STUDY OF ACOUSTIC PROPERTIES OF NASAL AND NONNASAL VOWELS IN TEMPORAL DOMAIN

Nishant Singhal, Pradip K. Das

Department of Computer Science and Engineering
Indian Institute of Technology Guwahati-781039, Assam, India.
{n.singhal,pkdas}@iitg.ernet.in

ABSTRACT

There has been considerable amount of work done in exploring the acoustic correlates of nasalized and non-nasalized vowels in the frequency domain. Nasalized vowels are characterized by the presence of extra pole-zero pairs near the first formant region and across the spectrum. Several other automatically extractable acoustic features have been proposed by researchers across the globe. This area has not been explored much in the temporal domain. In this study we have tried to find quantifiable differences/similarities between the nasal and non-nasal vowel /a/ in the temporal domain at the pitch synchronous level. The results show significant differences between nasalized and non-nasalized vowel /a/.

KEYWORDS

Nasal; non-nasal; pitch synchronous; temporal domain.

1. INTRODUCTION

Nasalization is the production of a sound while the velum is lowered, so that some air escapes through the nose during the production of the sound by the mouth. When this happens, the oral cavity is still the major source of output but the sound gets a distinctively nasal characteristic. An archetypal nasal sound is /n/.

In most languages, vowels that are adjacent to nasal consonants are produced partially or fully with a lowered velum in a natural process of assimilation and are therefore technically nasal, though few speakers would notice. This is the case in English: vowels preceding nasal consonants are nasalized, but there is no phonemic distinction between nasal and oral vowels. In some languages, by contrast, nasal vowels are phonemes distinct from oral vowels. In these languages, there are words that differ mainly in the nasal or oral quality of a vowel. Examples of such languages include French and Portuguese.

For our study, we have selected specific pairs of words each containing the cardinal vowel /a/, such that one word contains the non-nasal vowel /a/ while the other contains the same vowel but in the nasal context. The words have been selected from the Hindi language. In Hindi, every

consonant letter by itself automatically includes a short /a/ vowel sound unless otherwise specified. This short /a/ sound is like the /a/sound in English words "about" or "career." In linguistics, this sound has a special name "schwa". Each vowel has two forms: the dependent form is used to indicate that the vowel (other than schwa) is present as a diacritic attached to a consonant. The independent form is used when the vowel occurs alone, at the beginning of a word, or after another vowel. In other words, the independent form is used whenever there is no consonant for the vowel to attach to. Hindi vowels can be nasalized, that is, a nasal quality can be added to the vowel sound. Nasalization in vowel sounds is indicated with a small dot placed on top of the letter as in (IPA:)meaning "I". In some words containing long vowels such as /a/ and /u:/, the nasalization dot is accompanied by a small arc, called the "Chandra bindi" as in (IPA:) meaning "am".

A lot of research has been conducted in the field of analysis, synthesis, perception and recognition of nasalized vowels. Researchers have come up with several acoustic correlates and parameters that can be used to characterize nasalized vowels. House and Stevens found that as coupling to the nasal cavity is introduced, the first formant amplitude reduces and its bandwidth and frequency increase [1].Another interesting study was conducted by Dickson [2]. Fant (1960) reviewed the acoustic characteristics of nasalization pointed out in the literature until then, and from his own observations confirmed the reduction in the amplitude of the first formant due to an increase in its bandwidth, and the rise in the first formant frequency [3]. From the sweep-tone measurement of the vocal tract, Fujimura and Lindqvist observed a shift in the frequency of the first formant towards higher frequencies and the introduction of pole-zero pairs in the first (often below the first formant) and third formant regions [4].

As far as perception of nasality is concerned, several studies have shown close correspondence between the properties mentioned above and the perception of nasalization. These include the experiments conducted by House and Stevens (1956), Hattori et al. (1958) and Hawkins and Stevens (1983) to name a few. The study by Hawkins and Stevens reported that nasalization was introduced by inserting a pole-zero pair in the vicinity of the first formant [5]. The degree of nasalization was varied by changing the spacing between the pole and zero. Wider spacing of the pole-zero pair was found to be necessary for the perception of nasality.

