'avion ou de la

1. DANS CETTE SCÈNE de réalité virtuelle, un sujet (à droite), équipé d'un visiocasque et d'un gant muni de capteurs, voit un oiseau et entend le bruit de ses ailes. Il peut alors le repousser ou l'attirer de la main. L'expérimentateur (à gauche) peut visualiser la scène. Ce type d'expérimentation permet de tester l'influence des modes de perception sur l'action et le comportement.

L'ESSENTIEL

- ✓ La réalité virtuelle est une méthode d'immersion dans des environnements artificiels.
- ✓ Les mondes virtuels sont généralement privés de paramètres sonores réalistes alors que l'audition est la seule modalité sensorielle qui donne des informations sur tout l'espace environnant.
- ✓ La modalité sensorielle auditive, combinée à la vision, est utilisée dans des thérapies par réalité virtuelle.

Les applications de la réalité virtuelle

✓ La réalité virtuelle est utilisée dans de nombreux domaines : exploration de données scientifiques, géosciences, industries manufacturières, travail en milieu contaminé, robotique, architecture, archéologie, création artistique, architecture, entraînement à des manœuvres militaires, etc. Son utilisation médicale date des années 1990.

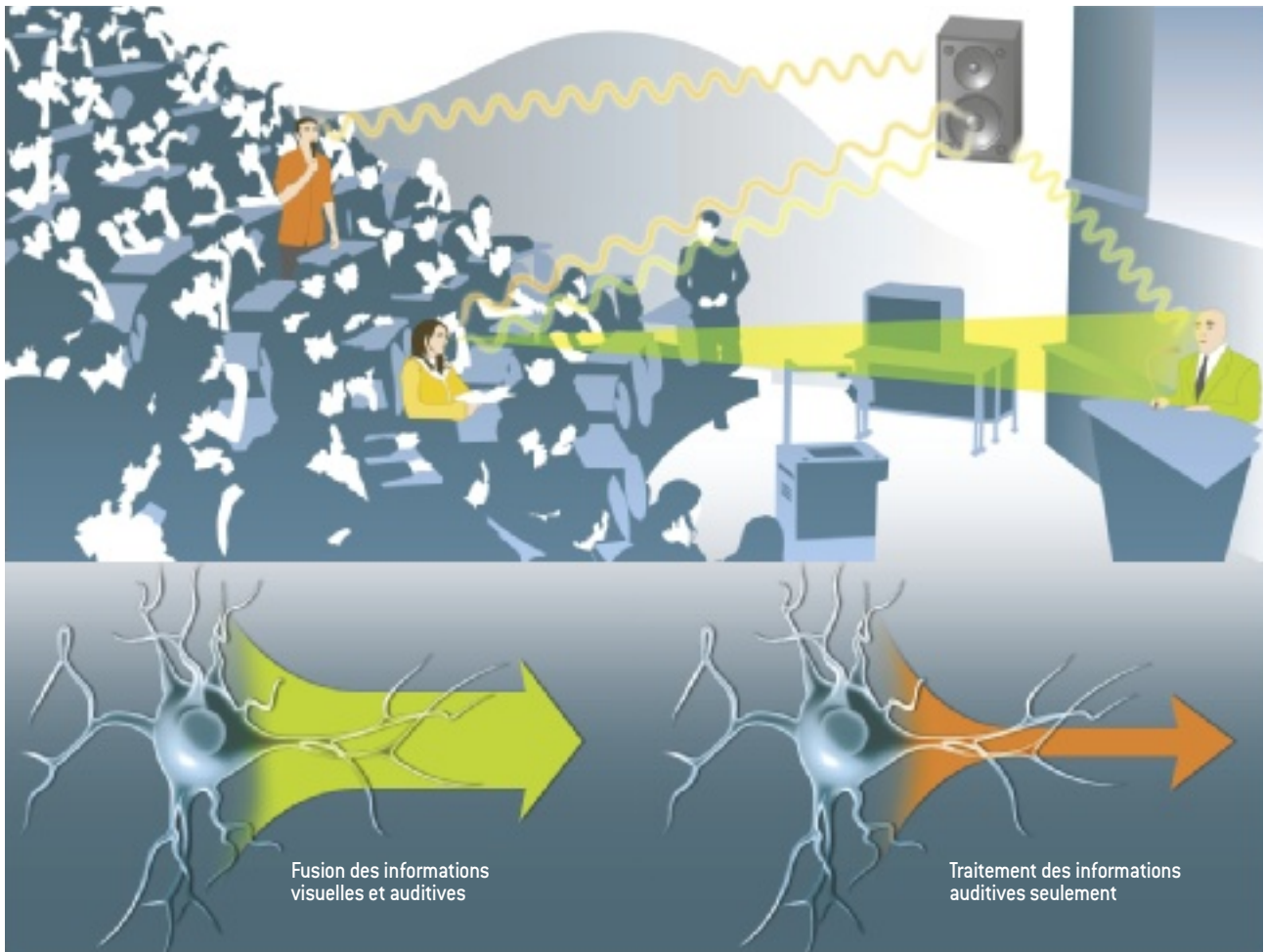
conduite, la phobie sociale (peur du regard des autres), la claustrophobie (peur de lieux clos), l'agoraphobie (peur des lieux ouverts et des foules) et l'arachnophobie (peur des araignées) sont couramment traitées de cette façon dans des centres médicaux aux États-Unis et dans certains pays européens. Autres indications : le stress post-traumatique, qui suit un accident, un attentat, une guerre, etc., ainsi que certains troubles des conduites alimentaires et diverses addictions au tabac, à l'alcool, aux drogues et au jeu.

En France, par exemple, l'équipe de Patrick Légeron, à l'Hôpital Sainte-Anne, à Paris, a testé ce type de traitement sur la phobie sociale, dans le cadre du projet européen VEPSY (*Virtual Environments for Clinical Psychology*). Un projet hospitalier de recherche clinique sur le traitement

de l'agoraphobie, coordonné par Jean Cottraux, a été mené en collaboration avec le Collège de France, et testé à l'Hôpital de la Salpêtrière à Paris, au Centre hospitalier de Luxembourg et à l'Hôpital universitaire de Lyon. Ces deux projets s'accordent avec les autres études pour conclure à l'efficacité des thérapies en réalité virtuelle contre les troubles anxieux.

L'intégration des sens

Aujourd'hui, les chercheurs en neurosciences étudient comment le cerveau intègre les différentes modalités sensorielles, c'est-à-dire réussit à créer une représentation cohérente d'une scène constituée par exemple d'éléments visuels, auditifs, odorants et tactiles : si vous tenez une



Delphine Baillly

2. LE COUPLAGE ENTRE UN SON ET UNE IMAGE est observé dans des neurones dits multisensoriels. Lors d'une conférence, la voix de l'intervenant est amplifiée par des haut-parleurs situés derrière ou sur le côté. Pourtant, on sait très bien que la voix entendue sur le côté est celle de la personne que l'on voit parler devant soi. Les informations visuelles et auditives sont fusionnées. Dans ces conditions, les neurones multi-

sensoriels sont très actifs (*en bas à gauche*). En revanche, quand un auditeur prend la parole dans la salle, on ne parvient pas à le situer du fait que sa voix est diffusée par les haut-parleurs et que l'on ne dispose pas des informations visuelles correspondantes. Quand une seule modalité sensorielle est activée, les neurones multisensoriels sont peu actifs (*en bas à droite*).

orange dans votre main, des éléments visuels, auditifs, odorants et tactiles contribuent à vous indiquer qu'il s'agit d'une orange ; et pourtant chacune des informations est de teneur bien différente. Le cerveau perçoit généralement un même objet par plusieurs canaux sensoriels. Des recherches ont démontré que la fusion de deux informations peut donner lieu à une interprétation perceptive qui ne se résume pas à une simple sommation des informations, mais qui produit des illusions. Par exemple, la perception d'informations visuelles et auditives incohérentes se traduit par l'effet McGurk, découvert par les Britanniques Harry McGurk et John MacDonald en 1976 : lorsque l'on voit une bouche prononcer la syllabe /ga/ en même temps qu'on entend un stimulus composé de la syllabe /ba/, on perçoit la syllabe /da/. Il s'agit d'un son intermédiaire du point de vue phonétique. Cette illusion reflète une stratégie employée par le cerveau pour combiner les informations visuelles et auditives afin d'aboutir à une perception unique et cohérente.

Entendre et voir

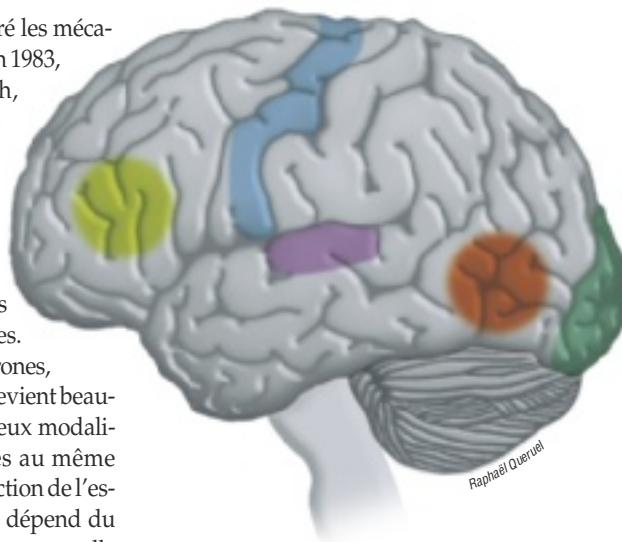
Imaginons maintenant une salle de conférence. Vous écoutez un séminaire donné par un conférencier équipé d'un micro (voir la figure 2). Alors qu'il n'y a pas de correspondance spatiale entre la direction de sa voix, émise par des haut-parleurs derrière vous ou sur le côté, et sa position dans l'espace, vous savez qu'il est bien l'orateur. Les informations auditives sont couplées automatiquement par votre cerveau aux informations visuelles. C'est l'effet du ventriloque : la voix du marionnettiste et la poupée en mouvement, bien que spatialement disjoints, sont fusionnées. Vient le temps des questions, diffusées par les mêmes haut-parleurs. Une personne dans la salle prend la parole, mais vous ne la voyez pas ; vous êtes incapable d'identifier qui parle puisque votre cerveau ne peut établir de correspondance spatiale entre la direction de la voix et un indice visuel. La personne se lève et c'est alors seulement que vous l'identifiez comme étant l'orateur, grâce aux informations visuelles liées à son mouvement ; vos informations auditives vous indiquent pourtant une autre direction. Comment le cerveau trouve-t-il la correspondance auditive et visuelle appropriée pour déterminer qu'un son et une image proviennent de la même source ?

Plusieurs études ont éclairé les mécanismes en jeu. En particulier, en 1983, Barry Stein et Alex Meredith, alors à l'Université de Virginie, ont observé que les neurones d'une région cérébrale où se projettent certains axones du nerf optique, le colliculus supérieur, répondaient aussi bien à des informations visuelles qu'à des informations auditives. L'activité électrique de ces neurones, qualifiés de multisensoriels, devient beaucoup plus importante si les deux modalités sont stimulées à peu près au même moment et selon la même direction de l'espace. L'amplitude de cet effet dépend du degré de coïncidence spatiale et temporelle entre les informations.

Le système visuel détecte avec précision la direction d'un stimulus, qui peut être représentée par deux coordonnées : l'azimut, c'est-à-dire la direction dans le plan horizontal, assimilable à un angle, et l'élévation, autrement dit la hauteur par rapport au plan horizontal (la verticale). L'œil détermine cette direction avec précision, parce que la direction sous laquelle un rayon de lumière entre dans l'œil est indiquée par le point qu'il stimule sur la rétine. En conséquence, le système visuel peut distinguer de très petits changements de direction. La plus petite différence perceptible entre deux positions visuelles est à peu près de cinq à dix secondes d'arc.

Le système auditif n'a pas cette capacité, parce que la direction du stimulus doit être calculée d'après une multitude d'indices. Un son émis sur la gauche atteint l'oreille gauche quelques microsecondes avant qu'il n'atteigne l'oreille droite, créant ainsi une différence « interaurale » de temps. De la même façon, il existe des différences interaurales d'intensité à cause de l'« effet d'ombre » exercé par la tête : un son provenant de la gauche arrive plus atténué à l'oreille droite qu'à l'oreille gauche. L'azimut d'une source sonore est déterminé grâce à ces différences interaurales de temps et d'intensité, mais avec une finesse de perception bien moindre que dans le cas de la vision. Quant à l'élévation de la source sonore, le système auditif l'estime avec une précision de l'ordre d'une dizaine de degrés.

Comme le système visuel est bien mieux équipé que le système auditif pour juger de la direction d'un stimulus, on pense généralement que les différences



3. L'INTÉGRATION SENSORIELLE. Trois aires sensorielles primaires du cortex cérébral humain sont représentées ici : l'aire de la vision (*en vert foncé*), l'aire somesthésique, ou aire de la sensibilité (*en bleu*) et l'aire auditive (*en violet*). D'autres régions corticales assureraient l'intégration des informations visuelles et auditives provenant du même objet : le cortex frontal inférieur (*en vert clair*), la partie postérieure du sillon temporal supérieur et le gyrus temporal moyen (*en orange*) et le cortex ventral occipitotemporal (*près de la limite avec le cervelet, non représenté*).

La psychothérapie par réalité virtuelle

✓ Utilisée pour traiter un trouble anxieux, la réalité virtuelle agit le plus souvent comme les thérapies cognitives et comportementales (TCC). En complément d'autres techniques, comme la relaxation, il s'agit d'exposer progressivement le patient aux conditions ou situations qu'il redoute ou qui provoquent son anxiété : les araignées, la foule, les hauteurs, les avions, etc. Cette exposition graduée amène le patient à contrôler les manifestations qui sont associées à son trouble. La thérapie en réalité virtuelle accomplit la même fonction mais en plaçant le patient dans des situations virtuelles dont le réalisme émotionnel lui fait vivre pleinement la situation qui déclenche les symptômes.

dans la précision de localisation entre les deux systèmes sensoriels provoquent une capture visuelle : la position d'un stimulus visuel dicte souvent la position apparente d'un stimulus auditif correspondant. Par exemple, quand on voit une moto à 100 mètres de soi dans un flot de voitures, on lui associe automatiquement le bruit qu'elle émet alors qu'il serait plus difficile de déterminer sa localisation les yeux fermés, rien qu'à son bruit.

Cependant, une étude récente de David Alais et Dave Burr, de l'Université de Sydney et de l'Institut des neurosciences de Pise, a suggéré que la capture visuelle d'un son ne dérive pas d'une règle rigide dans laquelle l'estimation visuelle détermine systématiquement la perception résultante. Le cerveau serait capable de pondérer les informations sensorielles selon leur qualité. Pour parvenir à cette conclusion, D. Alais et D. Burr ont demandé à des sujets d'estimer la localisation de sources sonores, puis de stimulus visuels. Ils ont ensuite proposé des stimulus visuels et auditifs de même position, ou de positions légèrement différentes. Les stimulus visuels variaient en taille, rendant leur localisation plus incertaine. Quand le stimulus visuel était très précis, les chercheurs observaient une dominance visuelle; lorsqu'il était de grande taille, la perception auditive l'emportait. De plus, quand les stimulus auditifs et visuels étaient tous deux

aussi précis, les sujets discriminaient plus finement leurs positions qu'avec un seul sens. Cela indique que l'intégration des informations spatiales visuelles et auditives améliore l'estimation de la position du stimulus. Ces observations s'accordent avec l'hypothèse que le cerveau utilise une règle statistique de combinaison optimale, reposant sur la fiabilité relative des informations sensorielles, pour combiner les signaux provenant des différents sens et pour percevoir la localisation d'un objet multisensoriel.

