

IMAGE ET REALITE VIRTUELLE

Parole et langage

G. Bailly

Sujet d'examen du 23 février 2002 - 2 heures - Cours et documents autorisés

1. Acoustique – 5pts

On considère un conduit vocal simplifié à 3 tubes de la figure ci-contre (A_i et L_i figurent respectivement les aires et les longueurs des tubes élémentaires). On prendra la vitesse du son égale à $c = 250$ m/s. On rappelle que

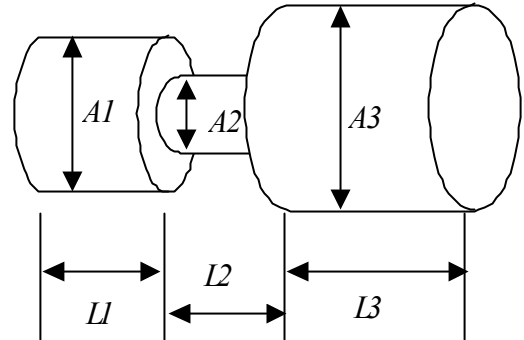
- les résonances d'un tube élémentaire de longueur L ouvert sont :

$$F_i(\text{Hz}) = (2 \cdot i + 1) \cdot c / 4L \text{ avec } i = 0.. \infty$$

- les résonances d'un tube élémentaire de longueur L fermé par un tube plus petit (le bec) sont :

$F_i(\text{Hz}) = i \cdot c / 2L$ avec $i = 1.. \infty$ et que la fréquence la plus basse - dite de Helmholtz - dépend des dimensions (aire a et longueur l) du bec et du volume V du tube élémentaire suivant la formule :

$$F_0(\text{Hz}) = \frac{c}{2p} \sqrt{\frac{a}{lV}}$$



On considère une configuration des tuyaux avec $A_1=5\text{cm}^2$, $L_1=5\text{cm}$; $A_2=0.5\text{cm}^2$, $L_2=5\text{cm}$; $A_3=10\text{cm}^2$, $L_3=10\text{cm}$

- Calculer les résonances de ce tuyau inférieures à 5000 Hz en considérant les résonances de chaque tuyau séparément
- Donner l'allure du spectre de puissance de ce son avec les formants bien visibles
- Quelle voyelle devrait-on percevoir ?
- Que se passe-t-il sur ce spectre si on multiplie toutes les aires de ce tuyau par 2 ?
- On excite ce tuyau par une source périodique de fréquence fondamentale f_0 égale à 300Hz. Tous les formants seront-ils bien audibles ?

2. Perception de la parole – 5pts

- La parole téléphonique est limitée à la bande de fréquence 300-3300 Hz. Quels sont les phonèmes qui sont le plus distordus par cette limite? Donnez des exemples de confusions entre mots que l'on peut attendre lors d'une conversation téléphonique ?
- Avec un niveau de bruit modéré, quels sont les phonèmes qui ont le plus de chance d'être confondus avec le son [b]? Entre [a], [i] et [u], quel est le contexte vocalique dans lequel cette confusion a-t-elle le plus de chance de se réaliser ? Pourquoi ?
- De quelles informations additionnelles pourriez-vous disposer dans une conversation face-à-face pour éviter ces confusions ? Expliquez comment la perception compose ces informations avec celles issues du signal acoustique pour fabriquer un percept unique plus robuste.

3. Synthèse de la parole – 5pts

Soit un système de synthèse opérant par concaténation de demi-syllabes. Une demi-syllabe est le segment sonore allant du milieu d'une voyelle au milieu de la voyelle suivante

- Donnez une idée du nombre de demi-syllabes nécessaires à la synthèse du japonais (en première approximation considérez une structure syllabique unique CV et comptez 5 voyelles et 18 consonnes)

- Donnez une idée du nombre de demi-syllabes nécessaires à la synthèse du français (en première approximation considérez une structure syllabique unique CV et comptez 14 voyelles et 19 consonnes). Que se passe-t-il si on considère toutes les autres structures syllabiques possibles (donnez-les) ?
- Quels sont les problèmes posés par les frontières de syllabes et de mots ? Donnez des exemples ?
- Comment proposez-vous de tenir compte des frontières de syllabe et de mot ? Comment gérer les pauses éventuelles dans l'énoncé ?

4. **Reconnaissance de la parole – 5pts**

Soit une calculatrice élémentaire «permettant les quatre opérations de base » pilotée à la voix utilisant un système de reconnaissance.

- Combien de mots de vocabulaire le système de reconnaissance doit-il reconnaître au minimum ?
- Quels seront les paires de mots qui engendreront le plus de confusions ? Pourquoi ?
- Quels sont les phénomènes qui apparaissent si on ne fait pas de pause entre les mots de l'opération ? Comment les prendre en compte dans le système ?
- L'utilisateur utilise une micro-oreillette et vous donnez le résultat en synthèse de parole. Comment utiliseriez-vous cette synthèse pour que l'utilisateur soit sûr que le résultat est bien celui de l'opération qu'il a dictée ?