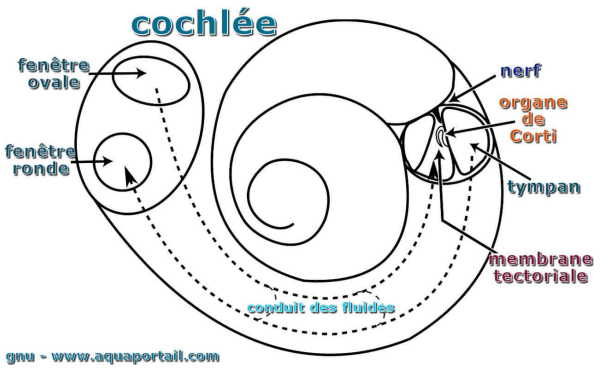
CS

1. **Proposer un schéma de la cochlée, et décrire son fonctionnement. Qu'est-ce qu'une bande critique ?**



La cochlée est une partie importante de l'oreille interne qui joue un rôle crucial dans le processus de l'audition. Voici un schéma simplifié de la cochlée:

1. **Membrane basilaire et membrane tectoriale:**
   - La cochlée est constituée de deux membranes principales - la membrane basilaire en bas et la membrane tectoriale en haut. Ces membranes s'étendent le long de la cochlée.

2. **Canal cochléaire:**
   - Le canal cochléaire est la cavité centrale de la cochlée où le fluide appelé endolymphe circule. C'est ici que se produit la conversion des vibrations sonores en signaux électriques.

3. **Limbe osseux et limbe membraneux:**
   - Le limbe osseux est une structure osseuse à la base de la membrane basilaire, et le limbe membraneux est une partie de la membrane basilaire elle-même. Ces structures aident à soutenir la cochlée.

4. **Étrier (stapes) et fenêtre ovale:**
   - Les vibrations du son sont transmises de l'oreille moyenne à l'oreille interne par une chaîne d'osselets, le dernier étant l'étrier. La fenêtre ovale est la connexion entre l'oreille moyenne et l'oreille interne, où les vibrations sont transmises au liquide de la cochlée.

5. **Organe de Corti:**
   - L'organe de Corti est situé sur la membrane basilaire et est l'endroit où les cellules ciliées, responsables de la conversion des vibrations en signaux électriques, sont localisées.

6. **Cellules ciliées internes et externes:**

- Les cellules ciliées internes et externes sont des cellules sensorielles spécialisées qui convertissent les vibrations sonores en signaux électriques. Ces signaux sont ensuite transmis au cerveau par le nerf auditif.

7. **Nerf auditif (cochléaire):**
   - Les signaux électriques générés par les cellules ciliées sont transmis au cerveau par le nerf auditif, permettant au cerveau de percevoir et d'interpréter les sons.

2. **On écoute deux sons sinusoïdaux superposés : le premier (A) à 125Hz possède un niveau de 60 dB(SPL); le second (B) à 2kHz possède un niveau de 50 dB(SPL). Lequel de ces sons semble le plus fort ?**

   Le niveau de pression acoustique (SPL, Sound Pressure Level) est une mesure de l'intensité d'un son et est exprimé en décibels (dB). La relation entre le niveau de pression acoustique (dB SPL) et la perception humaine du volume sonore n'est pas linéaire, mais plutôt logarithmique. Cela signifie qu'une augmentation de 10 dB SPL est généralement perçue comme approximativement deux fois plus forte.

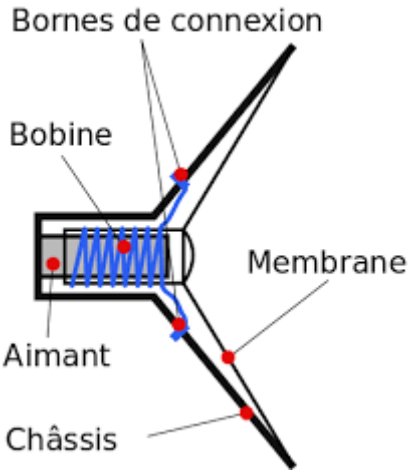
   Dans votre exemple :
   - Le son A a un niveau de 60 dB SPL à 125 Hz.
   - Le son B a un niveau de 50 dB SPL à 2 kHz.

