



HAL
open science

Du spectrogramme au cochléogramme - Intérêt pour la clinique orthophonique

Etienne Sicard, Anne MENIN SICARD, Manon Germain

► **To cite this version:**

Etienne Sicard, Anne MENIN SICARD, Manon Germain. Du spectrogramme au cochléogramme - Intérêt pour la clinique orthophonique. 2025. <hal-05048910>

HAL Id: hal-05048910

<https://hal.science/hal-05048910v1>

Submitted on 28 Apr 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Du spectrogramme au cochléogramme – Intérêt pour la clinique orthophonique

Etienne Sicard^{1,2}, Manon Germain¹, Anne Menin-Sicard^{1,3}

(1) Laboratoire LURCO/UNADREO, équipe de recherche ERU 46

(2) INSA-GEI, 135 Avenue de Rangueil, 31077 Toulouse

(3) Formatrice voix et parole pour orthophonistes, 81390 PUYBEGON

Résumé : Le spectrogramme est un assistant précieux, voire indispensable pour la prise en charge clinique des pathologies de la voix et de la parole. Cependant le calcul basé sur la conversion temps/fréquence ne reproduit que d'une manière imparfaite les caractéristiques spécifiques de la cochlée. Le but de cet article est de mettre en évidence les phénomènes non linéaires de la perception fréquentielle de l'oreille humaine et les reproduire sur une version modifiée du spectrogramme que nous appelons cochléogramme. Nous discutons de l'utilité d'une telle représentation en pratique clinique, notamment pour le bilan et la prise en soin des pathologies de la voix et la parole.

Mots clé : orthophonie, évaluation de la parole, altération de la voix, altération de la parole, pathologies de la voix, spectrogramme, temps réel, cochlée, cellules nerveuses, sensibilité, cochléogramme, VOCALAB, DIADOLAB.

Introduction

Parmi les outils objectifs les plus emblématiques d'aide à la prise en soin des pathologies de la voix et de la parole figurent le spectrogramme. Cet outil est notamment disponible pour la clinique orthophonique grâce aux logiciels VOCALAB [Gerip 2020], DIADOLAB [Menin Sicard 2019], PRAAT [Boersma 2018]. Une caractéristique commune de ces outils est de proposer une représentation de la production sonore sous forme de spectrogramme, noyau fondamental d'un véritable écosystème d'analyse de la voix et de la parole.

Le spectrogramme est une représentation permettant de voir l'ensemble de la décomposition spectrale de la voix et de la parole sur une même représentation graphique. Cet outil se révèle précis, informatif et fiable pour analyser les caractéristiques de la production sonore. Le spectrogramme a beaucoup intéressé la communauté scientifique ces dernières années comme un champ d'application possible de la reconnaissance des formes, en vue de l'aide au diagnostic. L'analyse des contrastes, profils, changements de formes dans les spectrogrammes permet par exemple d'identifier la dysarthrie parkinsonienne [Zahid 2020], le *Down Syndrome* [Jeffery 2018], ou plus généralement les troubles spécifiques du langage [Sharma 2020],.

Nous avons consacré un article spécifique sur l'intérêt du spectrogramme pour l'orthophonie [Sicard 2021]. Nous avons rappelé les bases de la décomposition temps/fréquence/énergie du signal sonore d'origine, qui reproduit de manière simplifiée le fonctionnement de la cochlée. Le spectrogramme fournissant une représentation visuelle plus ou moins comparable à l'influx nerveux circulant dans le nerf auditif de la cochlée vers le cortex. Toutefois, plusieurs paramètres importants ne sont pas pris en compte dans la création du spectrogramme : la répartition non linéaire des fréquences, une courbe de sensibilité de l'oreille qui varie en fonction de la fréquence, une stimulation particulièrement sensible aux variations d'intensité et non simplement proportionnelle à la pression acoustique, etc.

Dans cet article, nous nous intéressons à la construction d'une représentation du spectrogramme intégrant différents phénomènes physiologiques afin de s'approcher de ce qui est véritablement transmis au cortex. La représentation est significativement différente du spectrogramme d'origine, nous l'appelons donc cochléogramme. Dans la section II, nous rappelons les principes du passage temps/fréquence, puis nous décrivons les phénomènes qui amènent à modifier les calculs initiaux, et donnons les courbes ou ordres de grandeur de ce qui doit être modifié. Nous proposons à la section III une implémentation du cochléogramme destiné à être utilisé pour la clinique orthophonique en tant qu'outil d'imagerie de la phonation non invasif. Nous comparons le spectrogramme et le cochléogramme pour montrer en quoi ce dernier peut être intéressant dans le cadre de l'évaluation et de la prise en soin des pathologies de la voix et de la parole.

Passage temps fréquence

Le passage d'un signal temporel $x(t)$ à sa décomposition fréquentielle $X(f)$, encore appelée conversion temps/fréquence, repose sur l'équation dite de « Transformée de Fourier » (TF) qui s'écrit de manière théorique sous la forme de l'équation (1).

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-2j\pi ft} dt \quad (\text{Eq. 1})$$

En pratique, l'implémentation de cette équation pour le traitement du son issu du microphone repose non pas sur un signal infini mais sur une portion de signal échantillonné de quelques centaines à quelques milliers de points, d'une durée de l'ordre de 10-100 ms. L'intégrale est approximée par une somme et la partie exposant, la plus difficile à calculer, bénéficie d'un algorithme d'accélération appelé « *Fast Fourier Transform – FFT* ». Ces approximations et astuces permettent de calculer les composantes fréquentielles en temps réel à l'aide d'un ordinateur ordinaire.

