7th WSEAS Int. Conf. on MATHEMATICAL METHODS and COMPUTATIONAL TECHNIQUES IN ELECTRICAL ENGINEERING, Sofia, 27-29/10/05 (pp346-354)

A Review on Bangla Phoneme Production and Perception for Computational Approaches

SYED AKHTER HOSSAIN
Department of Computer Science and Engineering,
East West University
BANGLADESH

M LUTFAR RAHMAN
Department of Computer Science and Engineering,
University of Dhaka
BANGLADESH

FARRUK AHMED
Department of Computer Science and Engineering,
North South University
BANGLADESH

Abstract:- Bangla, a language of nearly 300 million people around the world, begun 11 century AD and originated from a dialect commonly known as Prakrit. Bangla Phoneme production and perception plays a central role in computer speech analysis, synthesis and recognition of Bangla. It is worth noting that there has not been much study accomplished for Bangla Computational Phonetics. In this paper we have discussed speech production mechanism along with the linguistics classification in contrast to English and emphasized on Bangla phoneme processing and classification criteria for computer analysis and synthesis of Bangla speech. The distinction between vowel and consonant is also discussed both from the context of linguistics as well as computer processing point of view. The phoneme perception plays an important role in the classification of phonemes. Besides, the paper also covers discussion on the phonemes and their variations in contextual speech production.

Key-Words:- Speech Processing, Formants, Linguistics, Voiced, Unvoiced, Phoneme

1 Introduction

Bangla is a language of about 300 million people in the eastern region of Indian subcontinent i.e. Bangladesh, Indian states of West Bengal, Trippura and around the world. The history of Bangla begun in the early centuries of the present millennium and before that there was only a family of dialects commonly known as Prakrit [1]. In linguistic relationship, Bangla is closer to Assamese then to Oriya and then to Hindi. The general structural pattern resembles close to the Dravidian language of south India. About sixty percent of the word types in formal Bangla are classical Sanskrit; the rest contains British English, Persian, Portuguese and other south Asian language [2]. The script is historically derived from ancient Indian Brahmi, itself a modification of ancient southern Arabic [1].

We have attempted a study on the Bangla linguistics along with the identification of phonemes from the perception based on computer processing of speech. In comparison to English vowels and consonants and the relevant phonetic features, Bangla linguistics classification of vowels and consonants are identified and acoustic features are analyzed to reveal the manner and position of articulators in the Bangla phoneme production.

Our goal in this paper is elaborate phonetic classification of Bangla in contrast to English from the perspectives of phoneme production and perception. In particular, we have applied computational approaches to extract features for phoneme identification and classification.

We have also elaborated general speech production with role of various articulators in the phoneme production along with the perception both from linguistics and computational point of view.

2 Speech and Phonetics

2.1 Speech Production

The speech signal consists of variations in pressure, measured directly in front of the mouth, as a function of time. The amplitude variations of such a signal correspond to deviations from atmospheric pressure caused by traveling waves. The signal is non-stationary and constantly changes as the muscles of the vocal tract contract and relax. Speech can be divided into sound segments, which share some common acoustic properties with one another for a short interval of time. Sounds are typically divided into two broad classes: (a) vowels, which allow unrestricted airflow in the vocal

tract, and (b) consonants, which restrict the airflow at some points and are weaker than vowels.

Speech is generated by compression of the lung volume causing airflow which may be made audible if set into vibration by the activity of the larynx. This sound source can then be made into intelligible speech by various modifications of the supralaryngeal vocal tract. The process of speech production involves the following:

a. Lungs provide the energy source - Respiration
b. Vocal folds convert the energy into audible sound - Phonation
c. Articulators transform the sound into intelligible speech - Articulation

An overview of the vocal tract showing structures that are important in speech sound production and speech articulation is shown in the Figure 1.

The human speech production mechanism consists of lungs, trachea (windpipe), larynx, pharyngeal cavity (throat), buccal cavity (mouth), nasal cavity, velum (soft palate), tongue, jaw, teeth and lips as shown in a simplified tube model in Figure 2. The lungs and trachea make up the respiratory subsystem of the mechanism. These provide the source of energy for speech when air is expelled from the lungs into the trachea. Speech production can be viewed as a filtering operation in which a sound source excites a vocal tract

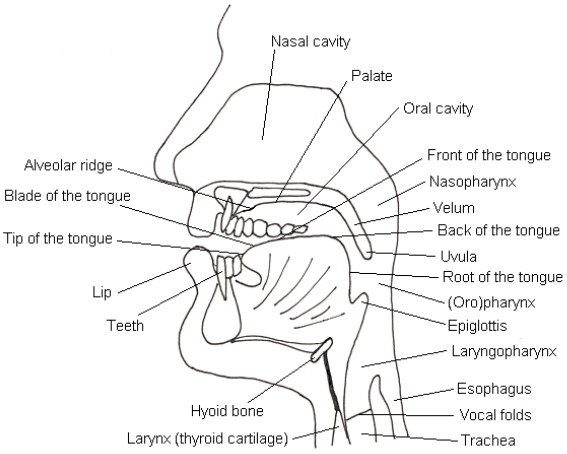


Fig. 1: Structure of the Vocal Tract

filter. The source is periodic, resulting in voiced speech or aperiodic, resulting in unvoiced speech as shown in Figure 2. The voicing source occurs at the larynx at the base of the vocal tract, where airflow can be interrupted periodically by the vocal folds.