Réalité virtuelle multisensorielle

Peut-on utiliser ces connaissances pour améliorer les applications médicales de la réalité virtuelle? La variété des systèmes sensoriels est actuellement peu représentée dans les mondes virtuels utilisés en médecine. Les sons y sont peu présents, et dépourvus des indices qui permettraient leur localisation et optimiseraient l'intégration cérébrale des informations visuelles et auditives. Ils représentent le plus souvent l'équivalent d'images bidimensionnelles apparaissant dans une simulation visuelle, comme la silhouette d'un animal qui surgirait devant un bâtiment et autour de laquelle on ne pourrait tourner. Cette pauvreté des indices auditifs est regrettable,

cité 2008
des sciences & de l'industrie
LES CONFÉRENCES
le mercredi à 18 h 30

LES CELLULES SOUCHES

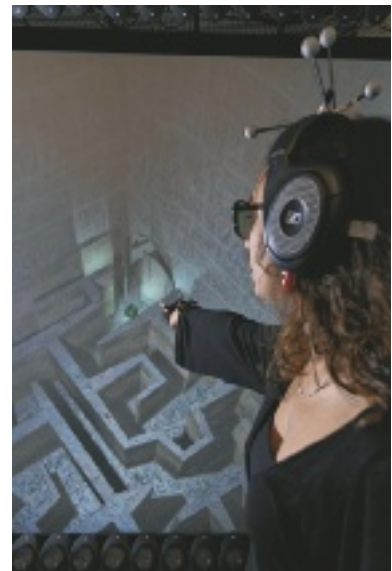
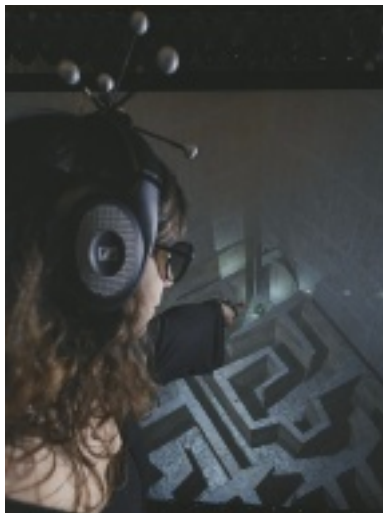
Avec Nicole Le Douarin,
professeur honoraire
au Collège de France

- 12 novembre
Les êtres qui régénèrent
- 19 novembre
Les cellules souches de l'adulte
- 26 novembre
La vie saisie par les biotechnologies

Accès libre

SCIENCE
FUTURA - SCIENCES
Cité

collège de la cité
Information 01 4005 35 96
cite-sciences.fr/collège
Porte de la Villette



© CNRS Bibliothèque F. Vignaud

Un sujet est équipé d'un capteur manuel et d'écouteurs. En bougeant la main, il contrôle une sphère à laquelle est relié un son qui reproduit celui de l'acouphène, une perception sonore spontanée, permanente ou intermittente, souvent handicapante. La scène comporte d'autres « objets sonores » dont la localisation dans l'espace est asservie aux mouvements de la tête du sujet. Cela enrichit son environnement virtuel auditif et favorise son impression d'être immergé dans l'espace. En déplaçant la sphère, le sujet apprend progressivement à maîtriser la localisation de son acouphène. Cette impression de le maîtriser fait qu'il perd son caractère agressif. L'environnement visuel représenté ici a été développé par des étudiants de l'Université de technologie de Vienne, en Autriche, pour le développement d'un jeu.

car l'audition est la seule modalité sensorielle qui nous mette en contact avec l'espace à 360 degrés autour de nous, un espace perçu bien plus large que celui du champ de vision.

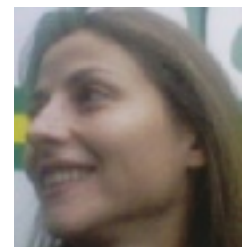
Puisque la modalité auditive fournit constamment des informations sur le monde environnant et la façon dont nous y évoluons, son ajout aux mondes virtuels ne peut être que bénéfique. Effectivement, lors du travail avec des patients anxieux, cet apport permet d'améliorer leur sentiment d'être présents dans l'environnement virtuel et non plus dans l'environnement réel. Pour cela, on crée des scènes sonores virtuelles en 3D en reproduisant les indices de perception sonores d'une situation réelle : ceux qui permettent la perception directionnelle horizontale, la perception directionnelle verticale et la distance du son. Une application en cours concerne le traitement de la phobie des chiens (voir l'encadré page 114).

Cependant, des problèmes peuvent surgir si la cohérence subjective des signaux sonores et visuels est difficile à obtenir. Dans ce cas, la nouvelle modalité sensorielle, au lieu d'enrichir l'intégration senso-

rielle, perturbe les sujets, car des informations non cohérentes créent une « surcharge cognitive ». De plus, les patients psychiatriques présentent souvent une hypersensibilité au bruit, alors que leurs audiogrammes sont généralement normaux. Il est ainsi nécessaire de comprendre de quelles façons la modalité auditive peut être utilisée efficacement.

Ainsi, afin de tester les réactions émotionnelles et comportementales de patients sensibles à l'espace et de sujets normaux, nous avons comparé leurs performances de navigation dans une ville virtuelle dans deux conditions d'immersion : visuelle d'une part, bimodale, c'est-à-dire auditive et visuelle, d'autre part. Le monde visuel comme le monde auditif modélisaient un environnement urbain réaliste. Cette expérience a confirmé l'importance du son 3D pour la construction d'un espace virtuel. La plupart des participants ont rapporté que l'immersion était plus convaincante lorsque la navigation était faite en condition bimodale. Cependant, plusieurs patients anxieux ont signalé qu'ils percevaient alors l'environnement visuel et l'environnement auditif séparément plutôt

L'AUTEUR



Isabelle VIAUD-DELMON est chargée de recherche CNRS à l'IRCAM et au Centre Émotion de l'Hôpital de la Pitié-Salpêtrière (UMR 7593 du CNRS), à Paris.

L'intégration d'informations visuelles et auditives permet à un sujet de maîtriser ses réactions émotionnelles lorsqu'il est exposé progressivement à des environnements où il peut rencontrer des chiens. Le sujet est équipé de lunettes polarisantes pour voir en trois dimensions et d'un capteur de position de la tête. Les informations de position de sa tête permettent d'ajuster la stéréo visuelle et les indices de localisation des sons en fonction de ses mouvements. Le sujet doit naviguer dans plusieurs

scènes virtuelles, où il peut voir ou entendre surgir des chiens. Les sources sonores fournissent des indices précis sur la localisation, quelle que soit la position du sujet dans la pièce expérimentale. Il peut par exemple percevoir des chiens arriver en courant derrière lui. Une étude est menée sur ce thème à l'IRCAM en collaboration avec l'équipe REVES de l'INRIA, à Sophia Antipolis, et le CNRS, dans le cadre du projet européen CROSSMOD. On saura prochainement si les personnes ayant été exposées à des « chiens virtuels » ont moins peur des chiens... réels.



© CNRS/IRCAM/INRIA

✓ SUR LE WEB

Cyberpsychologie, Québec
<http://w3.uqo.ca/cyberpsy/>

Projet FET Open Crossmod,
<http://www.crossmod.org>

✓ BIBLIOGRAPHIE

E. Klinger *et al.*, *Applications de la réalité virtuelle aux troubles cognitifs et comportementaux*, Traité de la réalité virtuelle, P. Fuchs et G. Moreau [sous la dir. de], 3^e éd., vol. IV, pp. 119-145, Presses de l'École des Mines, Paris, 2006.

I. Viaud-Delmon, *Corps, action et cognition : la réalité virtuelle au défi des sciences cognitives*. Intellectica, vol. 45, pp. 37-58, 2007.

B. E. Stein et T. R. Stanford, *Multisensory integration : current issues from the perspective of the single neuron*, Nature Rev. Neuroscience, vol. 9, pp. 255-266, 2008.

D. Alais et D. Burr, *The ventriloquist effect results from near-optimal bimodal integration*, Current Biology, vol. 14, pp. 257-262, 2004.

qu'unifiés, en dépit de la richesse équivalente des deux canaux sensoriels. Cette perception décalée serait liée à une surcharge sensorielle provenant de la grande sensibilité des patients anxieux aux informations multisensorielles.

Nous avons en outre observé que des patients anxieux peuvent parfaitement naviguer dans un environnement uniquement auditif, en l'absence d'informations visuelles. Ils souhaitent alors rester longtemps dans cet environnement qu'ils jugent réaliste, afin de l'explorer longuement. Quand on leur demande de replacer sur une carte différentes sources sonores qu'ils ont croisées durant leur navigation virtuelle, ils réussissent sans erreur. Ces résultats suggèrent qu'il pourrait être préférable, pour le traitement de certains cas pathologiques, non de copier la réalité le plus fidèlement possible, mais au contraire de créer un environnement simplifié qui permette au sujet de s'engager dans la tâche dont dépend le succès du traitement.

Une autre application de l'intégration visio-auditive concerne le traitement des troubles perceptifs, tels les acouphènes. Un acouphène est une perception sonore permanente ou intermittente (sifflement, bourdonnement) qui surgit indépendamment d'un son extérieur. Le plus sou-

vent, on en ignore la cause et on sait mal le traiter. Près de 15 pour cent des Français en seraient atteints, et on estime à 300 000 le nombre de personnes pour lesquelles ce son, quand il devient permanent, est intolérable.

Une collaboration entre l'IRCAM et le service d'ORL de l'Hôpital européen Georges Pompidou, à Paris, est en cours afin de développer un monde virtuel grâce auquel les patients apprendraient à mieux maîtriser cette perception (voir l'encadré page 113). Le sujet est immergé dans une scène virtuelle visuelle simple. Dans un premier temps, on détermine les caractéristiques sonores de l'acouphène du patient, c'est-à-dire l'amplitude et les fréquences du son perçu. Un son identique est alors recréé à partir de ces caractéristiques. Dans la scène virtuelle, il est localisé dans l'espace et relié à un objet. Le patient apprend à le déplacer par l'intermédiaire d'un capteur manuel. En déplaçant l'objet associé au son représentant l'acouphène, le sujet réussit progressivement à l'éloigner et à maîtriser sa localisation. Le but théorique est de rendre « évident » le caractère purement « illusoire » de la perception en lui faisant perdre ainsi son caractère agressif. Si cette méthode démontrait son efficacité, elle constituerait une percée dans le domaine du traitement des acouphènes. ■

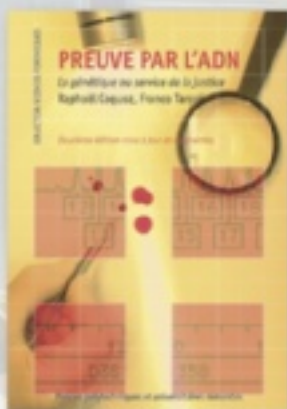


Collection publiée sous la direction du professeur Pierre Margot,
directeur de l'Institut de Police scientifique et de criminologie (Université de Lausanne)



ISBN 2-88074-697-3

Comment dater un corps grâce aux insectes amateurs de cadavres... La référence francophone dans le domaine !



ISBN 2-88074-696-5

Un cheveu suffit pour confondre un criminel... Un ouvrage destiné à tous, néophytes comme praticiens.



ISBN 2-88074-532-2

Comment à partir de saisies, identifier les réseaux de distribution, les sources et les méthodes de production ?



ISBN 2-88074-531-4

Accident, crime ou arnaque à l'assurance ? Les carcasses automobiles calcinées livrent leurs secrets.



ISBN 2-88074-511-X

La référence en matière de collecte et de protection des indices sur une scène de crime.



ISBN 2-88074-500-4

Balistique, identification des armes, prélèvement et analyse des résidus de tir sur les mains des suspects et sur les cibles... La référence dans ce domaine clé des sciences forensiques.

A paraître en 2007

Exploitation et expertises des traces de souliers

SCIENCES FORENSIQUES

Tous vos ouvrages de référence

www.ppur.org · ouvrages disponibles en librairies

L'organisation de l'illusion

Daniel Pressnitzer

Les sons d'un orchestre symphonique ne font qu'un, alors que certaines suites de Bach comportent plusieurs mélodies... jouées par un même instrument : les illusions sonores éclairent la façon dont le cerveau analyse une scène auditive.

Imaginez-vous confortablement assis dans une salle. Le concert n'a pas encore commencé et vous entendez par-dessus le bruissement des conversations autour de vous l'orchestre s'accorder. Vous distinguez successivement ou même simultanément divers instruments tels un hautbois, un violon, une contrebasse... Et tout à coup, après les premières mesures, cette collection disparate de sources sonores se transforme en un orchestre homogène, avec toute la section des cordes jouant une même mélodie.

En fait, vous êtes confronté à l'un des défis qui occupe les spécialistes de la perception auditive depuis de nombreuses années : comment, à partir d'une masse d'informations acoustiques entremêlées, déterminer le nombre et la nature des sources sonores réellement présentes ? Pour les chercheurs, il s'agit d'étudier ce que l'on nomme l'organisation des scènes auditives.

Au-delà de la musique, du point de vue de l'évolution, l'importance d'une telle faculté est évidente. Décider correctement, à l'écoute d'un bruit, s'il s'agit seulement du vent dans les arbres ou s'il dissimule aussi le bruit des pas d'un prédateur, confère un avantage capital en termes de survie.

Le problème à résoudre est pourtant complexe. Reprenons l'exemple de la salle de concert, mais avec une perspective physique. Des actions mécaniques variées – le souffle d'un instrumentiste sur l'anche du hautbois ou le mouvement de l'archet sur la corde du violon – font vibrer des objets. Ces vibrations excitent les modes de vibration d'autres objets – la colonne d'air du hautbois ou la table d'harmonie du



des sons, à la perception



Shutterstock/Jean Schweitzer

violon – et s’en trouvent modifiées. Puis elles se propagent dans l’air et se superposent les unes aux autres.

Cette superposition est une particularité de l’information auditive. Alors que les objets visuels révèlent leur présence instantanément, pour l’audition, l’information relative à l’objet physique est étalée dans le temps – c’est une onde acoustique – et elle est transparente : deux ou plusieurs ondes acoustiques peuvent se superposer dans le milieu de propagation. Il est alors impossible de décider combien de sources sont présentes à partir de la valeur de la pression acoustique à un instant donné.

Ces ondes acoustiques parviennent ensuite aux tympans, qui captent deux copies presque identiques des variations de pression au cours du temps. Ces seules informations nous permettent de distinguer combien de types d’instruments de musique sont présents, de les identifier et de suivre leur évolution dans le temps pour reconnaître une mélodie. L’étude de l’organisation des scènes auditives consiste à tenter de comprendre cette faculté.

Nous évoquerons les nouveaux outils utilisés pour comprendre les fondements cérébraux de l’organisation des scènes auditives, l’une des capacités les plus utiles et les moins bien comprises de la perception auditive. Ces outils reposent sur des illusions perceptives. Puis nous verrons que la musique est étroitement liée à l’analyse des scènes auditives ; plusieurs instruments peuvent se fondre dans une masse orchestrale ou, au contraire, chacun peut se démultiplier en plusieurs mélodies.

Pour aborder une faculté aussi complexe que l’analyse de scènes auditives, les neurosciences cognitives de l’audition combinent diverses méthodes fondées entre autres sur les observations comportementales et neurophysiologiques. L’objectif est de décrire le lien entre le phénomène physique, ici l’analyse acoustique de la scène sonore, et les percepts de l’auditeur,

L’ESSENTIEL

✓ **L’homme sait distinguer et reconnaître la nature des sons, comme une voix familière au milieu d’un brouhaha. Le cerveau analyse et organise les sons en utilisant divers indices auditifs.**

✓ **Mais des illusions auditives, telles les polyphonies virtuelles – plusieurs mélodies qui semblent jouées par un seul instrument –, peuvent leurrer la perception.**

✓ **Les musiciens utilisent ces ambiguïtés pour surprendre l’auditeur ; et les chercheurs s’en servent pour comprendre comment le cerveau organise une scène auditive.**

COMMENT ENTENDONS-NOUS ?

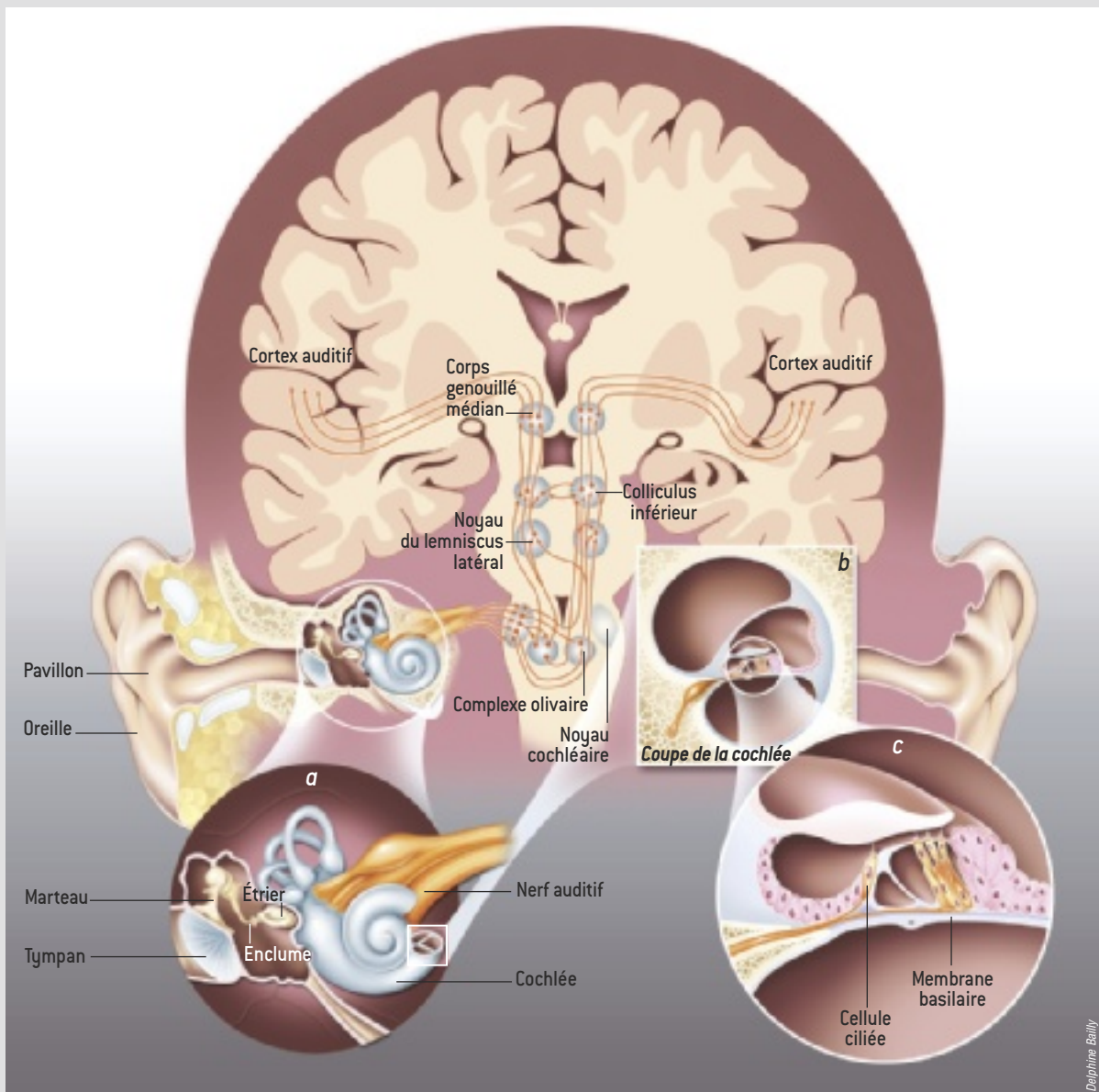
Chaque son parcourt un chemin type quand il arrive à nos oreilles. Il entre dans le pavillon et se faufile dans le conduit auditif où il rencontre le tympan (a). Les vibrations de l'onde sonore mettent en mouvement le tympan lié à trois petits os nommés osselets (le marteau, l'étrier et l'enclume). Ces premières étapes amplifient les vibrations qui atteignent la cochlée, l'organe récepteur de l'audition.