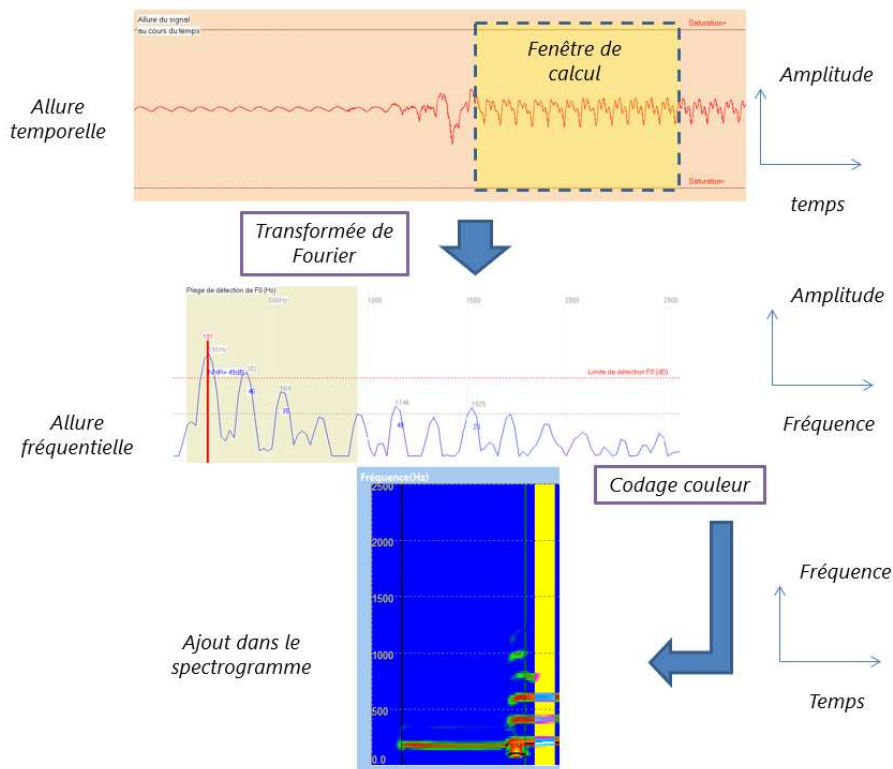


Figure 1 : Principe de construction du spectrogramme d'après le signal temporel de la parole capturé par le microphone [Sicard 2017]

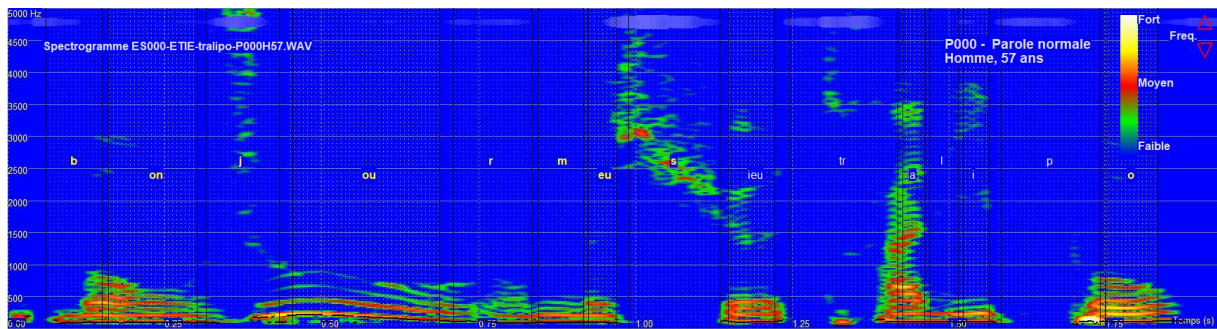


Figure 2 : Bonjour Monsieur Tralipau, spectrogramme linéaire en temps (Axe horizontal), en fréquence (Axe vertical) de 0 à 5000 Hz, et logarithmique en intensité (Echelle de couleurs).

Le dessin du spectrogramme repose sur le calcul itératif des composantes fréquentielles de portions de son, encore appelé « spectre », et ce jusqu'à convertir en une seule image l'ensemble de la production sonore. Chaque conversion porte sur une petite portion du signal sonore, que l'on décale progressivement dans le temps. Le résultat de chaque conversion est un tableau de valeur d'énergie fonction de la fréquence, que l'on reporte sous forme d'une barrette verticale de couleurs (Figure 1).

Environ 1000 conversions temps/fréquence ont été nécessaires pour calculer le spectrogramme de la figure 2. L'énergie importante est associée à une couleur chaude (jaune-blanc), une faible énergie à une couleur froide (vert-bleu). Le spectrogramme de la parole utilise une échelle de couleurs non linéaire, de type logarithmique. L'échelle en log de la palette des couleurs (du bleu au blanc) reproduit la sensibilité de l'oreille à la pression acoustique qui est aussi en log, permettant ainsi de s'approcher de la sensibilité de l'oreille en termes de perception de la puissance sonore.

Construction du cochléogramme

Répartition logarithmique des fréquences

Le premier défaut de la représentation du spectrogramme de la figure 2 est l'échelle linéaire utilisée pour les fréquences. La perception de l'oreille de la notion de hauteur (axe vertical du spectrogramme) n'est en effet pas linéaire mais logarithmique. La conversion temps/fréquence de la cochlée repose sur le principe d'une membrane résonante à une fréquence variable selon la position de la base (fréquences hautes) à l'apex (fréquences basses) par un changement de propriétés mécaniques, du fait de la largeur et épaisseur variable de la membrane. La répartition logarithmique des fréquences de résonance est illustrée à la figure 3 : par portion de 5 mm environ le long de la membrane en partant de la base, on divise la fréquence de résonance par 2.

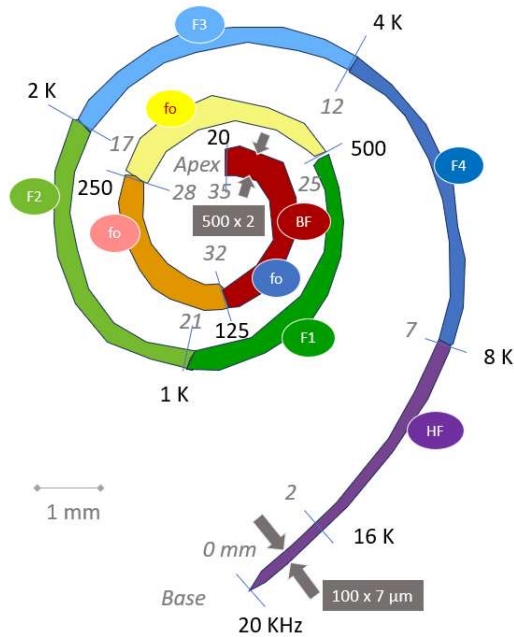
Nous avons ajouté à la figure initiale inspirée des images tri-dimensionnelles de [Li 2021] la zone approximative des formants et des zones de détection de fo pour hommes, femmes et enfants. On constate une remarquable concordance de dimension des différentes composantes de la parole, ce qui se traduit par une répartition assez semblable des fibres du nerf cochléaire pour chaque zone de fo et chaque formant.

Fréquences de résonance de la membrane basilaire dans la cochlée, vue de dessus

Distance de la base (mm)
Fréquence de résonance (Hz)

500 x 2 Largeur x épaisseur (µm) membrane basilaire

- F4 Formant F4
- F3 Formant F3
- F2 Formant F2
- F1 Formant F1
- fo enfant
- fo femme
- fo homme



D'après Li, H., (2021). Three-dimensional tonotopic mapping of the human cochlea based on synchrotron radiation phase-contrast imaging. Scientific reports, 11(1), 4437.

