

Domaine d'étude : émetteurs et récepteurs sonores
Mots-clefs : voix, acoustique physiologique, reconnaissance vocale

**LA VOIX ET L'OREILLE HUMAINE
APPLICATION À LA RECONNAISSANCE VOCALE**

CONTEXTE DU SUJET

La voix peut être considérée comme le plus ancien des instruments de musique et l'oreille comme le plus ancien récepteur sonore. En outre, la communication est depuis toujours un des fondements de nos sociétés, et aujourd'hui plus que jamais étant donnés les moyens de communication innovants apportés par les nouvelles technologies.

Comment le son de la voix est-il produit ? Comment l'oreille capte-t-elle ces sons et comment des dispositifs innovants permettent-ils de remplir cette même fonction ?

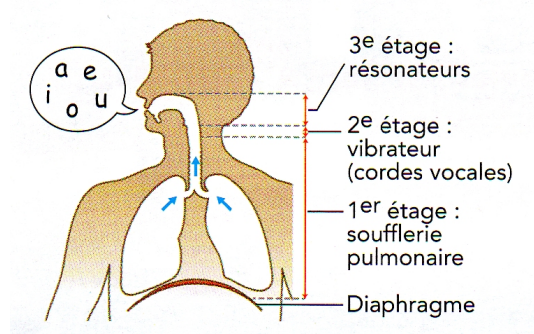
DOCUMENTS MIS À DISPOSITION

Document I : Comment naît la voix ?

La formation de la voix est déterminée par le jeu de plusieurs structures (systèmes) :

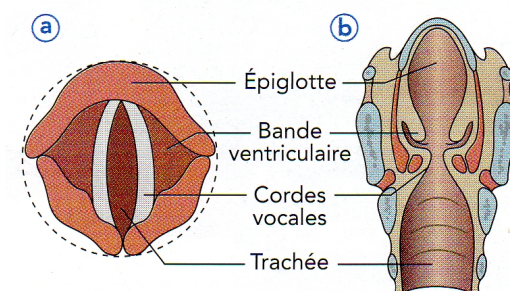
● **L'appareil respiratoire producteur de souffle**

Le souffle sera maîtrisé différemment pour l'émission de la voix.



● **Le vibreur laryngé producteur du son de la voix**

L'indispensable passage du souffle entre les deux cordes vocales, rapprochées de façon adéquate, les fait vibrer. C'est l'origine du son de la voix. Les deux cordes (plis vocaux) sont placées horizontalement dans le larynx (à hauteur de la pomme d'Adam).



Système laryngé :

(a) : coupe transversale
(b) : coupe longitudinale

● **Les espaces d'amplification et de résonance**

Ces espaces ou cavités (gorge, nez, bouche) changent de volume et de forme grâce au jeu des muscles du voile du palais, de la langue, des lèvres, ... Ainsi peuvent se modifier les qualités de la voix.

D'après :

Bureau International d'Audio-Phonologie
www.biap.org – *Recommandation 27*

Document II : Description de la voix

La voix peut être caractérisée par quatre paramètres principaux.

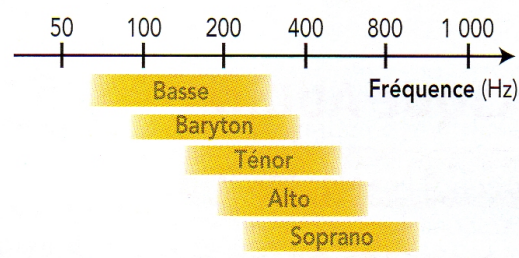
- La **hauteur** est la sensation auditive liée à la fréquence des vibrations des cordes vocales (son grave, son aigu). L'unité de mesure est le hertz, de symbole Hz.
- Le **niveau sonore** est la sensation auditive liée à l'amplitude des vibrations des cordes vocales (son fort, son faible). L'unité est le décibel de symbole dB.
- Le **timbre** est la sensation auditive liée aux harmoniques présents (son sombre, son clair).
- La **tenue** est la sensation auditive liée à la durée des vibrations des cordes vocales (son long, son court).