Great amount of work has also been done in the field of detection and automatic recognition of nasalized vowels. The problem of detection of nasalized vowels is motivated by its usefulness in speech recognition, speech enhancement and clinical assessment of nasal speech. In 1985, J. R. Glass and Victor W. Zue proposed six acoustic measures of nasalized vowels and performed recognition experiment using those measures [6, 7]. Acoustic parameters such as reduction in the first formant amplitude (A1) [1], the relationship between A1 and the amplitude of the first harmonic (H1) [8] and the difference between A1 and amplitude of the extra nasal poles P0 (the one below the first formant at around 250-450 Hz) and P1 (the one above the first formant at around 1000 Hz) [9] have been extensively used in recognition experiments. T. Pruthi proposed several automatically extractable acoustic parameters and among them selected nine acoustic parameters with the best discrimination capability through Analysis of Variance [10, 11]. The results of the recognition experiments using these acoustic parameters were remarkable. Another interesting study in this regard was conducted by Yuang and Liberman [12]. A modification of [13] based on fuzzy techniques can be extended to recognition of nasalized vowels.

Several speaker dependent features have been identified from the nasalized vowels that can be used for speaker identification and verification purposes [14, 15, 16]. The majority of these features/parameters are in the frequency domain.

When it comes to the temporal domain, there has not been much contribution about characteristics of nasalized and non-nasalized vowels, which formed the motivation for this study. The work mentioned in [17] and [18] has been a great inspiration for this study. The rest of this paper is organized as follows: Section II presents a brief description of the experiments carried out in the frequency domain. Section III presents a detailed description of the experiments carried out in the temporal domain and the observations. Section IV concludes the paper.

2. NASALIZATION IN FREQUENCY DOMAIN

This section covers a brief description of our study on nasalized vowels in the spectral domain. The main motivation behind carrying out the study in spectral domain was to verify the common theoretical consideration that nasalization introduces extra pole-zero pair in the vicinity of the first formant. In the higher frequencies, nasalization may introduce shifts in formants, modification of formant amplitudes and additional spectral peaks. However, these effects are not as consistent as those in the vicinity of the first formant. Our aim of studying nasalization characteristics in the spectral domain was towards the verification of these already known theoretical notions rather than exploration of some new characteristics.

2.1 Methodology

For this study, we selected words that form minimal pairs, kata/kanta, raat/raam and para/pani. The recordings consisted of 20 utterances of each of these words. The whole analysis has been conducted on a HP DV4 laptop. The recordings were done using the Cool Edit Pro 2.1 software. Conversion to spectrum and LPC smoothing was done using the features available in the Praat software. The speech segments corresponding to the nasalized and non-nasalized vowels were extracted manually from the word pairs. The non-nasalized vowel was extracted from its steady state region and the nasalized vowel was extracted from the onset of nasalization in the pronunciation of the word. These vowel segments were then converted to spectrum and LPC smoothing was applied to the spectrum to retain only the prominent frequency peaks. These smoothed spectrums were plotted both for the nasalized vowel and the non-nasalized vowel and examined. Fig. 1 presents the smoothed spectrums for non-nasalized and nasalized /a/ from two of the three word pairs.

