

LES TRACES D'UN SON

Laurent Demany

Laboratoire de Neurophysiologie, UMR CNRS 5543
Université Victor Segalen, 146 rue Leo-Saignat, 33076 Bordeaux
Laurent.Demany@psyac.u-bordeaux2.fr

En l'an 2200, on trouvera peut-être sur le bureau de tout bon compositeur de musique contemporaine un épais *Atlas de la Mémoire Auditive Humaine*. Imaginons par exemple qu'un de ces futurs compositeurs affectionne les bruits blancs et désire que le premier mouvement de sa nouvelle œuvre contienne un "thème" constitué par un segment arbitraire, mais fixe, de bruit blanc. Il faut qu'à l'écoute de l'œuvre, une répétition du thème soit reconnue comme telle. Quelle est donc la période *maximale* que peut avoir un bruit "blanc" – mais en fait périodique – que l'on *entend* comme périodique ? Le compositeur consultera l'Atlas pour le savoir, et verra au chapitre X que la réponse est (d'après Guttman et Julesz, 1964) : environ 1.5 seconde.

Dans le second mouvement de l'œuvre, il serait bon d'être moins austère ; pourquoi ne pas utiliser des "gammes de Shepard" ou des "glissandi de Risset", ces troublantes progressions de hauteur qui donnent l'illusion d'une montée sans fin, ou d'une descente sans fin ? On sait que les signaux acoustiques correspondants sont en réalité cycliques, autrement dit périodiques. Pour que se produise l'effet perceptif voulu – l'illusion d'une montée ou descente sans fin – il faut cette fois que la période soit *trop longue* pour être détectée auditivement. Quelle est donc la période *minimale* que l'on peut utiliser ? Là encore, espérons-le, l'Atlas donnera la réponse (qui n'est pas 1.5 seconde encore, mais davantage).

Pour l'heure, un tel Atlas n'existe pas et rien ne permet de prévoir qu'il existera bientôt. Dans leur écrasante majorité, en effet, les psychologues de la mémoire ne s'intéressent pas à la mémoire *sensorielle* (visuelle ou auditive) mais à la mémoire *verbale* (symbolique). Et parallèlement, l'écrasante majorité des psychoacousticiens ne s'intéresse pas à la *mémoire* auditive en tant que telle. C'est cependant un domaine de recherche privilégié de l'équipe de psychoacousticiens que je représente à ces Journées. Voici donc une synthèse de nos travaux sur la mémoire auditive à court terme.

Proposition 1 : il existe plusieurs "modules" ou "registres" de mémoire auditive à court terme.

Le cerveau humain dispose-t-il d'une seule mémoire auditive à court terme ? Mon équipe a fourni des arguments expérimentaux s'opposant à une telle conception. Nous avons attaqué le problème en exploitant un phénomène découvert et analysé auparavant par Diana Deutsch (voir Deutsch, 1999, pour une synthèse de son propre travail sur la mémoire

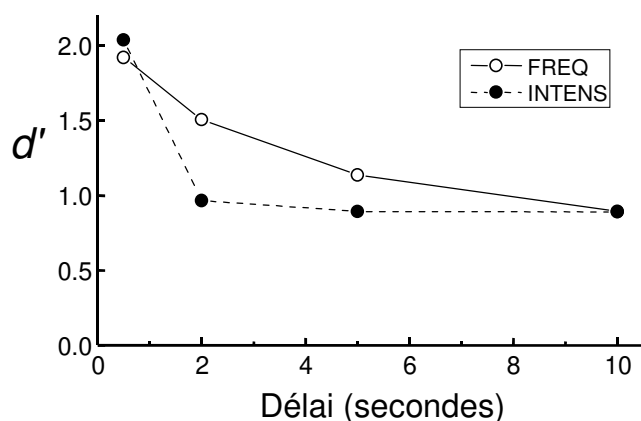
auditive). Dans plusieurs expériences de Deutsch, les sujets devaient juger si deux sons purs séparés par cinq secondes étaient identiques ou différents l'un de l'autre. Quand ces deux sons-tests différaient (ce qui arrivait en moyenne une fois sur deux), la différence consistait en un léger décalage de *hauteur* (fréquence). Au cours du délai de cinq secondes étaient présentés six autres sons purs, que le sujet était prié d'ignorer. Parfois, ces sons interposés étaient tous éloignés en hauteur des deux sons-tests ; mais parfois également, au moins un des sons interposés était proche en hauteur des sons-tests. Deutsch observa que ses sujets commettaient beaucoup plus d'erreurs dans le second cas que dans le premier. Partant de ce résultat, nous avons posé la question suivante : que se passe-t-il dans une telle expérience lorsque les sons interposés diffèrent très largement des sons-tests par le *timbre* ? Une prédiction plausible était qu'une large différence de timbre suffirait à faire disparaître l'effet néfaste d'une proximité de hauteur. Mais nous avons trouvé en fait que le timbre des sons interposés, contrairement à leur hauteur, a une influence très faible sur la performance des sujets (Semal et Demany, 1991, 1993). Ainsi, par exemple, lorsque les deux sons-tests possèdent chacun huit harmoniques d'intensités égales et de rangs 1-8, peu importe pour la performance que les sons interposés aient eux aussi cette structure spectrale ou qu'ils soient des sons purs : le facteur "hauteur des sons interposés" a un effet similaire (et important) dans les deux cas. De même, lorsque les deux sons-tests constituent un mot monosyllabique, la performance n'est pas moins bonne si les sons interposés sont eux aussi des mots que si les sons interposés ont un profil spectral plat (quelques harmoniques seulement) et ne ressemblent nullement à des signaux de parole (Semal *et al.*, 1996). Nous avons observé par ailleurs que *l'intensité relative* des sons interposés n'a pas plus d'importance que leur timbre (Semal et Demany, 1993). Ainsi, nos données dans leur ensemble ont suggéré que le système mnésique retenant le premier son-test – et particulièrement sa hauteur puisque, dans nos expériences comme celles de Deutsch, c'était seulement par cet attribut que les deux sons-tests différaient éventuellement – était sourd à tout attribut perceptif autre que la hauteur.

