IMAGE ET REALITE VIRTUELLE

Parole et langage — G. Bailly

Sujet d'examen du 2 février 2007 - 2 heures - Cours et documents autorisés

1. Structure du signal de parole (7 pts)

1. *Formants*. Quelle est la gamme de variation standard des deux premiers formants pour une voix d'homme ? Quels organes permettent de contrôler leurs valeurs ? Pourquoi l'espace de réalisation des deux premiers formants est-il un triangle ?
2. *Fréquence fondamentale*. Quelle sa gamme de variation standard chez un homme ? Quel organe permet de contrôler sa valeur ? Sa valeur peut-elle dépasser celle du premier formant ? Si oui, en quelles circonstances ? Quelles informations linguistiques et paralinguistiques les variations de fréquence fondamentale permettent-elles de véhiculer ?
3. La bande téléphonique étroite est de 300Hz-3300Hz. Comment expliquer que les variations de fréquence fondamentale soient clairement audibles au téléphone?
4. Quelle est la différence entre formants et résonances du conduit vocal ? Comment la fréquence fondamentale influence-t-elle cette différence ?
5. Des recherches menées dans les années 80 montrent que l'on peut estimer la taille et le poids d'une personne à partir de sa voix avec une incertitude de ±1.79 pouces et ±1.88 livres [Lass, N. J., C. A. Hendricks and M. A. Iturriaga (1980). "The consistency of listener judgements in speaker height and weight identification." Journal of Phonetics **8**(4): 439-448]. Quelle est l'influence supposée de l'âge et du poids sur la voix permettant de telles performances ?

2. Reconnaissance de parole multimodale (8 pts)

Vous devez mettre en œuvre un système de reconnaissance de parole silencieuse permettant de dicter un SMS ou de composer un numéro de téléphone sur un mobile sans clavier en articulant des sons sans les prononcer à haute voix.

1. Quelles techniques peuvent-elles fournir l'information nécessaire pour caractériser tout ou partie des mouvements articulatoires produits?
2. Quels sont les avantages et inconvénients d'utiliser une petite caméra vidéo filmant le visage de l'usager? Quels sont les paramètres essentiels que le système de vision devra extraire pour caractériser l'articulation visible ? Quel serait l'intérêt éventuel d'une vision stéréoscopique ?
3. Quels sont les avantages et inconvénients d'utiliser des électrodes de surface permettant de collecter l'activité électromyographique de certains muscles faciaux ? Où placeriez-vous ces électrodes ? (faites un croquis)
4. De ces deux dispositifs précédents, lequel est susceptible de récupérer de manière robuste des informations sur des articulateurs comme la langue ou le larynx ? Pourquoi et comment ?
5. Pourquoi ce système aura-t-il des difficultés à détecter l'activité vocale réelle de l'usager? Comment distinguer les mouvements de parole destinés au système de reconnaissance des autres ? Peut-on imaginer des mouvements faciaux spécifiques (instructions données à l'usager) signalant début et fin d'activité vocale ou des indices supplémentaires signalant cette activité (notamment utilisant le regard)?

Tourner la page svp ↳

3. Transcodage de parole multimodale (5 pts)

Vous devez mettre en œuvre un système de post-synchronisation du mouvement des lèvres d'un personnage animé sur un signal acoustique préenregistré. La bande son enregistrée par Tom Cruise ne peut être modifiée et le studio d'animation qui vous emploie doit donc faire articuler le héros virtuel du jeu de manière à ce que son et articulation soient les plus synchrones et cohérents possibles.

1. Vous décidez de capturer les mouvements des lèvres d'un locuteur qui va répéter la bande son. Expliquez comment un alignement par programmation dynamique peut vous permettre de remplir la tâche qui vous a été confiée. Quels sont les évènements importants qu'il serait intéressant de repérer sur les deux signaux acoustiques afin de contraindre l'alignement ? Quels sont les problèmes posés par cette solution ?
2. Vous disposez d'un ensemble de modèles HMM audiovisuels entraînés sur notre vedette ! Expliquez comment vous pourriez les utiliser ? Quel est l'intérêt de cette ressource ? Quels sont les problèmes posés par cette solution ?
3. Peut-on envisager d'utiliser cette dernière ressource pour doubler un dessiné animé produit par votre employeur avec la voix de Tom Cruise? Pourquoi faut-il imposer la chaîne phonétique complète ? Quelles informations doit-on encore fournir pour générer une voix de synthèse acceptable ?