Figure 3 : Répartition non linéaire de la résonance fréquentielle de la membrane basilaire d'après [Li 2021]

Nous modifions l'échelle verticale correspondant aux fréquences, qui au lieu d'être linéaire, devient logarithmique comme illustré figure 4. Cette représentation que nous appellerons « spectrogramme mélodique » permet de mieux apprécier le contour prosodique de fo (1^{er} trait harmonique horizontal en partant des basses fréquences) et de compacter l'énergie du bruit haute fréquence (comme le /s/ au temps 1.0s).

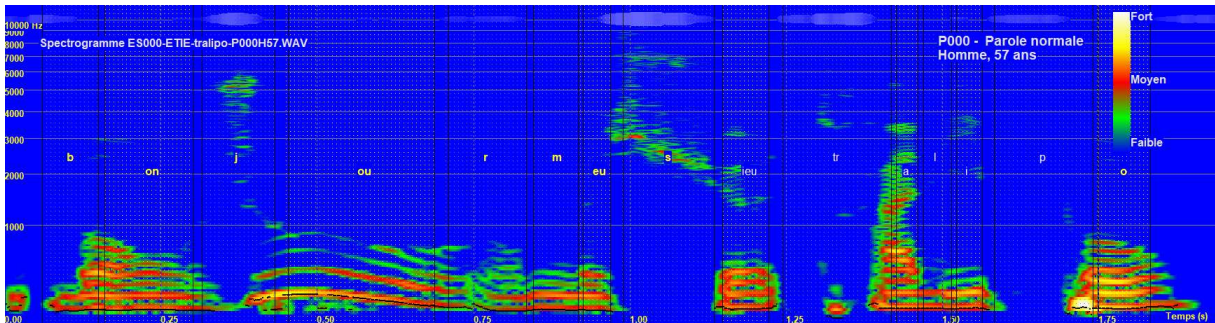


Figure 4 : Bonjour Monsieur Tralipau, spectrogramme mélodique de 0 à 10000 Hz, homme 57 ans normo-phonique. L'échelle temporelle est linéaire (Axe des X), logarithmique en fréquence (Axe des Y) et en logarithmique en intensité (Echelle de couleur).

Courbe de sensibilité de l'oreille

Les courbes isosoniques représentent des niveaux de pression en dB provoquant la même sensation d'intensité sonore pour l'oreille humaine, en faisant écouter des sons purs [Wikipedia 2024]. La fréquence varie de 20 à 15 000 Hz, pour des pressions acoustiques de 0 à 120 dB. La sensibilité du système auditif humain est très faible à 20 Hz, remonte graduellement jusqu'à 300 Hz, atteint un plateau relativement constant jusqu'à 1000 Hz (Figure 5). Du fait de la résonance du canal auditif, la sensibilité est maximale autour de 3000 Hz, puis diminue graduellement avec la fréquence.

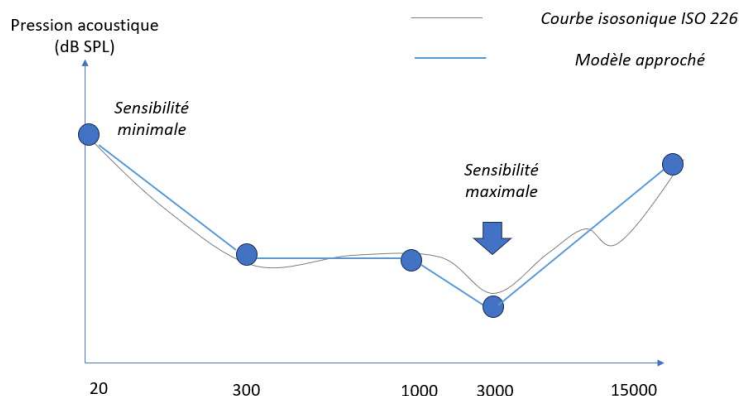


Figure 5 : courbe isotonique liant la pression acoustique d'un son pur et sa fréquence pour une perception de puissance sonore équivalente [ISO 226].

Nous avons implémenté une loi reproduisant la courbe de sensibilité de l'oreille telle que définie dans le standard ISO 226, atténuant ainsi les très basses et très hautes fréquences, et rendant légèrement plus proéminente la zone autour de 3000 Hz. Ces modifications influent cependant à la marge l'aspect général du spectrogramme, même si l'on voit mieux apparaître dans le spectrogramme certaines zones d'énergie peu visibles (1000-5000 Hz) comme certains composants du /on/ (Figure 4).

Sélectivité fréquentielle

Les cellules nerveuses que l'on trouve dans la cochlée et dans le nerf auditif ont des aspects et des comportements distincts. On considère en particulier 3 types de cellules, *bushy*, *chooper* et *small cells*. Les *bushy cells* ont des dendrites courtes avec de nombreuses petites ramifications d'où le surnom « buisson » ou « touffe ». Ces ramifications pourraient permettre de se connecter aux informations fréquentielles des fibres adjacentes, et par effet d'inhibition d'améliorer l'accord fréquentiel. Cette propriété composerait ainsi la nature relativement peu sélective de la vibration mécanique de la membrane basilaire [Meaud 2010], qui ne correspond pas à un point singulier mais à une zone relativement large à l'échelle des cellules cillées. Un gain de 50 dB et une allure relativement étroite de la résonance sont illustrés dans [Cochlea.eu].

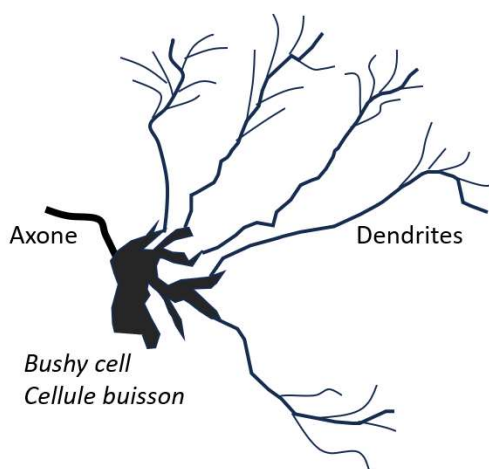


Figure 6 : Exemple de cellule buisson (Bushy cell) d'après [Oertel 2000]

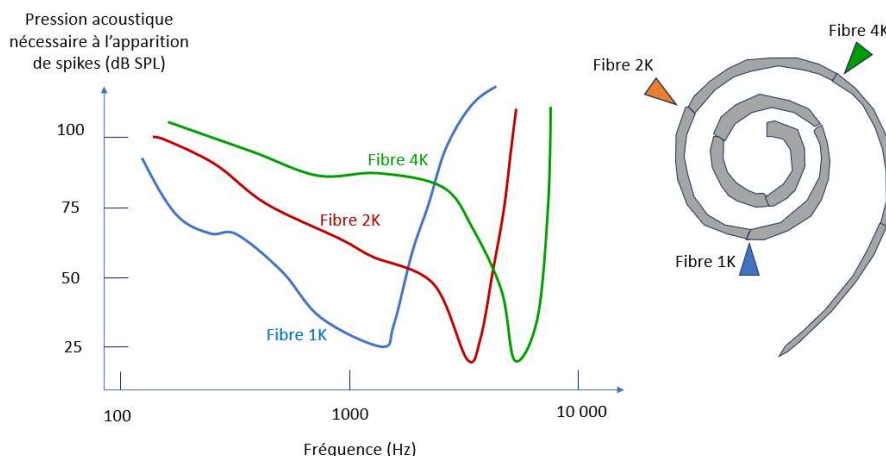


Figure 7 : sélectivité fréquentielle des fibres nerveuses, d'après [Heil 2015]

