Recognition of Nasalized & Non-nasalized Vowels

Bilal A. Raja, Carol Y. Espy-Wilson, and Tarun Pruthi
Speech Communication Lab, ECE Dept and Inst. Of Systems Research, University of Maryland, College Park, MD 20742

MERIT 2006 FAIR

Introduction

The presence of nasal consonants (/m/, /n/, /ng/) adjacent to vowels leads to nasalization of at least some part of the vowel. Nasalization is known to affect the perceived quality of vowels, and hence, their recognition accuracy. The objective of this project is to verify if it is beneficial to first recognize nasalization in vowels and treat the three groups of vowels (those occurring before nasal consonants, those occurring after nasal consonants, and those that are not in the context of nasal consonants, that is, oral vowels) separately rather than collectively for recognizing vowel identities.

What is Nasalization?

- Production of a sound while the velum is lowered.
- Some air escapes through the nose during the production of the sound by the mouth (see Figure 1).
- The resulting effect is as if an /n/ sound is being produced simultaneously with the oral sound.
- Common nasalized sounds are the nasalized vowels.
- Common spectral characteristics of nasalization are depicted in Figure 2.

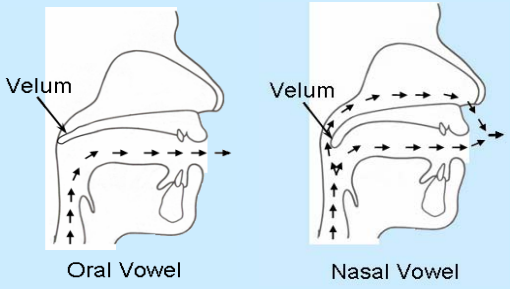
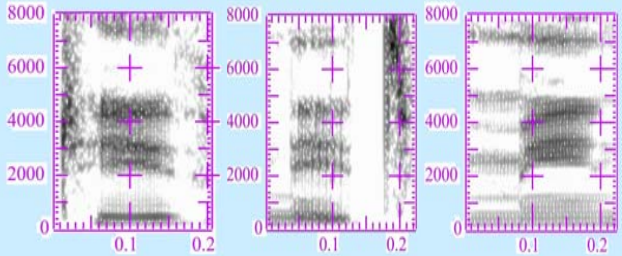


Figure 1: Production of oral and nasal vowels.



(a) OV CASE iy in “Teeth” (b) NV CASE iy in “Need” (c) VN CASE iy in “Mean”

Figure 2: Spectrogram of three words spoken by the same speaker with the same vowel ‘iy’ occurring in different contexts. Reduction in the amplitude of the first resonance, more peaks, and the flatter spectrum at low frequencies in (c) characterizes nasalization.

Expected Results

- Bond (1975) suggested that the recognition rates for vowels is worse when the vowels are excerpted from a context in which the vowels are affected by nasality.
- According to Bell-Berti (1993), carry over coarticulation is usually much smaller than anticipatory coarticulation.
- Keeping in mind the above mentioned results we expected the following:
  1) Vowels with the nasal consonants before them are not nasalized to a large extent. Therefore, their identification rates should be comparable to oral vowels.
  2) Vowels with the nasal consonants after them are nasalized strongly and should have worse identifiability.

Method

- Mel-Frequency Cepstral Coefficients (MFCCs) were used in a Hidden Markov Model (HMM) framework for all experiments in this project (see flowchart in Fig 3).
- HMM Tool Kit (HTK) was used to carry out the tests.
- All the vowels in TIMIT (1990) database were extracted from speech files by using TIMIT transcriptions.
- Vowels were divided into three categories:
  1) Oral Vowels (OV)
  2) Vowels before the nasal consonants (VN)
  3) Vowels after the nasal consonants (NV)

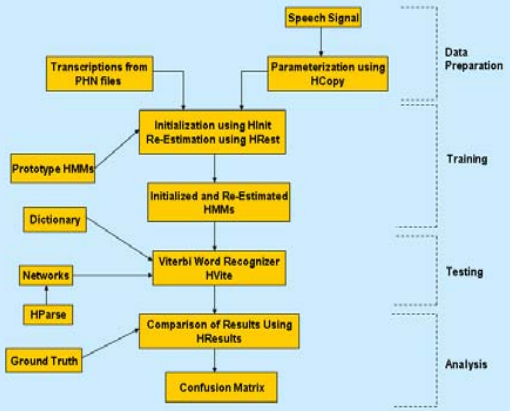


Figure 3: HTK Processing Stages.

Results

Experiment-1 Training was done using vowels from all the categories. Testing was done first on all vowels, and then on each individual category to find the contribution of each category to errors. Table 1 summarizes the results.

Table 1: Percentage Recognition Accuracy of different Categories

Category	Recognition Accuracy
ALL	52.92%
OV	55%
NV	58.25%
VN	39.75%

Experiment-2 Each category was individually trained and then tested. Table 2 summarizes the results. An example confusion matrix obtained through HTK is shown in Figure 4.

Table 2: Percentage Recognition Accuracy of Different Categories

Category	Recognition Accuracy
ALL	52.92%
OV	56%
NV	58.04%
VN	42.86%

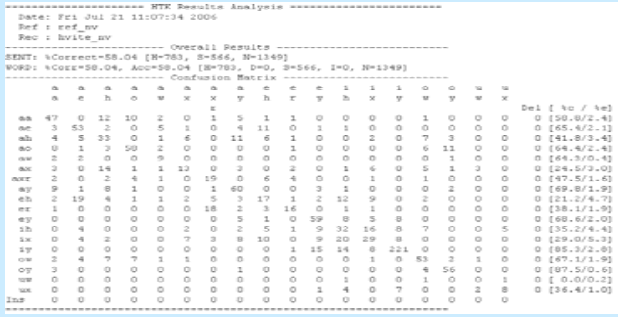


Figure 4: Confusion Matrix for the NV category showing the percentage accuracy of each vowel and the overall recognition accuracy.

Conclusion

The results show that when the system is trained on vowels in general, the recognition of nasalized vowels is 17% below that of oral vowels. The recognition accuracy for the VN category is improved by 8% when different vowel models are created for each category. This result suggests that automatic recognition of vowels can be improved by first detecting nasalization and then using different models.

References

[1] Bond, Z. S, “Identification of vowels excerpted from neutral and nasal contexts,” JASA, Vol. 59, May 1976.
[2] Bell-Berti, F., “Segmental Effects,” in “Phonetics and Phonology,” Eds: Huffman, M.K. and Krakov, R.A., Academic Press Inc, 1993.
[3] TIMIT, “TIMIT acoustic-phonetic continuous speech corpus”, NIST, October 1990.