   Le son A à 125 Hz est à un niveau de pression acoustique plus élevé que le son B à 2 kHz. Cependant, en raison de la sensibilité fréquentielle variable de l'oreille humaine, la perception du volume sonore peut être différente pour ces deux fréquences.

   Normalement, l'oreille humaine est plus sensible aux fréquences moyennes, autour de 1 kHz à 4 kHz. Cela signifie que même si le niveau de pression acoustique du son A à 125 Hz est plus élevé, le son B à 2 kHz pourrait être perçu comme plus fort en raison de la sensibilité accrue de l'oreille humaine à cette fréquence.

   Cependant, pour une comparaison plus précise, on pourrait utiliser une courbe de pondération fréquentielle spécifique, comme la courbe A ou la courbe B, qui tiennent compte de la sensibilité fréquentielle de l'oreille humaine. Ces courbes de pondération sont souvent utilisées pour mesurer le niveau de pression sonore équivalent (Leq) dans le domaine de l'audiométrie.

3. **Proposer un schéma d'un microphone acoustique**



1. **Membrane (Diaphragme) :**
   - Le diaphragme est une fine membrane généralement fabriquée à partir de matériaux légers et flexibles. Il est sensible aux variations de pression acoustique.

2. **Bobine mobile :**
   - Attachée au diaphragme, la bobine mobile est généralement constituée de fil conducteur. Lorsque le diaphragme vibre en réponse aux ondes sonores, la bobine mobile se déplace également.

3. **Aimant fixe :**
   - Un aimant fixe est positionné autour de la bobine mobile. La bobine mobile se déplace dans le champ magnétique de l'aimant, générant un courant électrique proportionnel aux mouvements du diaphragme.

4. **Support mécanique :**
   - Un support mécanique maintient le diaphragme en place tout en permettant des mouvements libres en réponse aux changements de pression acoustique.

5. **Boîtier :**
   - Le boîtier entoure les composants internes du microphone et offre une protection mécanique ainsi qu'une isolation acoustique.

6. **Connecteur :**
   - Le connecteur est l'interface électrique permettant de connecter le microphone à un préamplificateur ou à un autre équipement audio.

Le fonctionnement global du microphone acoustique est basé sur le principe électromagnétique. Lorsque des ondes sonores frappent le diaphragme, la bobine mobile bouge dans le champ magnétique, générant un courant électrique proportionnel aux variations de pression acoustique. Ce signal électrique peut ensuite être amplifié et traité pour produire un signal audio utilisable.

4. **A quelle condition sur la longueur d'onde un haut-parleur de diamètre D peut-il être approché par une source sonore isotrope ?**

   Lorsqu'un haut-parleur de diamètre \(D \) est suffisamment petit par rapport à la longueur d'onde du son qu'il émet, on peut considérer que la source sonore est isotrope pour des fins de modélisation. Lorsque le diamètre du haut-parleur est petit par rapport à la longueur d'onde, l'onde sonore émise par le haut-parleur peut être approximée comme émanant uniformément dans toutes les directions.

   La condition pour que cela soit vrai peut être formulée en termes de la comparaison entre le diamètre du haut-parleur D et la longueur d'onde λ du son émis. Mathématiquement, cela peut être exprimé comme suit :

   $$D \gg \lambda$$

   En d'autres termes, le diamètre du haut-parleur doit être beaucoup plus petit que la longueur d'onde du son pour que l'approximation d'une source isotrope soit valable.

   Cette condition est souvent rencontrée dans des situations où la fréquence du son émis par le haut-parleur est relativement élevée, car la longueur d'onde est inversement proportionnelle à la fréquence. Ainsi, pour des fréquences élevées, la longueur d'onde est plus courte, et il est plus probable que la condition $D \gg \lambda$ soit satisfaite.