Chaque vibration, caractérisée par sa fréquence et son intensité, circule dans la spirale de la cochlée (b) où elle est captée par les cellules ciliées de la membrane basilaire (c). Cette dernière réagit à toutes les fréquences audibles et transmet ses déplacements aux cellules ciliées qui transforment la vibration en message électrique, circulant dans le nerf auditif. L'information se dirige ensuite vers le cortex cérébral, via plusieurs relais : le noyau cochléaire, le complexe olivaire, le noyau du lemnicus latéral, le colliculus inférieur et le corps genouillé médian.

Tout au long du trajet, le message subit des transformations dues aux caractéristiques de l'activité des neurones.

Chaque parcelle de la membrane basilaire n'est activée que par un ensemble limité de fréquences et donc chaque fibre du nerf auditif ne transmet que quelques fréquences. Cette organisation dite tonotopique se retrouve dans tout le système auditif jusqu'au cortex primaire. En outre, plus le son est intense, plus l'amplitude de vibration de la membrane basilaire est grande ; il s'ensuit une augmentation de la décharge des neurones du nerf auditif.

Toutefois, la fréquence et l'intensité ne sont que deux indices parmi d'autres qui doivent être utilisés pour l'analyse d'une scène auditive mêlant un ensemble d'ondes acoustiques. Cette analyse se ferait tout au long du système auditif, et même au-delà, pour utiliser les informations liées aux autres sens, à l'apprentissage ou au contexte.



c'est-à-dire ce qu'il perçoit, par exemple le nombre de sources qu'il détecte dans une scène. Entre les deux, divers traitements cérébraux éclairent le passage de l'acoustique à la perception. Commençons par quelques éléments de l'architecture du système auditif.

Le système auditif : des voies parallèles

Le récepteur sensoriel permettant de traduire les vibrations acoustiques en impulsions nerveuses, ou potentiels d'actions, est la cochlée (*voir l'encadré page ci-contre*). Les vibrations acoustiques recueillies par le tympan sont transmises aux fluides de la cochlée, qui mettent ensuite en mouvement une membrane élastique nommée membrane basilaire. Disposées le long de cette membrane se trouvent des cellules dites ciliées, qui, comme leur nom l'indique, possèdent des cils. Quand ces derniers sont défléchis par une vibration mécanique, ils modulent le potentiel de membrane de la cellule. Cela engendre des potentiels d'action dans le nerf auditif.

Une caractéristique essentielle de cette étape de codage est que chaque endroit de la membrane basilaire, et donc chaque groupe de cellules ciliées, n'est activé que pour un ensemble limité de fréquences de vibration. Ainsi, un son de fréquence aiguë ne stimule pas les mêmes cellules qu'un son de fréquence grave. Chaque fibre du nerf auditif, connectée par une synapse avec une cellule ciliée particulière, est donc spécifique d'un ensemble limité de fréquences. Ce dispositif réalise une analyse de type spectrotemporelle de l'information acoustique, autrement dit il décompose l'ensemble des fréquences perçues en fonction du temps.

Cette organisation de l'information en bandes de fréquences, établie à la périphérie, se retrouve dans la plupart des étapes ultérieures du système auditif. Sa fonction est toujours sujette à débat : certains y ont vu le mécanisme de codage des fréquences composant les sons, mais c'est aujourd'hui controversé. D'autres théories actuelles proposent qu'une telle décomposition est utile à l'analyse des scènes. Elle permet en effet une première séparation des signaux, dont les composantes sont souvent concentrées dans des bandes de fréquences étroites, par rapport aux bruits aléatoires qui recouvrent une large bande de fréquences.

Une particularité du système auditif, comparé à d'autres modalités sensorielles, est qu'il présente plusieurs centres de traitement entre le récepteur sensoriel et le cortex. Le nerf auditif est d'abord connecté à une structure du tronc cérébral nommée noyau cochléaire, divisé en parties dorsale et ventrale. Ces divisions initient une divergence des voies auditives, pour un traitement parallèle des différentes informations contenues dans les ondes acoustiques. Une première voie projette sur le complexe olivaire où les informations des deux oreilles sont comparées, probablement pour exploiter les indices de latéralisation des sons dans l'espace. En effet, si un son est

latéralisé vers la droite, le temps d'arrivée à l'oreille droite est légèrement plus court que celui à l'oreille gauche (de quelques dizaines de microsecondes).

L'autre voie de traitement projette sur le colliculus inférieur, divisé en régions centrale, dorsale et externe. Elle participe au traitement des caractéristiques spectrotemporelles du signal. La diversité des réactions du colliculus inférieur est vaste. Quelques neurones sont sélectifs pour certaines fréquences, d'autres combinent l'information sur de larges plages de fréquences.

L'information auditive s'achemine ensuite vers le corps genouillé médian, le relais thalamique de l'audition, puis vers le cortex auditif qui se divise en aires primaire et secondaire, chacune ayant des réactions susceptibles de changer rapidement selon le contexte ou la tâche à réaliser. Dès le colliculus inférieur, mais surtout dans les régions corticales, les informations d'autres modalités perceptives, tels la vision et le toucher, peuvent moduler les réactions auditives.

Cette brève présentation suggère que pour un son donné, la localisation, l'intensité, la hauteur (grave ou aigu) et le timbre impliquent des caractéristiques

LES SONS SE MÉLANGENT : séparer et reconnaître les sons d'une scène est l'une des tâches principales de l'audition.

différentes de l'onde acoustique, sur des échelles de temps allant de la dizaine de microsecondes à la seconde et avec des résolutions fréquentielles variées. La mise en valeur d'un type d'indice ne peut se faire qu'au détriment d'un autre. Il se peut alors que la diversité des traitements sous-corticaux reflète les diverses caractéristiques d'un son.

Résumons notre schéma du traitement auditif : imaginez un ensemble de voies parallèles qui extrait, avant le cortex, des caractéristiques auditives d'une complexité croissante, et qui fait ensuite interface avec des analyses impliquant des informations non auditives ou contextuelles. En ce qui nous concerne, la question posée est alors : à quel endroit et sous quelle forme intervient l'analyse des scènes auditives ?

Analyser tous les indices auditifs

Une scène auditive typique crée au premier niveau d'analyse, le nerf auditif, une activité spectrotemporelle complexe (*voir la figure 1*). À chaque instant et pour chaque région fréquentielle, l'activité neuronale traduit la présence d'énergie dans le signal acoustique, au sein duquel les différentes sources s'entremêlent. Comment identifier l'activité due à une source et la séparer de celles liées aux sources concurrentes ? Quand on

est confronté à une scène auditive, on prend sans y réfléchir plusieurs décisions pour séparer de nombreuses informations. Chaque composante spectrotemporelle doit être attribuée à une source, mais chaque source ne peut être identifiée qu'à partir de ses composantes spectrotemporelles. Un ensemble de « paris » est donc fait afin d'obtenir la solution la plus plausible.

Ces inférences inconscientes, pour reprendre l'expression du physiologiste et physicien Hermann von Helmholtz (1821-1894), reposent en fait sur presque tous les indices auditifs qu'on est capable de percevoir. Ainsi, les composantes provenant d'un même endroit ont tendance à être assimilées à une même source, ce qui semble raisonnable. Mais la localisation est loin de déterminer seule notre perception des sources.

Cet indice est même assez faible étant donné que, dans l'environnement, la localisation d'une source peut être perturbée par les multiples réverbérations du son. De plus, on peut distinguer divers instruments composant une mélodie même quand on l'écoute avec une seule oreille; d'autres indices sont donc exploités.

L'un de ces indices est la synchronie : si deux composantes débutent en même temps, il est vraisemblable qu'elles proviennent de la même source sonore, et on a tendance à les grouper. Un autre indice est l'harmonicité. Les composantes fréquentielles entretenant une relation d'harmonicité – elles sont des multiples entiers d'une fréquence fondamentale – sont souvent regroupées. Ceci est lié à la physique : les sons périodiques peuvent se décomposer mathématiquement en harmoniques. Or beaucoup de sons (par exemple, ceux des instruments de musique qui produisent des notes bien définies) sont périodiques.

Notre perception se serait donc adaptée aux contraintes physiques de la production des sons, en utilisant tous les indices auditifs disponibles, de façon

à proposer une organisation qui correspond aux sources sonores dans l'environnement. Mais ces indices d'organisation peuvent entrer en conflit. Dans ce cas, la décision perceptive – y a-t-il une source sonore ou plusieurs ? – dépend de l'ensemble des indices en présence et de la vraisemblance de chacun d'eux.

Reprenons l'exemple initial : l'ensemble des cordes de l'orchestre joue à l'unisson une même mélodie. Les indices spatiaux indiquent plusieurs sources : tous les violons et altos ne sont pas au même endroit. Mais cet indice est faible, et la synchronie entre les notes et leur relation d'harmonicité suggèrent plutôt une seule source : on n'entend effectivement qu'une seule source hybride, constituée des divers instruments (même si on sait, grâce à l'information visuelle, qu'il existe plusieurs sources, on n'est pas capable de les dénombrer à l'oreille). Voilà un premier exemple d'illusion auditive.

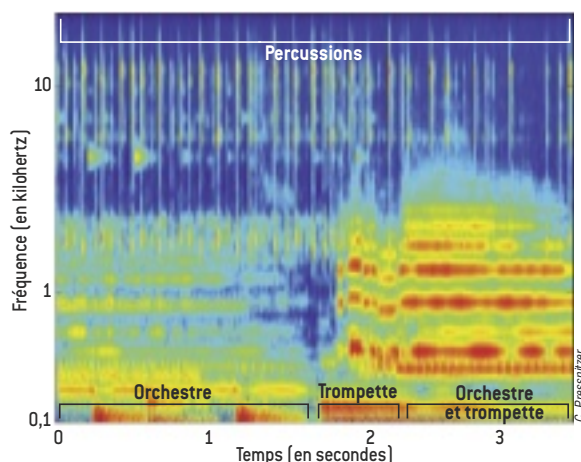
Des trompe-l'oreille

Une autre illusion est celle obtenue avec ce que l'on nomme les flux auditifs. Un flux auditif reflète l'attribution de composantes sensorielles à une même source, au cours du temps. On peut ainsi isoler la voix d'un chanteur dans un flux, distinct d'un autre flux formé par son accompagnement au piano. Pour étudier la formation de ces flux, un exemple simplifié est souvent employé : des sons purs de fréquences A et B qui alternent selon un motif répété ABA-ABA-etc. Si la différence entre les fréquences A et B est faible, les auditeurs perçoivent ce stimulus en un seul flux où tous les sons semblent découler d'une même source ABA-ABA-etc. En revanche, si la différence de fréquences est grande, le percept se sépare en deux flux distincts, comme si deux sources étaient présentes simultanément : A-A-A et B-B-B.

À nouveau, la décision perceptive est la plus vraisemblable. En effet, il est plausible que des fréquences proches et successives proviennent de la même source, mais cette interprétation devient de moins en moins réaliste quand les fréquences s'éloignent. Plusieurs indices peuvent être mis en compétition : la proximité fréquentielle, mais aussi temporelle ou spatiale des sons, donne lieu à des situations ambiguës ou illusoires.

En 2006, nous avons étudié ce qu'il advenait de la formation de flux auditifs quand les indices disponibles sont ambiguës. Par exemple, que se passe-t-il lorsque les indices de proximité fréquentielle et temporelle indiquent, de façon équiprobable, une seule source ou deux sources différentes ? Nous avons fait juger le nombre de flux à un groupe de volontaires, pour des séquences sonores assez longues (de quatre minutes) et ambiguës.

Loi de décider d'une interprétation et de s'y tenir, tous les participants alternaient spontanément entre un percept et un autre, c'est-à-dire entre un flux et



1. SIMULATION DE L'ACTIVITÉ au niveau du nerf auditif. L'ensemble des fréquences au cours du temps forment une représentation dite spectrotemporelle. Retrouver les sources est difficile : dans cet exemple [*Concerto de Aranjuez*, Joaquín Rodrigo], la trompette de Miles Davis se fond dans l'orchestre grâce aux subtils arrangements de Gil Evans.



2. DES FLUX sont créés par l'auteur dans les expériences de bistabilité auditive. Quand toutes les notes sont graves (*a, en vert*), l'auditeur entend un flux; quand des notes graves et aiguës alternent (*b, en vert et en violet*), il peut percevoir deux mélodies.

deux. Ces changements spontanés de perception semblaient aléatoires, et leur déroulement temporel spécifique à chaque personne. On pourrait résumer cette illusion ainsi : le stimulus physique ne change pas, mais la perception consciente qu'en ont les auditeurs alterne aléatoirement au cours du temps.

En collaboration avec Jean-Michel Hupé à Toulouse, nous avons comparé les caractéristiques de cette illusion auditive à celles observées pour une illusion similaire très étudiée dans le domaine de la perception visuelle. Dans cette modalité aussi, un stimulus visuel qui ne change pas peut provoquer des basculements spontanés de percept. Une telle succession de percepts, stables à chaque instant, mais variables au cours du temps, est nommée bistabilité.

La bistabilité visuelle

Cela fait plus de 250 ans que l'on étudie la bistabilité visuelle avec des stimulus variés. L'exemple le plus connu est la rivalité binoculaire; on présente deux images différentes de façon dichoptique, c'est-à-dire par exemple une maison dans l'œil droit et un visage dans l'œil gauche. Les deux images entrent alors en compétition pour parvenir à la conscience de l'observateur. Le résultat est une bistabilité perceptive: l'observateur voit à certaines périodes le visage, à d'autres, la maison. Autre exemple: dans le cas du canard-lapin créé par le psychologue Joseph Jastrow (1863-1944), deux interprétations sont possibles à partir de la même image fixe. Lors d'une observation prolongée, ces perceptions alternent dans l'esprit de l'observateur.

Ces exemples de bistabilité visuelle ont un point commun: l'information sensorielle disponible pour l'observateur est ambiguë. Il est aussi plausible de voir un canard qu'un lapin. Cependant, les observateurs ne déclarent jamais voir une créature mi-canard mi-lapin. La perception fait un choix, mais ce choix peut changer d'un instant à l'autre de façon à explorer les diverses possibilités. En fait, ce comportement serait l'expression des inférences inconscientes que nos systèmes perceptifs effectuent en permanence pour organiser l'information sensorielle.

Dans une scène naturelle, il existe toujours un certain degré d'incertitude. Mais en général, les indices disponibles sont concordants pour résoudre l'ambiguïté et déterminer s'il y a ou non, par exemple, des bruits de pas en plus du vent dans les branches. Ces

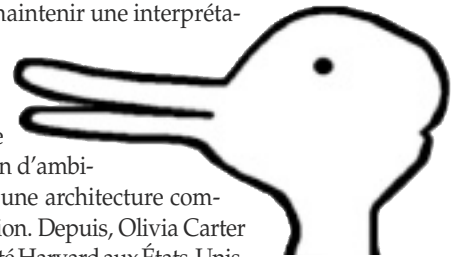
mécanismes de résolution d'ambiguïté se doivent d'être rapides et performants pour nous permettre d'évoluer dans le monde réel. En laboratoire, on peut cependant les mettre en défaut en produisant une ambiguïté insoluble.

Nous avons ainsi montré que les bistabilités visuelle et auditive sont semblables. Dans les deux cas, les percepts incompatibles sont traités successivement, et non mélangés. De plus, les durées des percepts successifs sont aléatoires, mais suivent une distribution statistique bien particulière. Enfin, les auditeurs ne peuvent pas maintenir une interprétation dans leur esprit, même s'ils peuvent parfois provoquer des transitions.

Nous pensons donc que les mécanismes de résolution d'ambiguïté perceptive partagent une architecture commune à l'audition et à la vision. Depuis, Olivia Carter et ses collègues, de l'Université Harvard aux États-Unis, ont établi l'existence de phénomènes de bistabilité pour le toucher. En conséquence, l'organisation d'informations sensorielles selon leur cause probable est un problème commun à toutes les modalités perceptives.

Alors la source de la bistabilité est-elle commune aux différents sens? Il serait possible d'envisager la présence d'une aire cérébrale dédiée à la résolution d'ambiguïté perceptive et située au-delà des régions consacrées au traitement de chaque sens. Mais plusieurs éléments indiquent que ce n'est pas le cas. Tout d'abord, nous avons remarqué que les caractéristiques des bistabilités visuelle et auditive, semblables en moyenne, changent d'un individu à l'autre. Certaines personnes produisent des alternances de percept très rapidement, tandis que d'autres restent fixées sur un percept plus longtemps. Toutefois, les personnes rapides pour la vision ne le sont pas forcément pour l'audition. De plus, lorsque deux stimulus bistables sont présentés simultanément en vision et en audition, les alternances de percepts sont indépendantes entre les modalités sensorielles, ce qui serait difficile à réaliser avec une aire unique de résolution d'ambiguïté.

