

PSYCHOACOUSTIQUE

L'imparfait du dispositif



Claude Bailblé

Lorsque la musique est bien jouée, dans une acoustique appropriée, l'auditeur en recueille les moments les plus émouvants, se les rejoue en mémoire, sans pour autant les retrouver totalement. L'enregistrement sur disque lui apporte la possibilité de les revivre auditivement, et même de se les rejouer plusieurs fois, si le cœur lui en dit. Mais la restitution stéréophonique ne peut égaler l'écoute directe en concert, pour de multiples raisons. Et l'on se prendrait à rêver d'un dispositif moins imparfait, capable de mieux redonner la musique.

L'écoute directe

Dans la salle de concert, l'auditeur se trouve placé à bonne distance (une quinzaine de mètres) de l'orchestre et, cependant, l'écoute est claire, esthétiquement plaisante. Les timbres sont suffisamment fondus, sans être noyés, les solistes se détachent de l'accompagnement, la réverbération donne sa plénitude au son. Et pourtant, à une telle distance, les deux tympans reçoivent principalement les ondes venues des parois de la salle et, secondairement, l'onde première, minoritaire, venue du podium. Le mélange tympanique est ainsi dominé par les ondes réfléchies, en retard sur l'onde directe. On ne devrait entendre qu'un son cafouilleux, brouillé, fortement réverbéré et flou. De fait, le *grossissement* (G), qui mesure le rapport entre l'énergie directe et l'énergie indi-

recte (réactive) rabattue par les murs, est très petit.

$$G = \frac{\text{direct}}{\text{indirect}} = \frac{13,8 V}{4\pi c d^2 T}$$

avec V le volume de la salle en m^3 ; T le temps de réverbération en secondes; d la distance aux pupitres en mètres; c la vitesse du son.

Ainsi, un auditeur placé à 12 mètres des solistes dans la *Grosser Musikverein* de Vienne ($T_r=2,2$ s... $V=14.600 m^3$) reçoit 5 fois plus d'énergie indirecte (retardée) que d'énergie directe en provenance des sources ($G=0,2$). Les derniers pupitres, placés à 20 mètres, ne lui donnent en droite ligne qu'un vingtième de l'énergie reçue ($G=0,05$). Le fauteuil idéal se trouve donc placé bien au-delà de la *distance critique* pour laquelle onde directe et ondes indirectes

égalisent leurs énergies en donnant un grossissement moyen ($G=1$). La balance y est objectivement très floue, le mélange très disproportionné. Il en va de même dans de nombreuses salles de concert, toutes aussi réputées. Mais qui s'en rend compte?

L'auditeur continue d'écouter la musique sans se soucier de cette incroyable disproportion, au demeurant peu accessible à son entendement. Le traitement auditif, tel le battement de cœur, se fait tout seul, sans passer - cela se saurait - par la conscience. Le résultat s'impose comme naturel, directement issu des sources instrumentales.

Certes, le dispositif scénique, l'acoustique de la salle, la disposition des fauteuils (en gradins) jouent fortement avec l'appareil auditif en s'accordant exactement à ses possibilités. L'auditeur le pres-

sent, même s'il ne peut se l'expliquer tout à fait. Cela devient évident dès que l'acoustique, la disposition du podium ou des musiciens ne conviennent plus à la musique : salle trop petite ou trop grande, réverbération trop forte ou trop faible, instruments trop bruyants ou effectifs insuffisants. Cela s'entend immédiatement, et l'on vient à penser que toute musique exige un réglage pointu quant à sa diffusion : un lieu parfaitement adapté, en termes d'acoustique architecturale, en termes de lisibilité auditive (1).

En premier lieu, la scène orchestrale est placée face à l'auditeur-spectateur : il voit et entend les instrumentistes devant lui. C'est en effet en immobilisant continûment l'attention "droit devant" que se suspend la motricité corporelle, condition de tout spectacle ; c'est également dans le plan de front que l'œil réalise sa meilleure discrimination (une minute d'angle) et cela vaut pour le concert comme pour l'écran de cinéma, la scène de théâtre, le tableau des peintres ou l'affiche publicitaire. C'est aussi frontalement que l'audition est la plus précise, la plus discriminante (2° d'angle). Von Békésy a dénommé *cône de présence* la région de l'espace où l'oreille localise et sépare le mieux lorsqu'elle s'associe à la vue.

Ce cône de présence - le cadrage auditif - contient et englobe l'aire de balayage du regard, obtenue tête immobile quand l'œil explore le plan de front, tout en se donnant une image correcte. La fusion des deux images rétinienne accordées point à point, exige en effet la superposition exacte des pixels momentanément appariés ; ceci implique un réglage musculaire très pointu et, ce faisant, une obliquité modérée des deux axes visuels. Un regard trop de biais détériorerait la qualité de l'image fusionnée, aussi le balayage oculaire reste limité à un secteur compatible avec la précision requise par le double poin-

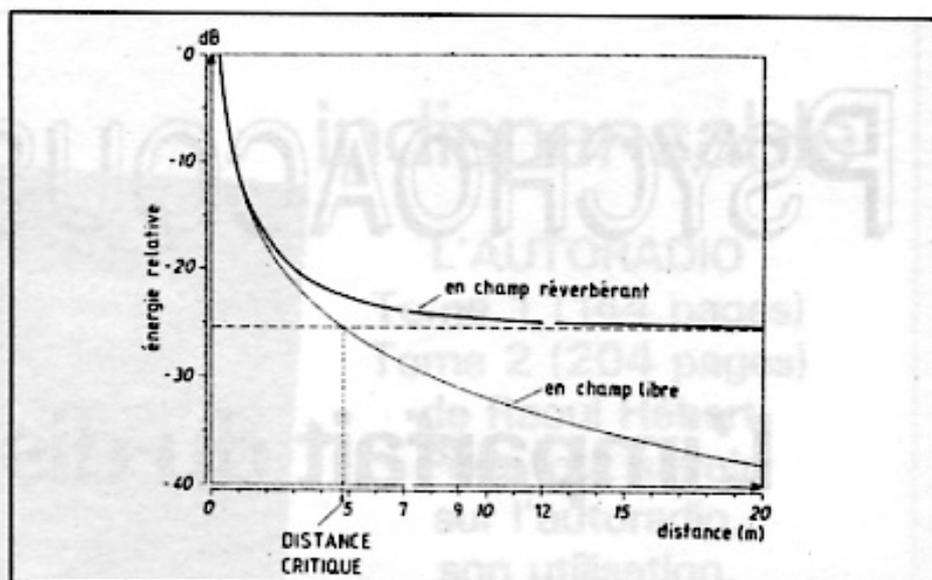


Fig. 1 : La distance critique. Niveaux d'énergie directe et d'énergie indirecte dans une salle de concert, en fonction de la distance à la source [d'après Meyer].

tage rétinien. Le cadrage auditif recouvre finalement la zone optiquement bonne, dessinée par un balayage oculaire préservant la qualité visuelle. S'il en est besoin, la tête tourne, entraînant dans un même mouvement les cadrages auditif et visuel.

L'inscription du regard à l'intérieur du cône de perception auditive indique clairement un champ d'appréhension focalisé vers l'avant calqué *de facto* sur le champ corporel : la vision, l'audition, la locomotion, la préhension font face aux objets qu'elles visent. Ainsi, l'attention visuelle et l'attention

auditive fonctionnent en parallèle de manière frontale et coaxiale et cela se lit dans les champs récepteurs (le territoire couvert par les neurones) des cortex de la vue et de l'ouïe. L'écoute droit devant, aidée par le regard, bénéficie alors d'une sélectivité accrue, d'un pouvoir séparateur augmenté.

