

Le son

# L'audition humaine au-delà du modèle informatique

J.-J. A.

**De la cochlée au cortex auditif, le cerveau humain semble traiter les sons selon une organisation hiérarchique qui rappelle celle des systèmes informatiques. Cette proximité conceptuelle a longtemps nourri les neurosciences comme l'intelligence artificielle. Mais les découvertes récentes sur la perception auditive révèlent une réalité bien plus complexe : l'oreille interprète déjà les sons, tandis que le cerveau les anticipe avant même de les percevoir.**

## L'audition humaine au-delà du modèle informatique

À mesure que l'intelligence artificielle progresse dans l'imitation des capacités humaines, une vieille question refait surface : le cerveau fonctionne-t-il comme un ordinateur ? De l'enthousiasme cybernétique des années 1950, quand Alan Turing, Norbert Wiener ou John von Neumann imaginaient leurs « machines à penser » en les calquant sur le cerveau, nous pensions pourtant être revenus. Depuis, nous avons en effet entendu dire - et à juste titre - que « le cerveau n'est pas un ordinateur ». Cette critique s'appuie notamment sur d'autres modèles cognitifs, comme celui des systèmes dynamiques, selon lequel l'activité cérébrale ne suit pas un simple schéma entrée-sortie, mais dépend d'états internes en évolution permanente. Ou encore sur l'énaction, qui considère que la cognition naît des interactions entre le cerveau, le corps et l'environnement, plutôt que d'un traitement abstrait de symboles. Toutefois, force est de constater que, si ces propositions théoriques avaient bonne presse, c'est qu'à l'époque, l'incapacité des ordinateurs à égaler les capacités humaines était flagrante : jusqu'au début des années 2020, ils peinaient encore à reconnaître un chat, comprendre une blague, ou distinguer un « la » d'un « le » dans le bruit d'un café bondé. Or, aujourd'hui, ils y parviennent. Dès lors, la question se pose à nouveau, de manière plus troublante : si une machine peut faire (presque) tout ce que fait un cerveau, n'est-ce pas finalement la preuve que le cerveau est, d'une certaine manière, une machine ?

## INTUITIF ET MODULAIRE, MAIS QUI A UNE FONCTION LIMITÉE

Dans le domaine qui est le mien - les neurosciences de la perception du langage et de la musique -, je constate que cette métaphore informatique (on dirait plus justement « cognitiviste ») colle encore, et peut-être même paradoxalement de plus en plus, à la peau de nos modèles.

En matière de traitement cognitif du son, un modèle fait consensus parmi les spécialistes : il le décrit comme un processus organisé hiérarchiquement. Tout commence dans la cochlée.

Cet organe situé dans la partie la plus profonde du système auditif périphérique, l'oreille interne, semble opérer un « encodage » de l'onde acoustique et la convertir en une série d'impulsions électriques. Les signaux, à travers toute une chaîne de relais dits « sous-corticaux », sont ensuite transmis au cortex auditif primaire (A1). Située en arrière (environ 2 à 3 cm) et en hauteur par rapport à l'oreille, cette zone du cerveau conçoit alors ce que nous, les neuroscientifiques, imaginons être une « représentation auditive » : une sorte de carte multidimensionnelle où les populations de neurones s'activent d'autant plus que le son possède certaines propriétés de base (plus ou moins aigu, plus ou moins brillant, etc.). Sur la base de cette représentation, on pense que le reste du cortex - en pratique, un réseau dit « temporo-frontal » qui progresse, comme son nom l'indique, de la tempe jusqu'aux côtés du front, dans les deux hémisphères du cerveau, avec certaines spécificités d'un côté ou de l'autre - construit des représentations de plus en plus élaborées, en combinant les propriétés de base pour en faire des catégories : ce son-là, c'est un /i/ prononcé par une femme; celui-ci, un /o/ grave, probablement produit par un homme.

Acquisition des données, traitement hiérarchique couche par couche, production de sens... Difficile de ne pas voir de ressemblances entre ce modèle hiérarchique et un pipeline informatique : le *hardware* passif - de l'oreille au cortex auditif primaire - fournit une information de base, que le *software* - tous les étages suivants - apprend ensuite à catégoriser. Intuitif et modulaire, ce modèle s'est d'ailleurs révélé extrêmement fécond, tant pour la compréhension du cerveau que pour le développement des technologies de reconnaissance vocale. Sauf qu'il ne fonctionne que jusqu'à un certain point.