5. **Qu'appelle-t-on distorsion d'un signal audio ? Quels éléments de la chaîne d'acquisition en sont à l'origine ?**

   La distorsion d'un signal audio se produit lorsqu'il y a une altération non désirée ou une déformation du signal d'origine. Cette altération peut se manifester de différentes manières, et la distorsion est généralement considérée comme indésirable car elle peut altérer la qualité sonore perçue. Voici quelques types courants de distorsion dans le contexte audio :

   1. **Distorsion harmonique :**
   - Elle est caractérisée par l'apparition de fréquences harmoniques non présentes dans le signal d'origine. Elle peut être causée par des non-linéarités dans les composants de la chaîne d'acquisition, tels que des amplificateurs.

   2. **Distorsion d'intermodulation :**
   - Elle se produit lorsque deux ou plusieurs fréquences différentes sont présentes dans le signal, créant des combinaisons de fréquences qui n'étaient pas présentes à l'origine. Cela peut être causé par des non-linéarités dans les composants électroniques.

3. **Distorsion de phase :**
- La distorsion de phase provoque un décalage temporel indésirable entre les différentes fréquences du signal. Elle peut résulter de l'utilisation de filtres ou de composants qui modifient la phase des différentes composantes fréquentielles.

4. **Distorsion de surcharge (clipping) :**
- Elle se produit lorsque le niveau du signal dépasse la capacité maximale de traitement d'un composant, tel qu'un amplificateur. Cela conduit à une coupure (clipping) des crêtes du signal.

Les éléments de la chaîne d'acquisition qui peuvent contribuer à la distorsion comprennent :

- **Amplificateurs :** Des amplificateurs défectueux ou fonctionnant à des niveaux élevés peuvent introduire de la distorsion.

- **Microphones :** Certains microphones peuvent présenter des distorsions, surtout à des niveaux de pression sonore élevés.

- **Convertisseurs analogique-numérique (CAN) et numérique-analogique (CNA) :** Des distorsions peuvent survenir lors de la conversion entre signaux analogiques et numériques.

- **Enregistreurs et traitements de signal :** Les processus d'enregistrement et de traitement du signal peuvent introduire de la distorsion s'ils ne sont pas réalisés correctement.

La minimisation de la distorsion est une considération clé dans la conception des équipements audio et dans l'ingénierie sonore pour assurer la reproduction la plus fidèle possible du signal d'origine.

6. **On souhaite encoder de manière optimale un signal échantillonné à 48kHz. Une trame d'analyse contient 512 échantillons. Le spectre d'une trame particulière comporte :**
    - **Une raie sinusoïdale à 80kHz, d'intensité 80dB(SPL)**
    - **Une raie sinusoïdale à 600Hz, dintensité 60dB(SPL)**
    - **Un bruit à bande étroite à 700Hz, d'intensité équivalente 70dB(SPL)**

   A quelles bandes critiques appartiennent chacun de ces signaux ? Quels sont les masqueurs a) tonal b) non-tonal ? Construire la courbe de masquage. En déduire le nombre de bits alloués pour cette trame et son taux de compression.

   http://www-reynal.ensea.fr/docs/audio-sia/cours/diapos_cours_traitement_signal_audio.pdf slide 65

bark	min	centre	max
1	0	50	100
2	100	150	200
3	200	250	300
4	300	350	400
5	400	450	510
6	510	570	630
7	630	700	770
8	770	840	920
9	920	1000	1080
10	1080	1170	1270
11	1270	1370	1480
12	1480	1600	1720
13	1720	1850	2000
14	2000	2150	2320
15	2320	2500	2700
16	2700	2900	3150
17	3150	3400	3700
18	3700	4000	4400
19	4400	4800	5300
20	5300	5800	6400
21	6400	7000	7700
22	7700	8500	9500
23	9500	10500	12000
24	12000	13500	15500

FIGURE 3 – Bandes critiques (en Hz, sauf colonne 1)

Pour déterminer les bandes critiques et les masqueurs tonals et non-tonals, nous devons d'abord calculer les fréquences des bandes critiques et les niveaux de masquage.