2.2 Inferences

The analysis of the spectrums of nasalized and non-nasalized vowels clearly verifies that there are indeed extra resonances present in case of nasalized vowel near the first formant region (at around 267 Hz). There is also a lowering in the amplitude of the first formant.
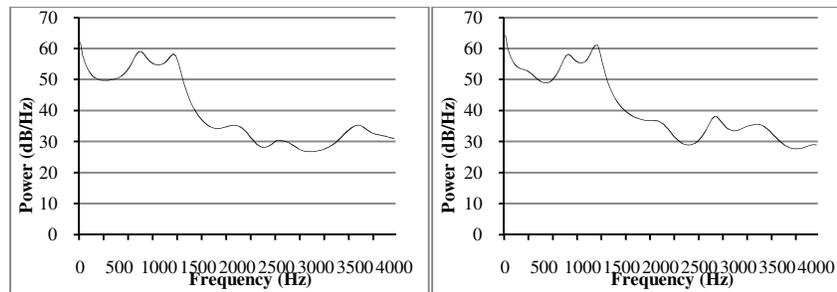
Fig. 1(a). Smoothed spectrum of non-nasalized /a/ (left) and nasalized /a/ (right) from the word pair /raat-raam/. For non-nasalized /a/, location of first formant is 762 Hz, second formant is 1187 Hz. For nasalized /a/, location of first formant is 812 Hz, second formant is 1170 Hz and the location of extra peak is at 287 Hz.
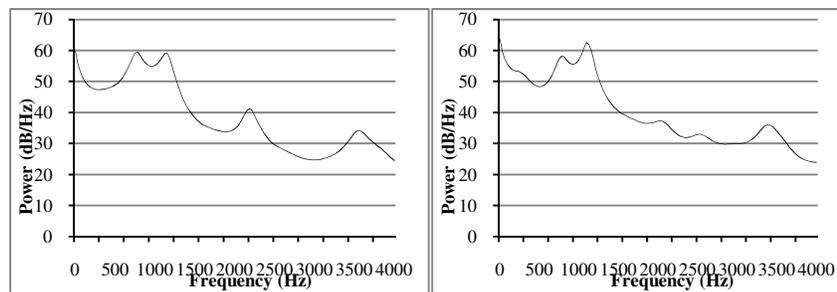
Fig. 1(b). Smoothed spectrum of non-nasalized /a/ (left) and nasalized /a/ (right) from the word pair /para-pani/. For non-nasalized /a/, location of first formant is 782 Hz, second formant is 1150 Hz. For nasalized /a/, location of first formant is 782 Hz, second formant is 1097 Hz and the location of extra peak is at 267 Hz.

3. NASALIZATION IN TEMPORAL DOMAIN

3.1 Methodology

The data consisted of recordings from a single male speaker. The word pairs were same as in the frequency domain analysis. The IPA representation of each of these words is given in Table 1.Each word pair was spoken 20 times. The recordings were done with the following parameters : sampling rate of 16000 Hz, mono channel and a bit resolution of 16 bits. All the 20 utterances were not recorded at the same time. The 20 recordings were done in groups of five at a time. There was a gap of at least 20-30 minutes between successive 5 recordings. From each of the 20 utterances for all the 3 word pairs, the corresponding nasalized and non-nasalized vowels were extracted manually.

From each utterance of the nasalized and non-nasalized vowel thus extracted (60 utterances of each nasalized and non-nasalized /a/), 3 consecutive pitch periods were extracted, starting from the 3rd pitch period up to the 5th pitch period, assuming that the initial two pitch periods might Consist of some effect of the phoneme preceding the vowel.

A pitch period detection program based on the autocorrelation method was developed to extract the pitch periods in the time domain [19]. A linear filter was applied to the speech signal to

enhance the signal and attenuate the noise components. The signal was then divided into sections of 300 samples, with adjacent sections overlapping by 200 samples. For a particular section, a clipping level is calculated. The clipping level is set at a value of 68 percent of smaller of the maximum absolute sample values in the first and the last one-third portions of the section. This speech section is then center clipped. After clipping, the autocorrelation function for that particular section is computed over a range of lags from 0 to 200 samples [20]. Although lags ranging from 20 to 200 are usually used in practice, but we are calculating these extra autocorrelation values just for the sake of observation.

