Automatic Speech Recognition: Identifying Nasalized Vowels

Shuo Chen

Introduction

In continuous speech, nasal sounds **/m/**, **/n/**, and **/ng/** can be substantially co-articulated with vowels – so much so, that at times, nasality is present only as vowel nasalization. The **accurate identification** of **nasalization** can **improve** the detection of nasals in general and is, hence, important to acoustic-phonetic **speech recognition** research. In addition, an established method of nasalization detection can also facilitate **speaker recognition** because degrees of nasality in speech vary uniquely from speaker to speaker. This research explored various, possible acoustic parameters (APs) that could be applied towards nasalization detection.

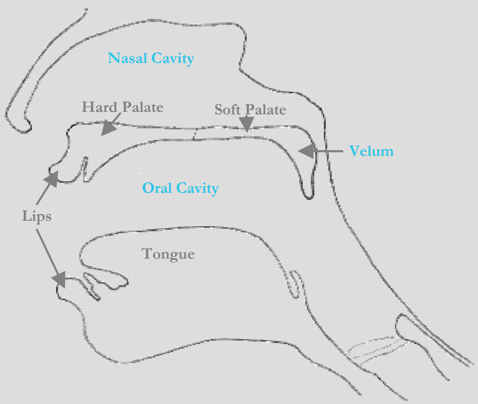

Figure 1. Illustration of Vocal and Nasal Passages

Physiological Cause

Nasalization occurs when the velum is partially lowered and the nasal cavity and oral cavity are coupled together. The addition of the nasal cavity to the vocal tract creates several **spectral changes**. These changes include the appearance of nasal resonances, nasal antiresonances, overall amplitude attenuation, and formant bandwidth broadening. All these changes can be **characterized** using **APs** and extracted from test waveforms to examine nasality.

Acoustic Parameters (APs)

1. Mean of **amplitude differences** between **A1 and P0.** Where A1 is the amplitude of the first formant and P0 is the amplitude of the first nasal resonance.
2. Mean of **amplitude differences** between **A1 and P1.** Where A1 is the amplitude of the first formant and P1 is the amplitude of the second nasal resonance.
3. Average **slope** of **spectrum envelope** between **0 and 1500 Hz.** Where slope was calculated in the frequency domain for the spectrum envelope underlying the voice harmonics.

4. Mean **area beneath** the **frequency spectrum.**
5. **Maximum** surface **area difference** beneath spectrum.
6. Average **center of mass** over **0-1000 Hz**
7. Average **standard deviation** of **energy levels** at **500 Hz from** the **center of mass.** Where range was established by adding and subtracting 500 Hz from previous AP's calculations.
8. Average **bandwidth** of **first formant**

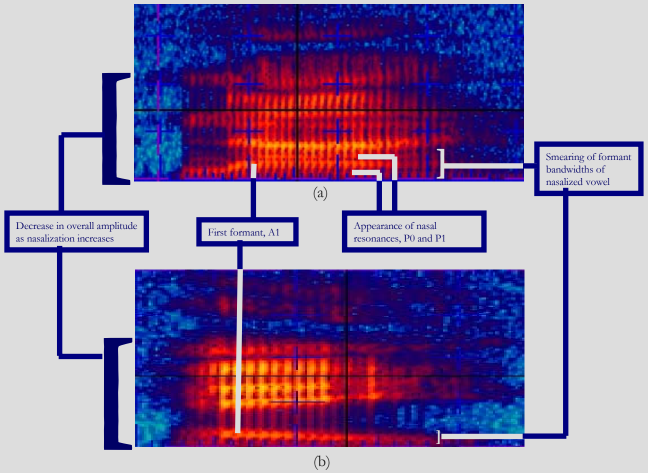

Figure 2a) Spectrogram of nasalized vowel. b) Spectrogram of non-nasalized vowel.

Experiment

The purpose of the experiment was to test the proficiency of the chosen APs at differentiating between nasalized and non-nasalized vowels. The accepted vowel set consisted of vowel sounds /iy/, /ih/, /eh/, /ey/, /ae/, /aa/, /aw/, /ay/, /ah/, /ao/, /oy/, /ow/, /uh/, /uw/, and /er/. Vowels followed by or buffered between nasals were considered nasalized; vowels not next to nasals were considered nonnasalized.

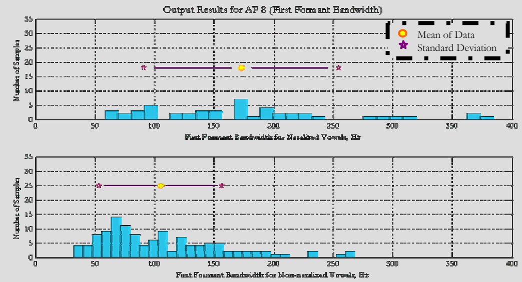

Figure 3. Histogram results of AP 8

Results and Conclusions

Although the APs were capable of separating some nasalized samples from non-nasalized samples, there is still **considerable overlap** between the nasalized and non-nasalized parameter values. This **error region** may result from **external factors** as well as internal faults:

- **Definition** of **nasalization not inclusive** of all examples of nasalization, leaving exclusions as non-nasalized samples
- Overlooked nasalization in vowels buffered from nasals by a semivowel.
- Overlooked pre-nasal nasalization in vowels buffered from nasals by another vowel.
- **Inadvertent nasalization** by speaker regardless of proximity to nasals.

With more refinements to the APs and adjustments to the definition of nasalization, future work with the proposed APs **may still yield positive, functional results**.