Ne pouvant examiner tout en même temps, la perception a donc divisé l'espace en deux parties : l'une *frontale* où s'activent l'écoute et le regard, l'analyse fine et détaillée, l'attention soutenue ; l'autre *périphérique*, un peu effacée, atténuée comme telle dans la

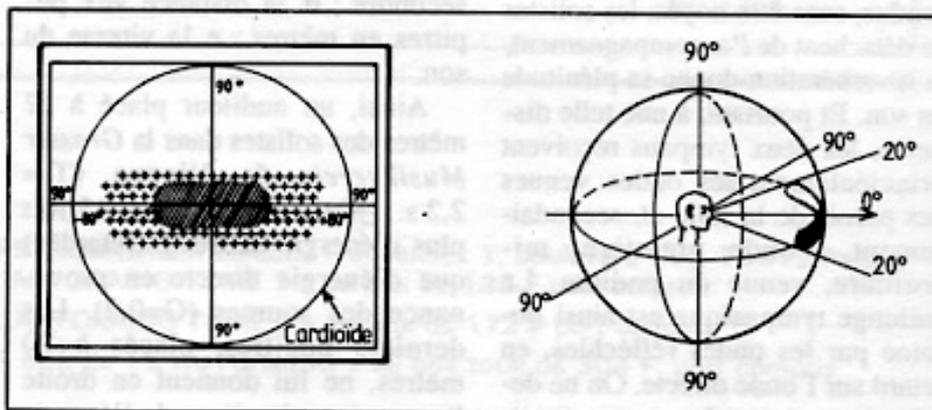


Fig. 2 : Le cône de présence. Le "cadrage auditif" serait préparé par une structure sous-corticale couvrant un large ovale (en +), tandis que le cône de présence s'étendrait à $\pm 20^\circ$ devant, dans le plan de front. Le microphone cardioïde couvre l'hémichamp frontal, à $\pm 90^\circ$ [d'après Buser et Imbert].

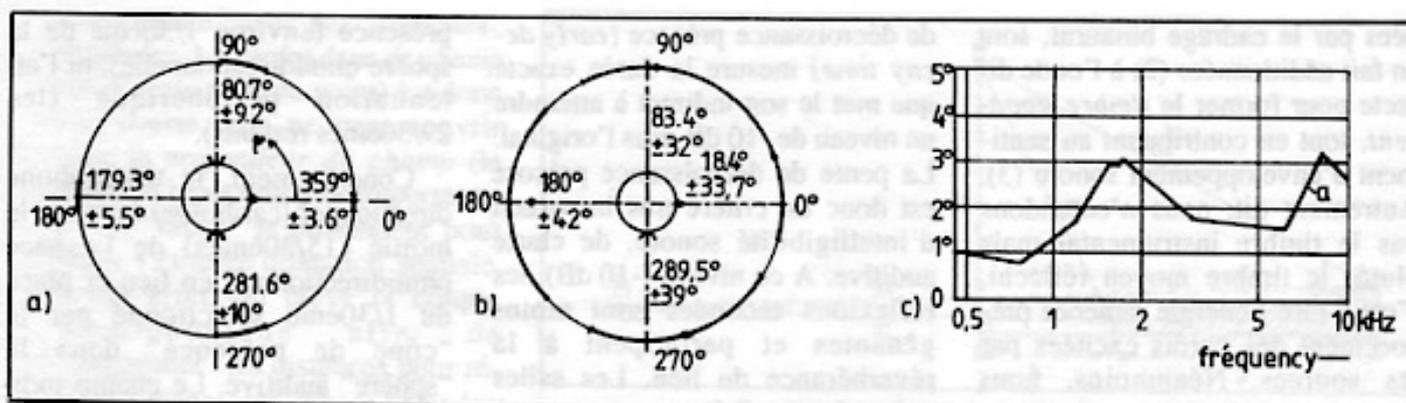


Fig. 3 : Le pouvoir séparateur de l'oreille. En a, pouvoir séparateur angulaire de l'appareil auditif dans le plan horizontal ; en b, le même, avec une seule oreille (les positions réelles des sources sont indiquées par des flèches ; les positions estimées par des cercles ; les segments donnent les écarts types ; en c, variation du MAA (minimum d'angle audible) dans l'incidence 0°, droit devant, selon la fréquence [d'après Blauert J. in "Spatial Hearing", M.I.T., 1983].

conscience, détachée comme ambiance peu présente. Ne s'y signalent que les mouvements, les bruits ou silences soudains, les sautes de lumière, les clignotements et autres scintillements. L'attention n'est attirée en fait que par les signaux transitoires présentant une dérivée non nulle, c'est-à-dire un changement, une fluctuation, une brusque modification. Quelle économie ! Les éléments inertes ou stationnaires - les plus nombreux - nous laissent de marbre, alors que tout ce qui bouge, craque, cliquette, grésille, tinte ou crépite, nous alerte fortement. Quelle efficacité aussi ! Toute variation alentour attire l'attention sur elle, déclenchant une réaction d'orientation dans sa direction. Tout événement sonore muni d'un front d'onde abrupt (dérivée importante) constitue donc un point d'appel irrépressible immédiatement collimaté

Cette répartition facilite le travail de l'attention frontale (neurones toniques, continûment actifs) sans pour autant abandonner le contrôle de l'environnement immédiat (neurones phasiques, omnidirectionnels, seulement sensibles aux changements, aux dérivés telles que dl/dt , dF/dt). La plupart des dispositifs scéniques - y compris la restitution stéréophonique - sont évidemment adaptés à la polarisation frontale du champ.

Tout son "hors-cône" (porte qui grince, cellophane qui crisse, gens qui toussent) déclenchent une distraction inutile et parfois agaçante parce que vide de sens, ou bêtement réflexe, non maîtrisable.

Ainsi se répartissent le *in* et le *off* dans la perception : le plan de front est à la fois très précis et très "présent" tandis que la périphérie reste floue, indistincte et relativement amorphe, peu "présente" dans la conscience. Cela veut dire qu'au concert, la réverbération réelle (périphérique) est fortement atténuée, qu'elle se signale peu à l'entendement (dérivée quasi-nulle, sauf dans le cas d'un écho détaché) alors que le podium (frontal) est investi des pouvoirs discriminants de l'oreille et de l'œil. Au total, le cadrage binaural privilégie la scène et pondère l'effet de salle.

Ainsi, l'intelligibilité se trouve augmentée par la vision directe des sources, par la sélectivité accrue du plan de front, tandis que le masquage son sur son est minimisé, en raison même de l'atténuation apparente du champ réverbéré, spatialement désinvesti.

En second lieu, l'acoustique des salles est soigneusement étudiée : elle se fait surtout entendre à travers les réflexions précoces, celles qui surgissent dans les quarante premières millisecondes, juste derrière le son direct. En effet, ces ré-

flexions - les plus énergétiques - ont la propriété de se contracter sur l'onde initiale, de fusionner avec elle, et donc de se fondre sur le lieu origine (le podium tout en participant grandement (à 90%, sinon davantage) à la sonorité. Ces réflexions hâtives, loin d'être inhi-

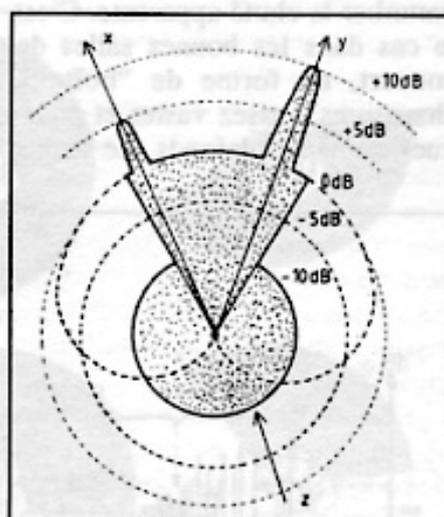


Fig. 4 : La cadrage binaural. En z, l'inhibition statique du champ périphérique. En x, hausse transitoire de la clarté. En y, hausse permanente de l'intelligibilité due au pointage visuel et auditif. Les gains (dB) ne sont que des ordres de grandeurs. (On se reportera utilement aux notions scientifiques de BMLD - binaural masking level difference - pour l'inhibition statique, et de BILD - intelligence level difference - pour le pointage visuo-moteur.) En pointillés, la courbe de réponse d'un microphone cardiode.

bées par le cadrage binaural, sont en fait additionnées (2) à l'onde directe pour former le *timbre apparent*, tout en contribuant au sentiment d'enveloppement sonore (3). Autrement dit, nous n'entendons pas le timbre instrumental mais plutôt le timbre moyen réfléchi, c'est-à-dire l'énergie ramenée précocement des parois excitées par les sources. Néanmoins, nous continuons à localiser selon leur provenance, à même le podium (loi dite du *premier front d'onde*; Cremer, 1948). C'est l'effet d'antériorité précisé par Haas en 1951.

