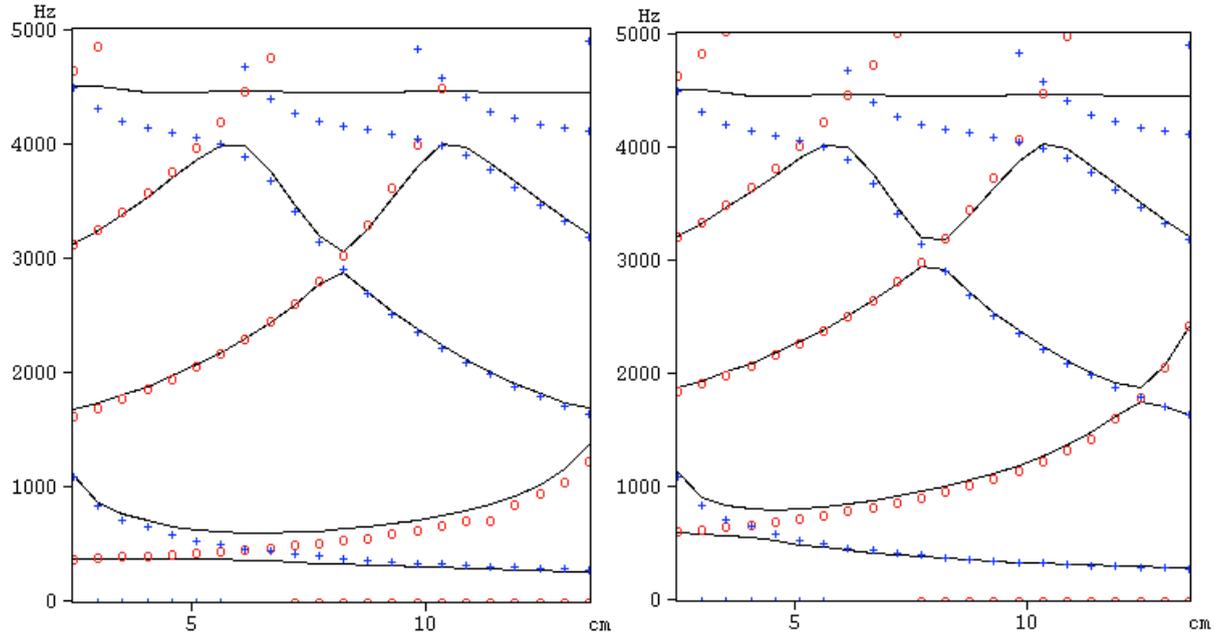


# IMAGE ET REALITE VIRTUELLE

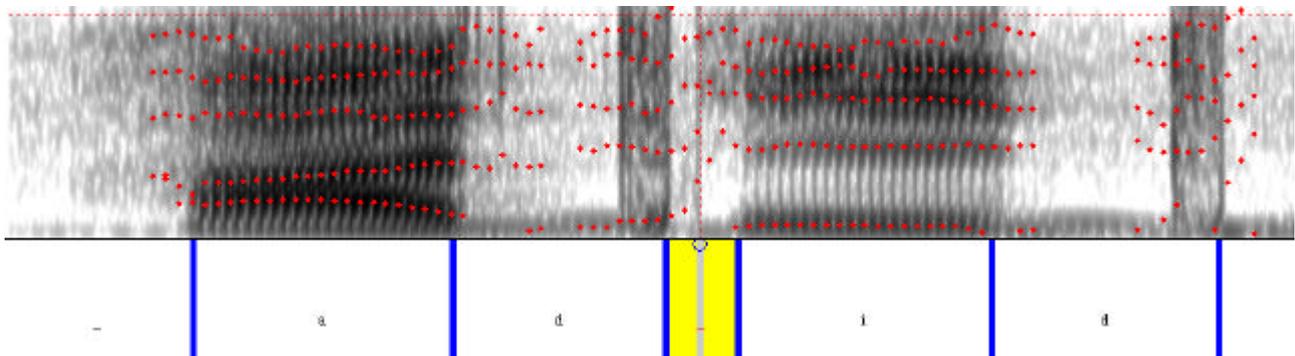
## Parole et langage

G. Bailly

Sujet d'examen du 17 février 2002 - 2 heures - Cours et documents autorisés



**Figure 1 :** Nomogrammes de conduits à 4 tubes (2 cavités + 2 constrictions). On fait varier le centre de la constriction de la glotte aux lèvres en maintenant les aires et les longueurs des constrictions constantes.