The autocorrelation function is given by the following equation:

$$R_x(m) = \frac{1}{N} \sum_{n=0}^{N-1-m} [x(n+1)][x(n+1+m)], 0 \leq m \leq M_0 \qquad (1)$$

whereN is the section length being analyzed, M0 is the number of autocorrelation points to be computed and l is the index of the starting sample of the section. In our case, the values of M0 and N are 200 and 300 respectively.

The autocorrelation function is then searched for its maximum value. If the maximum value is beyond a certain threshold, the section is classified as voiced in which case the pitch period is the location of that maximum value.

The program marks the beginning of a pitch period from the starting point of the first prominent peak on the positive side. The starting point of a peak is defined as the zero value just before that peak or the first positive sample value that marks the beginning of that peak. In this way, the pitch periods are aligned from peak to peak, so that a complete periodic unit is obtained. For each of the 3 word pairs, the 60 pitch periods (3 pitch periods from all the 20 utterances) for both the nasalized as well as the non-nasalized vowel were averaged. All these pitch periods were normalized so that the maximum and minimum amplitude lies in the range of +15,000 to -15,000.Since the pitch periods across different utterances had varying durations, all of these were time warped using the DTW method with respect to the longest pitch period among the 60 pitch periods. The averaging method used was simple unweighted mean. The averaged pitch period for both the nasalized as well the non-nasalized vowel were again time warped with respect to each other and plotted for analysis.

