Traits Prosodiques dans la Modélisation Acoustique à Base de Segment

Björn W. Schuller
Institute for Human-Machine Communication, Technische Universität München, Allemagne
schuller@tum.de

1. Introduction

À part les traits cepstraux habituellement employés pour la reconnaissance des phonèmes, la durée, les formants, le centre de gravité, l'hauteur et l'intensité sont aussi bien connues de pouvoir retenir d'information de valeur. Ce travail enquête sur leurs performances dans la modélisation acoustique à base de segment. Nous utilisons CFSS pour l'optimisation spatiale. Toutes les calculs sont réalisés avec SVM sur le libre corps IFA, qui contient 47 phonèmes hollandais segmentés et étiquetés manuellement de 4 femmes et 4 hommes, entre l'âge de 15 à 66 ans [1].

Mots-clés: Modélisation Acoustique, Traits Alternatifs

2. Expériments et Résumé

Nous considérons un 1er jeu de 13 x 3 des contours pour les traits cepstraux et un 2ème jeu de 6+4 pour les traits prosodiques et spectraux, présentés dans tab. 1. Pour les traits prosodiques et spectraux il existe 4 contours en version lissée (formants et hauteur). Moyenne et écart-type ne sont appliqués que pour l'hauteur et les traits cepstraux. 429 caractéristiques statiques par segment pour les traits cepstraux et 92 pour les traits prosodiques et spectraux provenant de l'application d'analyse statistique aux contours. Nous employons 3-pli validation par orateur [2].

Table 1. Contours et fonctionnels utilisés pour la construction des espaces de traits

Contour	Analyse
MFCC (0-12)	Min., Max.
ΔMFCC (0-12)	Val. au début du seg.
ΔΔMFCC (0-12)	Val. au 1/4 du seg.
Intensité	Val. au 2/4 du seg.
Hauteur	Val. au 3/4 du seg.
Pos. de Formant 1-3	Val. à la fin du seg.
Centre de Gravité Spectrale	Rel. pos. min./max.
	Moyenne, Écart-type

Le tab. 2 montre les résultats respectivement pour oratrices et orateurs. Le nombre d'instances valides est indiqué. En général les variantes de trait crues sont préférées à celles qui sont lissées. Grosso modo, l'exactitude 76% peut être annoncée, en employant seulement les traits cepstraux, contrairement aux taux de traits non ordinaires, prosodiques et spectraux, qui sont d'environ 60%. La combinaison des jeux ajoute 0.91% absolument.

Table 2. Exactitude pour la reconnaissance des phonèmes, CFSS, IFA, SVM, 3-pli validation

Exactitude [%]	Fém.	Mâle	F+M
#Instances	116091	61993	178084
Prosod. + spect.	60.79	53.32	58.19
MFCC	76.86	73.44	75.67
MFCC + autres	**78.08**	**73.77**	**76.58**

Ensuite, la distribution des traits après CFSS nous révèle qu'en moyenne 27% des traits sont non cepstraux. Au niveau pertinence, le type de trait MFCC vient en premier lieux, suivit par ΔMFCC en seconde position et ensuite vient comme prévu ΔΔMFCC. On note aussi l'importance de la valeur au centre du segment et la haute contribution de ses voisins immédiats à l'intérieur du segment, en particulier pour les traits MFCC. D'une façon intéressante, les positions des fonctionnels Min. et Max. à l'intérieur du segment sont plus pertinentes que les valeurs eux-mêmes, ceux-ci est plus discernable chez les MFCC. Dans la distribution des traits prosodiques et spectraux trouvés dans la sélection de trait, tout d'abord viennent centre de gravité spectrale et intensité. Ensuite, surtout F1 et F2 montrent une haute contribution. D'une façon intéressante, aussi F3 a un poids remarquable du point de vue du nombre total de traits. Enfin ça semble remarquable que l'hauteur a été aussi choisi.

Dans les travaux futurs nous visons l'analyse dans un cadre SVM/HMM.

Références

[1] R. van Son, D. Binnenpoorte, H. van den Heuvel, L. Pols, "The IFA corpus: a Phonemically Segmented Dutch 'Open Source' Speech Database," Proc. Eurospeech, Aalborg, Danemark, 2001.

[2] B. Schuller, X. Zhang, G. Rigoll: "Prosodic and Spectral Features within Segment-based Acoustic Modeling", Proc. 9th INTERSPEECH 2008, Brisbane, Australie, 2008.