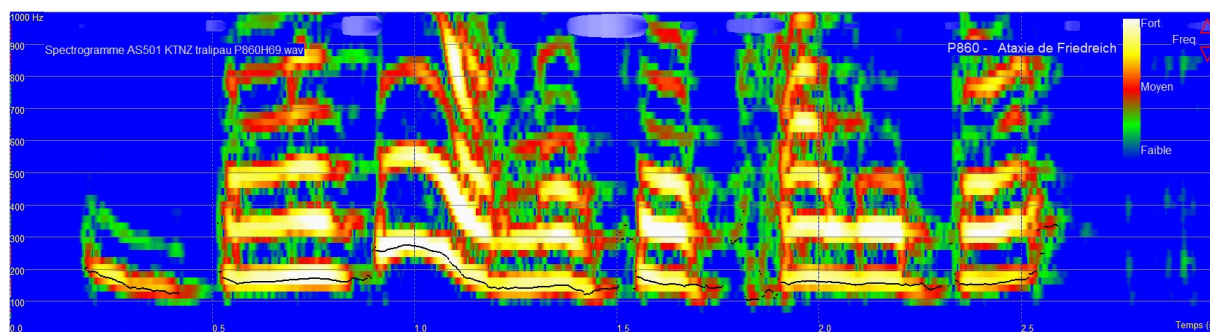


Figure 8 : extraction de f_0 avec une précision de l'ordre du Hz sur la base d'une moyenne pondérée des composantes issues de la FFT, d'une résolution de 20Hz environ, donnant à f_0 une trajectoire continue (trait noir superposé à l'harmonique la plus basse en fréquence).

Le lien entre la cochlée et le cerveau est établi à l'aide d'environ 32 000 fibres nerveuses myélinisées qui composent le nerf cochléaire [Spoendlin 1989], chacune ayant une certaine sélectivité fréquentielle, à laquelle une faible pression acoustique (autour de 25 dB) suffit à générer l'apparition de spikes. Les courbes de réponse sont plutôt sélectives, surtout à partir de 2 KHz, et présentent une certaine asymétrie : les pentes de descente sont plus faibles que les pentes de remontée de part et d'autre de la fréquence de sensibilité maximale.

Les spectrogrammes tels que ceux calculés par VOCALAB, DIADOLAB ou PRAAT disposent d'environ 1000 composantes fréquentielles sur une plage comparable à la sensibilité de la cochlée (10 Hz-10 KHz), soit une sélectivité 30 fois moins élevée que le nerf auditif. Ceci permet de limiter la lourdeur des calculs et de permettre un affichage en temps réel des composantes de la parole, un atout déterminant pour la prise en soin. Par contre, le spectrogramme est relativement grossier en termes de résolution fréquentielle, par rapport à la perception humaine.

Il existe cependant un intérêt clinique à connaître la fréquence fondamentale f_0 avec une précision de l'ordre du Hz, soit pour la situer par rapport à une norme [Sicard 2023], soit pour mesurer des progrès en lien avec le travail de l'intonation, du rendement vocal, des contours prosodiques et des accents toniques. Pour ce faire, nous faisons appel à une moyenne pondérée des contributions spectrales (dont la résolution fréquentielle est de l'ordre de 20 Hz), pour atteindre une résolution autour du Hz. Cette moyenne pondérée peut être vue comme la version informatisée du processus d'identification de résonance mécanique maximale notamment effectuée par les *bushy cells*, afin d'améliorer la sélectivité fréquentielle des fibres du nerf auditif.

Sensibilité aux variations de fréquence et d'intensité

Certains neurones tels que les *Small Cells* semblent ne servir qu'au simple transport de l'information sonore le long du nerf auditif [Hockley 2022]. On n'observe pas de déformation notable dans le transport du signal, grâce notamment aux axones myélinisés qui ont très peu de fuites. Certaines cellules nerveuses sont en revanche particulièrement sensibles à l'amorçage du signal ainsi qu'à ses variations, tandis qu'elles sont moins actives une fois le signal stable et établi. Leur comportement est donc fortement non linéaire.

Ceci est particulièrement vrai pour les *bushy cells* et *T-stellar/Chopper cells* (Figure 9 au centre et à droite), [Hockley 2022]. A noter que les *chopper cells* (Figure 9 à droite) ont la réponse la plus complexe à l'onde sonore incidente, qui est pourtant ici un simple signal pur d'amplitude constante, d'une durée de 50 ms.

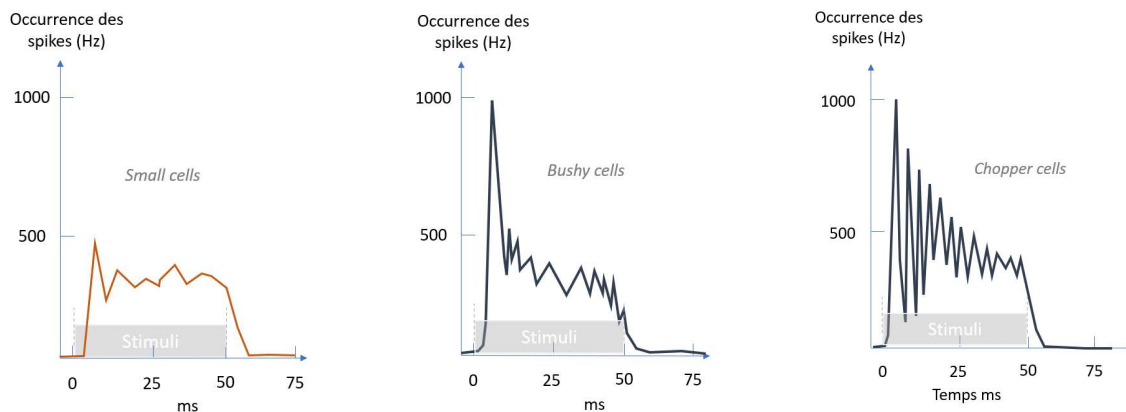


Figure 9 : Fréquence des impulsions en fonction du début et de la fin du son pur de 50 ms appliqué en entrée de la cochlée d'un cochon d'inde, selon le type de cellule analysée, d'après [Hockley 2022]

Les neurones dénommés "*T-Stellate*", "*Multipolar*" ou "*Chopper Cell*" constituent d'ailleurs l'une des principales voies auditives ascendantes de la cochlée vers le cortex. Leurs dendrites sont plus longues que pour les *Bushy cells*. On les appelle « Chopper – Hacheur », en référence à leur capacité à déclencher un train d'impulsions régulièrement espacées, à cadence élevée, comme illustré figure 10.