Les réflexions secondes (de 50 à 1 500 millisecondes), détachées comme telles, participeraient au contraire à la confusion et au masquage son sur son : une série d'images floues en surimpression sur l'image nette. Même si elles se trouvent atténuées par le cadrage binaural, il importe que leur niveau soit suffisamment bas pour ne pas perturber la clarté apparente. C'est le cas dans les bonnes salles de concert, en forme de "boîte à chaussures", assez vastes et pourvues de hauts plafonds. Le temps

de décroissance précoce (*early decay time*) mesure la durée exacte que met le son indirect à atteindre un niveau de -10 dB sous l'original. La pente de décroissance précoce est donc un critère très important d'intelligibilité sonore, de clarté auditive. A ce niveau (-10 dB), les réflexions secondes sont moins gênantes et participent à la réverbérance du lieu. Les salles polyvalentes (béton et moquette) sont souvent loin du compte.

Les réflexions tardives (très affaiblies) ajoutent à la musique un halo, un enveloppement diffus très agréable, une couleur qui correspond au spectre d'extinction acoustique de la salle. On peut représenter l'ensemble du champ indirect par un échogramme.

Le microphone bien tempéré

Que se passe-t-il lorsqu'on plante un couple de microphones au "fauteuil idéal"? Le timbre est bon, mais noyé dans une caverne, pris dans une réverbération excessive. La sélectivité spatiale du microphone ne recopie ni le cône de

présence (environ 1/30ème de la sphère omnidirectionnelle), ni l'atténuation périphérique (les 29/30èmes restants).

Concrètement, le microphone directionnel (cardioïde) favorise la moitié (15/30èmes) de l'espace omnidirectionnel, en lieu et place du 1/30ème sélectionné par le "cône de présence" dans la "sphère" auditive. Le champ indirect dû à la salle domine : il y a trop de halo, de réverbération.

Il faut alors rapprocher le micro du podium pour obtenir un plan apparent, une perspective proche de ce que l'on entend habituellement. Dans cette manœuvre, la balance de profondeur évolue favorablement. Le champ indirect se trouve réduit tandis que le champ direct est comme revitalisé : les ondes venues des instruments sont amenées au niveau des réflexions de la salle. Mais si l'on approche encore les micros, c'est le champ direct de quelques instruments qui va dominer, créant un gros plan partiel sur l'orchestre, une perspective au grand angle, en quelque sorte, avec une profondeur faussée, démesurée. Le pouvoir séparateur du microphone - par pure proximité avec telle ou telle source - va ruiner l'impression d'ensemble en favorisant un timbre ou des détails que l'oreille ne retient pas d'ordinaire, puisque masqués par l'éloignement ou par l'alliage des sonorités. On doit placer les capteurs suffisamment loin pour éviter ce type de distorsion. Mais ceci n'est possible que dans une salle acoustiquement adaptée.

Cependant, les premiers plans seront encore trop nets, les arrière-plans trop flous comme dans une perspective "accélérée". C'est que l'énergie directe décroît avec le carré de l'éloignement alors que l'énergie indirecte reste à peu près constante. L'étagement des plans sonores dans la distance suit effectivement la décroissance de l'énergie directe et cela entraîne un es-

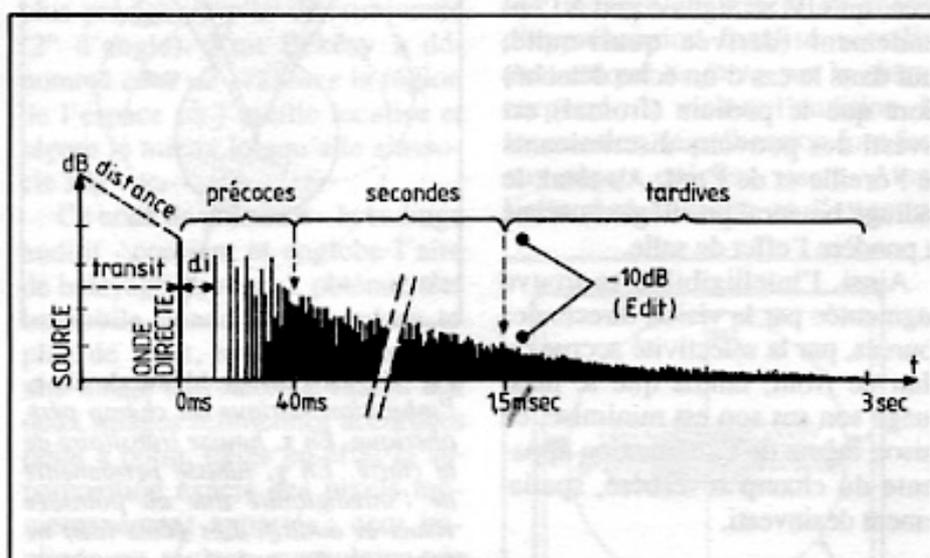


Fig. 5 : Echogramme d'une salle. Après un temps de transit, l'onde directe atteint l'auditeur (premier front d'onde); suit le délai avant la première réflexion précoce (délai initial) aussitôt complétée par ses semblables (jusqu'à 40 ms), toutes fusionnées sur le premier front d'onde. Viennent alors les réflexions secondes (jusqu'à 1.500 ms) suffisamment décroissantes (-10 dB, selon EDT) pour ne pas être gênantes et enfin, les réflexions tardives de la réverbération finissante et colorée (jusqu'à 3 s et plus).

pace "dilaté" avec des "fuyantes" exagérées. La *profondeur de champ* (ou d'égale mise au point) est donc très courte, elle ne peut couvrir toute la *profondeur du champ* (le podium) beaucoup plus importante. Il faut "tasser" la profondeur pour éviter l'étrange dilatation spatiale. A cet effet, on élève le couple microphonique afin de "comprimer" les distances pour réduire l'étagement sonore entre les premiers et les derniers pupitres.

On peut aussi pratiquer la prise de son *multimicros*, en proximité. La balance sera faite plus tard, au mixage, et on ajoutera de la réverbération naturelle (micros lointains) ou numérique (processeur temporel). Avec le risque de dénaturer tous les timbres et de les nimber d'espace artificiel. Le gros plan pose de toute façon de nombreux problèmes.

