

Protocole d'enregistrement d'un corpus audio

1.Introduction :

Pendant la reconnaissance automatique de la parole, les corpus sont conçus pour répondre à des défis méthodologiques et scientifiques posés par la grande variabilité du signal acoustique de la parole, ensuite pour couvrir au mieux un type d'application .

2.Défis posés par le signal de parole :

La parole, étant un signal physique continu, pose un nombre de problèmes avant de pouvoir être transformé en signe linguistique. Ces problèmes sont principalement liés à la variabilité de la langue, et parmi lesquels on trouve:

- 1.En générale, il n'y a pas de frontières détectables entre les mots et les frontières entre les phonèmes sont peu évidentes
- 2.Le signal de parole varie en fonction du locuteur (sexe, âge, émotions, accent...)
- 3.Le bruit de fond se superpose à la parole pendant l'enregistrement
- 4.Le style de la parole(lue, préparée, spontanée...), influe fortement sur la parole

3.Modélisation statistiques :

La transcription automatique de la parole vise à convertir le signal acoustique de parole en **suites de mots (phrases)** . C'est ainsi que, cette transcription repose sur une modélisation statistique de la parole .

3.1 Mécanisme de la modélisation :

Dans un premier temps, on doit mesurer la probabilité de toutes les suites de mots m possibles pour ce signal x **$P(m|x)$** . Ensuite selon la formule de Bayes , on a **$P(m|x) = P(m)*P(x|m)$** , où **$P(m)$** signifie la probabilité de la séquence de mots m (grâce à des modèles N-grammes), et **$P(x|m)$** est la probabilité d'observer le signal x connaissant les mots m . Donc, afin de calculer le terme $P(m|x)$, il suffit de savoir calculer les deux termes $p(m)$ et $p(x|m)$.

Ainsi, des suites de mots très fréquentes dans la langue telles que (*de la*) auront une probabilité $P(m)$ très élevée, or des suites de mots très rares dans la langue telles que (*la de*) auront une probabilité très faible.

3.2 Loi zipf :

La modélisation statistique nécessite un corpus de taille très grande pour ne pas observer des phénomènes équiprobables .

Selon la loi de Zipf, la langue possède peu de mots très fréquents et un très grand nombre de mots très rares . Or, l'étude linguistique fondée sur des grands corpus de représenter surtout des mots les plus fréquents. Alors, des techniques doivent être développées pour traiter les événements rares .

4.Etudes linguistiques :

4.1 Phonologie de corpus :

La notion phonologique est une notion très économique car elle permet d'associer à chaque mot une prononciation canonique. Or, la linéarité du signal acoustique (dans le temps) entraîne que les éléments à *priori* se présentent les uns après les autres (ils forment une *chaîne*).

4.2 Problème de texte spontané :

Dès lors, pour une bonne reconnaissance de parole on utilise la représentation phonologique sous forme de chaînes de phonèmes, et à cette représentation on associe des modèles de Markov cachés à mélange de gaussiennes. Pourtant, même si cette modélisation s'avère bien efficace pour une parole bien articulée (texte lu), elle pose des problèmes pour une parole spontanée : les prononciations observées peuvent avoir un contenu très différent de la forme canonique car des phonèmes et des syllabes entières peuvent disparaître .

4.3 Problème du message parlé :

Il apparaît dans la parole que tous les phonèmes ne sont pas articulés de la même précision. Cependant, ceci n'est pas toujours lié à la nature du phonème (comme pour le *schwa* en anglais), mais souvent à sa fonction dans le message parlé.

En outre, il apparaît dans les corpus que les mots à forte redondance sont peu articulés voire inexistent dans certains cas extrêmes, car, si l'information est donnée soit par le niveau syntaxique soit par le niveau pragmatique, alors le niveau acoustique n'a pas besoin d'être très complet. (à l'heure actuelle, nous manquons d'une description détaillée de ces phénomènes)