En premier lieu, on sait aujourd'hui que la perception auditive dépend fortement du contexte et des attentes de l'auditeur. Ainsi, si quelqu'un prononce à mon attention une phrase à voix haute, on sait qu'une partie de mon cerveau prédit quels sons je vais entendre (puisque j'ai appris sur des millions d'exemples ce que cela fait de prononcer le son /ta/, par exemple), et que cela va influencer mon écoute. Certaines catégorisations, plus prévisibles, vont se faire plus facilement, et à l'extrême, je peux même croire entendre un mot que j'ai prédit, même si le son réel était différent (lire l'encadré ci-contre). Pour l'expliquer, la théorie du « codage prédictif » propose que le cerveau fonctionne comme une machine à générer des hypothèses. Il comparerait en permanence les sons réellement reçus à ceux qu'il prédit.

## LE CERVEAU INTÈGRE SIMULTANÉMENT DES COMPORTEMENTS PERCEPTIFS

Cette idée est souvent formulée dans un cadre bayésien : la perception résulterait d'un compromis entre les informations sensorielles disponibles et nos attentes préalables sur ce qui est probable dans un contexte donné **(1)**. À première vue, cette logique peut rappeler celle des modèles d'intelligence artificielle actuels, notamment les grands modèles de langage -les fameux LLM -, eux aussi fondés sur la prédiction de séquences de mots ou de sons. Mais la comparaison atteint vite ses limites. Le cerveau humain semble capable d'intégrer simultanément le contexte social, l'expérience passée, les intentions supposées d'un interlocuteur ou encore l'état du corps lui-même. Il n'est pas évident que les modèles d'apprentissage profond reproduisent des comportements perceptifs aussi souples et aussi radicaux **(2)**.

L'oreille n'est pas un simple convertisseur analogique-numérique qui se contente d'enregistrer et de transmettre passivement les sons au cerveau : elle les interprète

## LA COGNITION EST DYNAMIQUE, INCARNÉE, CONTEXTUELLE

De plus, l'idée d'un *hardware* passif (de l'oreille au cortex) commence elle aussi à se lézarder. Il est connu, par exemple, que les sons dont l'intensité augmente rapidement, donnant ainsi l'impression qu'ils se rapprochent de l'auditeur, paraissent traités de façon très particulière par le cerveau : par rapport à des sons dont l'intensité diminue, ils sont détectés plus rapidement, sont perçus comme plus forts et, de manière générale, sont associés à de plus fortes activités corticales en imagerie cérébrale. Nous voilà, semble-t-il, dans un cas typique de *hardware-software*, pour reprendre la métaphore que nous avons déjà employée : on imagine une information d'intensité représentée au niveau de AI, et un détecteur de motif qui viendrait scanner cette représentation pour y détecter des intensités croissantes au cours du temps - on oserait presque l'analogie avec un antivirus, détectant des motifs anormaux ou dignes d'intérêt.

Même si une telle implémentation est possible, elle serait en réalité particulièrement inefficace, car elle mobiliserait d'importantes ressources corticales pour une tâche importante et récurrente. De fait, une de nos études récentes a montré, en modélisant l'activité du nerf auditif juste à la sortie de la cochlée, que la quasi-totalité du travail de traitement est effectuée dans l'oreille interne : en intégrant l'intensité sonore dans le temps, notamment par le jeu de la physique exquise de cellules sensorielles que l'évolution a précisément dimensionnées à cette fin, l'oreille amplifie d'elle-même les sons d'intensité croissante **(3)**. Autrement dit, l'oreille ne se contente pas d'enregistrer et de transmettre passivement les sons au cerveau : elle les interprète déjà. Elle n'est pas un simple convertisseur analogique-numérique, mais un capteur intelligent, à qui le système nerveux central a délégué une grande partie de sa capacité de traitement.

Ainsi, l'oreille ne se contente pas d'entendre -elle écoute -, et le cortex, lui, n'écoute pas - il entend autrement selon ce qu'il cherche.

On pourrait objecter que ces nuances ne remettent pas en cause le fond du modèle hiérarchique. Après tout, même si la cochlée fait plus que ce qu'on croyait, et même si le cortex auditif primaire est moins passif qu'on ne le pensait, l'information ne remonte-t-elle pas toujours des étages inférieurs vers les étages supérieurs ? C'est une vraie question.

