

Examen de Perception et Psychoacoustique

Session d'examens 2022-2023, tous documents autorisés

I. « A Task-Optimized Neural Network Replicates Human Auditory Behavior ». Analyse de l'article de Kell et al. (Neuron, 2018) [13 pts]

Un nombre croissant d'études dans le domaine de la psychoacoustique et des neurosciences s'intéressent à la simulation du fonctionnement du système auditif humain au moyen de « réseaux de neurones artificiels », un type particulier de modèle du système auditif central. En 2018, l'équipe de recherche d'Alexander Kell a comparé les performances de compréhension de mots pour un groupe de sujets humains et pour un réseau de neurones artificiels, les deux étant testés sur la même expérience psychoacoustique. Voici un résumé de la description du protocole fournie par les auteurs et autrices de l'étude :

Les performances de reconnaissance de mots ont été mesurées dans 26 situations différentes : 5 types de bruit masquant (musique, bruit ambiant d'un restaurant, brouhaha de 2 voix, brouhaha de 8 voix, bruit blanc) présentés chacun à 5 intensités (rapport signal sur bruit = -9, -6, -3, 0, +3 dB SNR), plus une condition sans bruit masquant (notée SB). La session expérimentale était divisée en une série d'essais successifs. Chaque essai consistait en un extrait d'un enregistrement de parole d'une durée de deux secondes masqué par un bruit de fond d'un type et d'une intensité donnée. Les participants et participantes devaient entrer au clavier le mot-cible situé au milieu de l'extrait (voir Figure 1). Au total, 587 enregistrements de parole ont été utilisés comme stimuli, ce qui fait de ce paradigme un choix forcé à 587 alternatives (587-AFC). Les performances des sujets humains et du réseau de neurones artificiels étaient caractérisées par la proportion moyenne de mots correctement identifiés (Figure 2).

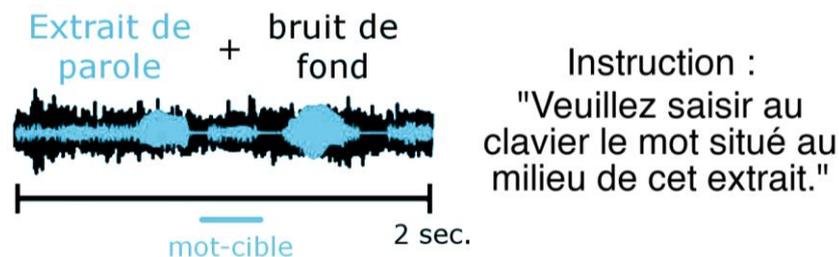
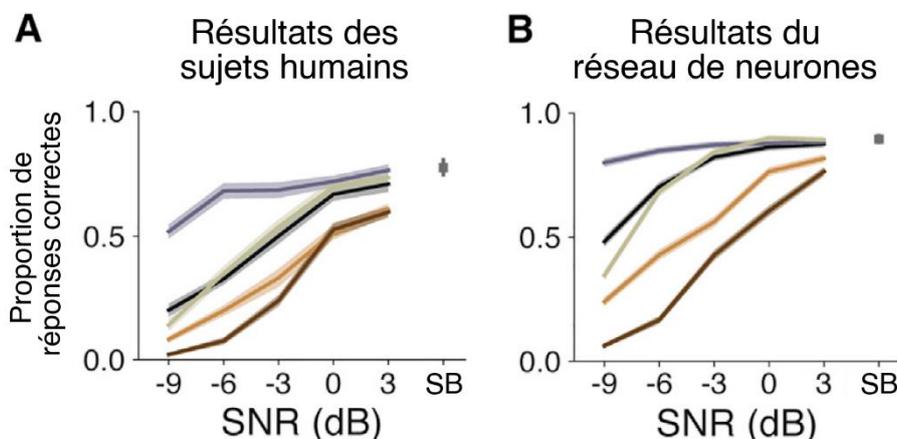


Figure 1. Exemple d'un essai présenté lors de l'expérience



Types de bruit masquant : musique, bruit ambiant, bruit stationnaire, brouhaha de 2 voix, brouhaha de 8 voix.

Figure 2. Résultats de l'expérience, pour les sujets humains (A) et pour le réseau de neurones artificiels (B).

Le barème des points est donné à titre indicatif. Veillez à justifier vos réponses de façon concise et à détailler votre réflexion en utilisant les concepts vus en cours.