Quand on s'intéresse aux causes neurophysiologiques d'une fonction mentale, on tente de mettre en regard le comportement lié à la fonction et l'activité cérébrale correspondante. Pour étudier la perception d'une ou deux sources dans l'environnement sonore, on ferait écouter aux auditeurs une source,



Du canard au lapin

✓ En observant cette image pendant un certain temps, on voit alternativement un canard ou un lapin, mais jamais les deux en même temps; la perception alterne entre deux images stables. On parle de bistabilité visuelle.

✓ BIBLIOGRAPHIE

D. Pressnitzer *et al.*, *Perceptual organization of sound begins in the auditory periphery*, *Current Biology*, vol. 18, pp. 1124-1128, 2008.

D. Pressnitzer et J. M. Hupé, *Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization*, *Current Biology*, vol. 16, pp. 1351-1357, 2006.

D. Huron, *Tone and voice: a derivation of the rules of voice-leading from perceptual principles*, *Music Perception*, vol. 19, pp. 1-64, 2001.

A. S. Bregman, *Auditory scene analysis*, Cambridge, MA, MIT Press, 1990.

puis deux, et on observerait les différences d'activité neuronale entre les deux situations. Un problème se pose toutefois: comment distinguer l'activité engendrée par le percept d'une source ou de deux sources et l'activité due aux conséquences physiques de la présence d'une ou deux sources ?

En effet, une scène contenant deux sources peut avoir une intensité plus forte ou une gamme de fréquences plus large qu'une scène ayant une seule source. Toutes les aires cérébrales spécifiques de l'intensité ou des fréquences seraient activées avec le percept d'une source ou de deux sources, sans que l'on sache si cette corrélation est due au signal ou à la perception.

L'ORGANISATION DES SCÈNES AUDITIVES recruterait tout le système auditif, au niveau cortical et sous-cortical.

La bistabilité auditive permet de lever cette difficulté. En effet, dans cette illusion, le percept d'une source ou de deux sources change subjectivement pour l'auditeur, sans que le stimulus physique ne soit modifié. Si l'on met en évidence une région cérébrale dont l'activité prédit le percept de l'auditeur, il est probable que cette région participe à la formation du percept lui-même.

Depuis peu, on utilise donc la bistabilité des flux auditifs pour dévoiler les fondements neuronaux de l'organisation des scènes auditives. Diverses techniques sont mises en œuvre. En imagerie par résonance magnétique fonctionnelle, Rhodri Cusack, du Centre de recherche médicale à Cambridge en Angleterre, a montré que le percept subjectif des auditeurs cor-

respond à l'activité cérébrale du sillon intrapariétal, une région corticale située au-delà du traitement auditif proprement dit et qui intègre des informations concernant différents sens (voir la figure 4).

Des études ultérieures, utilisant la magnétoencéphalographie ou l'imagerie par résonance magnétique, ont mis en évidence des corrélats neuronaux situés plus bas dans la voie de traitement des signaux auditifs, c'est-à-dire dans les cortex auditifs primaire et secondaire. Les diverses étapes du traitement cortical, auditives et multimodales – concernant plusieurs modalités sensorielles –, participeraient aux traitements des signaux visant à restituer l'organisation des scènes.

Les techniques d'imagerie pour l'homme ne permettent pas d'observer les étapes de traitement précoces (c'est-à-dire sous-corticales). Pourtant, l'organisation auditive pourrait commencer avant le cortex. En effet, les oiseaux sont capables d'isoler le chant d'un congénère dans le vacarme d'une forêt alors même qu'ils n'ont pas de cortex. Des régions sous-corticales participeraient-elles à cette organisation ?

En collaboration avec Ian Winter et Mark Sayles, de l'Université de Cambridge, et Christophe Micheyl, de l'Université du Minnesota aux États-Unis, nous avons cherché les traces de la bistabilité dans les premières étapes du traitement auditif. Chez le cobaye, nous avons enregistré les réactions neuronales dans le noyau cochléaire en situation de bistabilité auditive: des corrélats à l'organisation auditive se manifestent déjà, bien avant tout traitement cortical.

En résumé, nous avons vu que d'un point de vue comportemental, tous les indices auditifs peuvent influencer sur le choix de l'organisation perceptive. Au niveau neuronal, ils sont codés à divers endroits des voies du traitement auditif. De plus, l'organisation des scènes auditives recruterait l'ensemble du système auditif.

Peut-être vous demandez-vous quelle bistabilité auditive nous avons utilisée

L'AUTEUR



Daniel PRESSNITZER est chercheur au Laboratoire de Psychologie de la perception du CNRS, de l'Université Paris Descartes et de l'École normale supérieure, à Paris.

3. EXEMPLE d'une partition de Bach (Gigue de la Suite pour violoncelle seul n° 3 en do majeur) qui illustre les polyphonies virtuelles: plusieurs mélodies sont perçues, mais jouées par un seul instrument.



pour nos expériences ? Ce sont des enregistrements de sons conçus pour l'expérimentation (voir la figure 2). Mais nous aurions pu prendre de la musique.

Les musiciens ont placé la maîtrise de l'organisation auditive et des illusions au centre de leur art. Par exemple, les illusions dues à la formation de flux sont le fondement de ce que l'on nomme la polyphonie virtuelle. Dans cette technique d'écriture, plusieurs mélodies ou voix d'une polyphonie sont jouées par un seul instrument. Tout se passe comme si l'instrumentiste se dédoublait et l'auditeur peut, à loisir, porter son attention sur l'une ou l'autre des mélodies. La musique de Johann Sebastian Bach (1685-1750) pour instrument soliste, comme ses partitas pour violon ou suites pour violoncelle seul, présente des exemples frappants de cette technique (voir la figure 3).

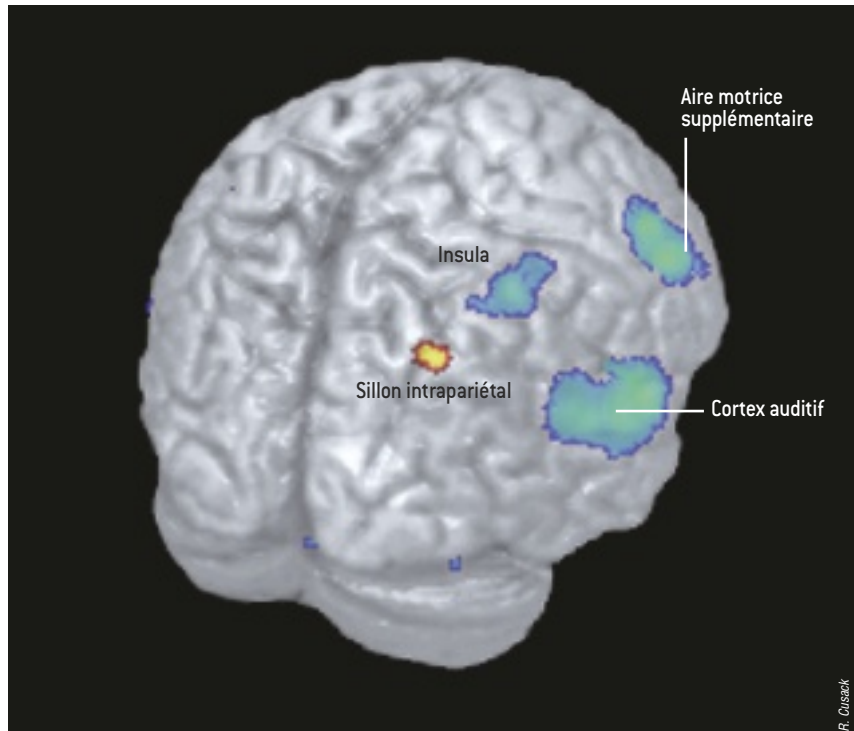
Et la musique dans cette analyse ?

L'utilisation des indices d'organisation auditive y est parfaitement maîtrisée. Lorsqu'une polyphonie virtuelle est voulue, elle se traduit par une faible distance entre les notes d'une voix, et de larges écarts entre voix. Si le tempo est bien choisi, sont alors présents les deux indices les plus importants pour la formation de deux flux distincts. Au-delà de la prouesse technique, rappelons que le percept de formation des flux est bistable, ce qui permet plusieurs écoutes très différentes de la même œuvre.

De nombreux compositeurs ont utilisé cette technique. Citons par exemple Ravel et son concerto pour la main gauche, écrit pour le pianiste virtuose Paul Wittgenstein qui avait perdu sa main droite pendant la Première Guerre mondiale. L'écriture, et ses sauts constants de registres, recrée de façon illusoire la main absente.

À l'inverse, l'orchestre symphonique réduit le nombre de sources perçues pour proposer à l'auditeur des masses orchestrales cohérentes. La technique pour fusionner des instruments suit aussi les principes de l'organisation auditive : des attaques synchrones et des intervalles entre instruments approchant les relations harmoniques.

Les musiciens ont donc compris certaines des règles de l'organisation des scènes auditives et s'en servent pour guider leur composition. Selon la psychologue et musicologue David Huron, à Toronto,



4. POUR UN SON QUI NE CHANGE PAS, mais qui est bistable, l'équipe de Rhodri Cusack a montré que le sillon intrapariétal postérieur (en orange) s'active davantage quand un auditeur perçoit deux sources que lorsqu'il en perçoit une seule. Cette aire participe à l'intégration de différentes modalités sensorielles, au-delà du système auditif. D'autres études ont montré l'implication de diverses régions cérébrales, corticales et sous-corticales, dans la résolution des illusions auditives. Les autres régions activées par la musique, un son ou même le langage sont le cortex auditif, l'insula et l'aire motrice supplémentaire (en vert).

il est ainsi possible de dériver nombre des principes du contrepoint classique en mettant en parallèle l'objectif musical (clarté des voix, contrôle des dissonances) et les règles d'organisation auditive.

Qui plus est, il montre avec une analyse statistique des partitions de Bach que les indices de formation des flux sont liés au contexte musical. Quand Bach veut faire entendre des voix distinctes, il évite les synchronies entre notes ; quand il cherche à les fusionner, l'analyse statistique indique qu'il synchronise parfaitement les notes.

Utiliser les illusions auditives

Notons que cette maîtrise des illusions d'organisation ne s'observe pas qu'en musique occidentale tonale. Les polyrythmes décrits par l'ethnomusicologue Simha Arom dans la musique d'Afrique centrale émergent uniquement parce que différents instrumentistes savent se fondre en un seul et même flux auditif. Enfin, certains compositeurs contemporains, tels

Györgi Ligeti et Jean-Claude Risset, se sont aussi emparés des illusions auditives pour en pousser l'exploration à l'extrême et en faire, parfois, l'un des principes de composition de leur musique.

Les illusions fascinent. Elles intéressent les philosophes depuis l'origine et sont à la base d'un argument fondamental sur la nature du lien entre notre perception et notre connaissance du monde. Pour les musiciens, elles servent d'inspiration pour jouer avec les oreilles de l'auditeur. Pour le chercheur, elles sont désormais un outil précieux permettant de révéler les inférences à la base de toute perception.

Dans le domaine de l'audition, ces nouveaux outils permettront peut-être de résoudre le problème de l'organisation des scènes auditives. D'autant que les applications sont nombreuses ; par exemple, l'une des conséquences majeures des pertes auditives est souvent une difficulté à suivre une conversation dans un environnement bruyant, ce qui suggère un déficit des capacités d'organisation auditive. ■



Barbara Tillmann

**Musiques d'Orient.
Musiques d'Occident.
Musiques d'Afrique.
Ou d'autres encore.
Le langage de la musique
est universel et on l'acquiert
de façon tout aussi
spontanée que l'on apprend
à parler. Même un auditeur
non musicien est un expert
de la perception musicale.**



La musique – comme le langage – sont des traits humains universels. Toutes les cultures produisent de la musique et y sont sensibles. Même sans être un musicien chevronné, chacun sait fredonner une mélodie. Qui plus est, cette capacité serait fort ancienne puisque les archéologues ont découvert un os d'ours percé de quelques trous datant de l'époque néandertalienne : la première flûte de l'humanité.

La musique serait-elle née avec l'espèce humaine ? À l'instar du langage, serait-elle un caractère inné et universel ? Pour aborder cette question, il convient d'examiner s'il existe des traits communs aux différentes musiques du monde – des universaux musicaux – et de rechercher comment les auditeurs perçoivent ces « invariants » musicaux.

Dans cette perspective, nous comparerons les caractéristiques des différents systèmes musicaux et observerons comment le cerveau les traite. Nous verrons que la comparaison de la perception musicale des adultes à celle des bébés permet d'étudier les processus innés et les processus acquis, lesquels dépendent de la culture. Enfin, en comparant le traitement de la musique à celui du langage, il est possible d'examiner s'il existe des circuits neuronaux spécifiquement impliqués dans la musique. Nous verrons ainsi que des invariants musicaux – des structures musicales présentes dans presque toutes les cultures – existent, et nous en dédui-

rons des invariants cognitifs, propres au cerveau de l'auditeur qui traite la musique.

Avant d'aborder cette étude des invariants cognitifs et de leurs substrats, partons à la recherche des invariants musicaux. Pour ce faire, rappelons brièvement les « fondamentaux » de la construction des musiques du monde et commençons avec la musique occidentale. Cette musique, dite tonale, repose sur les 12 notes de la gamme chromatique qui couvrent une octave.

Une gamme, des gammes

Les notes de la gamme se répètent d'une octave à la suivante, du grave à l'aigu. Les hauteurs des notes – leurs fréquences – sont fixées. Parmi les 12 notes, des sous-ensembles de sept notes définissent des gammes dites diatoniques. Par exemple, pour la gamme de *do majeur*, les sept notes sont les suivantes : *do ré mi fa sol la si*. Les deux notes les plus importantes de cette gamme sont le *do* et le *sol*, la première étant la tonique, la seconde la dominante. Le plus souvent, le *do* commence et finit la mélodie. Le choix de certaines combinaisons de notes et la séparation des notes en octave refléteraient des traits « naturels » respectant les lois de l'acoustique, d'une part, et celles de la physiologie du système auditif humain, d'autre part ; ils créeraient une bonne consonance acoustique.

Toutefois, quand on observe d'autres systèmes musicaux, on constate que cette

universel ?



organisation de gamme n'est pas universelle, mais qu'elle dépend de chaque musique du monde. Par exemple, dans la musique arabe ou dans la musique gamelan de Bali et Java, les types de gammes diffèrent ; dans la gamme orientale, on compte 24 notes organisées en différents sous-ensembles de sept notes. De plus, à Bali par exemple, la façon dont les notes sont accordées change selon les orchestres – les hauteurs spécifiques des notes sur lesquelles sont accordés les instruments dépendent du créateur de l'instrument ; elles peuvent donc changer d'un ensemble orchestral à un autre. En revanche, en musique occidentale, tous les orchestres s'accordent sur une même note (le *la* dont la fréquence est égale à 440 hertz).

Si la construction des gammes diffère, la musique est-elle vraiment un langage universel ? Où doit-on rechercher les universaux musicaux, s'ils existent ?

Quelques universaux

La musique est une information acoustique complexe organisée et structurée dans le temps. Les deux principales caractéristiques de cette organisation sont la hauteur des notes et la dimension temporelle – la durée des sons et leur distribution dans le temps.

En ce qui concerne la hauteur des sons, quel que soit le système musical, il existe certaines régularités musicales. D'abord, et nous l'avons évoqué, les notes sont orga-



nisées en gammes qui forment une progression discrète de hauteurs : l'ensemble des hauteurs n'est pas continu. Par ailleurs, un nombre réduit de notes (de cinq à sept) est choisi pour les sous-ensembles de la gamme (sept dans les exemples de l'encadré page 126). Les notes se répètent au fil des octaves selon une séquence cyclique du grave à l'aigu (pour la musique occidentale tonale, cette régularité est matérialisée par la séquence répétée des touches d'un clavier de piano).

En outre, les notes sont séparées par des intervalles inégaux ; par exemple, en musique occidentale tonale, les écarts entre les notes de la gamme diatonique (par exemple la gamme de *do majeur*), exprimés par rapport à la note initiale, sont les suivants : la deuxième est séparée de la première par deux demi-tons (matérialisés par une distance de deux touches sur le clavier de piano), la troisième de la première par quatre demi-tons, la quatrième par cinq demi-tons, la cinquième par sept demi-tons, la sixième par neuf demi-tons, la septième par 11 demi-tons et le *do* de l'octave suivante est séparé du *do* initial par 12 demi-tons. On dit alors que le patron de la gamme de *do majeur* en demi-tons est : 0-2-4-5-7-9-11-12. En le transformant en quarts de tons (0-4-8-10-14-18-22-24), on peut le comparer au patron de la gamme orientale nommée *rast* : 0-4-7-10-14-18-21-24.

Qui plus est, si, comme nous l'avons rappelé, toutes les notes n'ont pas la même

L'ESSENTIEL

- ✓ Les invariants musicaux sont des structures musicales communes à presque toutes les cultures.
- ✓ L'environnement musical dans lequel l'enfant grandit lui permet d'acquérir des connaissances implicites sur la musique, c'est-à-dire sans qu'il ait l'intention de les apprendre.
- ✓ Certains réseaux neuronaux seraient communs à la perception de la musique et à celle du langage.