En science comme ailleurs, les métaphores de pensée évoluent avec les technologies qui les inspirent. L'omniprésence actuelle des grands modèles de langage et des intelligences artificielles génératives ravive une illusion tenace : celle d'un esprit humain comparable à une machine, évalué à l'aune de sa productivité

ou de ses performances. Pourtant, cette analogie, déjà dépassée par l'histoire des sciences cognitives, occulte ce qui constitue l'essence même de la cognition : sa nature dynamique, incarnée et profondément contextuelle. L'étude de l'audition en témoigne avec force : réduire le cerveau à un simple pipeline de traitement, c'est méconnaître l'interdépendance constante entre perception et anticipation, entre le corps et son environnement, entre le biologique et le culturel. Alors que les discours dominants cherchent à aligner l'humain sur la machine, les neurosciences nous rappellent que l'intelligence ne se mesure pas à ce que les algorithmes savent reproduire, mais à ce qu'ils échouent encore à saisir.

(1)L. Aitchison et M. Lengyel, *Curr. Opin. Neurobiol.*, 46, 219, 2017.

(2)Ch. Caucheteux *et al.*, *Nat. Hum. Behav.*, 7, 430, 2023.

(3)S. Benghanem *et al.*, *Cortex*, 177, 321, 2024.

**Jean-Julien Aucouturier**

### **Quand le cerveau « entend » ce qu'il s'attend à entendre**

Un exemple spectaculaire du rôle des attentes dans l'audition est celui du *sine-wave speech* (parole en ondes sinusoïdales, en français). Dans ces expériences (voir « Pour en savoir plus », dans la marge p. 45), une phrase enregistrée est transformée par un algorithme qui ne conserve que quelques variations fréquentielles simplifiées (1).

À la première écoute, le résultat ressemble à une série de sifflements électroniques presque incompréhensibles. Mais après avoir entendu la phrase originale, tout change : en réécoutant la version dégradée, la plupart des auditeurs comprennent soudain les mots sans difficulté, alors même que le signal sonore est rigoureusement identique. Ce qui a changé, ce n'est donc pas le son, mais l'état du cerveau de l'auditeur, désormais capable d'interpréter ces indices acoustiques minimaux comme de la parole. Cette illusion auditive bien connue illustre de manière frappante l'influence des attentes et des connaissances préalables sur la perception sonore.

(1)M. H. Davis et I. S. Johnsrude, *Hear. Res.*, 229, 132, 2007.

### **POUR EN SAVOIR PLUS**

Écouter d'abord le signal dégradé, puis la phrase originale; et réécouter le signal dégradé.

Signal dégradé : [tinyurl.com/signal-degrade](https://tinyurl.com/signal-degrade)

Phrase originale : [tinyurl.com/phrase-originale](https://tinyurl.com/phrase-originale)

### **Illustration(s) :**



([https://nouveau.europresse.com/WebPages/media.aspx?](https://nouveau.europresse.com/WebPages/media.aspx?Type=image&DocName=20260618SRE722044_01.jpg&Logo=sre3_small.gif)

[Type=image&DocName=20260618SRE722044\\_01.jpg&Logo=sre3\\_small.gif](https://nouveau.europresse.com/WebPages/media.aspx?Type=image&DocName=20260618SRE722044_01.jpg&Logo=sre3_small.gif)).

Cette image montre la cochlée d'une souris âgée de 6 jours, dont les cellules ciliées (plus de 3 000 ici, en vert) captent des fréquences différentes en fonction de leur emplacement et les transforment en signaux bioélectriques qui, par le biais des neurones (en rouge), sont transmis au cerveau, où ils sont décodés.

**JEFFREY HOLT, BOSTON CHILDREN'S HOSPITAL AND HARVARD MEDICAL SCHOOL**



([https://nouveau.europresse.com/WebPages/media.aspx?](https://nouveau.europresse.com/WebPages/media.aspx?Type=image&DocName=20260618SRE722044_02.jpg&Logo=sre3_small.gif)

[Type=image&DocName=20260618SRE722044\\_02.jpg&Logo=sre3\\_small.gif](https://nouveau.europresse.com/WebPages/media.aspx?Type=image&DocName=20260618SRE722044_02.jpg&Logo=sre3_small.gif)).

Jean-Julien Aucouturier NEURO-SCIENTIFIQUE, BESANÇON Directeur de recherche au CNRS, il travaille sur la modélisation informatique pour les neurosciences de l'audition à l'Institut Femto-ST (CNRS, université Marie-et-Louis-Pasteur).

**. JEAN-JULIEN AUCOUTURIER / DR**

## Alan Turing

Citation(s) : Les tentatives de création de machines pensantes nous seront d'une grande aide pour découvrir comment nous pensons nous-mêmes. - Conférence à la BBC - 15 Mai 1951 ...



© 2026 La Recherche. Tous droits réservés.

Le présent document est protégé par les lois et conventions internationales sur le droit d'auteur et son utilisation est régie par ces lois et conventions.

**news·20260618·SRE·722044**