Qu'observe-t-on maintenant lorsque ce n'est plus par la hauteur mais par le timbre que les deux sons-tests diffèrent éventuellement ? Starr et Pitt (1997) se sont penchés sur cette question. Dans leur étude, les deux sons-tests à comparer avaient toujours la même fréquence fondamentale ; ils étaient tous deux composés d'harmoniques iso-intenses, mais les rangs de ces harmoniques pouvaient différer partiellement du premier son-test au second, ce qui créait perceptivement une différence de timbre. Les sons interposés étaient soit proches soit lointains des sons-tests, au point de vue de la structure spectrale d'une part (rangs des harmoniques), et au point de vue de la fréquence fondamentale (donc de la hauteur) d'autre part. Il est apparu que la performance des sujets dépendait fortement de la structure spectrale des sons interposés, mais ne dépendait pas de leur fréquence fondamentale. Ainsi, les résultats de Starr et Pitt ont été très comparables aux nôtres : en somme, lorsqu'il s'agit de retenir un attribut donné d'un son – sa hauteur ou son timbre – il apparaît que l'influence d'un son subséquent sur la trace mnésique du premier dépend de la similarité des deux sons *au point de vue de l'attribut à retenir*, et seulement à ce point de vue.

Ces expériences (voir aussi Krumhansl et Iverson, 1992 ; Clarke, Adriani et Bellmann, 1998) permettent de supposer que le cerveau humain contient un "module" ou "registre" de mémoire auditive spécialisé dans la rétention de la hauteur tonale, ainsi qu'un autre module ou registre spécialisé dans la rétention du timbre (ou au moins certains aspects du timbre).

Cependant, n'est-ce pas une interprétation excessive des résultats ? Ne faut-il pas penser plutôt que le cerveau ne contient aucun module *prédéterminé* de mémoire auditive, et qu'en réalité les sujets se sont adaptés aux spécificités des expériences (i.e., à la nature de l'attribut auditif pertinent) en exploitant d'une manière variable un registre mnésique unique ? A l'encontre de cette hypothèse, notons d'abord que la tâche des sujets consistait seulement à juger si les deux sons-tests étaient identiques ou différents (à n'importe quel égard) ; on ne demandait pas de qualifier les différences ; dans nos propres expériences, d'ailleurs, les sujets n'étaient même pas informés initialement de la nature des différences à détecter. Un second argument en faveur de l'idée que les modules sont prédéterminés est que la rétention mnésique d'un attribut tel que la hauteur semble être réalisée par des processus neuronaux *automatiques* (insensibles à l'attention du sujet). Je justifierai ce dernier point un peu plus loin.