Questions :

- 1) Quelle est la tâche psychophysique demandée aux participants et participantes ? [0.5 pt]

Tâche d'intelligibilité car les participants doivent répéter un mot (et non indiquer sa présence ou son absence, ce qui aurait correspondu à une tâche de détection)

- 2) Quelle est la dimension de variation des stimuli ? Justifiez. [1 pt]

Beaucoup de caractéristiques des stimuli varient au cours de cette expérience, en particulier le type de bruit, mais il n'y a **qu'une dimension de variation : le SNR**. On peut s'en rendre compte grâce à la Figure 2 qui présente le résultat de l'expérience sous la forme de fonctions psychométriques en fonction du SNR. C'est donc bien que le SNR est le paramètre physique varié dans cette expérience.

- 3) Quelle méthode psychophysique a été utilisée dans cette étude ? [0.5 pt]

Cette expérience s'appuie sur la **méthode des stimuli constants**. En effet il s'agit bien d'une méthode non-adaptative puisque le stimulus présenté à l'essai n ne dépend pas de la réponse donnée à l'essai n-1. De plus, comme vu à la question 2, la performance est mesurée sur une gamme de niveaux de stims (SNR), et non sur un niveau unique. Finalement, le résultat obtenu pour chaque participant est une fonction psychométrique (Figure 2) – or seule la méthode des stimuli constants permet d'obtenir une caractérisation complète de la fonction psychométrique.

- 4) Concernant le paradigme expérimental, l'article le décrit comme un « choix forcé à 587 alternatives ». Cette formulation a néanmoins été critiquée par d'autres scientifiques. Que pouvez-vous en dire ? [1.5 pts]

Il est erroné d'envisager cette expérience comme un 587-AFC pour plusieurs raisons. Tout d'abord, elle est basée sur une tâche d'intelligibilité qui n'appelle donc pas de paradigme expérimental. Plus fondamentalement, un 587-AFC est une expérience dans laquelle le participant **écoute une série de 587 stimuli et doit en sélectionner un** – ce qui ne correspond pas à l'expérience décrite ici.

- 5) Comment nomme-t-on les courbes représentées en Figure 2 ? Quelles informations peut-on extraire de chacune de ces courbes ? [1 pt]

Il s'agit de **fonctions psychométriques**. Elles permettent de caractériser complètement la perception sur la tâche au moyen de deux valeurs : la pente (ou **sensibilité**) et le **seuil** à 50%. On peut également en extraire le seuil de perception pour n'importe quel pourcentage de réponse.

- 6) Donnez la valeur des seuils à 50% pour les sujets humains dans les différents types de bruit. Décrivez en une phrase la signification de cette mesure. [1.5 pts]

Il suffit de lire sur la Figure 2A à quel SNR chacune des courbes croise Proportion de réponse correcte = 0.5.

Musique : $SNR_{50\%} = -9$ dB

Bruit ambiant : $SNR_{50\%} = -3$ dB

Bruit stationnaire : $SNR_{50\%} = -3$ dB

Brouhaha 2 voix : $SNR_{50\%} = 0$ dB

Brouhaha 8 voix : $SNR_{50\%} = 0$ dB

Ces valeurs de SNR correspondent au niveau de bruit dans lequel les participant.e.s sont capables de réaliser la tâche en répondant correctement sur 50% des essais.

- 7) Pouvez-vous proposer une explication pour la différence observée chez les participants humains entre la perception des mots dans un bruit stationnaire (bruit blanc) et dans le brouhaha ? [1 pts]

Plusieurs explications étaient envisageables. Primo, il est possible que les participants **confondent certains mots présents dans le brouhaha** avec le mot-cible à détecter, ce qui n'est pas le cas avec le bruit blanc. Deuxio, les deux bruits **couvrent des fréquences différentes** et il est probable que le brouhaha qui a un spectre similaire à celui de la parole masque plus efficacement les mots-cibles. Pour confirmer cette interprétation, on pourrait mesurer les patterns d'excitation.

- 8) En supposant que les scientifiques souhaitent mesurer uniquement les seuils à 50%, proposez une modification du protocole visant à obtenir plus rapidement cette même valeur. [2 pts]

Il faut pour cela **changer la méthode psychophysique**. Au lieu d'une expérience en stimuli constants, on utilise un escalier psychophysique. Ceci consiste simplement à choisir le niveau SNR à l'essai n en fonction des réponses aux essais précédents, selon une règle précise. Pour viser le seuil à 50% la règle à utiliser est un « 1-up-1-down ».

- 9) Quelles sont les similarités entre les performances des humains et celles du réseau de neurones artificiels ? Quelle est la principale différence ? [2 pts]

Pour juger de la ressemblance entre la perception humaine et notre modèle, on se base sur la comparaison des fonctions psychométriques en termes de pente et de seuil à 50%. A première vue la pente de chaque courbe est relativement similaire entre les figures 2A et 2B, ce qui indique que le réseau de neurones a la **même sensibilité** que l'humain. En revanche, **les seuils à 50% sont bien meilleurs** pour le réseau de neurones que pour l'humain. Le réseau est donc capable de reconnaître les mots dans des niveaux de bruits globalement plus élevés. Finalement, **l'effet du changement de type de bruits sur la perception est identique** entre le réseau de neurones et l'humain (i.e., les courbes sont dans le même ordre sur les deux figures).