L'AUTEUR



Barbara TILLMANN mène ses recherches au Laboratoire Neurosciences sensorielles, comportement, cognition, CNRS-UMR 5020, à l'Université Lyon 1.

importance dans la gamme de la musique occidentale tonale (il existe une tonique et une dominante, mais aussi une note sensible ou encore une sous-dominante, etc.), il en est de même dans la musique indienne où le bourdon joue le rôle de la tonique, ou dans la musique gamelan où ce rôle revient au gong. En outre, l'organisation des notes dont les hauteurs montent et descendent dessine un contour mélodique, dont l'importance se retrouve dans les musiques de toutes les cultures.

En ce qui concerne la dimension temporelle de la musique, il existe trois caractéristiques importantes. D'abord, le rythme (qui définit la durée relative des notes dans un morceau). Il en existe d'innombrables, qui diffèrent entre les pièces musicales et selon les cultures. Quant à la mesure, l'unité de base d'une partition, elle impose une pulsation régulière sur laquelle les patrons rythmiques se superposent. On en compte plusieurs types : la mesure à deux temps ou celle à trois temps – de la valse – par exemple. D'autres, plus complexes, sont des mesures à cinq ou à sept temps qui mêlent des groupements

de deux et de trois temps par exemple. Enfin, le tempo représente la vitesse d'exécution d'une œuvre musicale (il est plus ou moins lent ou rapide).

Diverses structures musicales se retrouvent donc dans toutes les cultures. Pourquoi ? Certains invariants résultent vraisemblablement de la façon dont le cerveau traite les sons. Quelles sont donc ces relations ? Peut-on en déduire des invariants cognitifs ?

Premiers liens entre musique et cerveau

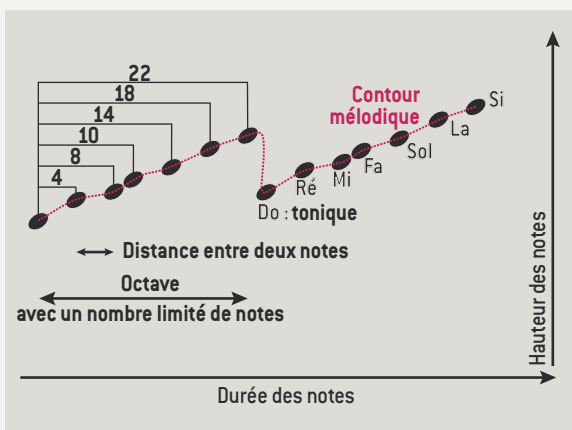
La musique, comme n'importe quel son, est traitée par le cerveau qui possède certaines propriétés d'organisation, d'apprentissage, de mémorisation et d'attention. Voyons quelques-unes de ces caractéristiques en relation avec la façon dont est construite la musique.

L'utilisation de notes de hauteurs discrètes permet de les catégoriser ; c'est une caractéristique cognitive de regrouper des événements différents dans une même catégorie selon leurs propriétés

LES POINTS COMMUNS DES MUSIQUES DU MONDE

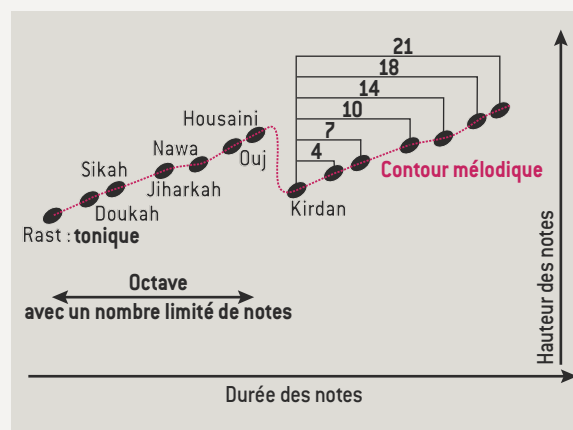
Certaines structures musicales sont présentes quelle que soit la culture ; on parle d'invariants musicaux. Il s'agit notamment de la distance variable entre les notes, du nombre de notes par octave qui est compris entre cinq et sept, de l'organisation des hauteurs des notes qui se fait par niveaux discrets. Ces invariants et d'autres existent dans toutes les musiques du monde. En revanche, le nombre de notes peut

changer et la façon dont elles sont accordées dépend de la culture. Nous prenons en exemple la gamme de *do majeur* de la musique occidentale tonale et celle nommée *rast* de la musique orientale pour présenter ces régularités musicales (sachant que certaines musiques ne s'écrivent pas sur des portées et, par conséquent, sont plus difficiles à présenter). Les invariants musicaux sont écrits en caractères gras.



Une gamme occidentale

Cette gamme de *do majeur* possède sept notes par octave : *do ré mi fa sol la si*. La distance entre les notes, exprimée en quarts de tons par rapport à la première note, représente le patron 0-4-8-10-14-18-22-24.



Une gamme orientale

Cette gamme *rast* possède aussi sept notes par octave : *rast doukah sikah jiharkah nawa housaini ouj*, la première note de l'octave suivante se nommant *kirdan*. La distance entre les notes, exprimée en quarts de tons par rapport à la première note, représente le patron 0-4-7-10-14-18-21-24.

AUTRES INVARIANTS

La fonction d'une note change selon le contexte ; ici le *do* est la tonique, la note la plus importante de la gamme, le *sol* la dominante et le *si* la note la moins importante de la gamme, la sensible. La note la plus importante de la gamme orientale est *rast*.

Le contour mélodique est le patron défini par les hauteurs des notes qui montent et qui descendent.

Le rythme est la durée des notes les unes par rapport aux autres.

La mesure est la pulsation de référence de la mélodie.

Le tempo est la vitesse d'exécution d'une mélodie.

similaires. Par exemple, la perception catégorielle des sons participe à la reconnaissance d'une mélodie même si les notes ne sont pas chantées justes. C'est aussi cette capacité qui aide à la reconnaissance des voyelles mal prononcées ou émises dans un environnement bruyant.

De même, on distingue deux notes de hauteurs différentes (à condition qu'elles soient séparées d'une différence minimale de hauteur), mais on associe deux notes identiques séparées par une octave (elles n'ont donc pas la même fréquence). Ces deux aspects – les contraintes de perception discriminative des hauteurs et le phénomène d'équivalence d'octave – reflètent des caractéristiques du système auditif.

Poursuivons la recherche de liens entre musique et cerveau. C'est le cas du nombre de notes par octave. Limiter ce nombre à cinq ou sept notes (et les répéter dans les différentes octaves) diminue le nombre de données que l'auditeur doit traiter. Cela refléterait les limitations cognitives de la perception catégorielle et de la mémoire à court terme qui ne pourrait stocker simultanément que quatre à neuf éléments. De plus, prendre des distances inégales entre les notes dans une gamme et leur attribuer des fonctions distinctes facilitent l'encodage et le stockage des informations mélodiques en mémoire (à court et à long termes) : les notes s'organisent autour d'un point de référence – la tonique par exemple – qui sert de point d'ancrage cognitif.

Des bébés universalistes

Pour préciser les universaux musicaux, les réactions des nouveau-nés ou des adultes n'ayant pas été exposés à certaines musiques permettent d'étudier des capacités innées. Certains travaux se sont intéressés à la mémoire et à la perception auditives des bébés, considérés comme des « universalistes », c'est-à-dire capables de percevoir la musique de toutes les cultures.

Dans des expériences dites de préférence, le bébé de 6 ou 12 mois est assis sur les genoux de sa mère et on lui fait écouter des mélodies sans que la mère ne les entende. On change ensuite une caractéristique musicale de la mélodie – la distance entre les notes, le contour mélodique ou la hauteur des notes, selon ce que l'on cherche à étudier – et on regarde quelle

chanson le bébé préfère écouter. Par exemple, un bébé regarde plus souvent le haut-parleur qui diffuse une mélodie familière et il ignore celui qui diffuse une musique qu'il ne reconnaît pas. On observe aussi s'il distingue des variations dans la structure musicale. Ainsi, on a montré que les bébés mémorisent mieux les notes quand elles sont séparées de distances inégales, et qu'ils préfèrent la consonance – des sons acoustiquement cohérents ou qui sont « accordés » – à la dissonance.

Qui plus est, les bébés prêtent davantage attention aux hauteurs relatives des

une « bonne forme » dans l'information fournie. Pour ce faire, il regroupe des données semblables pour y chercher une continuité. Les regroupements se font comme on le ferait pour des informations visuelles suivant leur similarité et leur proximité : ainsi, la suite XXXOO est perçue en deux groupes (XXX et OO) à cause de la similarité des lettres, et la suite XXX XX en deux groupes par proximité spatiale.

En analysant les structures mélodiques de différents systèmes musicaux, on a constaté que des lois similaires existent pour la perception auditive, puisque

ON RECONNAÎT UNE MÉLODIE même si elle est jouée à une hauteur différente.

notes, c'est-à-dire au contour mélodique, qu'à la hauteur spécifique des différentes notes de la mélodie. Ce comportement reflète aussi une caractéristique de la perception des auditeurs adultes.

L'importance cognitive de l'information relative s'observe de même pour la dimension temporelle de la musique : les bébés et les adultes focalisent leur attention sur l'organisation temporelle relative des notes, à savoir le rythme, plutôt que sur les durées absolues de chaque note. Ainsi, on reconnaît une mélodie grâce à son rythme, même si elle est chantée de façon différente – passant d'un registre aigu et d'un tempo rapide à un registre grave et un tempo lent par exemple.

En outre, on observe dans les musiques de différentes cultures que la mesure engendre un comportement de synchronisation : elle crée un cadre de référence temporelle qui permet de taper des mains ou du pied en cadence quand on écoute une mélodie à deux, trois ou quatre temps. C'est aussi grâce à elle que l'on danse sur la musique, que l'on chante et que l'on joue sur différents instruments ensemble. Cette caractéristique peut être décrite comme un invariant cognitif.

Pour trouver d'autres invariants cognitifs liés à la musique, on s'intéresse aussi aux travaux réalisés dans d'autres domaines. La psychologie cognitive a décrit des principes d'organisation pour la perception visuelle, notamment des règles de groupement. L'idée fondamentale est que l'observateur cherche à percevoir

l'auditeur groupe les séquences de notes selon une similarité de timbre ou d'intensité et une proximité (temporelle ou de hauteur). En 2002, Glenn Schellenberg, de l'Université de Toronto, et ses collègues ont étudié l'influence de la proximité de hauteur dans la perception des mélodies. Ils ont présenté aux participants des débuts de mélodies dont la dernière note variait en hauteur (donc en proximité avec la note précédente) et les participants devaient juger si elle convenait bien pour terminer la séquence. Dans une autre étude, les participants chantaient la suite des mélodies et les chercheurs analysaient la distance de hauteur entre la dernière note entendue et la première note chantée.

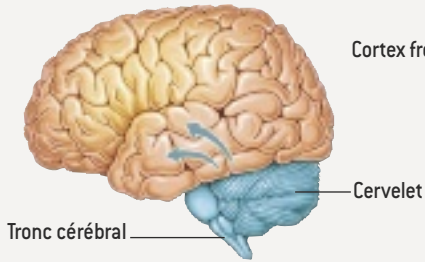
Les résultats montrent que les participants préfèrent produire ou entendre des notes proches en hauteur de la dernière entendue. Et ce, quels que soient leur âge (enfants ou adultes), leur expertise musicale (musiciens ou non-musiciens), leur origine culturelle (par exemple, Américains ou Chinois) et le style musical des mélodies (des chansons folkloriques britanniques, des chansons chinoises ou des musiques contemporaines).

Plusieurs études suggèrent donc que les invariants musicaux présents dans les différentes cultures reflètent les mêmes contraintes perceptives et cognitives. L'apprentissage des régularités qui existent dans les structures musicales est une autre façon d'étudier comment le cerveau traite la musique.

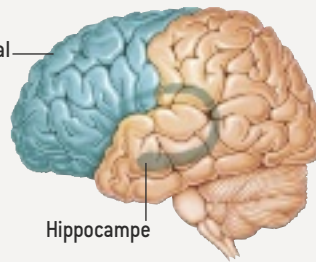
LE TRAITEMENT DE LA MUSIQUE PAR LE CERVEAU

Plusieurs régions cérébrales participent à la musique. Le son est d'abord traité par les structures de l'oreille et les régions sous-corticales et corticales propres au système auditif. Puis interviennent différentes parties du cerveau, impliquées dans la

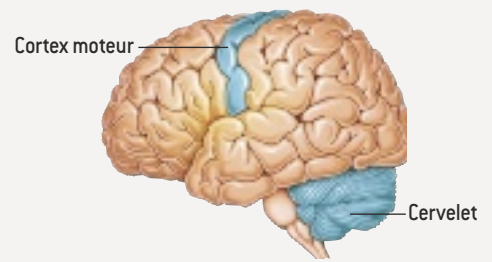
mémoire, les émotions, les mouvements ou d'autres modalités sensorielles. Certaines sont communes à la musique et au langage et d'autres seraient spécifiques à la musique. Voici quelques-unes de ces régions, cette liste n'étant pas exhaustive.



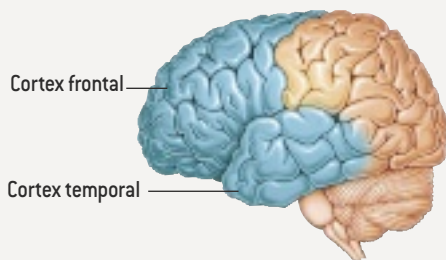
1 Écouter des sons active notamment le noyau cochléaire, le tronc cérébral et le cervelet. Puis l'information se déplace vers le cortex temporal où se trouvent les aires auditives primaires et secondaires.



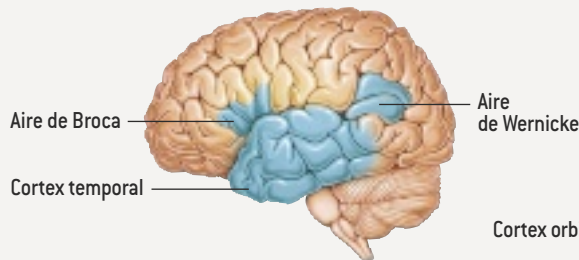
2 Écouter une musique familière active entre autres des régions impliquées dans la mémoire. Ce sont par exemple l'hippocampe et des aires du cortex frontal.



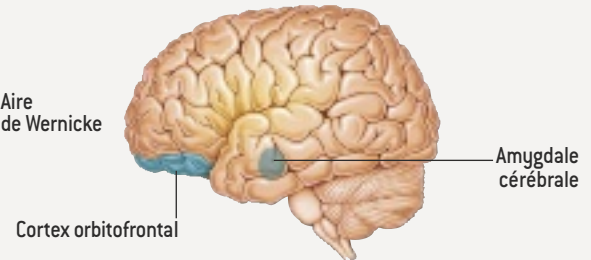
3 Battre la mesure avec le pied nécessite une synchronisation temporelle et implique le cervelet et les cortex moteur et frontal.



4 Inventer une musique, par exemple en chantant, met en jeu certaines régions situées dans les cortex frontal et temporal.



5 Écouter une musique et traiter ses structures impliquent des régions qui participent aussi au langage, telles les aires de Broca et de Wernicke, ainsi que d'autres régions du cortex temporal.



6 Les émotions ressenties à l'écoute musicale activent les structures participant aux émotions, tels l'amygdale cérébrale et le cortex orbitofrontal.

✓ BIBLIOGRAPHIE

F. Marmel, B. Tillmann et W. Dowling, *Tonal expectations influence pitch perception*, Perception & Psychophysics, vol. 70, pp. 841-852, 2008.

A. Patel, *Music, language and the brain*, Oxford University Press, 2008.

E. Hannon et S. Trehub, *Metrical categories in infancy and adulthood*, Psychological Science, vol. 16, pp. 48-55, 2005.

I. Peretz et K. Hyde, *What is specific to music processing? Insights from congenital amusia*, Trends in Cognitive Science, vol. 7, pp. 362-367, 2003.

S. Trehub, *The developmental origins of musicality*, Nature Neuroscience, vol. 6, pp. 669-673, 2003.

En psychologie cognitive, on a mis en évidence une capacité qui permet d'acquérir des connaissances sur des informations complexes par simple exposition, sans intention d'apprendre. Ce type d'apprentissage est qualifié d'implicite : étant exposé à des matériaux structurés, le cerveau extrait des régularités et devient sensible aux structures sans connaissances explicites. C'est ainsi que l'enfant acquiert des connaissances sur sa langue maternelle en étant exposé aux flots de paroles de son environnement. L'enfant (avant sa scolarisation et les cours de grammaire) ne peut pas expliquer des structures, des régularités ou des règles de la langue, mais il les comprend et peut deviner la fin d'une phrase, détecter des fautes de grammaire ou des irrégularités de structures.

De même, la capacité cognitive d'apprentissage implicite permet aux auditeurs

d'acquérir des connaissances sur le système musical de leur culture, notamment dans la vie quotidienne, par simple exposition à des pièces musicales (les berceuses, la musique à la radio, etc.). Et ce, sans formation musicale explicite. L'auditeur est face à la musique comme l'enfant est face à sa langue maternelle : il traite les structures et développe des attentes, sans être capable de les expliciter. On parle d'acculturation musicale : l'auditeur est un expert implicite de la perception de la musique de sa culture.