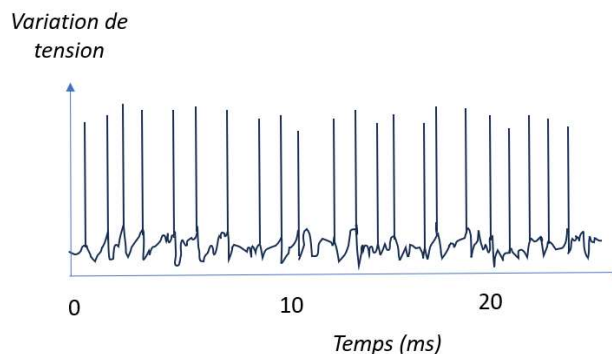


Figure 10 : Train d'impulsions très rapide à une fréquence d'environ 1000 évènements par seconde, soit un toutes les 1ms

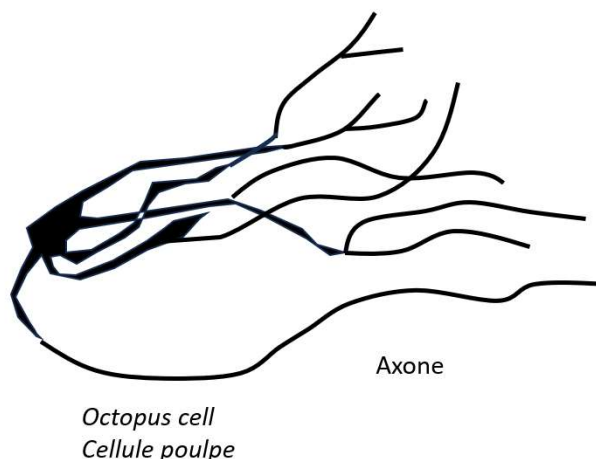


Figure 11 : exemple de cellule poulpe (Octopus cell) d'après [Oertel 2000]

Il existe aussi une catégorie de neurones qui s'activent principalement au démarrage d'une stimulation fréquentielle. Les cellules « Octopus » situées dans le noyau cochléaire ont des dendrites longues en forme de tentacule (Figure 11) qui ne réagissent qu'au début d'une stimulation sonore.

Pour implémenter ces différents effets, nous avons l'obligation de faire une modification fondamentale dans le spectrogramme : se focaliser plus sur les changements d'énergie, notamment au démarrage et moins prendre en compte les composantes stables. Autrement dit, nous passons d'une approche linéaire de la Transformée de Fourier à une approche non linéaire, en particulier à l'initialisation de l'onde sonore.

L'implémentation de la sensibilité aux variations peut se faire en comparant deux échantillons de composantes spectrales à des temps consécutifs t et $t-1$, pour une même fréquence f . On étalonnera la constante k de manière à produire un supplément d'excitation de l'ordre de 3 fois l'excitation continue telle que délivrée par les *small cells* (Voir figure 3).

$$spikes_{transitoire}(Hz) = k \times (dB_{SPL}[t, f] - dB_{SPL}[t - 1, f])$$

Nous avons donné une importance particulière aux changements de valeur d'énergie pour une fréquence donnée, entre deux moments consécutifs au cours du temps (axe horizontal). C'est ainsi que les phénomènes transitoires tels que les explosions ou les changements de fréquence en lien avec la prosodie sont mis en valeur, au détriment des portions stables en énergie et en fréquence, qui ont tendance à être négligées. La transition /b/on/ est particulièrement visible, de même que les composantes hautes fréquence du /j/ ou encore l'explosion du /p/ (Figure 12).

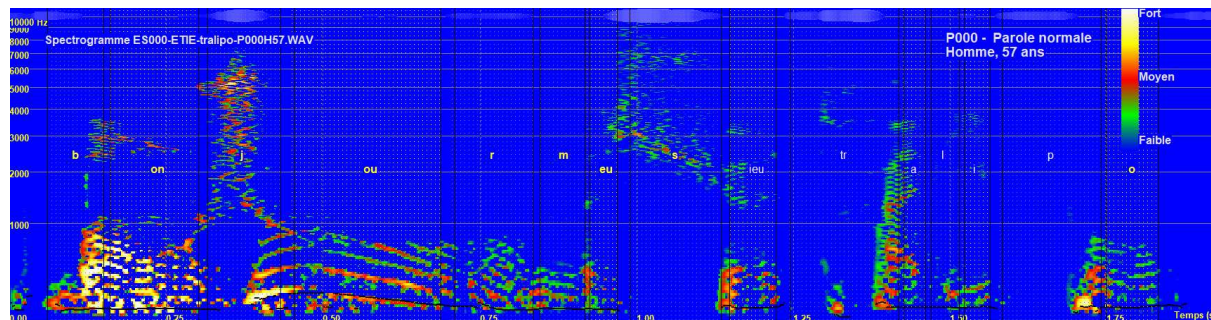


Figure 12 : Bonjour Monsieur Tralipau, cochléogramme de 0 à 10000 Hz

Conversion dB/spike

Nous nous posons enfin la question de la pertinence d’une représentation en échelle logarithmique de type dB, même si celle-ci semble proche de la perception de l’oreille humaine. Cela revient à se poser la question de la fréquence des spikes en fonction de la pression acoustique.