Auparavant, considérons l'idée suivante : s'il existe bien un module de mémoire auditive spécialisé dans la rétention de la hauteur, et que par exemple la *sonie* (l'intensité perçue) ne peut pas être retenue par ce module, alors on peut s'attendre à ce que la trace d'une sensation de hauteur ne s'efface pas dans le temps avec la même vitesse que la trace d'une sensation de sonie (en l'absence de toute perturbation des traces par des sons "parasites" tels que ceux utilisés dans les études décrites plus haut). Partant de cette idée, Clément, Demany, et Semal (1999) ont réalisé une expérience dont les sujets devaient, à chaque "essai", comparer deux sons purs séparés par un délai silencieux d'une certaine durée : 0.5, 2, 5, ou 10 secondes. Dans l'une des deux conditions expérimentales (baptisée "FREQ"), les deux sons de chaque essai différaient par la fréquence et les sujets devaient indiquer lequel était le plus aigu. Dans l'autre condition ("INTENS"), les deux sons différaient par l'intensité et les sujets devaient indiquer lequel était le plus intense. Dans chaque condition et à chaque essai, la fréquence et l'intensité du premier son étaient tirées au hasard dans des gammes s'étendant respectivement de 1000 à 2500 Hz et de 42 à 88 dB SPL. Les différences à détecter avaient au préalable été ajustées de façon à ce que, pour le délai le plus court (0.5 seconde), la



performance des sujets – mesurée par l'indice d' de la Théorie de la Détection du Signal (cf. Green et Swets, 1974) – soit à peu près la même dans les deux conditions. La Figure ci-contre montre comment d' a évolué en fonction du délai, dans la condition FREQ et dans la condition INTENS (ce sont les résultats moyens des quatre sujets). On voit que dans la condition INTENS, d' a fortement décliné quand le délai est passé de 0.5 à 2 secondes, mais s'est maintenu sur

un plateau pour des délais plus longs. L'évolution de d' dans la condition FREQ a été significativement différente : globalement plus graduelle, et surtout moins forte entre les délais de 0.5 et 2 secondes. La divergence observée entre ces deux délais courts révèle que *l'on oublie plus vite une sonie qu'une hauteur*. Vraisemblablement, dès que le délai atteignait 2 secondes, la trace sensorielle de la sonie du premier son était si dégradée que les sujets

n'utilisaient plus cette trace elle-même dans leurs jugements d'intensité ; nous supposons que c'était plutôt un *étiquetage catégoriel* de la sonie correspondante qui était conservé en mémoire (dans un registre non auditif mais amodal, peut-être verbal) ; on peut comprendre ainsi que, dans la condition INTENS, *d'* se soit maintenu sur un plateau au-delà de 2 secondes (voir à ce propos Durlach et Braida, 1969).

Notons que si de telles données étayaient l'hypothèse selon laquelle existe un module mnésique spécialisé dans la rétention de la hauteur, elles ne permettent aucune inférence sur la manière dont la sonie est, pour sa part, retenue (en mémoire *auditive*). Il se peut qu'elle aussi soit retenue dans un module spécifique, moins efficace que celui retenant la hauteur. Cependant, on peut trouver plus vraisemblable qu'elle soit retenue seulement dans un registre non spécialisé – une mémoire "échoïque" généraliste.

Proposition 2 : une trace mnésique purement auditive se maintient de façon automatique ; sa rétention ne dépend pas de l'attention.

Pour retenir un numéro de téléphone que l'on ne peut pas noter immédiatement, il est efficace de se répéter, à voix haute ou seulement mentalement, le numéro en question. Ceci montre que dans le domaine de la mémoire *verbale*, une trace mnésique peut être "entretenue" : elle se dégrade moins rapidement lorsqu'on y prête attention. En est-il de même pour la mémoire strictement *auditive* ? Peut-on freiner le déclin de la trace sensorielle d'un son, ou d'un attribut sonore, en maintenant son attention sur cette trace ? La question a été abordée dans quelques expériences sur la rétention de la hauteur d'un son pur (Keller et Cowan, 1994 ; Keller, Cowan, et Sauls, 1995), mais les réponses fournies par ces expériences ont été contradictoires.