Bandes critiques (Bark) :

Les bandes critiques sont des bandes de fréquence définies par l'oreille humaine. La relation entre la fréquence \(f \) en Hertz et la bande critique \(z \) en Bark est approximativement donnée par la formule de Zwicker :

$$z = 13 \cdot \arctan\left(0.00076 \cdot f\right) + 3.5 \cdot \arctan\left(\left(\frac{f}{7500}\right)^2\right)$$

Masqueurs tonals et non-tonals :
Les masqueurs tonals et non-tonals dépendent des niveaux de fréquence et sont souvent calculés en utilisant des modèles psychoacoustiques comme le modèle de Zwicker. Ces modèles prennent en compte la fréquence, le niveau sonore, et d'autres facteurs pour estimer la sensibilité auditive humaine.

Courbe de masquage :
La courbe de masquage combine les masqueurs tonals et non-tonals pour donner une estimation de la sensibilité auditive globale.

Calcul du nombre de bits alloués :
Le nombre de bits alloués dépend de la résolution souhaitée. Pour une résolution de N bits, le nombre de bits alloués serait $512 \times N$.

Calcul du taux de compression :
Le taux de compression dépend du nombre de bits alloués et de la fréquence d'échantillonnage. Il peut être calculé en utilisant la formule :
$Taux\ de\ compression = \frac{Fr[?][?]quence\ d'[?][?]chantillonnage}{Nombre\ de\ bits\ allou[?][?]s}$

Le calcul précis nécessite des valeurs numériques spécifiques pour chaque paramètre, et le modèle psychoacoustique choisi (comme le modèle de Zwicker) peut également affecter les résultats. En raison de la complexité du calcul, il peut être plus approprié de le faire à l'aide de logiciels spécialisés ou de bibliothèques de traitement audio.

Ces calculs sont généralement réalisés dans le cadre de l'audio perceptuel et de la compression audio telle que celle utilisée dans les codecs comme MP3.

7. **On souhaite calculer la transformée de Fourier d'un signal audio échantillonné à 48kHz, à partir d'une trame de 256 échantillons. Quelle résolution théorique peut-on espérer en Hz ? Si on utilise un fenêtrage de Hamming, on a $C_w = B_wL = 1.81$ où L est la durée d'une trame et B_w la largeur du lobe principal entourant chaque raie. Quelle est alors la résolution réelle ?**

http://www-reynal.ensea.fr/docs/audio-sia/cours/Cours_Transformation.pdf slide 15

La résolution en fréquence d'une transformée de Fourier dépend de la taille de la fenêtre temporelle utilisée pour le calcul. La résolution théorique Δf en Hertz est donnée par la relation :

$$\Delta f = \frac{1}{T}$$

où T est la durée totale de la fenêtre temporelle en secondes. Dans ce cas, la fenêtre temporelle est définie par le nombre d'échantillons N et la fréquence d'échantillonnage F_s :

$$T = \frac{N}{F_s}$$

Par conséquent, la résolution théorique peut être calculée comme suit :

$$\Delta f = \frac{F_s}{N}$$

Dans votre exemple, avec une fréquence d'échantillonnage $F_s = 48\,000 Hz$ et une trame de $N = 256$ échantillons, la résolution théorique serait :

$$\Delta f = \frac{48\,000}{256} \approx 187.5\,\text{Hz}$$

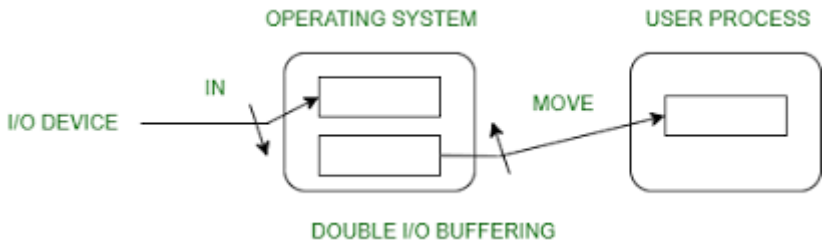
Cependant, lorsque vous appliquez une fenêtre, la largeur du lobe principal entourant chaque raie (le pic spectral correspondant à une fréquence spécifique) est élargie. La résolution réelle dépend de cette élargissement. La constante $C_w = 1.81$ est une approximation du rapport entre la largeur du lobe principal B_w et la durée de la fenêtre L pour une fenêtre de Hamming.