Fig. 2 shows the plot of all 60 pitch periods of nasalized and non-nasalized vowel /a/, for 2 of the 3 word pairs. Fig. 3 presents the same plots, but time warped with respect to the longest pitch period. Time warping was necessary to align the various peaks in the waveform so that a fine grained analysis of the properties of these peaks could be done.
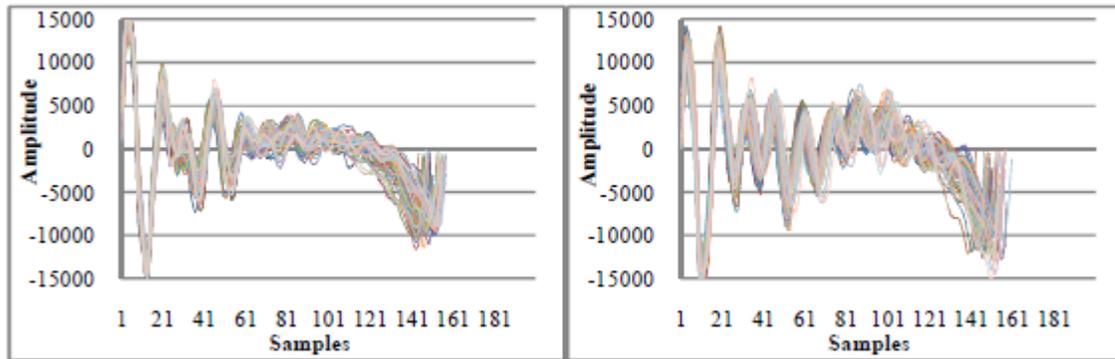
Fig. 2(a). Plots of all 60 pitch periods of non-nasalized (left) and nasalized (right) /a/ from the word pair /raat-raam/.
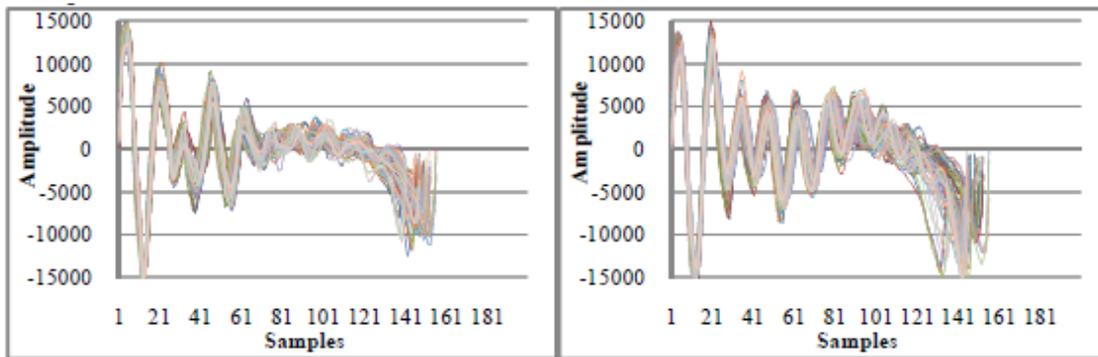
Fig. 2(b). Plots of all 60 pitch periods of non-nasalized (left) and nasalized (right) /a/ from the word pair /para-pani/.
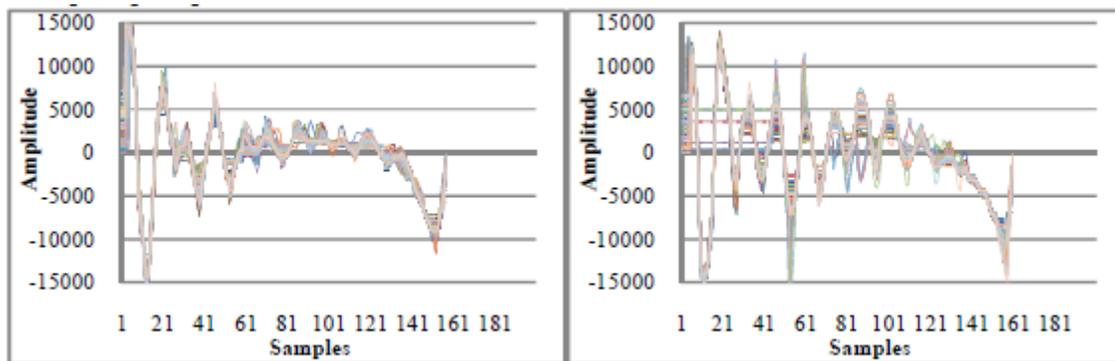
Fig. 3(a).Plots of all 60 pitch periods of non-nasalized (left) and nasalized (right) /a/ from the word pair /raat-raam/ after DTW.
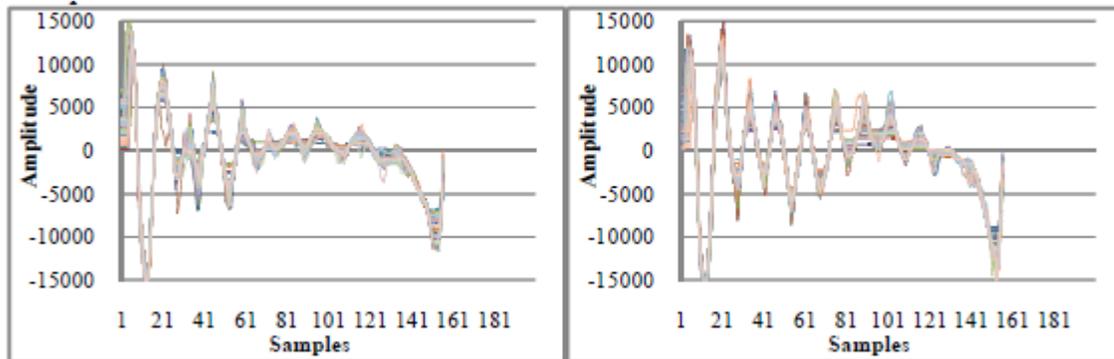
Fig. 3(b). Plots of all 60 pitch periods of non-nasalized (left) and nasalized (right) /a/ from the word pair /para-pani/ after DTW.

