Automatic Classification of Nasals and Semivowels

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

INTRODUCTION

In this paper, we discuss acoustic parameters and a classifier we developed to distinguish between nasals (/m/, /n/, /ng/) and semivowels (/r/, /l/, /w/, /y/). Based on the literature and our own acoustic studies, we use an onset/offset measure to capture the consonantal nature of nasals, and an energy ratio, a low spectral peak measure and a formant density measure to capture the nasal murmur.

DATABASE

The training tokens were chosen from 2586 'si' and 'sx' sentences spoken by 90 females and 235 males from the dialect regions 1-7 of the TIMIT training database. The test data consisted of 504 'si' sentences spoken by 56 females and 112 males from dialect regions 1-8 of the TIMIT test database.

METHOD

In our experiments, the TIMIT transcription was used to identify the nasal and semivowel boundaries, and to classify them as prevocalic (before a vowel), postvocalic (after a vowel) and intervocalic (between vowels). Nasal flaps (/nx/) and syllabic nasals (/em/ or /en/) were not included in this study. We use the following parameters:

•Energy Ratio E(0-358 Hz)/E(358-5373 Hz)

•An estimate of F1

•An indirect measure of the formant density

•Onsets and offsets

The four APs are then used for training three different Support Vector Machines (SVMs) (one each for prevocalic, postvocalic and intervocalic sonorant consonants) for binary classification of the test data.

RESULTS

An example of the APs extracted:

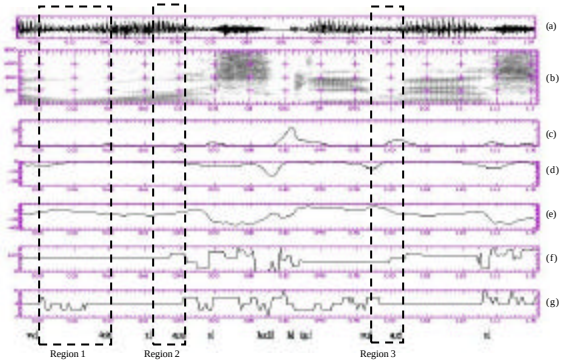

Figure 1. Acoustic parameters for an excerpt "wore a ski mask" from the file dr1.mgrl0.sx57.wav from TIMIT training database. (a) Waveform, (b) Wideband Spectrogram, (c) onsets, (d) offsets, (e) energy ratio, (f) F1 measure, (g) Formant density measure

Tables 1-3 give confusion matrices from the classification results for prevocalic, postvocalic and intervocalic sonorant consonants in the test database. Table 4 gives a comparison of our results and test database with those of earlier work.

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

	Nasals	Semivowels	% Correct
Nasals	250	28	89.93
Semivowels	117	875	88.21

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

	Nasals	Semivowels	% Correct
Nasals	912	51	94.70
Semivowels	19	376	95.19

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

	Nasals	Semivowels	% Correct
Nasals	355	46	88.53
Semivowels	87	399	82.10

Table 4. Comparison of Results and Test Database with earlier work

Author	Features	Results	Test Database
Glass [1]	Total energy, energy stability, % of time a resonance is below 350 Hz, E(0-500)/E(total), E(0-350)/E(350-1000)	88 % correct identification (Impostor class: semivowels and voice bars)	~200 words each spoken by 3M/3F, min. coarticulation, Training on 5 speakers, testing on 6 th
Chen [2]	Sum.diff, Sum.amp.diff, A1-A2, A1-A3, A2-A3, F1, A1-P0a, A1-P1a, Uses Liu's landmark detector first	88 % correct nasals 74 % correct non-nasals (on training set with LDA, 5 feature)	100 LAFF sentences, first 20 used in training also, 2 M/2F
Pruthi [3]	Onsets/offsets, E(0-358)/E(358-5373), F1 measure, Formant density measure, Will use our own landmark detector	90.1 % correct 92.4 % correct nasals 88.1 % correct semivowels	TIMIT test DB 504 'si', dr1-8, 112M/56F, Lot of coarticulation

FUTURE WORK

In future experiments, we will integrate additional phonetic features such as *lateral* and *rhotic* to help in this distinction as well. We will also integrate these features with the existing EBS and develop APs for the place features *labial*, *alveolar* and *velar* for nasals.

REFERENCES

[1] J.R. Glass, *Nasal Consonants and Nasalised Vowels: An Acoustical Study and Recognition Experiment,* M.S. and E.E. thesis, MIT, Cambridge, MA, 1984.

[2] M.Y. Chen, "Nasal Detection Module for a Knowledge-based Speech Recognition System," *Proceedings of the ICSLP 2000*, Vol. IV, pp. 636-639, Beijing, China, 2000.

[3] Tarun Pruthi, Carol Y. Espy-Wilson, "Automatic Classification of Nasals and Semivowels," *15th International Congress of Phonetic Sciences (ICPhS) 2003,* Barcelona, Spain, August 2003.

ACKNOWLEDGEMENTS

This work was supported in part by NIH grant 1 K02 DC00149-01A1.