En champ rapproché, le microphone capte surtout les ondes di-

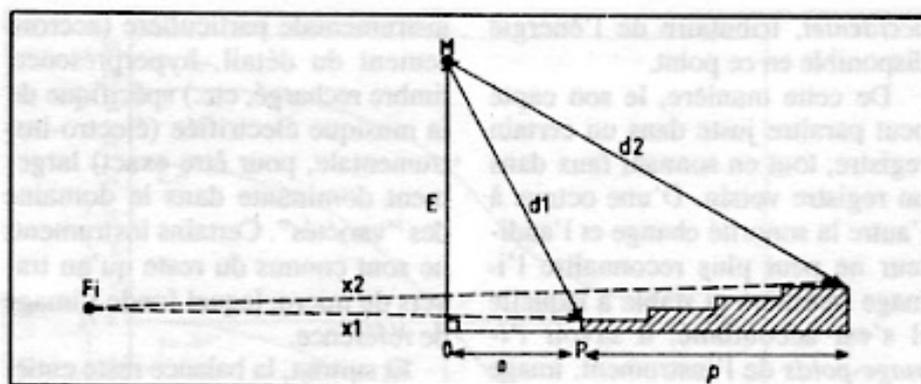


Fig. 6 : La perspective retrouvée. L'élévation E des micros (M) au-dessus du podium (p) réduit l'écart de distance ($d2-d1$) entre les premiers et derniers pupitres. Une implantation au "fauteuil idéal" donnerait une perspective au grand angle, tandis qu'une "vue de dessus" donnerait une perspective trop "tassée". La profondeur de champ (d'égale mise au point) est de toute façon très courte, à savoir : le 1/20ème de la distance critique... la bien nommée.

rectes émises par l'instrument. Dès lors, la captation est soumise au rayonnement inhomogène de l'émetteur sonore (rayonnement anisotrope). La plupart des instruments rayonnent en effet différemment leurs "formants" selon la note jouée, distribuant inégalement

l'énergie (grave, médium, aigu) dans les différentes directions de l'espace. Captées d'un seul point, certaines notes "aboient", deviennent "criardes" ou au contraire paraissent "en retrait", comme détimbrées. Le microphone - quasi-ponctuel - ne capte donc qu'un timbre

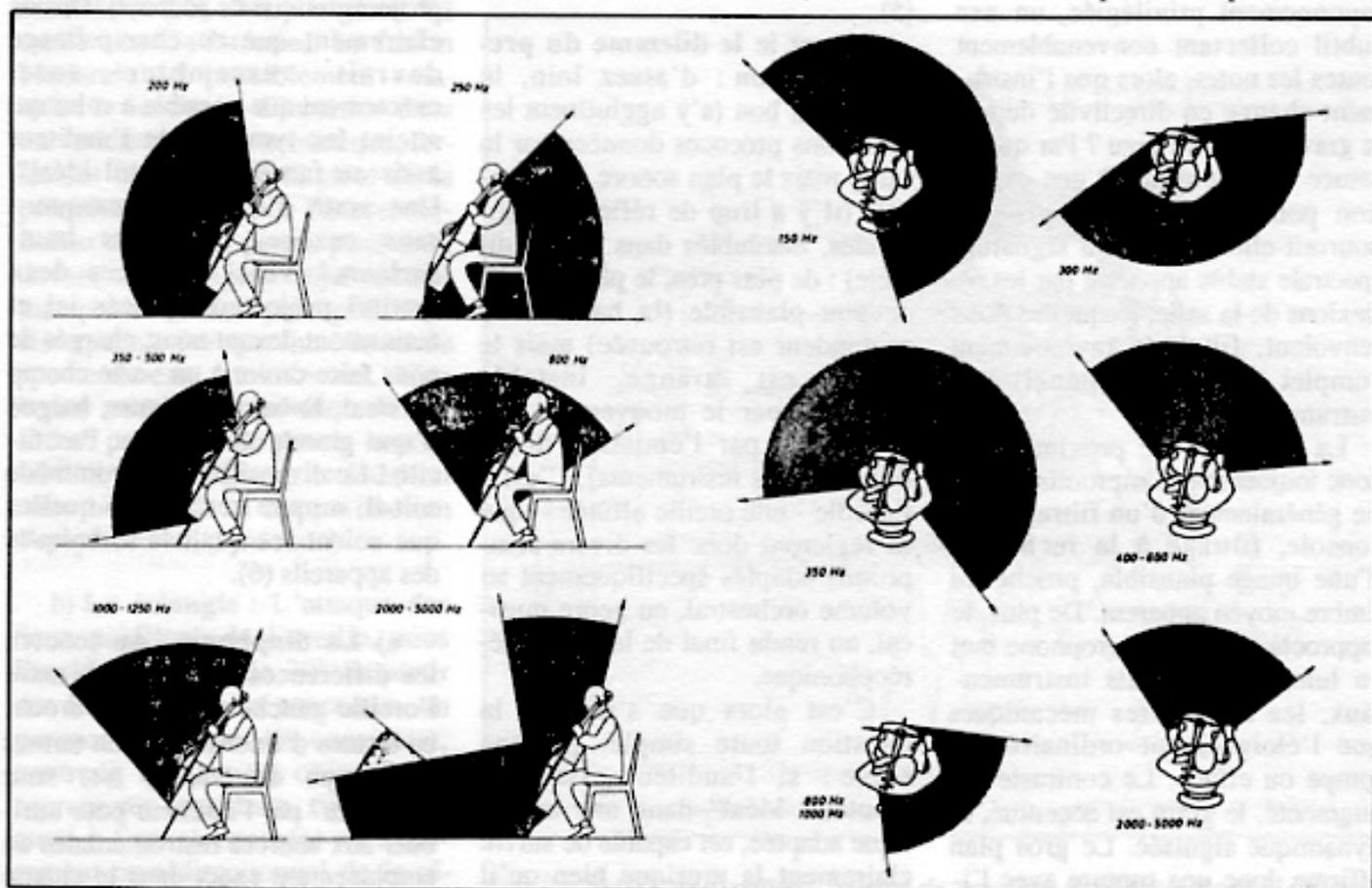


Fig. 7 : Le rayonnement (invisible) des instruments - ici le violoncelle - disperse inégalement les notes selon les hauteurs. Sous quel axe faut-il donc placer le micro pour recueillir un timbre plausible, une image de la bonne couleur ? [d'après Meyer].

accidentel, tributaire de l'énergie disponible en ce point.

De cette manière, le son capté peut paraître juste dans un certain registre, tout en sonnant faux dans un registre voisin. D'une octave à l'autre la sonorité change et l'auditeur ne peut plus reconnaître l'image familière et stable à laquelle il s'est accoutumé, à savoir l'image-poids de l'instrument, image venue - pour l'essentiel - des réflexions précoces de la salle. La dite salle recueille en effet l'énergie émise - de tous côtés - par les instruments, avant de la renvoyer - timbre moyen apparent (4) - aux auditeurs. Certes, on peut toujours espérer qu'en plaçant le microphone d'une manière très étudiée, il va peut-être se trouver un point de capture, un "angle d'attaque" favorable approchant le timbre de référence, mais rien n'est moins sûr. Pour quelle raison (magique, peut-être) devrait-il exister une zone de rayonnement privilégiée, un axe subtil collectant convenablement toutes les notes, alors que l'instrument change en directivité depuis le grave jusqu'à l'aigu ? Par quelle astuce (géométrique ?) une captation ponctuelle très rapprochée pourrait-elle intégrer la signature spectrale stable apportée par les réflexions de la salle, lesquelles nous renvoient, filtré, le rayonnement complet (multidirectionnel) des instruments ?

La captation de proximité est donc toujours un compromis, assortie généralement d'un filtrage à la console, filtrage à la recherche d'une image plausible, proche du timbre moyen apparent. De plus, le rapprochement du microphone met en lumière les bruits instrumentaux, les transitoires mécaniques que l'éloignement ordinaire estompe ou efface. Le contraste est augmenté, le grain est accentué, la dynamique aiguë. Le gros plan affirme donc une rupture avec l'image acoustique de référence et, pour certaines musiques, ce sera l'occasion d'exprimer une sonorité

instrumentale particulière (accroissement du détail, hyperprésence, timbre rechargé, etc.) spécifique de la musique électriée (électro-instrumentale, pour être exact) largement dominante dans le domaine des "variétés". Certains instruments ne sont connus du reste qu'au travers du micro, lequel fonde l'image de référence.