Mon équipe travaille actuellement sur ce sujet (Demany, Clément, et Semal, à paraître ; thèse de S. Clément en préparation). Dans l'une des expériences que nous avons achevées, les sujets devaient à chaque essai comparer deux sons complexes séparés par un délai silencieux. Ces deux signaux, variés d'un essai à l'autre, étaient des sinusoïdes modulées sinusoïdalement en amplitude, et chaque signal était défini par une conjonction de trois paramètres : la fréquence de la sinusoïde modulée, la fréquence de la modulation, et l'intensité globale. A chaque essai, les deux signaux présentés étaient différents mais ne différaient que par un seul des trois paramètres (et faiblement). La tâche du sujet était d'indiquer si, du premier au second signal, la valeur de ce paramètre était augmentée ou diminuée. Cependant, l'identité du paramètre pertinent (celui modifié au cours de l'essai) variait de façon imprévisible d'un essai à l'autre. Dans l'une des conditions expérimentales, les deux signaux n'étaient séparés que par 0.8 seconde et l'identité du paramètre pertinent était révélée au sujet avant le premier signal, par un indice visuel. Une seconde condition était identique à ceci près que l'indice visuel était présenté non pas avant mais *immédiatement après* le premier signal ; *pendant* le premier signal, par conséquent, le paramètre pertinent était inconnu et le sujet ne pouvait pas focaliser son attention sur lui autant que dans la première condition. Nous avons trouvé que la performance des sujets était un peu moins bonne dans cette seconde condition que dans la première. Ceci a montré que la *perception* et/ou la "*mise en mémoire*" du paramètre pertinent du premier signal était dépendante de l'attention. Afin de déterminer si la *rétention* de ce paramètre était également dépendante de l'attention, nous avons éloigné

temporellement les deux signaux – en faisant passer le délai de 0.8 à 4 ou 6 secondes – et comparé les performances des sujets dans des conditions où l'indice visuel était présenté : (i) immédiatement après le premier signal ; (ii) 0.8 seconde avant le second signal (donc plusieurs secondes après le premier). Dans un cas comme dans l'autre, la performance s'est avérée nettement moins bonne que lorsque les deux signaux n'étaient séparés que par 0.8 seconde. De plus et principalement, *la performance n'a pas été meilleure dans le premier cas que dans le second.*¹ Ce résultat négatif indique que la rétention du paramètre pertinent n'a pas bénéficié d'une focalisation de l'attention sur la trace correspondante.

Nous avons été surpris de ne pas observer, dans cette expérience, un effet de l'attention sélective sur la rétention. Cette absence d'effet était-elle imputable au fait que les indices visuels guidaient l'attention sur un attribut d'un seul "objet sonore", et non pas sur un objet sonore présenté en même temps que d'autres objets sonores ? Pour en juger, nous avons réalisé une nouvelle expérience où le premier signal de chaque essai était constitué par une somme de trois sinusoïdes modulées en amplitude. Chacun de ces trois sons complexes était clairement perçu comme un objet sonore séparé car il différait largement des deux autres par (i) la fréquence de la sinusoïde modulée, (ii) la fréquence de la modulation, et (iii) un attribut spatial (les trois sons étaient respectivement présentés à l'oreille gauche seule, à l'oreille droite seule, et diotiquement). Le second signal présenté à chaque essai était quant à lui formé d'une seule sinusoïde modulée en amplitude, répliquant en tout point l'une des trois composantes du premier signal (sélectionnée au hasard) à ceci près que la fréquence de la sinusoïde modulée était légèrement différente ; la tâche du sujet consistait à indiquer si ce décalage de fréquence était positif ou négatif. L'identité de la composante pertinente du premier signal était parfois révélée au sujet avant l'arrivée du second signal, par un indice visuel présenté soit immédiatement après le premier signal, soit plusieurs secondes plus tard, peu avant le second signal. La performance des sujets n'a pas été meilleure dans le premier cas que dans le second, à l'image de ce que nous avons observé dans l'expérience précédente. Ainsi, l'attention sélective ne semble pas plus utile pour la rétention d'un objet sonore parmi d'autres que pour la rétention d'un attribut donné d'un seul objet sonore.

Mémoire auditive et potentiels évoqués cérébraux

Depuis une quinzaine d'années, de nombreuses recherches concernant la mémoire auditive à court terme ont été réalisées par des psychophysiciens qui s'intéressent à un potentiel évoqué cérébral enregistrable chez l'Homme (au moyen d'électrodes posées sur le cuir chevelu) et connu sous le nom de *Mismatch Negativity* (MMN). Une MMN est typiquement évoquée par un son lorsque celui-ci a été précédé par la répétition d'un son *différent* [le son "déviant" (nouveau) et les sons précédents étant séparés par un délai qui peut atteindre plusieurs secondes]. On suppose que ce potentiel reflète l'activité de structures cérébrales qui ont pour fonction de détecter des changements dans l'environnement acoustique en effectuant des comparaisons – qui peuvent être subtiles – entre les traces mnésiques de sons successifs (voir les revues de question de Näätänen, 1990, et Schröger, 1997). Un fait très intéressant est qu'il est possible d'enregistrer une MMN chez un sujet qui ne prêche pas

¹ Dans un cas comme dans l'autre, cependant, la performance s'est avérée supérieure à celle mesurée dans une condition-contrôle où aucun indice visuel n'était présenté.