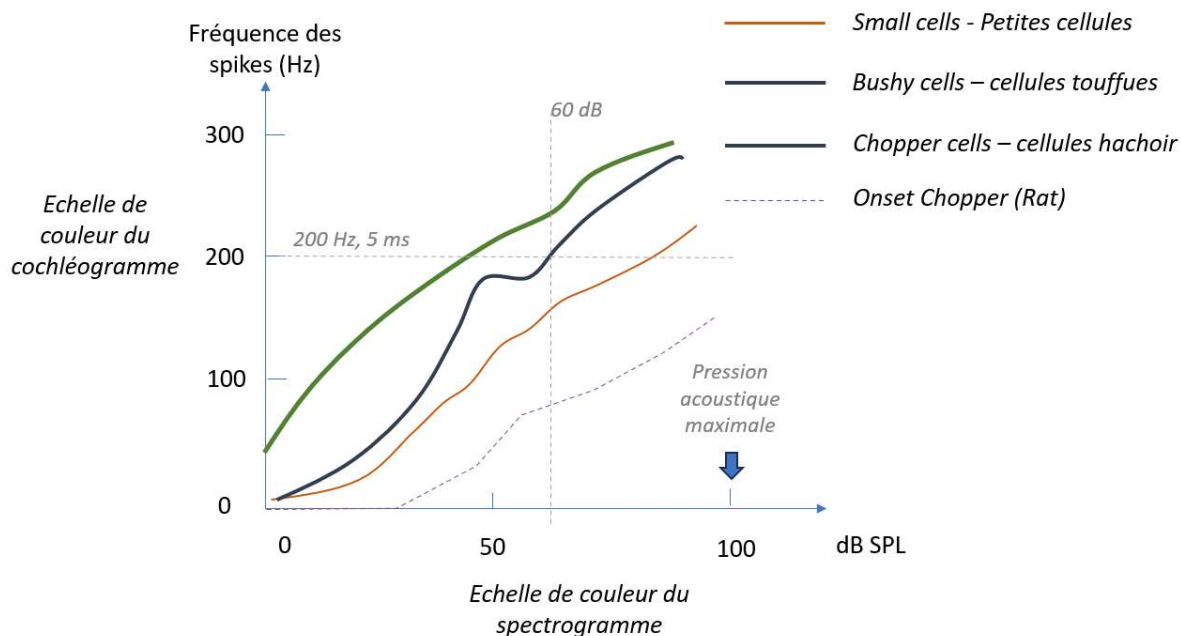


Figure 13 : fréquence des impulsions en fonction du niveau de pression acoustique en entrée de la cochlée d’un cochon d’inde, selon le type de cellule analysée d’après [Hockley 2022] ainsi que la cochlée d’un rat [Paolini 1999]

La réponse n’est pas la même selon la fréquence et la pression acoustique de l’influx auditif, mais la tendance est similaire : la fréquence d’apparition des impulsions tend à augmenter avec la pression acoustique en entrée de la cochlée, selon une loi relativement monotone (Figure 13) [Hockley 2022]. Certaines cellules sont actives même sans pression acoustique (*bushy cell*) tandis que d’autres cellules s’activent seulement à partir d’un certain niveau de pression acoustique (effet de seuil comme dans le rat [Parolini 1999]). A une puissance acoustique moyenne (60 dB, voix conversationnelle) correspond une occurrence moyenne de 200 impulsions par seconde, soit un délai entre deux spikes de 5 ms environ pour la *Chopper cell*.

Nous pouvons donc considérer que la palette de couleur du cochléogramme peut être déduite de l’échelle de couleur du spectrogramme par une simple loi linéaire entre dB acoustique et nombres d’impulsions, à laquelle on peut éventuellement ajouter un effet de seuil pour ignorer les valeurs d’énergie trop faible (*Small cells* et *Chopper cells* principalement).

A une fréquence d’occurrence de 300 impulsions par seconde, correspond une pression acoustique maximale, fixée à 100 dB, en régime continu d’après les courbes de la figure 13. La formule de conversion simplifiée est donc :

$$spikes_{Continu} (Hz) = 3 \times dB_{SPL}$$

Voies auditives descendantes

Les voies auditives descendantes suivent des chemins réciproques des voies ascendantes. Elles partent du cortex auditif et atteignent la cochlée en suivant un chemin proche des voies auditives montantes, dont on trouvera un schéma simplifié à la figure 14.

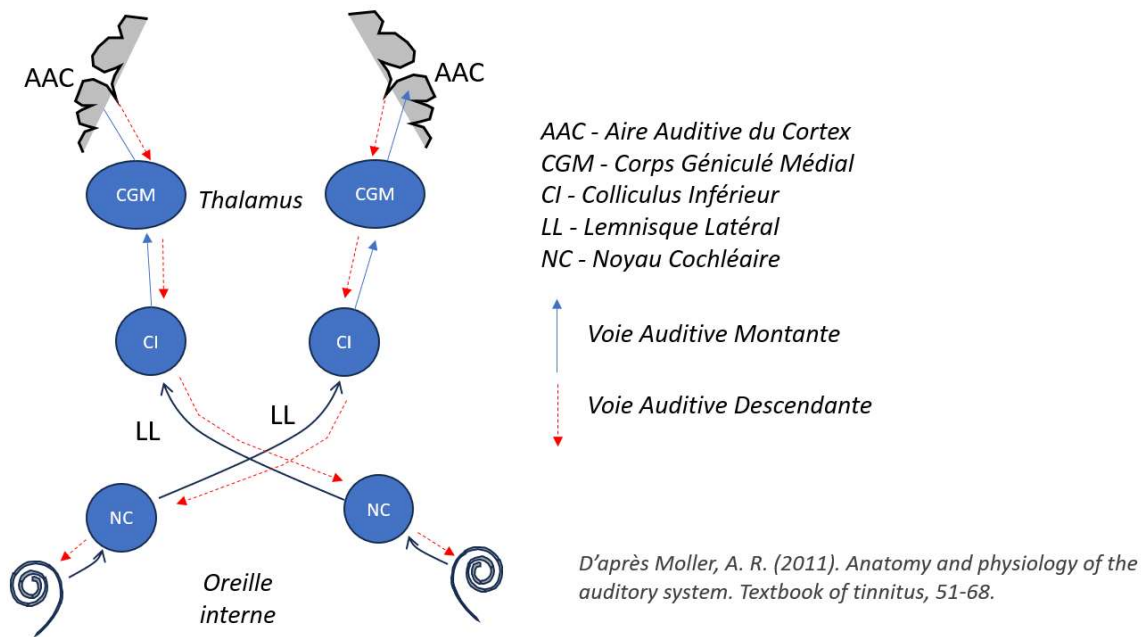
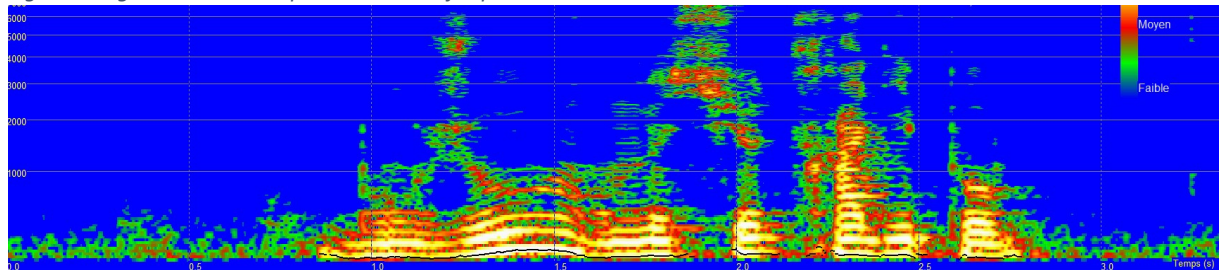


Figure 14 : Principe des voies auditives montantes et descendantes, d'après [Moller 2011]

Les voies descendantes peuvent influencer les propriétés mécaniques de la membrane basilaire, ce qui modifie la sensibilité auditive ainsi que la sélectivité en fréquence. Cette propriété permet par exemple de « masquer » certaines zones de fréquences, permettant par exemple de mieux discerner un auditeur dans le bruit. Une des actions de masquage les plus commune consiste à réduire le bruit ambiant, souvent appelé bruit en 1/f du fait de sa prédominance à basse fréquence (Figure 15).

