CONCLUSION

Le système de reconnaissance présenté dans ce document est constitué de deux modules. Le premier module développe les N meilleures solutions pour un mot (ou une phrase) prononcé et le second module recherche la solution la plus pertinente dans la liste proposée.

Nous avons présenté dans ce document un ensemble de méthodes de recherche des N meilleures solutions. La méthode retenue dans notre cas est déduite de celle de [Soong, 91]. Elle permet d'obtenir d'une manière optimale les N meilleures solutions et cela en deux passes. L'algorithme de Viterbi est utilisé dans la phase aller et l'algorithme A* dans la phase retour. Nous avons adapté la méthode de Soong à la recherche des N meilleures solutions dans un réseau acoustique afin d'obtenir la segmentation (phonétique) associée à chaque solution. On a présenté la manière d'introduire cette méthode de recherche et son amélioration pour avoir des solutions dans un temps raisonnable. Le temps de recherche est une fonction affine du nombre de solutions développées, pour N petit

Les résultats obtenus sur différents corpus de données testés montrent qu'on peut espérer réduire le taux d'erreur dans un rapport de 4, en utilisant un post-traitement approprié permettant de choisir la bonne réponse parmi les 5 solutions proposées.

Dans le cadre d'une application de reconnaissance de villes épelées, le post-traitement approprié est un post-traitement syntaxique. En effet, on dispose d'un dictionnaire spécifiant les noms de villes possibles. Pour un dictionnaire de 30000 villes le post-traitement syntaxique conduit à un taux de reconnaissance correct des épellations de 84%. Le taux d'erreur était très faible, de l'ordre de 3% seulement. Le taux de rejet, de l'ordre de 13%, peut ne pas être pénalisant dans le cadre d'une application s'il est possible de renvoyer les appels correspondants vers un opérateur.

Dans le cas où il n'y a pas de contraintes syntaxiques complexes (chiffres, Trégor et nombres à deux chiffres), le post-traitement syntaxique n'est plus adapté ; on doit donc s'orienter vers d'autres approches. L'approche segmentale dans ce cas est la plus appropriée. Elle suppose qu'on dispose de la segmentation de chaque solution en phonèmes. Elle consiste à calculer un nouveau score pour chaque solution. Ce score est une combinaison du score obtenu par le module fournissant les N meilleures solutions et du score segmental.

Nous avons mentionné auparavant, que la méthode généralement utilisée pour calculer le score segmental d'une solution, utilisait des réseaux de neurones. Nous avons proposé dans le cadre de cette thèse d'utiliser une approche statistique. Pour cela, il faut définir des modèles statistiques pour chaque segment de chaque solution.

La première modélisation développée, incluait des modèles uniquement appris sur des solutions correctement reconnues. Les résultats obtenus ne permettaient pas de réduire d'une manière

significative les taux d'erreur. En plus de cela nous avons constaté que cette manière de modéliser ne permettait pas de tenir compte de la nature des solutions développées (correcte ou incorrecte). De ce fait nous avons voulu réaliser une modélisation qui tienne compte du fait qu'une solution peut-être correcte ou incorrecte. Pour cela, nous avons développé une nouvelle approche utilisant conjointement des modèles corrects et des modèles incorrects. Le premier modèle décrit les statistiques d'un segment qui appartient à une solution correctement reconnue. Le second modèle représente les statistiques des segments qui appartiennent à des solutions incorrectement reconnues. Le score segmental d'un segment d'une solution est obtenu en combinant les scores fournis par ces deux modèles.

Ce qui il faut retenir des résultats présentés :

- Le post-traitement segmental seul obtenait des meilleurs performances avec des modèles corrects/incorrects qu'avec des modèles corrects uniquement. Le gain était dans ce cas de 40%.

- Sur les différentes configurations testés, la configuration la plus appropriée est : contexte dépendant, avec silence et durée relative. Même si ses résultats sont équivalents à ceux obtenus par une configuration contexte dépendant, sans silence et avec durée absolue ; il reste néanmoins que cette configuration nécessite moins de paramètres que la configuration contexte dépendant, avec silence et durée relative. La prise en compte des contextes droit et gauche est un apport considérable pour la modélisation.