D'après : Musica et Memoria – www.musimem.com

Document III : Les registres de la voix

On peut changer la fréquence de vibration des cordes vocales en modifiant leur tension, et surtout leur épaisseur. Cela a pour effet de faire varier la hauteur des sons émis par la voix. Ces modifications sont le fruit d'actions neurocérébrales.

Le **registre** d'une voix est l'étendue de son échelle vocale, de la note la plus grave à la note la plus aiguë. Pour la voix chantée, l'échelle des sons pouvant être émis de façon homogène est appelée **tessiture**, c'est une partie du registre d'un interprète.

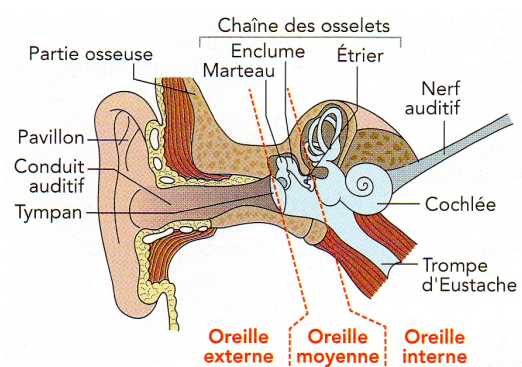


D'après : Wikipedia – www.wikipedia.org

Document IV : Acoustique physiologique

L'oreille comporte trois parties.

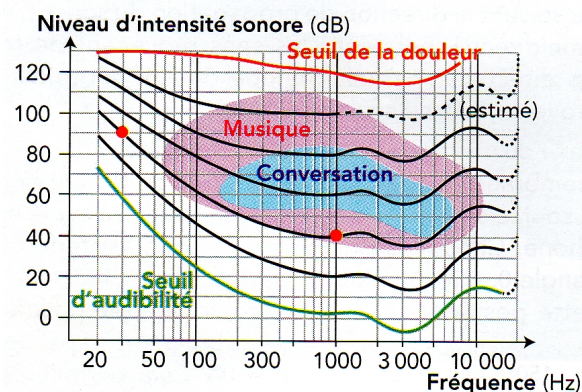
- L'**oreille externe** est la seule partie en communication directe avec l'extérieur. Elle est composée d'un pavillon et d'un conduit auditif. C'est une simple structure de transmission des sons vers le tympan.
- L'**oreille moyenne** tient le rôle de protection et de transmission mécanique. Les vibrations du tympan sont transmises au marteau, à l'enclume et à l'étrier.
- L'**oreille interne** est la partie la plus fragile de l'oreille. Elle est constituée de quelques milliers de cellules ciliées situées dans la cochlée et qui convertissent les vibrations mécaniques en signaux électriques. C'est notre « capital auditif ».



D'après : Un site web géré par l'INPES
www.ecoute-ton-oreille.com

Document V : Sensibilité de l'oreille

L'oreille perçoit convenablement les sons dont le niveau d'intensité sonore est compris entre le seuil d'audibilité et le seuil de douleur. Sa sensibilité varie en fonction de la fréquence. Cela peut être représenté par une série de courbes dites « d'égale sensation auditive ». Ces courbes montrent, par exemple, qu'un son de fréquence 1000 Hz dont le niveau d'intensité sonore est de 40 dB donne la même sensation d'intensité qu'un son de fréquence 30 Hz dont le niveau sonore est de 90 dB (voir les points sur le diagramme). Par ailleurs, la sensibilité de l'oreille diminue avec l'âge, ce qui conduit à des pertes auditives.



Document VI : Les dangers du bruit

Les sons deviennent nocifs lorsque leur intensité dépasse les possibilités de réception de l'oreille.

Le niveau d'intensité sonore est exprimé en décibel (dB). L'échelle va de 0 à 120 dB, mais certaines sources (avions, canons, fusées) émettent des sons d'un niveau supérieur. La réglementation limite à 100 dB le niveau de sortie des baladeurs et à 105 dB celui dans les lieux musicaux. La limite de nocivité est située à 85-90 dB.