Signal d'origine avec bruit important à basse fréquence



Réduction du bruit basse fréquence par filtrage sélectif

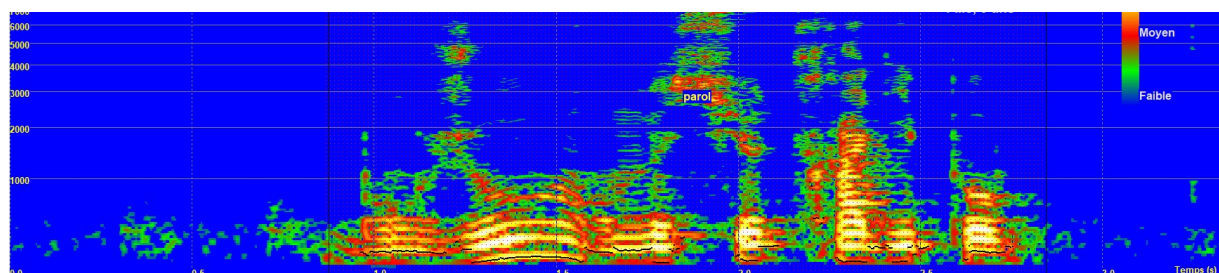


Figure 15 : Réduction du bruit basse fréquence par modification des caractéristiques fréquentielles de la cochlée

Discussion

La représentation tridimensionnelle temps/fréquence/énergie issue de la simple Transformée de Fourier ne donne pas une image fidèle de l'information qui est transmise au cerveau pour plusieurs raisons. Tout d'abord, la perception de l'oreille des fréquences est logarithmique et non linéaire. De plus, l'oreille a une courbe de sensibilité qui n'est pas constante, elle atténue en particulier les très

basses et très hautes fréquences. Enfin, l'influx nerveux n'est pas lié de façon linéaire à la notion d'énergie : il est plus important à l'initiation ou à l'arrêt d'un son.

Notre but a donc été de proposer une représentation des trois mêmes dimensions temps/ fréquence/ énergie, en tentant d'y introduire certaines caractéristiques de la cochlée, afin de faire évoluer la représentation vers ce que l'on peut appeler le « cochléogramme ».

Nous n'avons pas la prétention de reproduire fidèlement l'influx nerveux tel qu'il parvient au cortex, tellement le chemin ascendant et descendant de la cochlée au cerveau est complexe. Il s'agit simplement de proposer une représentation un peu plus fidèle du codage de l'information, en espérant qu'il puisse apporter une aide à la prise en soin orthophonique des pathologies de la voix et de la parole.

Parmi les aspects intéressants du spectrogramme mélodique, on peut noter la capacité à regrouper en un seul graphique l'ensemble des composantes, sans nécessité de « zoomer » sur des zones de fréquences en particulier. Par exemple, il sera désormais possible d'apprécier l'évolution des contours mélodiques (basses fréquences) et la constriction (hautes fréquences) sans changer d'échelle. Le spectrogramme mélodique est envisagé comme représentation par défaut dans DIADOLAB (prise en soin des pathologies de la parole).

L'un des aspects les plus intéressants du cochléogramme est de bien faire apparaître les éléments transitoires de la parole, notamment les plosives de faible énergie comme les phonèmes /b,d,g,l/. Cette représentation, bien que perturbante par rapport aux représentations communes des spectrogrammes, pourrait faciliter les tâches de segmentation manuelle et la mise au point d'outils de segmentation automatique en révélant la position précise des frontières.

Sur l'aspect reconnaissance automatique des événements sonores, le cochléogramme s'avère plus informatif que le spectrogramme linéaire ou mélodique, comme le montre Sharan [2019]. Une précision proche de 98% (Soit 2% d'erreur) dans la classification des événements est obtenue avec des données issues du cochléogramme, contre 95% avec le spectrogramme (5% d'erreur). L'étude ne portait cependant pas sur des phonèmes, mais sur un grand corpus d'environnements sonores.

Le cochléogramme a l'avantage de mieux séparer les motifs entre eux, dans la mesure où les zones d'incertitude ou d'ambiguïté entre les différentes classes de production sonore sont réduites [Gao 2014]. On peut donc espérer que les oppositions sourdes/sonores, nasales/non nasales, plosives voisées/non voisées, ou d'autres traits distinctifs sont mieux mis en valeur grâce au cochléogramme, en faisant apparaître des contrastes souvent difficiles à identifier sur les spectrogrammes conventionnels.

Conclusion

Après avoir rappelé les principes de base du passage temps/fréquence et la construction du spectrogramme de la voix, nous avons fait évoluer le principe de la distribution temps/fréquence/énergie vers une représentation un peu plus fidèle des stimulations générées par la cochlée vers le cerveau, en incluant notamment une représentation fidèle à la tonotopie, la prise en compte de la courbe de sensibilité de l'oreille et la sensibilité aux variations de fréquence et d'intensité. Le cochléogramme permet notamment de mieux mettre en valeur certaines composantes de la parole et peut améliorer la classification des phonèmes, que ce soit pour une approche automatisée en lien avec la segmentation ou la reconnaissance des phonèmes, ou une approche thérapeutique pour faciliter la prise en soin des confusions et imprécisions phonémiques.

Remerciements

Nous tenons à remercier les orthophonistes contributrices à l'ERU 15 « voix pathologique » et de l'ERU 46 « parole pathologique » du laboratoire LURCO ainsi que les orthophonistes qui ont accepté d'enregistrer des témoins et des patients ou de co-encadrer des mémoires en lien avec la voix, la parole et leurs pathologies.

Nous remercions l'ensemble des orthophonistes, étudiants et chercheurs qui nous ont accordé leur confiance en utilisant nos méthodologies, approches et outils dans le cadre de leur pratique clinique ou de leurs travaux scientifiques.

A propos des auteurs



Etienne SICARD est professeur à l'INSA Toulouse en électronique, informatique et traitement du signal. Il est directeur de recherches de l'équipe ERU 46 sur la parole pathologique au Laboratoire LURCO. Il a été chercheur associé à l'IRIT dans le cadre du projet ANR Voice4PD sur le diagnostic différentiel des

Syndromes Parkinsoniens. Etienne SICARD est co-auteur des logiciels VOCALAB et DIADOLAB pour le bilan et la rééducation orthophonique de la voix et la parole. Il a co-écrit une vingtaine d'ouvrages et plus de 300 publications dont 150 dans le domaine de l'analyse de la voix et parole pathologique. Il a été nommé en 2006 *Distinguished Lecturer* de la société savante IEEE.