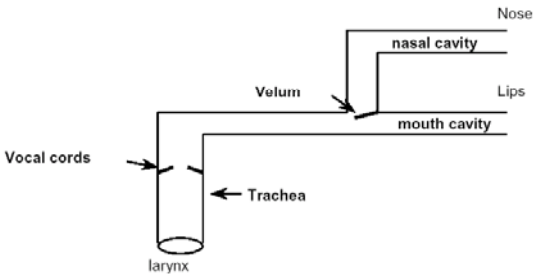


Fig. 2 A simplified tube model of the human speech production system

Velum, tongue, jaw, teeth and lips are known as the *articulators* (Figure 1). These provide the finer adjustments to generate speech. The excitation used to generate speech can be classified into *voiced, unvoiced, mixed, plosive, whisper* and *silence*. Any combination of one or more can be blended to produce a particular type of sound. A *phoneme* describes the linguistic meaning conveyed by a particular speech sound [3,4].

2.2 Larynx Structure and Function

The larynx is a continuation of the trachea but the cartilage structures of the larynx are highly specialized. The main cartilages are the thyroid, cricoid and arytenoid cartilages. These cartilages variously rotate and tilt to affect changes in the vocal folds. The vocal folds (also known as the vocal cords) stretch across the larynx and when closed they separate the pharynx from the trachea. When the vocal folds are open breathing is permitted. The opening between the vocal folds is known as the glottis. When air pressure below closed vocal folds (sub-glottal pressure) is high enough the vocal folds are forced open, the vocal folds then spring back closed under both elastic and aerodynamic forces, pressure builds up again and the vocal folds open again, and so on for as along as the vocal folds remain closed and a sufficient sub-glottal pressure can be maintained. This continuous periodic process is known as phonation and produces a "voiced" sound source [4,13].

2.3 Articulation and Coarticulation

Articulation is defined as the sound produced at the larynx and modified through the alteration of the shape

of the vocal tract above the larynx (supralaryngeal or supraglottal). The shape can be changed by opening or closing the velum (which opens or closes the nasal cavity connection into the oropharynx), by moving the tongue or by moving the lips or the jaw.

Coarticulation is defined as the movement of two articulators at the same time for different phonemes. Coarticulation can occur with or without a change in sound production. One example is for the word “two” in English or "dao" (দাও) in Bangla. Coarticulation can result in a “smearing” of segmental boundaries between phonemes, which can modify the characteristics of the phoneme.

3 Phoneme Perception in Linguistics

3.1 Distinction between Vowel and Consonant

The distinction between vowels and consonants is based on three main criteria as follows:

1. physiological: airflow / constriction
2. acoustic: prominence
3. phonological: syllabicity

Sometimes, it is necessary to rely on two or three of these criteria to decide whether a sound is a vowel or a consonant.

Physiological Distinction

In general, consonants can be said to have a greater degree of constriction than vowels. This is obviously the case for oral and nasal stops, fricatives and affricates. The case for approximants is not so clear-cut as the semi-vowels /j/ in English or "za" (য) in Bangla is very often indistinguishable from vowels in terms of their constriction.

Acoustic Distinction

In general, consonants can be said to be less prominent than vowels. This is usually manifested by vowels being more intense than the consonants that surround them. Sometimes, certain consonants can have a greater total intensity than adjacent vowels but vowels are almost always more intense at low frequencies than adjacent consonants [7,8,12].

Phonological Distinction

Syllables usually consist of a vowel surrounded optionally by a number of consonants. A single vowel forms the prominent nucleus of each syllable. There is only one peak of prominence per syllable and this is nearly always a vowel. The consonants form the less prominent valleys between the vowel peaks. This tidy picture is disturbed by the existence of syllabic consonants.

Syllabic consonants form the nucleus of a syllable that does not contain a vowel. In English, syllabic consonants occur when an approximant or a nasal stop follows a homorganic (same place of articulation) oral stop (or occasionally a fricative) in words such as "bottle" in English or “kalom” means Pen (কলম) in Bangla.

The semi-vowels in English play the same phonological role as the other consonants even though they are vowel-like in many ways. The semi-vowels are found in syllable positions where stops, fricatives, etc. are found (e.g. "pay", "may", and "say" versus "way") or “zabe” means will go (যাবে), “khabe” means will eat (খাবে) versus “khaiba” means will eat (খাইবা) which ends with “be” (বে) or “ba” (বা) in Bangla [1].