Words	IPA	Words	IPA
Kata	/ka:ʈa:/	Raam	/ɾa:m/
Kanta	/kã:ʈɑ:/	Para	/pa:ɾa:/
Raat	/ɾa:ʈ/	Pani	/pa:ni/

Table 1. IPA representation of the words used in the experiments.
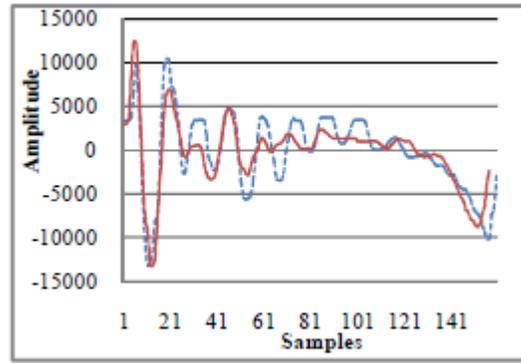
Fig. 4. Pitch periods of non-nasalized (solid line) and nasalized (dotted line) /a/ averaged across all the utterances of all 3 word pairs.

The warped graphs presented in Fig. 3 provide a clear picture of the nature of these waveforms, since warping has aligned together the peaks and reduced the corridor of variation among the pitch periods across different utterances. The final task in this experiment was to take the average of the 3 averaged pitch periods of nasalized /a/ obtained from /kata/, /para/ and /raat/ and the average of the 3 averaged pitch periods of nasalized /a/ obtained from /kanta/, /raam/ and /pani/. The graphs of both of these final averaged pitch periods of nasalized and non-nasalized /a/ are presented in Fig. 4.

3.2 Observations and Inferences

The graphs given in Fig. 4 were analyzed. It should be noted that these final averaged pitch periods are not normalized. Before carrying out the average at any step throughout this experiment, care was taken to examine the properties of each individual entity that was to be averaged, so that any outlier was rightfully taken into account. Analyzing the nasalized and non-nasalized vowel /a/ at the pitch synchronous level has yielded several interesting observations. The following properties and differences were observed between the pitch periods of the nasalized and non-nasalized vowel /a/. It may be mentioned here that none of these properties have been observed based only on the final averaged pitch periods presented in Fig. 4.

These properties have been observed throughout the experiment:

1. The properties of both nasalized and non-nasalized /a/ did not change significantly across 3 different nasal contexts, namely /m/, /n/, and /ŋ/. The waveforms for nasalized as well as non-nasalized /a/ are similar across the 3 different word pairs.
2. The pitch period of nasalized /a/ has a more cyclical nature as compared to that of the non-nasalized /a/. This is evident from the graphs in Fig. 2, 3 and 4.
3. In case of non-nasalized /a/, the first peak always overshoots the second peak by significant amplitude. In nasalized /a/, the first two peaks are almost comparable in amplitudes. In fact, in some cases, the second peak was found to rise above the first peak, but only by a small magnitude.
4. Third peak in non-nasalized /a/has a random nature. It is sometimes high and at times low. This randomness is clear from the graphs in Fig. 2 and 3.The third peak in nasalized /a/ has a well-formed nature. It usually has high amplitude than the third peak of non-nasalized vowel.
5. The nature of fourth peak is more or less the same in both cases. Beyond the fourth peak, the waveform for non-nasalized /a/ is distorted, without any proper structure, whereas in case of nasalized /a/, the waveform is still having a cyclical nature.
6. In nasalized /a/, from the third peak onwards, the peaks are steeper/taller than the corresponding peaks in non-nasalized /a/.
7. The dip towards the end of the pitch period is almost always greater in case nasalized /a/, as compared the non-nasalized /a/.

During the experiment, another interesting observation was made while performing the analysis for the word pair raat/raam, specifically for the word raam. The prolonged /a/ sound in the pronunciation of the word raam is not entirely nasalized. Only the ending portion of the vowel sound is nasalized due to effect of the succeeding nasal /m/. The initial /a/ sound remains non-nasalized. Analyzing the waveform of the entire vowel region in the word raam showed a clear visually perceptible transition of the non-nasalized /a/ into nasalized /a/. Fig. 5 shows the vowel region extracted from one of the utterances of raam and the transition of nonnasalized /a/ into nasalized /a/.