- En combinant les scores PTS et HMM les réductions du taux d'erreur par rapport à l'utilisation du HMM seul étaient de 15% à 23% suivant les corpus de données utilisés.

- En parallèle à ce travail, l'approche à base de réseaux connexionnistes a été testée par [Boiteau, 93-a]. Les performances obtenues sont similaires à celles de l'approche statistique.

A partir des résultats obtenus, des expériences effectuées et des remarques soulevées, nous présentons dans ce qui suit un ensemble de perspectives pour espérer plus d'amélioration. A la sortie du module markovien, on dispose du nom, du score et de l'alignement de chaque solution. En se basant sur ces informations, nous proposons d'explorer les domaines suivants :

- Vecteur des coefficients

Calculer les paramètres du vecteur acoustique non plus sur une fenêtre de 32 ms, mais sur une fenêtre de la largeur du segment. L'évolution temporelle sera déterminée à partir des vecteurs statistiques calculés sur les 2 moitiés du segments.

Il est aussi intéressant de rechercher dans le vecteurs des coefficients les paramètres les plus influants.

- Un second modèle de Markov pour améliorer la segmentation

Un premier modèle de Markov (HMM1) très performant va calculer les *N* meilleures solutions. Un second modèle de Markov (HMM2) sera utilisé pour générer la segmentation. La génération est effectuée en alignant chaque solution obtenu par le HMM1 sur le HMM2. Ainsi le calcul des paramètres des modèles associés à chaque segment va tenir compte des frontières entre segments fixées par HMM2. Des tests préliminaires ont été réalisés sur le corpus des chiffres isolés, avec un modèle de Markov à 3 densités de probabilité (4 états), pour tous les phonèmes. En modélisation par allophones, la première densité symbolise le contexte droit, la seconde la partie centrale, la troisième le contexte gauche. En comparant les résultats obtenus sur le post-traitement segmental seul, corpus de test, avec une configuration contexte dépendant, avec silence et durée relative (1 gaussienne), les résultats étaient de 3.4% au lieu des 4% fournis avec la segmentation résultante du premier modèle utilisé. Cependant la procédure de calcul du score segmental devient très lourde en temps de calcul, puisque à chaque fois qu'une solution est générée, une procédure d'alignement sur le second modèle, est effectuée. Pour contourner le problème, on a examiné le cas ou le HMM1 était appris avec une contrainte de segmentation issue du HMM2. Cette manière de procéder nous semblait efficace dans le sens où on gardait la même structure du réseau acoustique du HMM1 qui donnait les meilleurs performances en reconnaissance et en plus on dotait HMM1 d'une segmentation plus fine obtenue par le HMM2. Les résultats obtenus sur le corpus de test des chiffres montraient une amélioration significative sur le PTS seul, on est passé de 4% à 3.1%. Par contre le résultat obtenu en combinant HMM et PTS était moins bon puisque on est passé de 0.85% (HMM1 de départ) à 1.63%.....

Même si les résultats obtenus ne semblent pas encourageant, on pense qu'une étude plus approfondie sur l'utilisation d'une autre segmentation comme post-traitement est à approfondir.

- Bruit du canal de transmission

Des tests ont été réalisés sur les différentes corpus de données dans un but d'éliminer le bruit du canal [Mokbel, 93]. Les premiers résultats obtenus montrent une réduction significative du

taux d'erreur en réduisant au premier ordre l'effet du canal. Puisque l'approche segmentale est appliquée dans le module de post-traitement, on dispose donc de tout le signal de parole. On peut donc estimer le bruit lié au canal et le soustraire du signal de parole avant d'appliquer l'approche segmentale.

Malgré l'amélioration significative obtenue avec le système proposé dans cette thèse, la question de la reconnaissance segmentale persiste et des travaux dans cette direction sont à poursuivre.