attention aux sons qu'on lui présente : on en a même enregistré chez des bébés endormis et des patients en état de coma. Dans le passé, certains auteurs en ont déduit que la détection perceptive de changements acoustiques était entièrement "pré-attentionnelle". Mais aujourd'hui, il est bien établi que l'amplitude de la MMN dépend de l'attention que porte le sujet aux sons. Tout récemment, Szymanski, Yund, et Woods (1999) ont supposé que cet effet reflète principalement un affinage par l'attention de la *mémorisation* du son répétitif précédant le son déviant (plutôt que la *perception* de ces deux sons). Or les données expérimentales que j'ai résumées dans la section précédente suggèrent, on s'en souvient, que l'attention n'affecte pas la *réten*tion auditive elle-même mais soit la perception, soit la "mise en mémoire". Ces données sont conciliables avec la supposition de Szymanski *et al.* en admettant que ce sur quoi l'attention influe crucialement est la mise en mémoire.

J'ai défendu auparavant une hypothèse selon laquelle il existe plusieurs modules de mémoire auditive à court terme, spécialisés dans la rétention d'attributs sonores distincts. Cette hypothèse est-elle étayée par la littérature relative à la MMN ? La réponse est oui, dans une certaine mesure. En particulier, une étude franco-finnoise (Giard *et al.*, 1995) a récemment révélé que les MMN respectivement induites par une modification de la fréquence, de l'intensité, et de la durée d'un son pur donné avaient des topographies un peu différentes au niveau du cuir chevelu, et émanaient de populations neuronales au moins partiellement distinctes au sein du cortex auditif. Si la génération d'une MMN prend place là où sont stockées les traces mnésiques pertinentes, les résultats de cette étude signifient que des attributs différents d'un même son ont des représentations mnésiques *anatomiquement* différentes. N'est-ce pas spécialement fascinant pour un son *pur* – le plus simple des sons sur le plan physique comme sur le plan de la perception ?

Références

- Clarke, S., Adriani, M., & Bellmann, A. (1998). Distinct short-term memory systems for sound content and sound localization. *Neuroreport*, 9, 3433-3437.
- Clément, S., Demany, L., & Semal, C. (1999). Memory for pitch versus memory for loudness. *J. Acoust. Soc. Amer.*, 106, 2805-2811.
- Demany, L., Clément, S., & Semal, C. (à paraître). Does auditory memory depend on attention? In Houtsma, A., Kohlrausch, A., Prijs, V.F., & Schoonhoven, R. (Editors): *Common Issues in the Physiology and Psychophysics of Hearing*. Whurr, London.
- Deutsch, D. (1999). The processing of pitch combinations. In Deutsch, D. (Editor): *The Psychology of Music* (2nd edition), 349-411. Academic Press.
- Durlach, N.I., & Braida, L.D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *J. Acoust. Soc. Amer.*, 46, 372-383.
- Giard, M.H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Näätänen, R. (1995). Separate representation of stimulus frequency, intensity and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *J. Cognit. Neurosci.*, 7, 133-143.
- Green, D.M., & Swets, J.A. (1974). *Signal Detection Theory and Psychophysics*. Krieger, Huntington (New York).
- Guttman, N., & Julesz, B. (1964). Lower limits of auditory periodicity analysis. *J. Acoust. Soc. Amer.*, 35, 610.

- Keller, T.A., & Cowan, N. (1994). Developmental increase in the duration of memory for tone pitch. *Dev. Psychol.*, 30, 855-863.
- Keller, T.A., Cowan, N., & Saults, J.S. (1995). Can auditory memory for tone pitch be rehearsed? *J. Exp. Psychol.: Learn. Memory Cognit.*, 21, 635-645.
- Krumhansl, C.L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *J. Exp. Psychol.: Human Percept. Perform.*, 18, 739-751.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav. Brain Sci.*, 13, 201-288.
- Schröger, E. (1997). On the detection of auditory deviations: a pre-attentive activation model. *Psychophysiology*, 34, 245-257.
- Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *J. Acoust. Soc. Amer.*, 89, 2404-2410.
- Semal, C., & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *J. Acoust. Soc. Amer.*, 94, 1315-1322.
- Semal, C., Demany, L., Ueda, K., & Hallé, P.A. (1996). Speech versus nonspeech in pitch memory. *J. Acoust. Soc. Amer.*, 100, 1132-1140.
- Starr, G.E., & Pitt, M.A. (1997). Interference effects in short-term memory for timbre. *J. Acoust. Soc. Amer.*, 102, 486-494.
- Szymanski, M.D., Yund, E.W., & Woods, D.L. (1999). Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN). *J. Acoust. Soc. Amer.*, 106, 3492-3505.