The waveform in the lower half of Fig. 5 is the continuation of the waveform from the upper half. The initial pitch periods in the upper half correspond to the nonnasalized vowel /a/ and have a distorted structure beyond the prominent peaks and the first peak is dominating the second peak. As the vowel slowly begins to get nasalized, the amplitude of first peak reduces and becomes comparable to that of the second peak. The pitch periods towards the end of the lower half correspond to the nasalized vowel /a/ and have steeper peaks and terminate in a structured way. The overall amplitude of the vowel is also reduced.

We believe that the extended cyclical nature of the pitch periods in case of nasalized /a/ can be attributed to the nasal resonances. It can be the case that the combination of nasal and oral resonances in case of nasalized /a/ is responsible for steeper peaks in nasalized /a/. Based on the above observations, we expect that the other four vowels, namely /e/, /i/, /o/ and /u/ will have similar characteristics in nasal contexts.
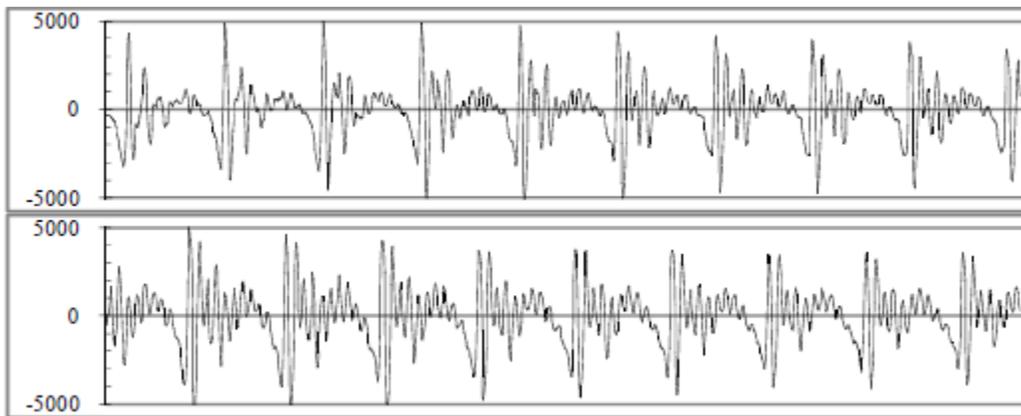
Fig. 5.Vowel region extracted from the word raam showing the transition of non-nasalized /a/ into nasalized /a/.

4. CONCLUSIONS

The paper reports the experiments carried out in the temporal domain to analyze the properties of nasalized and non-nasalized vowel /a/.The frequency domain acoustic correlates of nasalized vowels have also been verified as part of this study. The temporal domain analysis shows that there are indeed significant differences between the waveforms of nasalized and non-nasalized /a/. Nasalized /a/ has a cyclical nature and steeper/taller peaks which can be attributed to the combination of oral and nasal resonances. These results encourage us to carry out further experiments in this area.

The analysis can be extended to include the remaining four vowels /e/, /i/, /o/ and /u/. As this study involved data from a single male speaker, the generalization of the properties observed is not self-evident. A larger group of speakers would have to be investigated to claim more generalized properties. Nevertheless, 20 utterances for each of the three word pairs is a large data set for a single speaker and the results are encouraging.

As a possible future work, speaker dependent characteristics can be studied from the waveforms of nasalized vowels. Since different people have different nasal tract structure and the degree of nasalization varies from speaker to speaker, we intend to discover speaker-dependent features through this analysis. Such features, in conjunction with existing speaker-dependent features, may be used for developing more robust speaker recognition/identification systems. Another interesting activity that can be tried is to develop an automaton to convert nasalized /a/ into non-nasalized /a/ and vice-versa. Manual transformations have already been tried in this regard and were successful. However, developing an automaton for this can lead to development of algorithms for automatic recognition of nasalized and non-nasalized vowels in the temporal domain. Several temporal domain techniques have already been proposed for automatic recognition of oral vowels. Finding quantifiable properties of nasalized vowels in the temporal domain will allow us to extend those techniques to automatic recognition of nasalized vowels.