**Figure 2:** Spectrogram des séquences [ad] et [id].

### 1. Production de la parole – 8pts

Considérez Figure 1 les nomogrammes d'un conduit vocal à 2 cavités de 21cm.

- Reproduisez rapidement ces courbes sur votre copie et placez approximativement (par des barres verticales) les voyelles [a], [i], [u] et [y] sur ces nomogrammes selon l'emplacement de leur constriction linguale et leur aire aux lèvres.
- Que peut-on dire des affiliations entre les formants de ces voyelles et les cavités avant et arrière ?

Apico-dentales de la Figure 2

- Expliquez pourquoi dans [ad] le deuxième formant croit alors qu'il reste fixe dans [id].
- Que ce passerait-il si la consonne était la bilabiale [b] ?

Conduit neutre : considérez un conduit vocal neutre de  $L=8,5\text{cm}$  d'aire uniforme égale à  $4\text{cm}^2$ .

- Donnez les trois premiers formants (vitesse du son égale à  $c=360\text{m/s}^1$ ) et l'allure de l'enveloppe spectrale en amplitude.
- Supposons une fréquence fondamentale de 400Hz. Donnez la composition spectrale du son jusqu'à 3000Hz. Les amplitudes de ce spectre harmonique étant donné par l'enveloppe spectrale « vraie » ci-dessus, expliquez pourquoi il serait difficile de comprendre les sons émis par un enfant sur la seule base d'une estimation spectrale.
- Quels sont les connaissances auxquelles on doit faire appel pour estimer plus précisément l'enveloppe « vraie » à partir des simples raies harmoniques ?

## 2. Système de post-synchronisation – 7pts

Vous devez concevoir un système qui synchronise de manière optimale les mouvements originaux d'un acteur prononçant un texte en langue étrangère (vous disposez de la bande originale du film) avec la traduction du texte enregistrée par un acteur français. Vous devez opérer *a posteriori*, soit après l'enregistrement de l'acteur français, qui a déjà essayé de faire « au mieux ».

- Donnez trois stratégies possibles. Commentez leurs avantages et inconvénients.

Vous remplacez l'acteur français par un système de synthèse à partir du texte.

- Vous pouvez donc intervenir *a priori* sur la génération du doublage. Pouvez-vous en tirer parti ? Si oui, de quelle manière ?
- Divers choix de traduction sont possibles. Quelles métriques pourraient être invoquées pour décider quelle traduction assure une meilleure adéquation entre gestes originaux et son doublé ?
- En quoi les langues sources et cibles imposent des contraintes sur les choix lexicaux du traducteur ?

## 3. Reconnaissance audiovisuelle de la parole – 5pts

Vous devez concevoir un système qui utilise de manière optimale l'information audiovisuelle pour reconnaître la parole

- Quels sont les traits articulatoires (modes et lieux d'articulation) les plus difficiles à discriminer acoustiquement ? visuellement ?
- Expliquez pourquoi certains mouvements articulatoires peuvent être peu audibles ?
- Citez trois modèles possibles de fusion entre ces deux modalités ?
- Comment pondérer dynamiquement l'importance donnée à l'information délivrée par chaque modalité ? Pourquoi ces pondérations sont-elles nécessaires ? Comment ces pondérations peuvent-elles être estimées ?
- Quel est le modèle qui résistera le mieux à l'absence momentanée d'une modalité ? Pourquoi ?

---

<sup>1</sup> Rappel : mode de résonance quart d'onde :  $f = (2i-1)c/4L$ ,  $i$  entier strictement positif,  $L$  longueur de la cavité et  $c$  vitesse du son dans le milieu considéré.