Acoustic Parameters for Automatic Detection of Nasal Manner

Tarun Pruthi, Carol Y.Espy-Wilson, Speech Communication Lab, Institute for Systems Research, ECE Department
http://www.isr.umd.edu/Labs/SCL/

INTRODUCTION

Nasals are the only class of sounds with dominant speech output from the nasal cavity instead of the oral cavity. This gives them some very unique properties and makes nasals one of the hardest sounds to study. Very few people have attempted to find Acoustic Parameters (APs) for automatic detection of nasal manner before. Here, we focus on the development of APs for the linguistic feature *nasal*, particularly so that they can be extracted automatically and reliably in a speaker independent way.

ACOUSTIC PARAMETERS

The nasal sounds (/m/, /n/, /ng/) and the semivowels (/r/, /l/, /w/, /y/) share many acoustic properties and, therefore, may be confused with each other. Fig. 1 compares a typical nasal with a typical semivowel. The following four APs are used to capture acoustic differences between the sounds:

•*Energy Onsets/Offsets*: stronger for nasals.

•*Energy Ratio*: larger for nasals.

•*Spectral Peak Frequency*: lower for nasals.

•*Envelope Variance Measure:* less for nasals.

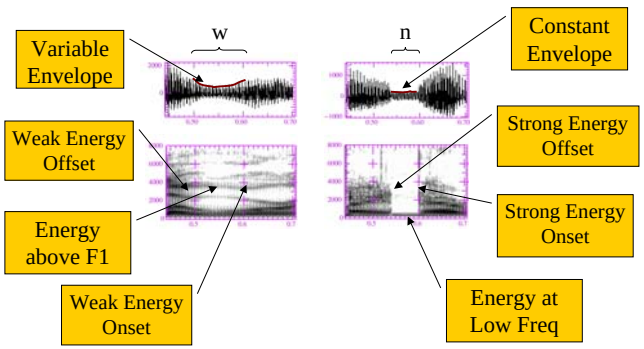

Figure 1. A typical semivowel /w/ is shown on the left and a typical nasal /n/ is shown on the right.

CLASSIFICATION EXPERIMENT

This experiment evaluated the efficiency of the proposed APs on a classification task between nasals (/m/, /n/, /ng/) and semivowels (/r/, /l/, /w/, /y/).

Database

Training: 2586 'si' and 'sx' sentences spoken by 90 male and 235 female speakers from dr1-dr7 of TIMIT training database. Testing: 504 'si' sentences spoken by 56 female and 112 male speakers from dr1-dr8 of TIMIT test database.

Results

Tables 1-3 give confusion matrices of the classification results for prevocalic, postvocalic and intervocalic sonorant consonants in the test database.

Table 1. Confusion matrix of the classification results for prevocalic sonorant consonants

	Nasals	Semivowels	% Correct
Nasals	259	19	93.17
Semivowels	114	878	88.51

Table 2. Confusion matrix of the classification results for postvocalic sonorant consonants

	Nasals	Semivowels	% Correct
Nasals	921	42	95.64
Semivowels	15	380	96.20

Table 3. Confusion matrix of the classification results for intervocalic sonorant consonants

	Nasals	Semivowels	% Correct
Nasals	368	33	91.77
Semivowels	75	411	84.57

Averaging across the three classes of sonorant consonants gives accuracies of 94.27 % and 89.11 % for nasals and semivowels respectively.

RECOGNITION EXPERIMENT

This was a digit recognition experiment with the APs as the front-end and HMM as the back-end. The purpose of this experiment was to quantify the reduction in confusions between English digits (particularly 1 and 9) by adding the APs proposed in this work to the original set of APs.

Database

TI46 database was used for this experiment.

Results

The results are shown in Tables 4 and 5 below.

Table 4. Confusion matrix with old set of 28 APs

	One	Five	Nine	Total	% Correct
One	250	0	5	255	98.0
Five	0	254	0	254	100.0
Nine	4	2	248	254	97.6

Table 5. Confusion matrix with new set of 30 APs

	One	Five	Nine	Total	% Correct
One	254	0	1	255	99.6
Five	0	254	0	254	100.0
Nine	0	1	253	254	99.6

FUTURE WORK

In future we would like to work on the development of APs for vowel nasalization. We believe this can give a substantial improvement in the detection of nasals because at times nasals might be articulated only as vowel nasalization. We would also like to work on extracting parameters for nasal place detection.

ACKNOWLEDGEMENTS

This work was supported in part by NSF grant BCS0236707.