Après exposition prolongée à un niveau proche de 100 dB, par exemple après une soirée en discothèque, on constate divers états auditifs que l'on peut classer par gravité croissante :

- aucun phénomène auditif particulier, c'est le cas le plus fréquent, mais à coup sûr, quelques cellules ont été fragilisées ;
- phénomènes temporaires tels que des bourdonnements ou des sifflements (acouphènes) ;
- phénomènes de type acouphènes persistants et irréversibles, baisse sensible de l'audition.

D'après : brochure « Nos oreilles on y tient » – Association Journée Nationale de l'Audition

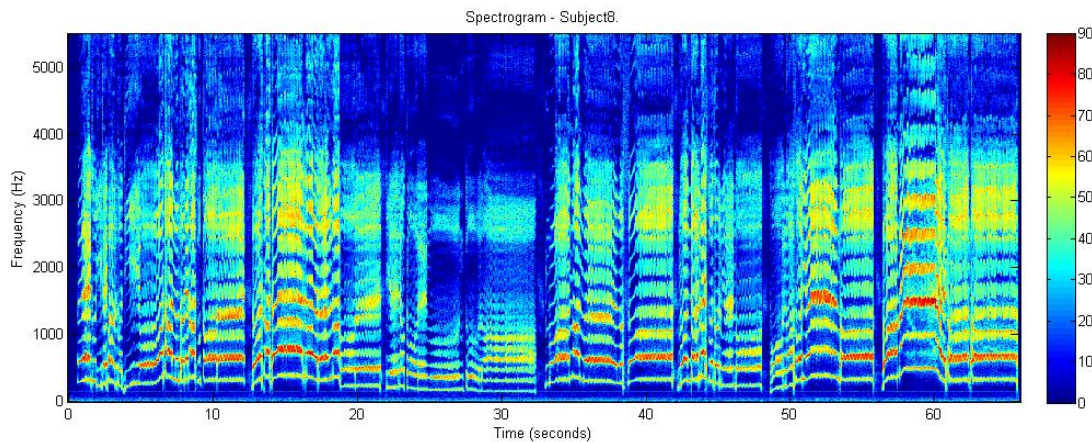
Document VII : Les techniques de reconnaissance vocale

Pour identifier le langage parlé, deux techniques peuvent être employées en fonction de l'application souhaitée.

- **L'approche globale** s'intéresse directement aux mots. Une image acoustique d'un certain nombre de mots est enregistrée dans une base de données et comparée à l'image acoustique du mot prononcé. Cette technique est efficace mais le nombre de mots est limité par la capacité de stockage et de traitement de la machine.
- **L'approche analytique** est basée sur la reconnaissance des phonèmes et syllabes. L'avantage est incomparable : tous les mots de la langue française peuvent être prononcés avec seulement 36 phonèmes. Pour pallier les différences d'accent, de timbre ou de rythme de parole, on « entraîne » le système de reconnaissance en faisant lire au locuteur des phrases choisies. La plus grande difficulté est que la machine doit interpréter la signification du discours pour ne pas confondre deux phrases phonétiquement identiques (comme par exemple : « la citerne est pleine d'eau » et « la scie terne est pleine dos »).

Document VIII : Les outils de la reconnaissance vocale

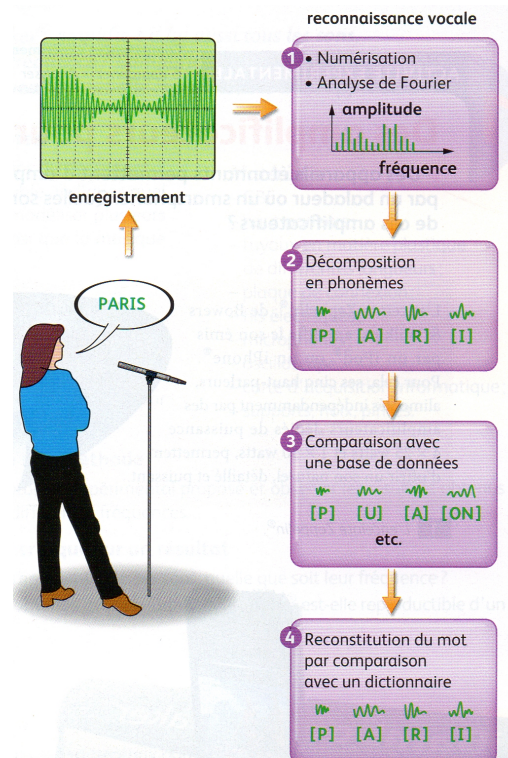
Pour identifier les sons (mots ou phonèmes), on utilise des spectrogrammes qui représentent l'évolution des différentes fréquences contenues dans le son au cours du temps. En effet, dans le langage parlé, les sons sont constitués d'une multitude de fréquences qui évoluent au cours du temps.