Et surtout, la balance reste entièrement à refaire. L'enregistrement multimicros brise *de facto* le mixage acoustique naturel, l'impression d'ensemble ; il faut donc reconstituer la scène musicale voulue par le compositeur avec ses alliages de timbres, tout en rendant justice à l'interprétation, avec ses nuances, sans trahir outre mesure l'acoustique donnée par la salle. Le mixage constitue donc une étape délicate où le sens musical, la sensibilité auditive entrent en jeu, au même titre que le savoir procédural, la mise en espace et en niveaux (5).

Tel est le dilemme du preneur de son : d'assez loin, le timbre est bon (s'y agglutinent les réflexions précoces données par la salle) mais le plan sonore est mauvais (il y a trop de réflexions secondes, insolubles dans l'onde directe) ; de plus près, le plan sonore devient plausible (la balance de profondeur est retrouvée) mais le timbre est étrange, instable (perturbé par le mouvement des musiciens, par l'émission directionnelle des instruments). C'est à l'oreille - une oreille affûtée - que se régleront donc les divers compromis adaptés spécifiquement au volume orchestral, au genre musical, au rendu final de la scène stéréophonique.

C'est alors que s'insinue la question toute simple, presque naïve : si l'auditeur, assis au "fauteuil idéal" dans une acoustique adaptée, est capable de suivre clairement la musique bien qu'il reçoive principalement de l'énergie indirecte retardée, c'est que le cadrage auditif, l'écoute et la vision

frontale lui redonnent une image claire de la scène orchestrale. Pourquoi n'en ferait-il pas autant devant la scène stéréophonique ? A-t-il donc perdu ses facultés auditives ? Ne peut-il restaurer, comme au concert, la clarté ordinaire de l'écoute ? Pourquoi alors s'embarasser de toutes ces précautions microphoniques ?

La restitution stéréophonique

Si l'écran stéréophonique ressemble à la scène du concert, c'est surtout grâce aux précautions prises pendant l'enregistrement ! Car, pour l'essentiel, le dispositif de restitution - très imparfait - est loin de ressembler à un orchestre. Et pour cause : la projection stéréophonique ne restitue qu'un champ-image (un projecteur pour chaque tympan) sans reconstituer le champ-objet origine en son entier (autant de projecteurs que de sources). Disons clairement que ce champ-image devrait ressembler aussi exactement que possible à celui qui atteint les tympans de l'auditeur assis au fameux "fauteuil idéal". Une sorte d'écoute au casque... sans casque, avec des haut-parleurs ! Voici donc ces deux (petits) projecteurs placés ici et maintenant devant nous, chargés de nous faire croire à un vaste champ musical, là-bas et ailleurs, baigné d'une grande acoustique. Pas facile ! Le dispositif, aussi commode soit-il, montre ses limites quelles que soient les qualités techniques des appareils (6).

a) **La diaphonie** : Au concert, les différences de captation entre l'oreille gauche et l'oreille droite, en termes d'intensité Δi , en termes de temps de transit Δt , sont "utilisées" par l'auditeur pour attribuer aux sources instrumentales un emplacement exact dans le champ frontal (l'orchestre, centré par le cône de présence) et à la réverbération un enveloppement atténué,

sans localisation précise.

Ici, l'image sonore venue du haut-parleur de gauche atteint aussi l'oreille droite, avec un léger retard ($-\delta t$) et une légère atténuation dans les aigus ($-\delta i$) ; idem pour le haut-parleur de droite. La diaphonie est alors masquée en exagérant les différences interaurales, en augmentant sur les micros (ou à la console) ΔI et ΔT , les deux données fondamentales de la spatialisation et, ce faisant, de l'impression de réalité.

Mais cet artifice laisse apparaître des résidus, des images fantômes, un manque de netteté. La discrimination sur la scène stéréophonique se fera à 5° près (et non plus à 2°). On ne peut espérer séparer, sur les 60° du plan de front stéréophonique, plus de 12 directions ou provenances instrumentales. Ce qui n'est déjà pas si mal !

Il n'empêche. Coexistent deux projections, une principale ciblant les bonnes oreilles (ΔI , ΔT), une secondaire ($-\delta i$, $-\delta t$) atteignant les oreilles contra-latérales. La fusion binaurale est possiblement entamée ; le sentiment d'insertion dans un véritable espace acoustique, fortement dilué ; on reste privé de la précision ordinaire, de l'accommodation sur l'objet perçu, posé droit devant soi. L'espace-objet restitué manque de confort perceptif, d'assise spatiale transparente, de rendu spontanément convaincant. Une bonne part de l'impression de réalité ne provient-elle pas de la possibilité de spatialiser, de localiser sans hésitation ni approximation l'emplacement d'un son ?

b) **Le triangle** : L'attaque des deux pavillons de l'oreille, sous l'incidence figée de 30° (à partir des deux projecteurs ponctuels - ou presque - que sont les enceintes), contredit l'étalement objectif des sources sur le podium, contrevient aussi à l'enveloppement nécessairement omnidirectionnel de la réverbération. Les pavillons externes de l'oreille ont en effet une courbe de réponse spatiale précise, à la-

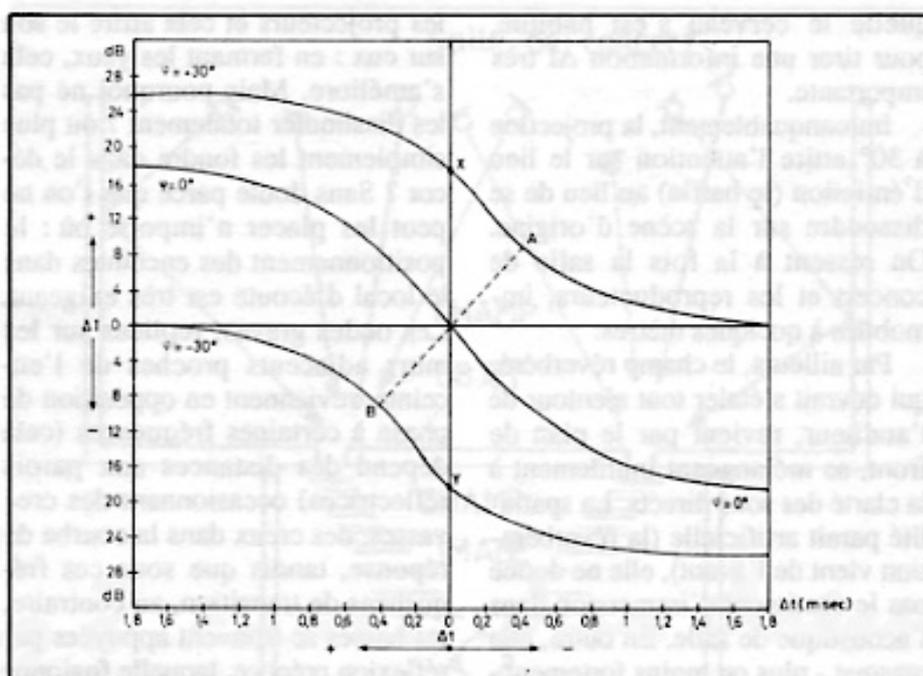


Fig. 8 : L'anamorphose. Alors que l'écoute binaurale directe localise à 30° un son présentant un ΔI de 4 dB combiné à un ΔT de 300 μs , l'écoute stéréophonique nécessite une combinaison d'écarts plus importants [d'après Carl Ceoen]. L'anamorphose à l'enregistrement est compensée lors de la restitution diaphonique.