- 10) Le réseau de neurone artificiel est un modèle du système auditif central. Contrairement aux participants humains, il n' « écoute » pas les stimuli présentés à chaque essai, mais effectue des calculs à partir d'une représentation numérique de ces stimuli, choisie par l'expérimentateur. Quel type de représentation numérique faudrait-il utiliser dans ce cas ? Justifiez. [2 pts]

Puisque le réseau de neurone artificiel est un modèle du système auditif central, il faut pour obtenir un modèle complet du système auditif lui adjoindre un modèle du système auditif périphérique. Cela peut être fait très simplement en fournissant en entrée du réseau de neurone la sortie d'un des modèles vus au cours 2, par exemple le modèle du « **spectrogramme auditif** » qui est le plus complet et permet notamment de représenter à la fois l'information temporelle et l'information tonotopique.

II. Perception du timbre de différents pianos viennois [7 pts]

Une équipe de recherche souhaite étudier les similarités ou les différences de timbre entre sept modèles de piano au moyen d'une expérience psychoacoustique. Elle dispose pour cela de nombreux enregistrements de chaque note isolée pour chaque piano. L'objectif est d'obtenir la mesure la plus objective et la plus indépendante possible des biais éventuels des sujets, même au prix d'une expérience relativement longue et coûteuse. Quelle **tâche**, quels **stimuli** et quel **paradigme expérimental** recommanderiez-vous dans ce cas ? Justifiez en présentant les avantages de votre protocole et en explicitant le déroulement d'un essai.

Les seules contraintes indiquées dans l'énoncé pour l'élaboration de cette expérience sont :

- On s'intéresse au timbre, il faut donc essayer d'isoler ce percept des autres attributs perceptifs.
- Rendre l'expérience aussi objective que possible, c'est-à-dire dans l'idéal « criterion-free » et implicite.

Ces trois points (isoler le timbre + indépendance du critère + instruction implicite) vont nous permettre de déterminer quels stimuli, quelle tâche et quel paradigme utiliser.

On se base naturellement sur les enregistrements de piano préenregistrés pour nos **stimuli**. L'expérience porte sur la perception du timbre. Il faut donc **s'abstraire des deux autres attributs auditifs que sont la hauteur et la sonie**, qui risquent d'interférer avec le timbre. Pour cela, une solution simple consiste à comparer toujours les différents pianos sur une même note et en normalisant l'enregistrement en puissance. Ainsi, pour chaque essai, on présentera des enregistrements de la même note à la même puissance ce qui permettra que la tâche ne repose que sur une différence de timbre. En revanche, d'un essai à l'autre, la note et la puissance peuvent être différentes.

On souhaite une mesure indépendante des biais éventuels du sujet, ce qui nous invite à choisir un **paradigme du choix forcé** qui est indépendant du critère. Ainsi par exemple l'éventuelle préférence du participant ou de la participante pour le son d'un piano particulier n'interférera pas avec la mesure.

De plus, on préférera une **tâche de discrimination** pour son caractère implicite : l'expérimentateur n'a pas besoin de mentionner le nombre ou le nom des pianos, ni de préciser ce qu'il veut dire par « timbre ».

Au final le déroulement d'un essai est donc le suivant : une note de la gamme et un niveau sonore sont sélectionnés (ils peuvent être différents pour chaque essai) et on présente 3 stimuli successifs, deux enregistrements issus d'un même piano et un enregistrement d'un second piano, dans un ordre aléatoire. Il est demandé au participant de « sélectionner le son qui est le plus différent des deux autres » ou plus simplement « sélectionner le son qui correspond à un piano différent ».

Ce protocole risque d'être long et fatiguant puisque chaque essai nécessite d'écouter 3 sons puis d'en choisir un – une tâche relativement complexe. Néanmoins d'après l'énoncé ceci n'est pas une contrainte dans notre cas. L'avantage majeur de ce protocole en revanche est qu'il est totalement indépendant du critère et implicite. Notre mesure ne sera pas impactée si les participants ont différentes conceptions de ce qu'est le « timbre » (puisque on ne leur demande pas explicitement de juger le timbre) ni s'ils ont des biais de jugement (par exemple une préférence pour le son d'un des pianos).

Notez que l'énoncé ne demandait pas de choisir une méthode psychophysique car il n'y a pas vraiment d'argument ici en faveur d'une méthode ou d'une autre.