Document IX : Les étapes de la reconnaissance vocale

La reconnaissance vocale, domaine scientifique en plein essor, peut se décomposer en quatre étapes :

- ① le traitement acoustique qui numérise le signal et le traite par analyse spectrale (analyse de Fourier) ;
- ② la décomposition du mot en phonèmes ;
- ③ l'apprentissage automatique qui compare des séquences enregistrées à une base de séquences connues ;
- ④ le décodage qui reconstitue le discours le plus probable en assemblant les séquences apprises précédemment.



1. Extraire et exploiter des informations issues de documents scientifiques

- 1.1. D'un point de vue physique, qu'est-ce que la voix ?
- 1.2. Identifier les trois organes ou parties du corps humain qui interviennent dans la production de la voix et préciser le rôle de chacun d'eux dans la phonation.
- 1.3. En s'appuyant sur un exemple précis d'instrument, indiquer en quoi la production de la voix peut être comparée avec la formation d'un son dans un instrument de musique ?
- 1.4. Sur un schéma prenant la forme d'une chaîne fonctionnelle, représenter les principales étapes de la production d'un son par la voix et de la perception de ce son par le cerveau.
- 1.5. Comment peut-on expliquer la différence de tessiture entre un baryton et une soprano ?
- 1.6. Comparer les niveaux d'intensité sonore permis par la réglementation avec les limites de nocivité et le seuil de douleur. Quel(s) autre(s) paramètre(s) convient-il de prendre en compte pour évaluer la nocivité d'un son ?
- 1.7. *Chaque individu a une voix différente, facilement reconnaissable à l'oreille, si bien que celle-ci peut être utilisée, par identification vocale, pour accéder à des zones sécurisées ou encore dans le cadre d'une enquête judiciaire.* Expliquer l'individualité de la voix humaine.
- 1.8. Quels sont les avantages et les inconvénients des deux techniques utilisées pour la reconnaissance vocale ?
- 1.9. Dans la reconnaissance vocale, quel type d'instrument remplace l'oreille ?
- 1.10. Dans un spectrogramme, quelle grandeur est portée en abscisse ? en ordonnée ? Préciser l'intérêt de ces diagrammes dans la mise en œuvre de la reconnaissance vocale.
- 1.11. Pourquoi, en dépit de tous les progrès actuels, la reconnaissance vocale d'un texte parlé dans la langue naturelle n'est-elle pas fiable à 100% ?
- 1.12. Citer des applications concrètes de la reconnaissance vocale dans la vie quotidienne, le monde médical ou le domaine scientifique.

2. Rédiger une synthèse argumentée

À l'aide de vos connaissances et des documents précédents, rédiger, en une vingtaine de lignes maximum, une synthèse argumentée répondant à la problématique suivante :

**En quoi la compréhension des mécanismes de la parole et de l'audition
a-t-elle permis l'émergence d'une nouvelle forme de communication
dans les sociétés modernes ?**

Pour cela, expliquer tout d'abord en quoi l'oreille est-elle adaptée à la voix. Préciser ensuite les caractéristiques des sons émis par la voix en indiquant celles qui sont utiles en reconnaissance vocale. Présenter alors les grandes étapes menant à la reconnaissance vocale et conclure en expliquant en quoi cette technique constitue une nouvelle forme de communication.