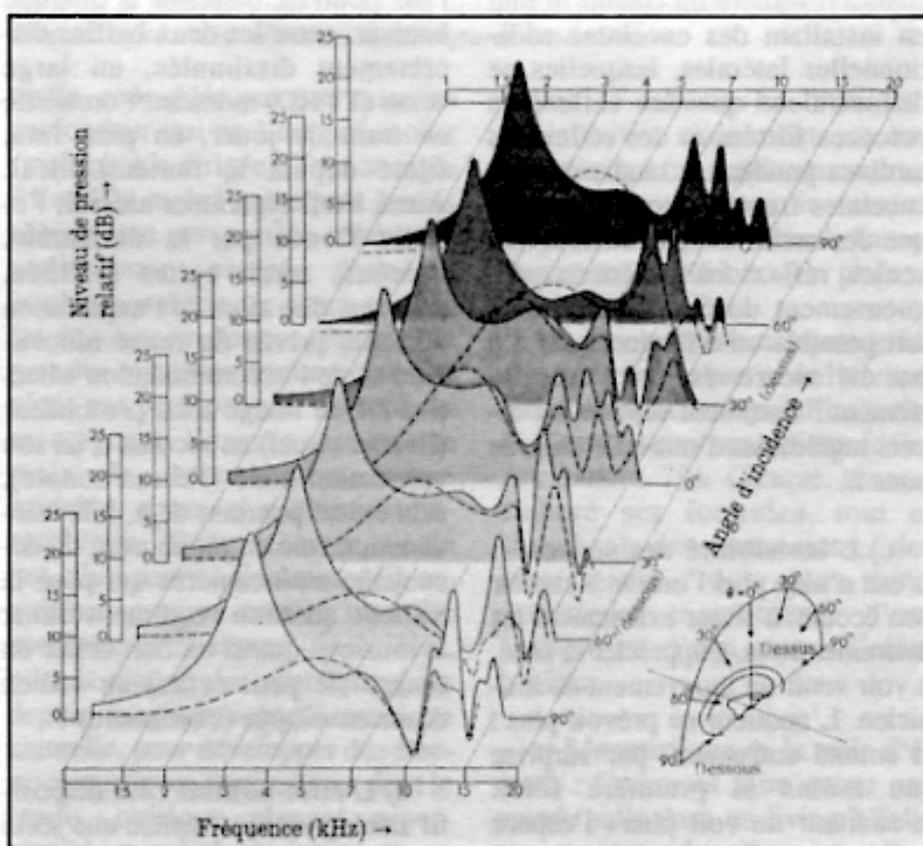


Fig. 9 : Action du pavillon. Niveau de pression acoustique dans le conduit auditif en fonction de la fréquence, pour diverses incidences [d'après Shaw et Teranishi] cités par Georges Canévet, in "Psychoacoustique et perception auditive", éd. INSERM, 1989.

quelle le cerveau s'est habitué, pour tirer une information ΔI très importante.

Immanquablement, la projection à 30° attire l'attention sur le lieu d'émission (le baffle) au lieu de se dissoudre sur la scène d'origine. On ressent à la fois la salle de concert et les reproducteurs, immobiles à quelques mètres.

Par ailleurs, le champ réverbéré, qui devrait s'étaler tout alentour de l'auditeur, revient par le plan de front, se mélangeant inutilement à la clarté des sons directs. La spatialité paraît artificielle (la réverbération vient de l'avant), elle ne donne pas le sentiment d'immersion dans l'acoustique de salle. En outre, elle masque - plus ou moins fortement - la lisibilité des sons instrumentaux. Tout se passe comme si l'on ajoutait, par dessus une image nette, une seconde image identique, mais floue... et légèrement décalée...

D'où l'idée de dissocier le champ réverbéré du champ frontal en installant des enceintes additionnelles latérales, lesquelles ne diffuseraient que des réflexions précoces filtrées et des réflexions tardives pondérées, tandis que les enceintes frontales ne donneraient que des ondes directes mêlées aux seules réflexions précoces, soigneusement dosées. C'est tout à fait possible en diffusion mais n'a pas été encore résolu à l'enregistrement ! Comment séparer les divers ingrédients d'une telle prise de sons ?

c) **L'invisibilité des sources** : L'œil n'aide plus l'oreille à choisir son écoute, à situer exactement les instrumentistes, à apprécier la salle, à voir venir un mouvement de musicien. L'auditeur ne prévoit plus : il entend autrement, par surprise (au moins la première fois). L'auditeur ne voit plus : l'espace musical est flou, moins sélectif, imprévisible. Les sons sont moins détachés, moins isolés les uns des autres.

Par contre, l'œil ne voit que trop

les projecteurs et cela attire le son sur eux : en fermant les yeux, cela s'améliore. Mais pourquoi ne pas les dissimuler totalement ? ou plus simplement les fonder dans le décor ? Sans doute parce que l'on ne peut les placer n'importe où : le positionnement des enceintes dans le local d'écoute est très exigeant. Les ondes graves, repliées sur les murs adjacents proches de l'enceinte, reviennent en opposition de phase à certaines fréquences (cela dépend des distances aux parois réfléchissantes) occasionnant des crevasses, des creux dans la courbe de réponse, tandis que sous ces fréquences de transition, au contraire, les basses se trouvent appuyées par réflexion précoce, laquelle fusionne parfaitement avec l'onde émise (7).

Idéalement, pour une émission hémisphérique régulière, les haut-parleurs devraient affleurer à la surface d'un mur de front, lui-même traité acoustiquement, et l'on pourrait disposer à quelque hauteur, entre les deux baffles discrètement dissimulés, un large écran (TVHD) montrant l'orchestre en train de jouer, en plan fixe, filmé depuis le fauteuil idéal. Ainsi, l'œil séparateur aiderait l'oreille flouée par la diaphonie. L'écoute serait certes facilitée, mais que dire alors de l'exploration visuelle, privée du relief binoculaire et de l'accommodation sélective ? Une image sans profondeur (l'écran visuel) au secours d'un son vaguement étalé (l'écran sonore), cela existe pourtant déjà, à la télévision, même en mono avec un découpage multicaméras qui pose la difficile question : comment filmer la musique quand on doit coller un timbre (le petit écran) au milieu d'une enveloppe (l'orchestre) ?

d) **L'effet-hublots** : Le dispositif stéréophonique réalise une sorte de projection paradoxale presque incroyable. C'est à la fois un projecteur ($2 \times 60 \text{ W}$?) et un écran sans support autre que l'air (entre les deux baffles). Depuis quand re-

garde-t-on une projection en se tournant vers l'objectif du projecteur ?

En réalité, vision et audition diffèrent totalement sur ce point.

Le véritable écran visuel est la rétine où s'imprime par projection conique - depuis l'iris - une image recueillie au fond du globe oculaire, image issue d'un écran relais, écran éclairé également par projection conique, depuis l'objectif d'un projecteur placé dans un local suffisamment obscur. La projection conique restitue donc la largeur et la profondeur du champ en étalant (verticalement et horizontalement) la lumière sur la mosaïque des cônes rétinien. L'espace est donc restitué à même le capteur, par la physique lumineuse de l'optique oculaire.

