

IMAGE ET REALITE VIRTUELLE

Parole et langage

G. Bailly

Sujet d'examen du 28 janvier 2008 - 2 heures - Cours et documents autorisés

1. Production de parole (4 pts)

Répondez de manière synthétique à cette série de questions de cours :

1. Quel est le mécanisme de base de l'oscillation des cordes vocales (responsable de la séquence séparation puis contact) ?
2. Qu'est ce qu'un résonateur de Helmholtz ? Donnez quelques exemples de voyelles présentant une telle cavité ?
3. Qu'est ce qu'un régime de phonation ? Donnez quatre exemples pour les consonnes.
4. Qu'est ce qu'un formant ? En quoi est-ce distinct d'un maximum spectral ?

2. Reconnaissance de parole (6 pts)

Vous faites partie d'une équipe de R&D d'un équipementier vendant des téléphones mobiles. La direction vous demande d'y intégrer une fonction de reconnaissance vocale permettant de numéroter automatiquement en prononçant le nom du correspondant. Les usagers ne devront faire aucun apprentissage préalable ni entrer aucune information additionnelle dans l'annuaire de leur carte SIM.

1. Dessinez un schéma bloc identifiant les divers modules de traitement (transcription orthographique-phonétique, décodage acoustico-phonétique, analyse lexicale, etc.) et les diverses ressources (annuaire, bases de données, modèles, etc.) nécessaires à la mise en œuvre de cette nouvelle fonction.
2. Identifiez les verrous technologiques. Quels sont les modules ou ressources posant le plus de problèmes de développement dans cette application ?
3. Comment détecter les erreurs de prononciation par défaut des noms qui a été calculée initialement par le système et comment les corriger semi-automatiquement ?
4. Que devrez modifier dans ce système initial pour prendre aussi en charge des commandes vocales simples comme l'activation/désactivation du mode silence, l'augmentation/diminution du volume sonore, la prise d'appel, etc. ? Quel est l'avantage de considérer des mots-clés comme accès à des menus et sous-menus ?
5. Comment adapter progressivement le modèle décodage acoustico-phonétique à la voix de l'utilisateur du téléphone ?
6. Que peut-on ajouter comme capteur ou comme module additionnel au système de traitement de la parole pour identifier l'utilisateur du téléphone et charger les bonnes ressources adaptées ?

3. Synthèse de parole (5 pts)

La ville de Marseille vient d'acquérir un centre d'appel équipé d'un serveur vocal qui doit filtrer automatiquement le maximum d'appels des usagers. Malheureusement la voix parisienne qui les accueille ne satisfait pas les clients et vous êtes chargés de remplacer le système de synthèse par concaténation d'unités stockées existant afin que la voix soit plus méridionale !

1. Dessinez un schéma bloc identifiant les divers modules de traitement (analyse linguistique, transcription orthographique-phonétique, etc.) et les diverses ressources (dictionnaires, règles de prononciation, etc.) exploitées par le système existant.

2. Quels sont les modules ou ressources que vous devriez modifier pour remplir l'objectif fixé ?
3. Le maire veut faire une surprise à ses concitoyens pour Noël et que le serveur parle avec sa voix. Quels sont les solutions qui s'offrent à vous ? Quels sont les avantages et inconvénients de ces solutions ?

4. Reconnaissance audiovisuelle du locuteur (5 pts)

BioSmart est une jeune start-up spécialisée dans la biométrie pour contrôle d'accès. Elle veut développer un nouveau produit par reconnaissance audiovisuelle du locuteur. Le système doit opérer à partir d'un portail vidéo à l'entrée de l'entreprise.

1. Comment envisagez-vous d'utiliser la modalité visuelle ? Quels sont les paramètres de variation d'apparence que vous devrez prendre en compte ?
2. Comment entraîner le système afin que les employés passent un temps minimal devant le portail vidéo ?
3. Comment envisagez-vous d'utiliser la modalité auditive? Comment choisir le texte à prononcer par l'employé afin de limiter les intrus ? Comment vérifier que le candidat n'est pas un intrus se faisant passer pour un employé ?
4. Comment vérifier que l'image et le son sont bien capturés in vivo et non préenregistrés ou synthétisés par un intrus potentiel ?