Site : <https://cv.hal.science/etienne-sicard>

Email : etienne.sicard@insa-toulouse.fr



Anne MENIN-SICARD est orthophoniste et titulaire d'un master recherche en Sciences du langage de l'université de Grenoble. Elle a exercé en libéral à Toulouse de 1994 à 2015. Anne MENIN-SICARD est co-auteur des logiciels VOCALAB et DIADOLAB.

Dans le cadre de AMS Formation, elle organise et anime des formations aux orthophonistes sur le bilan et la prise en charge de la voix et parole pathologique. Elle propose depuis 2019 des formations en ligne personnalisées à destination des orthophonistes. Chercheuse associée à l'ERU 46 du LURCO, elle développe des outils de mesures objectives et de rééducation de la parole en lien avec le logiciel DIADOLAB. Elle est auteur d'un ouvrage sur l'évaluation objective de la voix ainsi que de nombreuses publications dans le domaine de la voix et de la parole pathologique.

Email : anne.sicard2@orange.fr



Manon GERMAIN est orthophoniste en cabinet libéral à Paris, France. Après un Bachelier de logopédie à la Haute école de Liège, Belgique, elle se forme à la méthode ABA et la méthode PECS. En 2021, elle est certifiée par l'Université de

Liège sur l'Evidence-Based-Practice (EBP) au service du patient. Elle a été d'Avril 2021 à Février 2022 orthophoniste à la Fondation A. de Rothschild à Paris, au service des urgences neurovasculaires, ainsi que pour le suivi des pathologies neurodégénératives. Elle a aussi été journaliste scientifique pour la vulgarisation d'articles scientifiques, des interviews d'experts. En 2020, elle est lauréate du D.U. d'enseignement pratique pluridisciplinaire de la santé connectée, de l'Université de Paris. Lors de cette même année, elle travaille sur le développement d'une plateforme numérique facilitant l'inclusion scolaire des enfants en situation de handicap (*Handicap'Access*). En 2023, elle a été labélisée approche

métacognitive et objective (AMCO) sur les pathologies de la parole avec DIADOLAB par AMS-Formation par Anne MENIN SICARD. Elle oriente désormais sa carrière vers la gestion de projets numériques en santé.

Email : manon.germain17@icloud.com

Références

- [Allen 2017] Allen, E. J., (2017). Representations of pitch and timbre variation in human auditory cortex. *Journal of neuroscience*, 37(5), 1284-1293.
- [Bahmer 2006] Bahmer, A., & Langner, G. (2006). Oscillating neurons in the cochlear nucleus: I. Experimental basis of a simulation paradigm. *Biological Cybernetics*, 95(4), 371-379.
- [Bahmer 2006b] Bahmer, A., & Langner, G. (2006). Oscillating neurons in the cochlear nucleus: II. Simulation results. *Biological cybernetics*, 95(4), 381-392.
- [Bamiou 2003] Bamiou, D. E.,(2003). The insula (Island of Reil) and its role in auditory processing: literature review. *Brain research reviews*, 42(2), 143-154.
- [Cochlea.eu] <https://www.cochlea.eu/cochlee/fonctionnement-de-la-cochlee/>
- [Gao 2014] Gao, B., (2014). Cochleagram-based audio pattern separation using two-dimensional non-negative matrix factorization with automatic sparsity adaptation. *The Journal of the Acoustical Society of America*, 135(3), 1171-1185.
- [Hockley 2022] Hockley, A., Wu, C., & Shore, S. E. (2022). Olivocochlear projections contribute to superior intensity coding in cochlear nucleus small cells. *The Journal of physiology*, 600(1), 61-73.
- [Heil 2015] Heil, P., & Peterson, A. J. (2015). Basic response properties of auditory nerve fibers: a review. *Cell and tissue research*, 361, 129-158.
- [ISO 226] <https://www.iso.org/fr/standard/83117.html>
- [Jeffery 2018] Jeffery, T., (2018). Analyses of sustained vowels in down syndrome (DS): a case study using spectrograms and perturbation data to investigate voice quality in four adults with DS. *Journal of Voice*, 32(5), 644-e11.
- [Li 2021] Li, H., (2021). Three-dimensional tonotopic mapping of the human cochlea based on synchrotron radiation phase-contrast imaging. *Scientific reports*, 11(1), 4437.
- [Meaud 2010] Meaud, J., & Grosh, K. (2010). The effect of tectorial membrane and basilar membrane longitudinal coupling in cochlear mechanics. *The Journal of the Acoustical Society of America*, 127(3), 1411-1421.
- [Moller 2011] Moller, A. R. (2011). Anatomy and physiology of the auditory system. *Textbook of tinnitus*, 51-68.
- [Oertel 2000] Oertel, D., (2000). Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proceedings of the National Academy of Sciences*, 97(22), 11773-11779.

[Paolini 1999] Paolini, A. G., & Clark, G. M. (1999). Intracellular responses of onset chopper neurons in the ventral cochlear nucleus to tones: evidence for dual-component processing. *Journal of neurophysiology*, 81(5), 2347-2359.

[Sharan 2019] Sharan, R. V., & Moir, T. J. (2019). Acoustic event recognition using cochleagram image and convolutional neural networks. *Applied Acoustics*, 148, 62-66.

[Sharma 2020] Sharma, G., (2020, July). Screening and analysis of specific language impairment in young children by analyzing the textures of speech signal. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 964-967). IEEE.

[Spoendlin 1989] Spoendlin, H., & Schrott, A. (1989). Analysis of the human auditory nerve. *Hearing research*, 43(1), 25-38.

[Sicard 2021] Sicard, E., Menin-Sicard, A. (2021). Le spectrogramme et son application en clinique orthophonique. <https://hal.science/hal-03107434>

[Sicard 2023] Sicard, E., Menin-Sicard, A. (2023). Fréquence fondamentale de la parole – Repères pour la prise en soin orthophonique. <https://hal.science/hal-04073008>

[Sicard 2025] Sicard, E., Menin-Sicard, A., Borel, S. (2025). MADLEN, une aide à la prescription d'exercices d'après le bilan de la voix et de la parole : illustration sur des cas de dysarthrie Parkinsonienne et Ataxie de Friedreich. In "Les maladies neurodégénératives : état des pratiques et de la recherche", Ortho-Editions

[Wikipedia 2024] https://en.wikipedia.org/wiki/Equal-loudness_contour

[Zahid 2020] Zahid, L., & al. (2020). A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson's Disease. *IEEE Access*, 8, 35482-35495.