Tandis que le véritable écran auditif est le tympan où l'onde aérienne se transforme en vibration tangible. On devrait dire *les* ondes aériennes car il en vient de tous côtés, se résout en une seule équation vibrante, celle de la membrane tympanique. L'espace n'est donc pas reconstitué sur le capteur, mais beaucoup plus haut, après traitement cortical des disparités gauche-droite, dans une "rétine auditive" neuronale. Pour un tympan isolé, comme pour un microphone monophonique, il n'y a en effet qu'une seule dimension spatiale : la profondeur, laquelle s'apparente à une perspective sommaire réduite à la *profondeur de temps*. La dimension "largeur de champ" n'apparaît qu'après de nombreux traitements corticaux portant sur deux signaux différents mais corrélés. Ces différences et cette corrélation ne dépendent pour ainsi dire que de l'emplacement physique (immuable) des tympanes sur la boîte crânienne : la disposition et la forme des oreilles - mine de rien - installent un prétraitement "topologique" qui conditionne le travail neuronal subséquent. Pour ce qui est de l'écoute humaine, il n'y a en définitive qu'un seul sys-

tème de prise de sons : deux membranes écartées de 17 cm (ΔT) et orientées à 180° (ΔI).

Pour des raisons pratiques bien compréhensibles, on a inventé une restitution stéréophonique à deux projecteurs (il y a en effet deux tympans) alors que l'écran neuronal interne (le champ-image auditif) comporte des centaines de *champs récepteurs*, chacun en correspondance directe - comme les cônes de la rétine - avec une petite partie de l'espace extérieur (le champ-objet). Ces différents champs récepteurs dessinent la "coupole auditive", forme obligée et naturelle de la perception de l'espace sonore. Chacun d'entre eux nécessiterait - dans l'absolu - un haut-parleur simulant l'énergie venue de cette petite région de l'espace environnant. C'est tout à fait impossible, sauf peut-être en laboratoire (*holophonie*), mais à quel prix ! L'espace auditif doit donc se faufiler au travers de deux écouteilles, de deux hublots, et apporter aux tympans les deux signaux *ad hoc*, proches des originaux entendus au concert, mais corrigés ou compensés des défauts inhérents au système stéréophonique (diaphonie, triangle, etc.) (8).

Si la prise de son ne tient pas compte de la captation binaurale ($\Delta I/\Delta T$), si l'enregistrement (multimicros) ne cherche pas à restituer ces différences corrélées et compensées, alors l'entreprise acoustique devient périlleuse. Si l'on s'imagine que les deux haut-parleurs jouent comme des sources vraies placées dans un champ-objet, on fait de la *biphonie* ; il faut réécrire toute la musique pour deux instruments, sinon réduire le répertoire aux seuls *duos*, ou alors accepter de placer l'auditeur face à deux hublots creusés dans un mur, deux hublots donnant ouïe sur la scène d'origine. Impossible !

D'où le *pan-pot* qui, en stéréophonie d'intensité, cherche à rétablir l'illusion d'une scène continue, par le biais d'une largeur artifi-

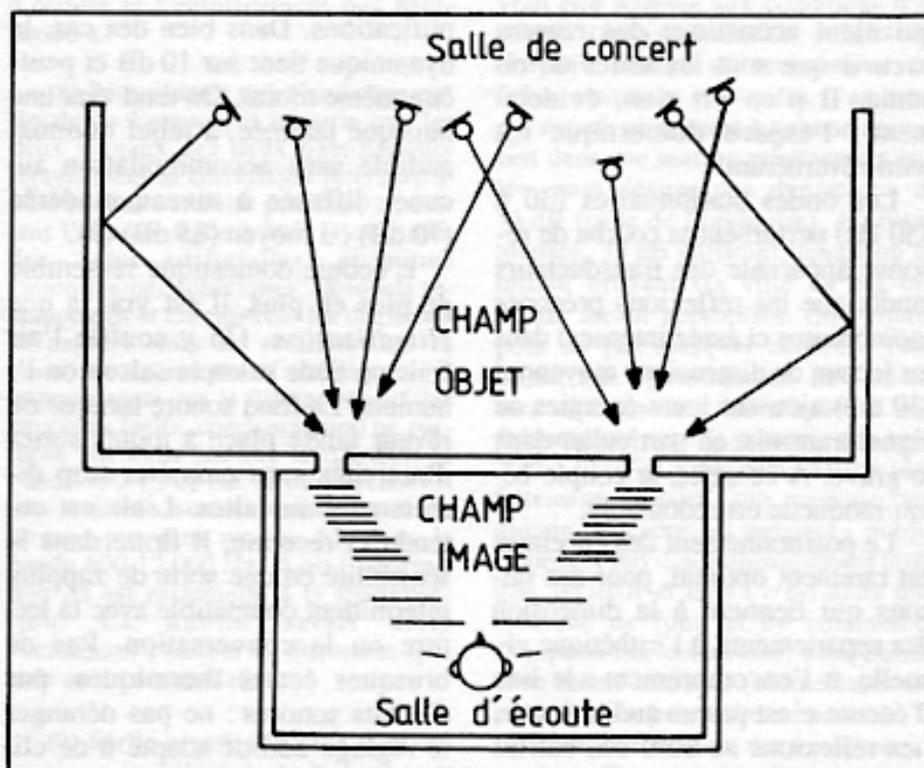


Fig. 10 : La biphonie. L'émission de deux signaux non corrélés (biphonie) réduirait la stéréophonie à deux goulots acoustiques en lesquels s'étranglerait toute l'étendue spatiale, soudain privée des $\Delta I/\Delta T$ indispensables à l'étalement de l'image, tant en largeur qu'en profondeur. Conglomerés entre eux, les sons se masqueraient les uns les autres par surimpression destructive, par impossibilité de se dénouer, de se séparer spatialement.

cielle, redoublée par une mise en profondeur qui ne l'est pas moins (multiphonie dirigée). Tout se règle à l'oreille, mais l'on ne peut jamais obtenir une perspective cohérente, auditivement centrée. Cela ne semble pas déranger outre mesure les éditeurs de disques, car la plupart des musiques commercialisées n'ont pas pour premier objectif de recomposer l'espace acoustique mais de donner une image agréable, voire volontairement accrocheuse, d'une scène musicale qui, d'un genre à l'autre, a déjà acquis ses propres conventions. De nombreuses musiques - sonorisées, pour la plupart - se sont échappées depuis longtemps de l'acoustique naturelle, pour développer des électro-acoustiques particulières dont le rendu - onirique, planant, superspatial, mirobolant, intimiste ou percutant - ne vise qu'à instaurer un espace imaginaire, délivré de l'impression de réalité orienté vers le rêve, la piste de danse ou l'atmo-

sphère de cabaret. La musique à effets ouvre clairement sur une spatialité qui fait reculer les murs et bouger les sons, en insérant le corps dans un lieu virtuel, lieu où chacun est invité à trouver place, si bon lui chante.

La dimension électrospatiale fait donc désormais partie de l'écriture musicale, surtout depuis l'invention de l'enregistreur multipistes (début des années 70). Chaque genre a élaboré ses formules, tout en cherchant des extensions, des innovations, des plus-values esthétiques contournant - d'une manière ou d'une autre - l'effet-hublots.

e) L'acoustique de la salle d'écoute : Comment faire entrer une grande salle dans un living ? Telle sera notre dernière question concernant la restitution stéréophonique. Idéalement, la salle d'écoute devrait être conçue comme une *camera muta*, c'est-à-dire l'é-

quivalent acoustique des *camera oscura* que sont les salles de cinéma. Il n'en est rien, évidemment : l'espace domestique est semi-réverbérant.

Les ondes stationnaires (30 à 350 Hz) perturbent la courbe de réponse apparente des transducteurs tandis que les réflexions précoces (nombreuses et assez intenses) dans les locaux de dimensions moyennes (30 m²) ajoutent leurs énergies au signal transmis, en particulier dans le grave. A ce sujet, le couple béton-moquette est redoutable.