La résolution réelle $\Delta f_{\text{r[?][?]elle}}$ peut être calculée comme suit :

$$\Delta f_{\text{r[?][?]elle}} = \frac{F_s}{C_w \cdot N}$$

En substituant $C_w = 1.81$, $F_s = 48\,000$, et $N = 256$, vous pouvez calculer la résolution réelle.

8. **Illustrer sur sur un schéma le mécanisme de *double buffering*.**



Le double buffering est une technique utilisée dans la programmation informatique pour éviter le scintillement ou les déchirures d'écran lors de l'affichage d'animations ou de graphismes en temps réel.

Voici comment vous pouvez représenter le mécanisme de double buffering dans un schéma simple :

1. **Image Primaire (Front Buffer) :**
   - Dessinez une boîte ou un rectangle représentant l'écran d'affichage. À l'intérieur, dessinez une image ou un graphique pour représenter le contenu actuellement visible à l'écran. Étiquetez cela comme "Front Buffer" ou "Image Primaire".

2. **Image Secondaire (Back Buffer) :**
   - À côté de l'image primaire, dessinez une autre boîte ou un rectangle avec un contenu similaire ou identique à l'image primaire. Étiquetez cela comme "Back Buffer" ou "Image Secondaire".

3. **Échange des Buffers :**
   - Utilisez des flèches pour montrer le processus d'échange entre l'image primaire et l'image secondaire. Cela représente le moment où les deux images sont échangées, ce qui se produit généralement à une fréquence régulière, comme à chaque trame d'animation.

4. **Mise à jour du Contenu :**
   - Ajoutez des flèches ou des annotations pour montrer comment le contenu de l'image secondaire est mis à jour avant l'échange, généralement par le biais de calculs ou de dessins.

5. **Répétition du Processus :**
   - Ajoutez une flèche indiquant que le processus d'échange et de mise à jour se répète continuellement pour créer une animation fluide.

L'idée est que pendant que l'image primaire est affichée à l'écran, le programme peut travailler sur la création de la prochaine image dans l'image secondaire. Lorsque cette nouvelle image est prête, elle est échangée avec l'image primaire, et le processus se répète. Cela évite les scintillements ou les déchirures d'écran, car l'image n'est pas mise à jour directement à l'écran, mais plutôt dans un tampon en arrière-plan.

9. **Pourquoi le calcul des chromas et des MFCC peuvent-ils être modélisés par un réseau de neurones ?**

Le calcul des chromas (chromagrammes) et des coefficients cepstraux de fréquence mel (MFCC) dans le domaine de l'audio peut être modélisé par un réseau de neurones en raison de la capacité des réseaux de neurones à capturer des relations

complexes et non linéaires entre les caractéristiques d'entrée et de produire des représentations abstraites utiles.

1. **Non-linéarités complexes :**
   - Les relations entre les caractéristiques audio, telles que les fréquences et les amplitudes, sont souvent complexes et non linéaires. Les réseaux de neurones, en particulier les réseaux de neurones profonds, sont capables de modéliser ces non-linéarités, ce qui les rend adaptés à la tâche de traitement audio.

2. **Extraction automatique de caractéristiques :**
   - Les réseaux de neurones peuvent apprendre automatiquement à extraire des caractéristiques discriminantes à partir des données brutes. Plutôt que de concevoir manuellement des algorithmes pour extraire des caractéristiques spécifiques, les réseaux de neurones peuvent apprendre à partir des données d'entraînement quelles caractéristiques sont importantes pour la tâche.