La plupart des travaux sur la cognition musicale se sont intéressés à la perception de la musique occidentale tonale par l'auditeur occidental et à ses principales régularités qui s'appliquent à une variété de styles musicaux (musique classique, pop, folk, jazz, etc.).

On sait qu'une même note peut remplir différentes fonctions musicales selon

son contexte d'utilisation. Par conséquent, une note peut être adéquate pour finir une mélodie donnée, mais pas pour en terminer une autre. Pendant l'écoute, un auditeur acculturé développe donc des attentes perceptives sur les notes futures qui diffèrent selon le début de la mélodie.

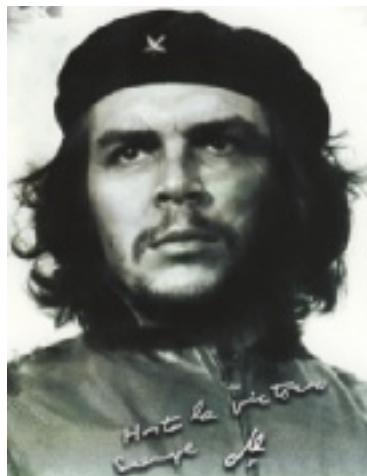
Ces attentes musicales sont semblables aux attentes face au langage: un même mot, *neige* par exemple, est plus attendu après *Le skieur glisse sur la* qu'après *Le chauffeur conduit sur la*. Les connaissances sur la langue française, les relations sémantiques entre les mots et leur fréquence d'association permettent de développer des attentes sur la suite d'une phrase; et ces attentes sont plus fortes pour un mot plus probable. Ainsi, lors d'une expérience de psychologie cognitive, un participant perçoit plus vite *neige* dans la première phrase que dans la seconde. Plus un événement est attendu, plus on le traite rapidement.

Cette expérience étudiant les connaissances de l'auditeur sur le langage a été transposée à l'étude des connaissances de la musique. On a proposé à des participants deux mélodies qui ne différaient que par une seule note (voir la figure 2). Musicalement, ce changement d'un demi-ton – une note proche en hauteur – laisse inchangé le contour mélodique, mais modifie la tonalité installée, de sorte que la dernière note de la mélodie est adéquate en note finale pour la première mélodie, mais ne l'est pas pour la seconde; cette dernière note prend la fonction de la tonique (la note la plus importante de la tonalité) dans la première mélodie, mais la fonction d'une sous-dominante (une fonction moins importante) dans la seconde.

Un auditeur non musicien, sans formation musicale explicite, fait-il cette différence, même si acoustiquement les deux mélodies sont presque identiques? Oui. Il traite plus rapidement la dernière note de la première mélodie. Comme il s'agit de la même information acoustique, cela suggère que l'auditeur a des connaissances sur le système musical et l'utilisation des notes, ce qui lui permet de distinguer les deux mélodies; en d'autres termes, les débuts des mélodies – qui diffèrent par une note – activent différemment les connaissances musicales et engendrent des attentes perceptives différentes qui influent sur la perception de la dernière note. Voilà une démonstration de l'acculturation tonale des auditeurs

DES INVARIANTS COGNITIFS

- ✓ Une préférence pour des notes proches en hauteur.
- ✓ Un traitement des sons par catégories.
- ✓ Une mémoire à court terme de cinq à neuf éléments.
- ✓ L'information relative (le contour mélodique et le rythme) est privilégiée.
- ✓ On associe la musique à d'autres comportements (la danse par exemple).
- ✓ L'exposition à la musique et son apprentissage implicite, sans intention d'apprendre, engendrent des connaissances et des attentes musicales.



1. CHE GUEVARA était atteint d'amusie congénitale. Ce déficit musical serait d'origine génétique, mais ses fondements neuronaux commencent seulement à être étudiés. Les personnes atteintes sont incapables de reconnaître une chanson sans paroles, ni de détecter une fausse note.

occidentaux *via* une exposition à la musique occidentale tonale.

Quelques études avec des musiques et des auditeurs d'autres cultures permettent de soutenir que l'acculturation par simple exposition est un invariant cognitif. Par exemple, en comparant la perception de la musique indienne par des auditeurs indiens – c'est-à-dire acculturés – et par des auditeurs américains – dits naïfs, c'est-à-dire n'ayant pas été exposés à ce type de musique –, on constate que les deux groupes sont sensibles aux caractéristiques acoustiques des notes entendues (leur durée et leur fréquence d'apparition), mais seuls les auditeurs indiens perçoivent les différences d'organisation fonctionnelle des notes (il existe différents sous-ensembles de notes utilisés dans les gammes).

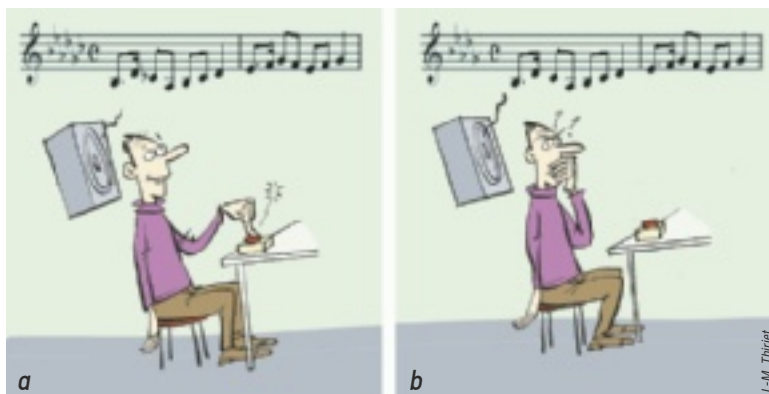
Apprendre en écoutant

On a obtenu des résultats similaires pour la perception de la musique arabe improvisée: les auditeurs européens (naïfs) et arabes (acculturés) sont sensibles à certains indices acoustiques, tels des pauses (aucun son n'est émis) ou des changements de registres (on passe de notes graves à des aiguës), mais seuls les auditeurs arabes perçoivent des changements subtils de modes (des modifications de gamme par exemple).

Un autre exemple porte sur la perception des structures temporelles. Plusieurs études avaient montré que les auditeurs occidentaux préfèrent les rythmes avec des mesures simples (par exemple, à deux ou trois temps). On pensait alors que les mesures plus complexes nécessitaient davantage de ressources cognitives.

En 2005, Erin Hannon et Sandra Trehub, de l'Université de Toronto, bousculent cette interprétation en suggérant l'importance de l'acculturation musicale, même pour les structures temporelles. Ils montrent que les auditeurs américains perçoivent facilement les mesures simples pour lesquelles ils sont acculturés, mais qu'ils rencontrent des difficultés pour des mesures complexes. En revanche, des auditeurs bulgares et macédoniens sont sensibles aux deux types de mesures: ils sont acculturés pour les deux, car la musique folklorique des Balkans contient des mesures complexes.

De plus, E. Hannon et S. Trehub ont montré que des bébés américains de



2. CES DEUX MÉLODIES se différencient par une seule note : la troisième. Musicalement, ce changement d'un demi-ton – une note proche en hauteur – ne modifie pas le contour mélodique, mais change la tonalité, de sorte que la dernière note est adéquate en ce qui concerne la note finale pour la mélodie *a* alors qu'elle ne l'est pas pour la mélodie *b*. Un auditeur, même non musicien, entendant les deux débuts des mélodies ne réagit pas de la même façon : il développe des attentes plus fortes pour la dernière note de la mélodie *a* que pour celle de la *b*. Ainsi, la perception de la note finale est facilitée pour la mélodie *a* : lors d'une expérience, quand on demande au sujet si la note finale est adéquate ou non, il appuie plus vite sur le bouton de réponse pour la mélodie *a* que pour la mélodie *b*. Les connaissances musicales de l'auditeur influent sur sa perception de la dernière note d'une mélodie.

six mois sont sensibles aux deux types de mesures ; cependant, dès l'âge de 12 mois, l'acculturation s'est mise en place et les bébés américains obtiennent les mêmes résultats que les adultes américains. Les bébés pourraient donc apprendre différentes structures musicales (ici temporelles), mais après une période restreinte, ils se spécialiseraient aux caractéristiques musicales de leur culture. Notons que cette période nécessaire à l'acculturation musicale correspond à celle qui permet à un bébé de se spécialiser pour la perception des sons de sa langue maternelle.

Il existe donc des invariants cognitifs liés à la perception de la musique : les auditeurs apprennent des informations sur le système musical de leur culture par simple exposition – ils reconnaissent les notes utilisées, leur combinaison en gammes, leurs organisations –, et ces connaissances musicales implicites influent sur le traitement des structures musicales, par exemple par la formation d'attentes perceptives. Alors quelles sont les aires cérébrales dédiées au traitement des structures musicales ? Existe-t-il des régions cérébrales et des capacités cognitives spécifiques à la musique ?

De nombreuses études sur les fondements biologiques de la musique ont montré un recouvrement des réseaux neuronaux impliqués dans le traitement de la musique et du langage (voir l'encadré page 128). Par exemple, certaines études ont appliqué à la musique des méthodes expérimentales utilisées pour le langage, notamment en introduisant un événement inattendu dans une séquence musicale et en comparant la réaction du cerveau à cette violation de structure avec sa réaction face à un événement qui respecte les structures musicales. Ainsi, par électroencéphalographie – une technique qui mesure l'activité électrique des neurones sur le scalp et qui a une très bonne résolution temporelle –, on a montré que le cerveau réagit rapidement après une violation musicale (en 200 millisecondes après le début du son). Cette réaction est semblable à celle observée après la violation d'une structure syntaxique dans des phrases.

En outre, en imagerie cérébrale fonctionnelle par résonance magnétique – une méthode de mesure

de l'activité des neurones dans des régions cérébrales précises –, on a mis en évidence l'importance du cortex frontal inférieur (qui comprend l'aire de Broca et son homologue dans l'hémisphère droit) dans la musique. L'aire de Broca n'est donc pas spécialisée dans le traitement du langage : sa fonction est plus générale. Elle serait impliquée dans les mécanismes nécessaires à l'intégration structurale des informations, notamment au cours du temps – qu'il s'agisse des notes d'une mélodie ou des mots d'une phrase. En 2003, Aniruddh Patel, de l'Institut de neurosciences à San Diego aux États-Unis, émet l'hypothèse que des réseaux neuronaux pour le traitement des informations structurales ou syntaxiques sont partagés entre la musique et le langage.

Des régions cérébrales spécifiques de la musique ?

Toutefois, Isabelle Peretz, de l'Université de Montréal, et d'autres chercheurs défendent l'hypothèse que certains réseaux neuronaux sont spécifiques à la musique et séparés de ceux du langage. Notamment, certains patients souffrant de lésions cérébrales – après un accident cérébral vasculaire ou un traumatisme – présentent des troubles du langage (nommés aphasies) sans troubles du traitement de la musique. Et d'autres patients ont les symptômes inverses : ils ne rencontrent aucune difficulté avec le langage, mais montrent des troubles de traitement de la musique (nommés amusie).

Un dysfonctionnement de la perception musicale permettra peut-être d'en apprendre davantage. Pour quelques personnes, la musique ne fait pas sens, elle sonne comme une langue étrangère – voire comme du bruit. Ces personnes sont atteintes d'amusie congénitale, un déficit musical qui ne s'explique pas par des lésions cérébrales, des pertes auditives, des déficits cognitifs, intellectuels et sociaux ou un manque d'exposition à la musique. Les personnes en souffrant comprennent normalement la parole et reconnaissent des voix et des sons de l'environnement. En revanche, elles sont

incapables de reconnaître une mélodie familière sans paroles ni de détecter une fausse note ou quelqu'un qui chante faux (y compris elles-mêmes).

On estime qu'environ quatre pour cent de la population souffrent d'amusie congénitale (le révolutionnaire argentin Che Guevara en était atteint, voir la figure 1), mais il est probable que la proportion réelle soit supérieure, car il est socialement peu acceptable de ne pas aimer la musique, donc certaines personnes n'avoueraient pas leur handicap musical. Ce dernier contraste avec les capacités musicales des bébés et les connaissances musicales des adultes non musiciens.

Les fondements neuronaux de l'amusie congénitale commencent seulement à être étudiés. Le déficit musical serait présent dès la naissance (par opposition à l'amusie acquise liée à des lésions cérébrales) et des études récentes suggèrent même qu'il se transmet génétiquement. L'étude de l'amusie congénitale améliorera notre compréhension des fondements comportementaux, neuronaux et génétiques de la cognition musicale. Toutefois, à l'heure actuelle, les corrélats neuronaux spécifiques de la musique restent l'objet des recherches en neurosciences.

D'autres universaux

En comparant différentes cultures musicales, on a trouvé des universaux en musique, telle l'utilisation d'un nombre limité de notes ou la façon dont elles sont organisées. Ces caractéristiques nous informent sur des contraintes perceptives et cognitives plus générales qui ont conduit à la construction de ces systèmes musicaux. En étudiant l'auditeur exposé à la musique de sa culture, on a ainsi constaté que les traitements cognitifs impliqués dans l'acculturation et la perception musicales sont communs au traitement d'autres structures, tel le langage.

Mais il existe d'autres universaux en musique : la présence des berceuses, l'émergence des émotions et l'association entre musique et mouvement (notamment dans la danse). Par exemple, il est possible de reconnaître une berceuse d'une autre culture (sans connaître les mots) ou certaines émotions dans des pièces musicales d'une autre culture. Ces invariants seraient communs au langage et à la voix. Les berceuses, à savoir la musique destinée aux bébés, mettent en œuvre des caractéristiques acoustiques comparables à celles utilisées quand on parle à un bébé : par exemple, un contour mélodique simple, l'utilisation de répétitions, une variété limitée de hauteurs. De même, les émotions suscitées par la musique ont des caractéristiques retrouvées dans la voix : par exemple, l'expression de la joie est souvent associée à un tempo rapide et à une large variété de hauteurs. Toutefois, certaines émotions, évoquées par les structures spécifiques d'un système musical, renvoient à l'invariant cognitif de l'acculturation musicale ■



L'Année Internationale de la Planète Terre

« Les géosciences au service de l'Humanité »

En France, plus de **427 projets** (expositions, animations scolaires, colloques, conférences, ouvrages, films documentaires, découvertes du territoire, excursions géologiques, sites Internet...) concernant l'ensemble des géosciences ont été soumis à labellisation dans le cadre de l'Année internationale.

En NOVEMBRE, un riche programme est proposé, incluant notamment plusieurs réunions scientifiques :




- Cycle de conférences sur **Les bassins sédimentaires français**, organisé par l'Union des Professeurs Agrégés, Paris, 3 et 4 novembre ;
- Cycle de conférences organisé par la Société Géologique de France : **Evolution ? Evolution !** Bordeaux, 7 novembre ;
- Colloque : **Enseigner l'évolution**, organisé le ministère de l'Éducation Nationale, à la **Cité de Sciences, au MNHN et au Collège de France, Paris**, 13 et 14 novembre ;
- Colloque : **L'eau dans tous ses états**, organisé par le Comité National Français de Géodésie et de Géophysique (CNFGG), **Unesco, Paris**, 17 au 19 novembre ;
- Colloque : **Sol, la vie de la Terre**, organisé par l'INRA, au **Muséum d'Orléans**, 21 novembre.

La Fête de la Science 2008 constituera un temps fort de l'Année Internationale de la Planète Terre en France (17 - 23 novembre). De nombreuses expositions, animations scolaires et visites sur le terrain sont organisées à cette occasion.



Plusieurs expositions ouvrent en novembre :

- **Mémoires de pierres**, Gap
- **L'or bleu, les sciences de la mer** (plusieurs expositions simultanées et nombreuses conférences), **Issy-les-Moulineaux**, 8 au 30 novembre ;
- **Le cycle naturel de l'eau**, **Villiers-Saint-Georges** ;
- **Environnement géologique des futurs citoyens**, **Montpellier**.

L'agenda détaillé des activités liées à l'Année et tous les renseignements utiles sur le site : www.anneeplaneteterre.com

Les émotions musicales sont reconnues instantanément, aussi vite qu'un signal de danger. Ainsi, la musique a une valeur adaptative : elle favorise notamment la cohésion sociale.

Emmanuel Bigand



Les émotions

La musique est la langue des émotions.
Emmanuel Kant

Qui n'a ressenti des frissons en écoutant le *Requiem* de Mozart ou *La jeune fille et la mort* de Schubert ? La musique exerce un effet profond sur l'être humain bien au-delà des sphères restreintes des mélomanes cultivés. Si l'on en croit de récentes statistiques économiques, la musique représenterait l'un des marchés les plus développés, bien avant l'industrie pharmaceutique. Les pratiques musicales ont une importance notable tant dans les sociétés industrialisées que dans les sociétés non occidentales. Ainsi, dans certaines cultures traditionnelles, on consacre plus de temps aux activités musicales qu'à celles dont dépend la tribu, la chasse notamment. Pourquoi la musique a-t-elle un tel impact ? C'est aujourd'hui l'objet d'études scientifiques novatrices en neurosciences et en psychologie cognitive.

Les processus cognitifs impliqués dans la perception musicale (comment le cerveau traite-t-il les informations musicales ?) sont étudiés depuis longtemps, alors que les réponses affectives à la musique (comment la musique déclenche-t-elle des émotions ?) ont été négligées jusqu'à présent. Pourtant, si nous écoutons de la musique, ce n'est pas pour le plaisir d'entendre des structures sonores bien construites, que notre cerveau redé-

ploierait lors de l'écoute, même si cette conception très formaliste correspond sans doute à une réalité pour certains mélomanes érudits. Pour beaucoup de compositeurs et pour la plupart des auditeurs, le propre de la musique est d'être expressive. La musique renvoie à autre chose qu'aux sons et aux architectures sonores qui la composent : elle nous plonge dans un état psychologique et physiologique spécifique, qui ne se confond pas avec l'excitation sensorielle produite par les signaux acoustiques et qui se différencie clairement de l'état psychologique déclenché par les autres stimulations sonores de l'environnement.

