

IMAGE ET REALITE VIRTUELLE

Parole et langage

G. Bailly

Sujet d'examen du 4 Février 2010 - 2 heures - Cours et documents autorisés

1. Phonétique (6 pts)

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

1. Un métaplasme désigne les altérations phonétiques du mot par adjonction, suppression ou déplacement de phonèmes ou de lettres. Tentez d'expliquer par quels mécanismes – disparition de phonèmes, désynchronisation articulatoire, changements de mode de phonation, etc – est-on passé (a) du grec *kamara* (vôte) au latin *kamera* [kamera] au français *chambre* [ʃãbr], (b) de *stella* [stela] à *étoile* [etwal] en passant par *estela* [estela], (c) de *circare* [tʃirkare] à *chercher* [ʃɛʁʃe] en passant par *cerchier* [ʃɛʁʃje] en ancien français.
2. Comment fait-on la distinction en anglais (a) entre la prononciation de *beat* et *bit*, (b) entre *permit* (nom permis) et *permit* (verbe permettre), (c) entre *live* (verbe vivre) et *live* (nom vie).
3. Transcrivez phonétiquement la phrase suivante : « le loup s'est jeté immédiatement sur la petite chèvre » lorsque vous la prononcez de manière claire. Quelles modifications cette chaîne phonétique pourrait-elle subir si vous prononcez cette phrase rapidement (élision de e-muets, assimilation de voisement, etc) ?

2. Reconnaissance de parole (6 pts)

1. Donner le schéma général d'un système de reconnaissance de parole
2. La probabilité que la séquence de mot M corresponde à la phrase S prononcée est : $p(M/S)$. Grâce à la loi de Bayes, relier cette probabilité à $p(M)$, probabilité de la séquence de mots et $p(S/M)$, probabilité que S soit bien la prononciation effective de ces mots.
3. Comment un modèle de bigrams ou de trigrams simplifie-t-il le calcul de $p(M)$?
4. Le modèle acoustique a prédit la chaîne phonétique suivante : [lamɔrsyrprãlɔpɔtitami]. Parmi toutes les suites de mots possibles, le modèle lexical prédit les trois solutions suivantes: (a) la morsure prend le petit ami; (b) la mort surprend le petit ami, (c) la mort surprend le petit tamis. Le modèle de langage est un modèle de bigrams de mots entraîné sur un ensemble de phrases - très limité – suivant : « Le soldat brave la mort. La mort surprend le plus souvent sans prévenir. Elle embrasse le petit ami de sa sœur. La morsure n'entraîne pas la mort. Cette morsure prend le passant au dépourvu. Il secoue son tamis. Elle n'a plus de petit ami. ».
 - a. Donnez le produit de probabilités conditionnelles qui doit être calculé pour chaque phrase hypothétique et calculer sa valeur numérique. Notez que, si un bigram n'apparaît pas dans le corpus d'apprentissage, sa probabilité est nulle.
 - b. Basé sur vos résultats et considérant une probabilité équiprobable de chaque mot, quelle phrase sera reconnue ?
 - c. Le locuteur a en fait inséré une petite pause syntaxique après [syr]. Comment cette information aurait pu être introduite dans le modèle de bigrams ?

3. Conversion de voix (8 pts)

Vous venez d'être recruté par SSI (Silence Speech Interfaces) une start-up qui veut exploiter un brevet japonais sur la télécommunication silencieuse. Ce brevet décrit un système de conversion de

voix permettant de convertir une voix chuchotée captée par un microphone stéthoscopique - en contact avec la peau et placé sous l'oreille – en voix normale. L'objectif est de permettre l'usage des téléphones portables dans les espaces publics sans nuisance sonore. L'utilisateur chuchote à voix très basse et son interlocuteur écoute la voix normale convertie/calculée par le système proposé

1. Donnez deux stratégies possibles permettant cette conversion entre chuchotement et voix normale ? Quels sont des avantages et inconvénients d'effectuer une reconnaissance phonétique de la parole ?
2. Quels traits phonétiques (voisement, nasalité, lieux d'articulations, etc) sont les plus dégradés en parole chuchotée ? Ceux qui seront les plus difficiles à restituer ?
3. Quels sont les types de bruits que le microphone stéthoscopique récupère et ceux dont il est relativement préservé si on le compare à un microphone-cravate ?
4. Comment peut-on améliorer la robustesse du système ? Peut-on envisager un retour auditif pour l'utilisateur ? Si oui, quels sont les problèmes ?
5. Un concurrent exploite un système basé sur une sonde ultrasonore (échographe) fixé sous le menton permettant de capter une partie des mouvements de la langue. Cette start-up veut carrément proposer un système sans production de son (même chuchoté), juste basé sur les mouvements articulatoires. Est-ce réaliste ? En quoi une collaboration peut-elle intéresser SSI ?