3. **Adaptabilité à la tâche spécifique :**
   - Les architectures de réseaux de neurones peuvent être adaptées spécifiquement à la tâche audio envisagée. Par exemple, une architecture de réseau de neurones convolutif (CNN) peut être utilisée pour capturer des motifs locaux dans un spectrogramme, tandis qu'un réseau de neurones récurrent (RNN) peut être utilisé pour capturer des relations temporelles dans une séquence audio.

4. **Apprentissage non supervisé :**
   - Les réseaux de neurones peuvent également être utilisés pour effectuer un apprentissage non supervisé, où le modèle apprend à représenter automatiquement les caractéristiques pertinentes sans avoir besoin d'étiquettes de classe spécifiques. Cela peut être utile dans des tâches où la tâche d'extraction de caractéristiques est complexe et mal définie.

5. **Capacité à modéliser des séquences :**
   - Les tâches liées à l'audio, telles que la reconnaissance de la parole, peuvent bénéficier de l'utilisation de réseaux récurrents ou de réseaux récurrents récurrents (LSTM) qui sont capables de modéliser des séquences temporelles.

En résumé, les réseaux de neurones offrent une flexibilité et une puissance de modélisation qui les rendent appropriés pour des tâches complexes d'analyse audio comme le calcul des chromas et des MFCC. Leur capacité à apprendre des représentations hiérarchiques complexes à partir des données brutes en fait des outils puissants dans le domaine du traitement du signal audio.

10. Soit x(t) un signal constitué du mélange de deux instruments, guitare et voix.
On note $X(f,t) \in \mathbb{C}$ la transformée de Fourier à court terme de x(t). On suppose qu'on dispose d'un modèle qui fournit un estimé du spectrogramme d'amplitude $S_g(f,t) \in \mathbb{R}^+$ de la guitare et de celui de la voix

$S_v(f,t) \in \mathbb{R}^+$ à partir de X(f, t). Comment obtenir un estimé des signaux de la voix $x_v(t)$ et de guitare $x_g(t)$ à partir de ces estimés de spectrogramme ?

Pour séparer les signaux de la voix $x_v(t)$ et de la guitare $x_g(t)$ à partir des estimés de spectrogramme $S_g(f,t)$ et $S_v(f,t)$, on peut utiliser des techniques de déconvolution ou de séparation de sources. Une approche courante est d'utiliser la méthode de la déconvolution basée sur la minimisation d'une fonction coût. Voici une approche possible en utilisant des masques de magnitude pour séparer les composantes du mélange :

1. **Calcul des masques de magnitude :**
   - On peut définir des masques de magnitude pour la guitare $M_g(f,t)$ et la voix $M_v(f,t)$ en comparant les spectrogrammes estimés $S_g(f,t)$ et $S_v(f,t)$ avec une certaine règle. Par exemple, on peut définir les masques comme suit :

$$M_g(f,t) = \frac{S_g(f,t)}{S_g(f,t) + S_v(f,t)}$$
$$M_v(f,t) = \frac{S_v(f,t)}{S_g(f,t) + S_v(f,t)}$$

2. **Application des masques :**
   - Utiliser les masques pour pondérer la transformée de Fourier à court terme du signal mixte \(X(f, t)\) :

$$X_g(f,t) = M_g(f,t) \cdot X(f,t)$$
$$X_v(f,t) = M_v(f,t) \cdot X(f,t)$$

3. **Transformée de Fourier inverse :**
   - Appliquer la transformée de Fourier inverse pour obtenir les signaux dans le domaine temporel :

$$x_g(t) = \mathrm{IFT}\{X_g(f,t)\}$$
$$x_v(t) = \mathrm{IFT}\{X_v(f,t)\}$$

Ainsi, $x_g(t)$ et $x_v(t)$ sont des estimations des signaux de la guitare et de la voix, respectivement, obtenues à partir des spectrogrammes estimés et des masques de magnitude.

Il est important de noter que la qualité de la séparation dépend de la précision des spectrogrammes estimés et des masques de magnitude. Des approches plus avancées peuvent également utiliser des algorithmes de séparation de sources tels que la décomposition en valeurs singulières (SVD), l'analyse en composantes indépendantes (ICA) ou les réseaux de neurones dédiés à la séparation de sources audio.