Un analyseur sujet aux émotions

L'étude de la perception de la musique auprès des enfants malentendants en est une démonstration. Lorsque le signal acoustique, tout appauvri qu'il soit par la surdité, est perçu comme une production musicale, l'émotion devient manifeste dans le regard de celui qui l'écoute. Ce n'est d'ailleurs pas un hasard si le second souhait le plus fréquemment exprimé par les personnes atteintes d'une surdité profonde est de pouvoir retrouver la perception de la musique (qui vient juste après celui de mieux comprendre le langage). La musique ouvre sur un monde

L'ESSENTIEL

✓ La musique déclenche un état psychologique et un état physiologique que n'entraînent pas les sons non musicaux.

✓ On reconnaît qu'une musique est triste ou gaie aussi vite que l'on identifie un danger.

✓ Toutes les émotions musicales sont constituées à des degrés divers des quatre émotions fondamentales : la colère, la sérénité, le désespoir et la gaieté.



© Getty Images/John Lund, Paula Zacharias

musicales

sensible où émotions, expressions et sentiments se côtoient.

Les sciences des activités mentales ont longtemps dominé les sciences de l'affectivité, y compris dans le domaine de la musique. À l'image d'un ordinateur qui traiterait froidement les informations du monde extérieur, l'auditeur est le plus souvent présenté comme un « analyseur » de signaux acoustiques mettant en œuvre des méthodes précises. On ne songe pas à s'intéresser à ses réactions sensibles ni à l'influence de ces réactions sur sa façon d'écouter la musique.

Il est vrai qu'il est plus facile de trouver des indicateurs comportementaux et neurophysiologiques des traitements de l'information qu'effectue le cerveau que de trouver des paramètres associés aux réactions émotionnelles. Ainsi, on mesure des temps de réponse, on analyse des mouvements oculaires, on enregistre une activité cérébrale, mais on n'a pas identifié de paramètres fiables de mesure des réactions émotionnelles. Autrement dit, en quoi les paramètres « objectifs » nous renseignent-ils sur l'état émotionnel, sur le ressenti de l'auditeur ?

Ces difficultés n'ont pas empêché la psychologie des émotions de se développer dans de nombreux domaines, mais elles ont longtemps paru insurmontables dans le cas de la musique. Contrairement à certains sons qui peuvent évoquer un

danger (un sifflement de serpent, un aboiement de chien, un bruit de pas dans la nuit, etc.), la musique n'a pas de conséquences immédiates : la survie d'un individu ne dépend pas de sa sensibilité à la musique. Nous écoutons de la musique pour le plaisir qu'elle nous procure, mais ce plaisir est libre de prendre des formes variées, lesquelles dépendent seulement du vécu de l'auditeur et de son état au moment où il écoute de la musique.

Aujourd'hui, les émotions musicales sont au cœur de nombreuses recherches, et les résultats obtenus sont souvent surprenants et contribuent à mieux comprendre la place de la musique dans les sociétés humaines. Plusieurs idées reçues ont ainsi été remises en question. C'est par exemple le cas de celle voulant que les réponses émotionnelles à la musique soient uniquement subjectives. Certes, les émotions suscitées par certaines œuvres peuvent être déterminées par des éléments non musicaux, par exemple des événements associés « accidentellement » à l'œuvre. Ainsi, telle œuvre évoquera la tristesse pour une personne, car elle est associée à la disparition d'un être cher, et à la gaieté pour une autre parce qu'elle évoque la rencontre de l'être aimé. Ces associations émotionnelles extramusicales existent, mais ne représentent qu'une part minime de nos expériences.

Si l'émotion dépendait uniquement des contextes d'écoute, une œuvre donnée évoquerait autant d'expressions différentes qu'il y a d'auditeurs et chacun en aurait une expérience particulière. Lors d'un concert, son voisin de droite pleurerait, tandis que celui de gauche se délecterait joyeusement de la soirée. Or il est évident que cela ne se produit jamais. Les œuvres musicales ont une structure expressive suffisamment puissante pour imposer des états émotionnels communs à un grand nombre d'auditeurs. La musique peut mettre à l'unisson émotionnel une foule entière.

Ce pouvoir lui confère une force de cohésion sociale essentielle dans la plupart des cultures du monde. Il s'exerce déjà chez le nourrisson par l'intermédiaire des comptines qui lui sont chantées. Le bébé est d'ailleurs plus fasciné par la voix de sa mère quand elle chante que quand elle parle. Le choix approprié des comptines permet de moduler ses états émotionnels, et il existe manifestement des universaux expressifs puisque les comptines du monde entier partagent de nombreux traits structuraux.

Un facteur de cohésion sociale

La fonction de cohésion sociale s'exerce ensuite tout au long de la vie, et plus particulièrement au moment de l'adolescence. À ce stade, la musique traduit les états émotionnels traversés par les adolescents, ce qui facilite les regroupements par affinité. Les réponses émotionnelles à la musique sont également plus stables que l'on pouvait imaginer pour un auditeur donné. L'émotion que procure une œuvre diffère sans doute d'un jour à l'autre, en fonction de l'humeur et du contexte, et elle évolue tout au long de la vie. Une telle variété est souhaitable, sinon l'expérience musicale deviendrait très vite répétitive.

Toutefois, ces variations restent centrées autour d'une même expérience émotionnelle. S'il n'en était ainsi, choisir un disque dans sa discothèque relèverait plus du jeu de hasard que d'un choix volontaire. Or il est rare que nous fassions de nombreux essais avant de trouver le morceau qui correspond à l'émotion recherchée. Lorsque nous connaissons bien une base de données musicales, nous savons très précisément quel type de musique s'ajuste le mieux à l'état psychologique du

moment. Ce savoir n'est possible que dans la mesure où les émotions musicales obéissent à des régularités.

Les recherches ont confirmé la stabilité des réponses émotionnelles. Lorsque l'on utilise des œuvres connues et bien caractérisées du point de vue expressif qui évoquent des émotions de gaieté (le *Printemps* des *Quatre saisons* de Vivaldi), de colère ou de peur (*La Nuit sur le mont chauve* de Moussorgski), de tristesse (*l'Adagio* d'Albinoni) et dans une moindre mesure de sérénité, les réponses sont très reproductibles d'un auditeur à l'autre. Cette régularité est mise en évidence lors d'études où l'on demande à des auditeurs d'écouter des pièces qu'ils ne connaissent pas et d'exprimer leurs émotions.

Des réactions émotionnelles stables

Dans l'une de ces expériences, nous avons demandé à des auditeurs musiciens et non musiciens d'écouter 27 extraits musicaux et de regrouper ceux qui exprimaient des émotions similaires. On ne leur demandait pas d'exprimer un jugement verbal. L'analyse de ces groupements permet de définir une distance émotionnelle entre les œuvres : si deux œuvres ne sont jamais regroupées, c'est qu'elles déclenchent des émotions très différentes, et que leur distance émotionnelle est importante.

Qui plus est, dans ce type d'expériences, les distances émotionnelles évaluées à partir des réponses sont quasi identiques lorsque les auditeurs refont l'expérience trois semaines plus tard. Plus surprenant encore, ces distances sont très similaires pour des groupes d'auditeurs différents et elles ne changent presque pas avec l'expertise musicale (que le sujet soit familier ou non de la musique). Suzanne Filipic et Philippe Lalitte, dans notre laboratoire, ont récemment exposé des auditeurs n'ayant pas de formation musicale et des instrumentistes spécialistes de musique contemporaine à des œuvres de ce type de musique. Ils ont comparé leurs réactions émotionnelles (tristesse, gaieté, anxiété, etc.). Malgré la difficulté stylistique de ces œuvres, ils ont observé des résultats tout à fait comparables : les émotions sont similaires.

Les émotions perçues sont-elles réellement vécues ? Autrement dit, les auditeurs identifiaient bien les émotions exprimées par les œuvres les ressentent-



1. L'INFLUENCE DE LA MUSIQUE est évaluée au moyen d'une expérience où l'on commence par faire entendre à un sujet une musique soit triste (a), soit gaie (c). On présente ensuite au sujet un idéogramme qui n'a pas de sens pour lui et dont on pourrait attendre qu'il le qualifie de neutre, n'évoquant pas d'émotion particulière. On demande au sujet d'appuyer sur une touche indiquant s'il trouve le caractère présenté sur l'écran *Triste*, *Gai* ou *Neutre*. Or après avoir entendu une musique triste, le sujet indique qu'il trouve l'idéogramme triste (b) et après une musique gaie, il le trouve gai (d). La musique influence sur notre état d'esprit.

ils vraiment ? Les auditeurs pourraient identifier les mêmes émotions sans pour autant les ressentir de la même façon. Cette différence entre émotions « perçues » et émotion « vécue » reste un sujet de débat. Pour certains, on peut reconnaître le caractère triste d'une musique que l'on écoute sans devenir triste ni même ressentir la moindre tristesse. De même, il est possible de voir des personnes tristes, tout en étant gai soi-même.

Émotions perçues, émotions vécues

Toutefois, de tels décalages – de telles dissociations – ont leurs limites, c'est-à-dire qu'ils sont normalement de courte durée, sauf dans des cas pathologiques ou dans un contexte expérimental. Antonio Damasio, à l'Université de Californie, à Los Angeles, a montré que les patients qui ne sont plus capables d'identifier les émotions à la suite d'une lésion cérébrale ne ressentent en général plus d'émotions non plus.

Par ailleurs, Paula Niedenthal et ses collègues de l'Université Blaise Pascal, à Clermont-Ferrand, ont empêché des sujets de ressentir l'émotion qu'ils devaient identifier en utilisant un artifice expérimental approprié : ils forçaient le sujet à rire (en projetant, par exemple, une image humoristique) alors même qu'ils lui présentaient un visage triste et lui demandaient de préciser quelle émotion traduisait le visage. Ils ont ainsi montré que le sujet éprouve plus de difficultés à identifier l'émotion du visage lorsqu'il est perturbé. Dans un contexte naturel, l'émotion identifiée dans une œuvre ne reste pas longtemps dissociée de l'émotion ressentie : si l'écoute d'une musique triste pendant quelques minutes ne nous rend pas tristes, celle de musiques sinistres pendant une heure influence sur notre état d'esprit, même si nous étions joyeux au début.

On peut donc admettre que l'émotion identifiée est liée à l'émotion ressentie. Une expérience le met en évidence : on demande à des sujets d'évaluer le caractère émotionnel de stimulus non musicaux, qui n'évoquent pas d'émotions particulières (tels que des lettres chinoises sans signification pour un sujet occidental). On fait précéder la projection du stimulus par un extrait musical auquel le sujet n'est pas invité à prêter attention, mais qui est soit gai,

soit triste. On demande ensuite au sujet de dire s'il trouve le dessin projeté (théoriquement neutre) gai, triste ou neutre. On constate que le caractère émotionnel attribué au stimulus neutre est influencé par l'émotion déclenchée par le morceau qui précède (voir la figure 1). Un tel effet est qualifié d'amorçage affectif et suggère que la musique modifie l'état affectif du sujet, ce qui le conduit à projeter l'émotion musicale ressentie sur le stimulus neutre. Dans ce type d'études, l'émotion musicale influe sur le comportement du sujet sans qu'il n'ait à exprimer ce qu'il a ressenti.

L'analyse des réponses physiologiques à la musique constitue un moyen supplémentaire pour s'assurer que les émotions musicales sont bien « vécues ». Nos émotions sont suffisamment fortes pour entraîner de nombreuses modifications physiologiques, telles que le rythme cardiaque, le rythme respiratoire ou encore la conductance de la peau (une émotion fait transpirer, ce qui modifie la capacité que présente la peau de conduire un infime courant électrique). Le frisson dans le dos ou la chair de poule mentionné au début seraient une traduction physiologique spécifique (mais non exclusive) de l'émotion musicale.

Les aires cérébrales des émotions musicales

L'imagerie par résonance magnétique fonctionnelle, IRMf, confirme l'implication des réseaux neuronaux émotionnels quand on écoute de la musique. Ann Blood et Robert Zatorre, de l'Université McGill, à Montréal, au Canada, ont demandé à des auditeurs volontaires d'écouter dans un scanner les musiques qui leur procurent habituellement de fortes émotions. Ils ont constaté que la musique active les mêmes régions cérébrales que les stimulus ayant une forte implication biologique, tels que la nourriture et les stimulations sexuelles (et aussi certaines drogues).

Les résultats obtenus en imagerie cérébrale suggèrent que des liens anatomiques et fonctionnels se sont créés entre les systèmes cérébraux anciens (liés aux émotions) et les régions corticales (plus récentes et liées au raisonnement et autres processus cognitifs supérieurs). C'est ce qui nous permet de ressentir des émotions en pré-

sence de stimulus abstraits et culturels, notamment la musique. Les stimulations les plus intenses sont provoquées par des musiques généralement qualifiées de tristes ou mélancoliques.

Eckart Altenmüller, de l'Institut de musique de Hanovre, a cherché à identifier la signature électrophysiologique des « frissons dans le dos ». En présentant des extraits du *Requiem* de Mozart à des auditeurs allongés dans un scanner, il a associé cette émotion à des passages précis de l'œuvre. L'analyse des indices électrophysiologiques montre que la musique ne provoque pas simplement des sentiments abstraits, mais qu'elle déclenche des changements d'activité au cœur même du cerveau.

Certaines émotions semblent plus fréquemment associées à des modifications physiologiques spécifiques. Selon Carol Krumhansl, de l'Université Cornell, aux États-Unis, la gaieté entraînerait une accélération du rythme respiratoire et une respiration plus profonde, alors que la tristesse se manifesterait par des changements du rythme cardiaque, par une augmentation de la pression sanguine et une diminution de la conductance de la peau.

D'autres études d'imagerie cérébrale suggèrent que les deux hémisphères ne contribueraient pas de façon identique aux émotions musicales : l'hémisphère gauche semble plus actif lors de l'écoute de musique gaie, et l'hémisphère droit, lors de l'écoute de musique triste. Toutefois, même si l'imagerie cérébrale nous permet de localiser les aires impliquées dans l'émotion musicale,



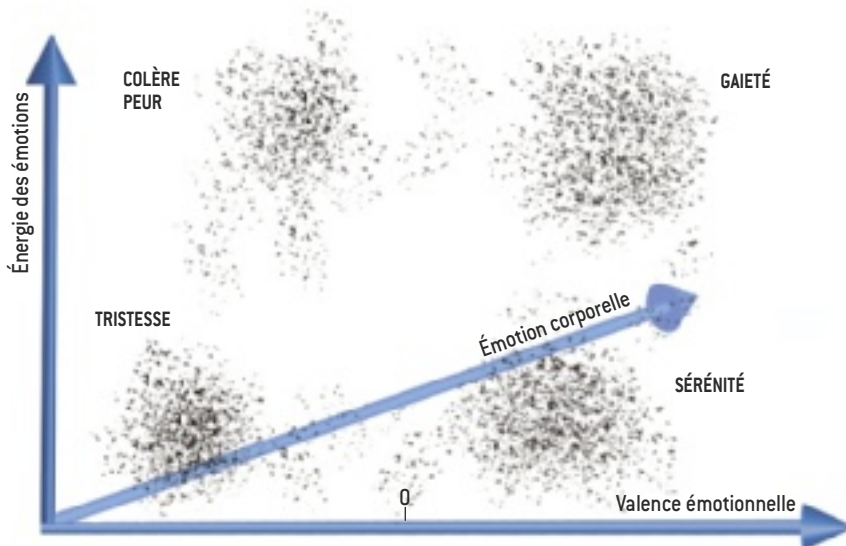
Robbie Jacky Corbis

L'AUTEUR



Emmanuel BIGAND, professeur de psychologie cognitive, membre de l'Institut universitaire de France, dirige le Laboratoire d'étude de l'apprentissage et du développement, UMR 5022, à l'Université de Bourgogne, à Dijon.

2. LES CHANTEURS D'OPÉRA accentuent les émotions véhiculées par la musique en « jouant » la comédie. Costumes, mise en scène, expressions et jeux des comédiens permettent aux spectateurs de percevoir encore mieux si l'émotion associée à la musique est la gaieté, la tristesse ou l'inquiétude. Dans cette scène de la *Flûte enchantée* de Mozart, la mystérieuse Reine de la nuit.



3. DES EXTRAITS MUSICAUX sont diffusés à des sujets qui doivent dire s'ils les trouvent gais, tristes, sereins ou inquiétants. L'analyse des résultats montre qu'on peut les décomposer selon trois paramètres : la valence émotionnelle, l'émotion corporelle (ou corporalité) et l'énergie. On constate que les extraits sont regroupés en quatre zones que l'on associe à la tristesse (valence négative, corporalité faible et énergie faible), la sérénité (valence positive, corporalité faible et énergie faible), la gaieté (valence positive, corporalité élevée et énergie élevée) et la colère ou la peur (valence négative, corporalité élevée et énergie élevée).

et la chorégraphie de leur activation, cela ne nous suffit pas pour comprendre comment naît une expérience émotionnelle et en quoi elle consiste. Dès lors, il faut avoir recours à la psychologie expérimentale.