Le positionnement des enceintes est rarement optimal, pour des raisons qui tiennent à la dimension des appartements, à l'esthétique visuelle, à l'encombrement : le lieu d'écoute n'est pas un auditorium... Les réflexions au bord des baffles ajoutent donc un filtrage en peigne dans le haut-grave, tout en renforçant les basses, au-dessous d'une fréquence limite qui est souvent déterminée par la distance haut-parleur-plafond.

Par ailleurs, la directivité des haut-parleurs se resserre dans les octaves supérieures, en sorte que l'énergie réfléchi (et fusionnée) diminue du grave à l'aigu. S'ajoute à cela une réverbération plus grande dans le bas que dans le haut du spectre, les aigus sont facilement absorbés par les tapis, moquettes, livres (etc.), tandis que les basses se réfléchissent plus longuement, faute d'absorbants efficaces. La balance spectrale apparente s'en ressent.

Encore n'a-t-on pas abordé d'une part le problème du bruit de fond urbain - lequel oblige à remonter les *pianissimi* à un niveau confortable - et, d'autre part, le problème du bon voisinage - lequel incite à modérer les *fortissimi*, à pondérer l'excursion vers les forts niveaux. La compression de dynamique est inévitable : un rapport de 50 dB apparaît comme largement suffisant, voire trop important, pour la plupart des musiques enregistrées, diffusées sur de petites am-

plifications. Dans bien des cas, la dynamique tient sur 10 dB et peut-être même moins. On tend vers une musique laminée, à relief minimal audible sans accommodation aucune, diffusée à niveau modérée (70 dB) ou moyen (85 dB) (9).

L'écoute domestique ressemble de plus en plus, il est vrai, à une *climatisation*. On y souffle l'air frais ou tiède selon la saison ou l'humeur. Le fond sonore tonique ou rêveur laisse place à toutes sortes d'activités sans empiéter trop directement sur elles. L'air est entendu et reconnu, il flotte dans la sensibilité en une sorte de zapping intermittent compatible avec la lecture ou la conversation. Pas de brusques écarts thermiques, pas d'éclats sonores : ne pas déranger le réglage auditif adapté à ce climat. Eviter les contrastes, rester sur une fourchette compatible. Pourquoi pas ? De fait, l'écoute purement musicale se fait sans doute plus rare. Elle nécessiterait un investissement auditif et émotionnel particulier, reléguant au rang de bruit de fond les autres activités de la maison.

Or, cette écoute aurait besoin d'une dynamique adéquate aux fins de préserver l'esthétique musicale, l'étagement des plans sonores, le dégagement des nuances, sans lesquels l'émotion se perd. On remarquera en passant qu'une forte compression aplatit la "lisibilité" musicale, entame l'esprit ou l'interprétation d'une œuvre... Le masquage entre instruments est accentué, les alliages et mélanges de timbres se multiplient, le phrasé perd de sa vigueur, de son élan. La vie musicale y perdrait probablement beaucoup si l'on devait contraindre les enregistrements à n'être que de pâles copies - certes bruyantes mais sans relief - et, ce faisant, des ersatz, des sons-témoins "climatisés". Heureusement, beaucoup résistent encore à ce ramollissement, à cet aplatissement dommageable.

La restitution stéréophonique a ses limites. L'ingénieur du son les anticipe, les contourne, par un savoir-faire instrumental, une écoute attentive et maîtrisée du rendu sonore. La prise de son est donc un art dans le sens où, luttant contre l'imparfait du dispositif, elle rétablit - en attendant de nouveaux progrès techniques - le présent de la musique.

La chaîne la plus fidèle, c'est peut-être finalement celle qui relie le facteur d'instruments, l'architecte, le compositeur, les interprètes, l'ingénieur du son, le constructeur d'appareils et le mélomane. Tous ont leur part dans la création ou la recréation de l'émotion musicale.

(1) Cf. article in *L'Audiophile* n°10 et 12 "Musique et architecture" à propos des qualités requises pour les salles de concert : l'impression de volume spatial, le relief, la clarté, l'amortissement, la texture, la diffusité et la couleur sonore.

(2) L'addition "transparente" de ces réflexions sur l'onde directe évite la prise en compte de nombreuses sources fantômes renvoyées par les murs environnants et, de ce fait, localisées sur les parois. 40 ms constituent donc pratiquement "l'épaisseur du présent" auditif, épaisseur sans laquelle nous serions des "chauves-souris" assaillies d'échos rapprochés et spatialement détachés.

(3) Il n'y aurait rien de pire qu'une chambre sourde pour écouter un instrument : le timbre y paraîtrait squelettique, dépourvu de corps, une sorte de figure noire et blanche posée sur un fond vide.

(4) Cf. les travaux d'Arthur Benade, mentionnés in *L'Audiophile* n°23 "La musique sur un triangle" (V)^b : p. 62 et sq.

(5) Cf. "La musique sur un triangle" (IV) in *L'Audiophile* n°21. Il faut re-

trouver les équilibres de niveaux, de timbre, de présence, voulus par le compositeur et assemblés par les inter-prètes.

(6) La perception coplanaire (2D) ne peut égaler la perception en volume (3D). Jamais peinture ne sera sculpture. "On a constaté qu'avec deux voies, écrit Günther Thiele (dans la Revue de l'UER, n°241-242, juin-août 90), les améliorations de l'image spatiale se réduisent essentiellement à une meilleure présentation bidimensionnelle de la perspective dans le plan image, entre les haut-parleurs de droite et de gauche. Il est en principe impossible d'obtenir une impression réaliste d'espace et de profondeur..." L'auteur ajoute que "l'environnement acoustique de l'auditeur est exclusivement déterminé par les réflexions et la réverbération du local d'écoute"... p. 100.

(7) Voir *L'Audiophile* n°11, 12, 13, les articles de Roy Allison sur le local

d'écoute et l'emplacement des haut-parleurs.

(8) Pour obtenir une simulation optimale de l'espace, il importe que les deux signaux des deux haut-parleurs représentent la corrélation interauriculaire la plus naturelle possible, en évitant l'attaque figée à 30°. D'où l'idée des "têtes artificielles" et autres "microphones sphériques", lesquels reconstituent et enregistrent les disparités de capture entre les oreilles, sans se contenter du simple ΔI classique... Il importe aussi que le filtrage réactif des enceintes (selfs, condensateurs) ne perturbe pas outre mesure le ΔT dans la bande 150-1.500 Hz, là où l'oreille est très sensible aux mouvements de la phase; pas plus que le ΔI , dans la bande 600-12.000 Hz, là où l'information ΔI est essentielle à la spatialisation. Le hublot serait alors "déformant", facteur de distorsion spatiale ou de décorrélation.

(9) Sur le niveau d'écoute et la dynamique, voir l'article de Pierre Loyez in *L'Audiophile* n°16 et 17. Idéalement, la dynamique du programme de-

vrait être adaptée aux conditions d'écoute (environnement calme ou bruyant, niveau d'écoute élevé ou faible, voisinage et horaire). L'auditeur qui écoute un concert à volume important dans une maison relativement isolée, peut désirer une dynamique de 55 dB (celle de l'orchestre). A niveau modéré, 35 dB suffiront, tandis qu'en milieu bruyant (la ville en est un) 25 dB seront appropriés. En voiture, pour ne pas masquer certains bruits utiles, une dynamique de 10 à 15 dB sera raisonnable. Cela n'a rien à voir évidemment avec la compression type "radio libre" à 5 dB, compression parfaitement utilitaire: une musique "audimate" encadre les annonces publicitaires et sert d'accompagnement tonique, incitatif et lisse; c'est alors une musique de fond et de commerce. Compression, compresse, cataplasme, cette même musique (beauté laminée) peut atténuer (ou rendre supportable) la violence du travail ou des transports, et servir de *Tranxène* à la fuite des jours, à la rage "automobilique" des bou-chons.