REFERENCES

[1]   House, A. S., & Stevens, K. N.: Analog studies of the nasalization of vowels. Journal of Speech and Hearing Disorders, 22(2), 218-232(1956)

[2]   Dickson, D. R.: An acoustic study of nasality. Journal of Speech and Hearing Research, 5(2), 103-111 (1962)

[3]   Fant, G.:Acoustic Theory of Speech Production. Netherlands: Mouton and Co., 's- Gravenhage, ISBN: 9027916004(1960)

[4]   Fujimura, O., & Lindqvist, J.: Sweep-Tone measurements of the vocal tract characteristics. Journal of the Acoustical Society of America, 49(2), 541-548 (1971)

[5]   Hawkins, S., & Stevens, K. N.: A cross-language study of the perception of nasal vowels. Paper presented at the 105th meeting of the Acoustical Society of America, Cincinnati, Ohio (1983)

[6]   Glass J. R.: Nasal Consonants and Nasalized Vowels: An Acoustic Study and Recognition Experiment. SM. Thesis, Massachusetts Institute of Technology (1984)

[7]   Glass, J. R., & Zue, V. W.: Detection of nasalized vowels in American English. In: Proceedings of ICASSP, 1569–1572 (1985)

[8]   Huffman, M. K.: Implementation of nasal: Timing and articulatory landmarks. UCLA WPP, 75, 112-143 (1990)

[9]   Chen, M. Y.: Acoustic correlates of English and French nasalized vowels. Journal of the Acoustical Society of America, 102(4), 2360–2370 (1997)

[10] Pruthi, T.: Analysis, Vocal-tract modeling and automatic detection of vowel nasalization. Ph.D. dissertation, University of Maryland (2007)

[11] Pruthi, T., & Espy-Wilson, Carol Y.: Acoustic Parameters for the Automatic Detection of Vowel Nasalization. Interspeech, 1925-1928 (2007)

[12] Yuang, J., & Liberman, M.: Automatic measurement and comparison of vowel nasalization across languages. In: Proceedings of ICPhS XVII, 2244-2247 (2011)

[13] Tripathy, H. K., Tripathy, B. K., & Das, P. K.: A Knowledge based Approach Using Fuzzy Inference Rules for Vowel Recognition. Journal of Convergence Information Technology, 3(1), 51-56(2008)

[14] Amino, K., & Osanai, T.: Speaker characteristics that appear in vowel nasalization and their change over time. Acoustical Science and Technology, 33(2), 96-105.

[15] Pfeifer, L. L.: Contributions of Vowel and Nasal Sounds to Speaker Identification. Journal of the Acoustical Society of America, 55(2), 462-462 (1974)

[16] Su, Lo-Soun, Li, K. P. & Fu, K. S.: Identification of speakers by use of nasal coarticulation. Journal of the Acoustical Society of America, 56(6), 1876-1882

[17] Stokes, M. A.: The Waveform Model of Vowel Perception and Production, Boca Raton, FL: Universal-Publishers, ISBN: 1599428881 (2009)

[18] Ha, S., & Kuehn, D.: Temporal characteristics of nasalization in children and adult speakers of American English and Korean during production of three vowel contexts. Journal of the Acoustical Society of America, 120(3), 1622-1630 (2006)

[19] Rabiner, L. R., Cheng, M. J., Rosenberg, A. E., & McGonegal, C. A.: A Comparative Performance Study of Several Pitch Detection Algorithms. IEEE Trans. on Acoustics, Speech, and Signal Processing, 24(5), 399-418(1976)

[20] Tan, L., & Karnjanadecha, M.: Pitch Detection Algorithm: Autocorrelation Method and AMDF. In: Proceedings of the 3rd International Symposium on Communications and Information Technology, 551-556 (2003)