Quatre grandes catégories d'émotions semblent prédominer en musique : la gaieté, la colère (ou la peur), la tristesse et la sérénité, qui seraient identifiées sans difficulté dès l'âge de cinq ans. La musique peut-elle déclencher des émotions plus subtiles ? Tout mélomane sera tenté de répondre oui, et l'on pourrait même penser que la musique suscite des émotions spécifiques – mais cela n'est pas démontré. En 2005, en collaboration avec l'équipe de Stephen McAdams de l'IRCAM, nous avons voulu tester diverses méthodes pour approfondir cette question. Une des méthodes, déjà évoquée, consiste à présenter sur un écran d'ordinateur un ensemble de 27 extraits musicaux, suffisamment longs pour installer un climat expressif précis, et suffisamment courts pour que ce climat reste stable tout au long de l'extrait.

Ces extraits étaient choisis en collaboration avec le musicologue François Madurell, de l'Université Paris IV-Sorbonne, pour représenter une large variété de styles musicaux, de formations instrumentales et d'expressions. Le dispositif expérimental permettait aux sujets d'écouter ces extraits autant de fois qu'ils le voulaient ; ils devaient ensuite grouper par deux, trois ou davantage ceux qui, selon

eux, évoquaient une même émotion. Une fois réalisée avec un nombre suffisant de sujets, l'expérience fournissait une matrice de co-occurrence, qui indiquait combien de fois tel ou tel morceau avait déclenché la même émotion. C'est une matrice dite de proximité émotionnelle des extraits.

Les trois coordonnées des émotions

Cette démarche est facile à réaliser, plaisante pour les sujets, et n'exige pas de formuler ses émotions. Elle permet d'identifier les dimensions psychologiques qui organisent les distances émotionnelles entre les différents extraits. Si, par exemple, la matrice contenait les distances en kilomètres séparant différentes villes de France (à la place des morceaux de musique), l'analyse statistique permettrait d'extraire les dimensions physiques qui structurent l'espace des distances entre les villes. Trois dimensions seraient nécessaires, correspondant aux axes Nord-Sud, Est-Ouest, et à l'altitude. De même, l'analyse des matrices émotionnelles permet d'isoler les dimensions psychologiques qui structurent l'espace de nos émotions musicales. Trois dimensions principales ont été observées : l'axe de l'énergie des émotions ressenties, l'axe de valence émotionnelle et l'axe de la dynamique.

L'axe de l'énergie des émotions ressenties va des émotions de grande énergie, intenses (gaieté, colère) à des émotions de faible énergie, reposantes (dépression, sérénité). Le second axe oppose des émotions positives (gaieté ou sérénité) aux émotions négatives (colère ou désespoir). Cet axe ne retrace pas le caractère agréable – l'agréabilité –, généralement rapporté dans les études sur les émotions. En effet, l'une des spécificités de la musique est de pouvoir être jugée plaisante même si l'émotion induite est triste, c'est-à-dire même si elle a une valence négative. Enfin, nous avons pu dégager un dernier axe qui est lié au mouvement corporel induit par la musique, l'axe de l'émotion corporelle ou corporalité.

Comme l'avait noté le psychologue Robert Francès dès 1956, la musique réveille des schémas sensorimoteurs qui renforcent l'expérience affective du sujet : il y a une interaction entre les schémas purement cognitifs, l'affect et le corps. L'émotion ressentie par l'auditeur active des schémas moteurs – des mouvements – mémorisés qui, en retour, agissent sur la

✓ BIBLIOGRAPHIE

E. Bigand et B. Poulin-Charronat, *Musical Pitch, Handbook of Music Cognition*, sous la direction de I. Cross, Oxford University Press, à paraître.

E. Bigand et S. Filipic, *Cognition et émotions musicales, Intellectica*, vol. 48-49, pp. 37-50, 2008.

E. Bigand, *Musiciens et non-musiciens perçoivent-ils la musique différemment ? Le Cerveau Musicien*, sous la direction de B. Lechevalier, F. Eustache et H. Platel, De Boeck, 2006.

E. Bigand et al., *Multidimensional scaling of emotional responses to music: the effect of musical expertise and excerpts « duration »*, *Cognition & Emotion*, vol. 8, pp. 1113-1139, 2005.

S. McAdams et E. Bigand, *Penser les sons*, PUF, 1994.

perception de la musique écoutée. Autrement dit, cette troisième dimension indique que nos émotions musicales ne sont pas purement intellectuelles : elles puisent dans l'ensemble des affects liés aux expériences corporelles que le sujet a déjà vécues. Dans notre étude, cette dimension de l'espace oppose des œuvres dont le processus musical est continu (par la forme de la courbe mélodique, le rythme, ou le type de progression harmonique) à des œuvres dont les processus sont plus fragmentés (arpèges brisés, rythme irrégulier, modulations subites).

Les œuvres que nous écoutons engendrent des émotions qui prennent des positions spécifiques sur ces dimensions d'énergie, de valence et de corporalité, en fonction de leurs caractéristiques structurelles. Ces dimensions étant continues, il y a une infinité de façons de combiner ces valeurs, laissant ainsi place à un nombre quasi infini d'émotions subtilement différentes. Cette approche rend mieux compte de la complexité et de la richesse des émotions en musique qu'une approche sous forme de catégories.

Les quatre grands types d'émotions (gaieté, désespoir, sérénité, colère) existent bien, mais ils ne représentent qu'un tout petit aspect de nos émotions. Les pièces musicales « gaies » déclenchent des émotions de grande dynamique et de forte positivité qui sont localisées en haut et à droite de cet espace (voir la figure 3). Selon que ces pièces ont ou non un caractère dansant (émotion corporelle plus ou moins élevée), elles ont une place différente sur le troisième axe. Les pièces colériques suscitent des émotions de forte énergie, mais de valence négative. De même, les pièces tristes ou dépressives se situent dans la région de l'espace émotionnel où la dynamique est faible et la valence négative. Les pièces sereines, telles que les comptines pour enfants, engendrent des émotions de faible dynamique et de valence positive. Cette représentation laisse transparaître la richesse possible des émotions induites par la musique.

Toutefois, le problème reste entier : comment déterminer ce qui, dans les œuvres, module l'intensité de nos émotions sur chacun de ces axes ? Cette question est à l'intersection de la musicologie, de la psychologie et des neurosciences cognitives. La psychologie isole les traits musicaux porteurs d'expression, les neurosciences cognitives observent comment

ces traits affectent l'auditeur. La musique étant un objet de culture, certains éléments ne seront jugés expressifs que par les membres de telle ou telle culture.

Ainsi tout un ensemble d'organisations musicales liées à la tonalité a un pouvoir expressif qui peut être confirmé par des méthodes directes auprès d'auditeurs occidentaux (on demande au sujet d'indiquer quel passage dans un extrait musical lui semble le plus expressif par exemple), ou indirectes (on mesure les modifications physiologiques de l'auditeur qui écoute un trait spécifique). On peut par exemple constater qu'une modulation, c'est-à-dire le passage d'un ton à un autre (de la gamme de *do* à la gamme de *sol*) ou d'un mode à un autre (majeur à mineur) et surtout la vitesse à laquelle cette modulation est réalisée provoquent des réponses émotionnelles.

Une partition dans le cerveau

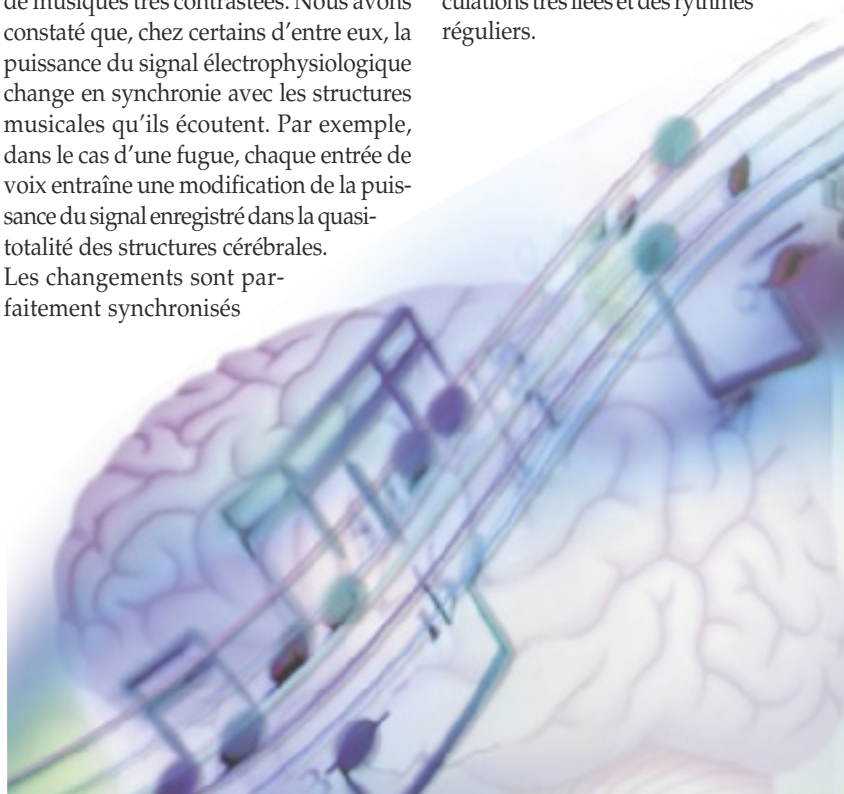
C'est ce qui a été vérifié lors d'une étude récente, conduite à l'Hôpital La Timone, à Marseille, en collaboration avec Catherine Liegeois-Chauvel, Stéphanie Khalfa, Michel Paindavoine et Charles Delbé. Nous avons enregistré en temps réel, dans différentes aires cérébrales (dont le cortex auditif et l'amygdale cérébrale), les réponses émotionnelles d'auditeurs à différents types de musiques très contrastées. Nous avons constaté que, chez certains d'entre eux, la puissance du signal électrophysiologique change en synchronie avec les structures musicales qu'ils écoutent. Par exemple, dans le cas d'une fugue, chaque entrée de voix entraîne une modification de la puissance du signal enregistré dans la quasi-totalité des structures cérébrales.

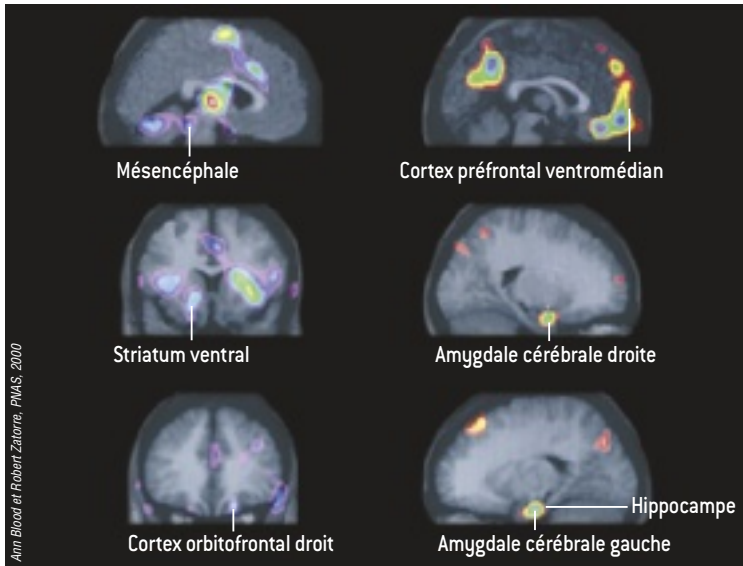
Les changements sont parfaitement synchronisés

avec la structure musicale. Pour certains de ces auditeurs (qui n'étaient pas des musiciens), le signal électrophysiologique prenait des allures de partition musicale !

D'autres changements ont des effets sur les émotions et ce quelle que soit la culture. Ainsi, Simone Dalla Bella et Isabelle Peretz, de l'Université de Montréal, au Canada, ont montré que les changements de mode et de tempo ont des effets systématiques sur la valence émotionnelle des morceaux et que ces effets sont perceptibles dans plusieurs cultures. Il ne s'agit bien sûr pas de prétendre que les musiques en mode majeur sont systématiquement perçues comme plus gaies que les musiques en mode mineur, mais un même morceau semblera plus gai lorsqu'il est joué en mode majeur plutôt qu'en mode mineur.

De même, un morceau donné semblera plus gai lorsque son tempo augmente. Le mode et le tempo constituent deux caractéristiques musicales essentielles pour faire varier les émotions selon les dimensions de valence émotionnelle et de dynamique. D'autres paramètres liés aux multiples qualités acoustiques des sons contribuent également à faire varier l'expression des extraits musicaux. La sérénité et la douceur sont généralement déclenchées par des sons joués à faible amplitude, avec des harmonies très consonnantes, des timbres acoustiquement pauvres, des articulations très liées et des rythmes réguliers.





4. L'IMAGERIE CÉRÉBRALE révèle les aires activées par différents types de musiques. Ici, le sujet écoutait une musique très plaisante. On a constaté que dans certaines aires – le mésencéphale, le striatum et le cortex orbitofrontal droit –, le débit sanguin augmentait, tandis qu'il décroissait dans le cortex préfrontal ventromédian, l'amygdale cérébrale et l'hippocampe.

La colère sera plus facilement évoquée par des sons détachés, puissants, ayant des enveloppes d'amplitude instable, des spectres harmoniques très riches, des harmonies dissonantes et des rythmes et tempos irréguliers. Il est évident que la façon dont ces différents aspects du langage musical occidental (modulation, tonalité, ornements, etc.) sont combinés au sein d'une œuvre avec les autres paramètres sonores offre au compositeur une infinité de moyens de définir des structures sonores expressives.

L'émotion véhiculée par la musique est reconnue instantanément

L'œuvre musicale regorge donc d'éléments susceptibles d'émouvoir l'auditeur. Ces éléments peuvent être immédiatement perceptibles lorsqu'ils sont liés aux qualités des matériaux sonores ou, au contraire, ils peuvent nécessiter des traitements cognitifs complexes liés à l'analyse de la syntaxe musicale. Ainsi, il est possible que plusieurs processus émotionnels cohabitent à différentes échelles de temps lorsque nous écoutons une œuvre, certains étant immédiats, d'autres plus lents. Comprendre le décours temporel de ces processus constitue un autre aspect important des études actuelles.

Les premiers résultats sont surprenants. Dans la vie quotidienne, les réponses émotionnelles peuvent être extrêmement rapides, et même précéder l'analyse cognitive, ce qui est compréhensible lorsque les stimulus rencontrés ont une très forte implication biologique et adaptative (la vue d'un serpent déjà évoquée). En revanche, on pensait que les processus

émotionnels en musique étaient plus lents. Ne dit-on pas qu'il faut « entrer » dans l'œuvre pour la ressentir ? Pourtant, les études révèlent le contraire. I. Peretz et ses collègues ont montré qu'il suffit de 500 millisecondes de musique pour différencier des musiques gaies de musiques tristes. Toutefois, les musiques gaies étant généralement jouées à des tempos plus rapides que les musiques tristes, on s'est demandé si les sujets de l'étude n'avaient pas fondé leurs réponses sur les différences de rythmes, perceptibles en 500 millisecondes. Nous avons repris cette recherche en présentant les 27 extraits de musique classique de l'expérience précédente, mais, à la différence de l'étude initiale, nous n'avons fait entendre que les 500 premières millisecondes des extraits. On demandait ensuite aux sujets de regrouper les extraits qui semblaient évoquer des émotions similaires.

Les résultats obtenus avec des extraits aussi courts sont très similaires à ceux obtenus avec des extraits de 20 secondes. Cela suggère que 500 millisecondes de musique suffisent pour que l'auditeur perçoive l'expression de la pièce. Dans une variante de cette étude, nous avons présenté des extraits de musique (classique ou populaire) dont nous savions par des études préalables qu'une moitié était jugée très expressive par les auditeurs et l'autre moitié comme relativement neutre. Nous avons cherché à savoir combien de millisecondes de musique étaient nécessaires pour que les auditeurs différencient ces deux catégories d'extraits. Les résultats indiquent que 250 millisecondes de musique suffisent pour que l'auditeur ait une intuition juste du caractère émotionnellement riche ou neutre de l'extrait musical qu'il écoute.

Ce résultat est important, car il indique que le cerveau humain répond émotionnellement aussi vite à la musique qu'à un stimulus biologiquement pertinent (une menace pour la vie de l'individu). Il suggère également qu'il y aurait, pour les émotions musicales, une voie très rapide essentiellement fondée sur les caractéristiques spectrales du son et une voie plus lente, qui analyserait les structures plus abstraites des œuvres. Il est bien sûr évident que l'émotion musicale change au fil de l'écoute et s'enrichit à mesure que l'œuvre se déploie.

L'ensemble de ces observations contribue à mieux comprendre la place que la musique revêt dans nos sociétés. Le fait que nos réponses émotionnelles à la musique soient aussi riches, reproductibles pour le même individu, mais aussi d'une personne à l'autre, rapides et si profondément enracinées dans notre cerveau, suggère que la musique a un statut bien spécifique pour l'espèce humaine, même si une vaste question reste ouverte : quelle est sa valeur adaptative ? Aurait-elle été sélectionnée en même temps que le langage, autre voie de communication universelle ? Ou a-t-elle été sélectionnée parce qu'elle « adoucit les mœurs » et que les sociétés dont les membres étaient sensibles à la musique se sont davantage multipliées ? ■