3.2 Phoneme and Allophone

Linguistic units, which cannot be substituted for each other without a change in meaning, can be referred to as linguistically contrastive or significant units. Such units may be phonological, morphological, syntactic, semantic etc. Logically, this takes the form as shown in the table 1.

Table 1: Linguistics Units

	IF	Unit X	In context A	GIVES meaning 1
	AND IF	Unit Y	In context A	GIVES meaning 2
	THEN	Unit X AND unit Y		belong to separate linguistic units
e.g.	IF	Sound [k]	In context [æt]	GIVES meaning "cat"
	AND IF	Sound [m]	In context [æt]	GIVES meaning "mat"
	THEN	Sound [k] and sound [m]		belong to separate linguistic units
e.g.	IF	Sound "da" (দ)	In context "ao" (াও)	GIVES meaning "dao" (দাও)
	AND IF	Sound " kha" (খ)	In context " ao" (াও)	GIVES meaning " khao" (খাও)
	THEN	Sound " da" (দ) and sound " kha" (খ)		Belong to separate linguistic units

Phonemes

Phonemes are the linguistically contrastive or significant sounds (or sets of sounds) of a language. Such a contrast is usually demonstrated by the existence of minimal pairs or contrast in identical environment (C.I.E.). Minimal pairs are pairs of words which vary only by the identity of the segment (another

word for a single speech sound) at a single location in the word (e.g. [mæt] and [kæt]) or "dao" (দাও) and "khao" (খাও) for Bangla. If two segments contrast in identical environment then they must belong to different phonemes. A paradigm of minimal phonological contrasts is a set of words differing only by one speech sound. In most languages it is rare to find a paradigm that contrasts a complete class of phonemes (eg. all vowels, all consonants, all stops etc.) [9,10,11].

The Bangla stop consonants could be defined by the following set of minimally contrasting words:

i) "nim" / নিম/ vs "din" / দিন/ vs "tin" / টিন/ vs "pin" / পিন/

Only "i" / ি / does not occur in this paradigm and at least one minimal pair must be found with each of the other 4 stops to prove conclusively that it is not a variant form of one of them.

ii) "paan" / পান / vs "dhaan" / ধান / vs "maan" / মান / vs "taan" / তান /

Again, only four stops belong to this paradigm. A syntagmatic analysis of a speech sound, on the other hand, identifies a unit's identity within a language. In other words, it indicates all of the locations or contexts within the words of a particular language where the sound can be found.

Allophones

Allophones are the linguistically non-significant variants of each phoneme. In other words a phoneme may be realized by more than one speech sound and the selection of each variant is usually conditioned by the phonetic environment of the phoneme. Occasionally allophone selection is not conditioned but may vary form person to person and occasion to occasion.

A phoneme is a set of allophones or individual non-contrastive speech segments. Allophones are sounds, whilst a phoneme is a set of such sounds.

Allophones are usually relatively similar sounds, which are in mutually exclusive or complementary distribution (C.D.). The C.D. of two phonemes means that the two phonemes can never be found in the same environment (i.e. the same environment in the senses of position in the word and the identity of adjacent phonemes). If two sounds are phonetically similar and they are in C.D. then they can be assumed to be allophones of the same phoneme.

In many languages voiced and voiceless stops with the same place of articulation do not contrast linguistically but are rather two phonetic realizations of a single phoneme.

In other words, voicing is not contrastive (at least for stops) and the selection of the appropriate allophone is in some contexts fully conditioned by phonetic context (e.g. word medially and depending upon the voicing of adjacent consonants), and is in some contexts either partially conditioned or even completely unconditioned (e.g. word initially, where in some dialects of a language the voiceless allophone is preferred, in others the voiced allophone is preferred, and in others the choice of allophone is a matter of individual choice).

4 Phoneme Perception

4.1 Computational Approach

Speech production can be viewed as a filtering operation in which a sound source excites a vocal tract filter. The source is periodic, resulting in voiced speech or aperiodic, resulting in unvoiced speech as shown in Figure 2.

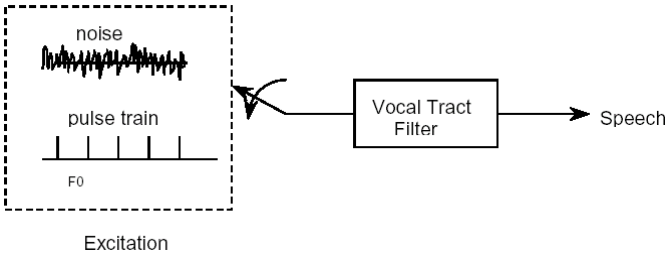


Fig. 2 Generation of Voiced and Unvoiced Speech

The voicing source occurs at the larynx at the base of the vocal tract, where airflow can be interrupted periodically by the vocal folds. The velum, tongue, jaw, teeth and lips are known as the *articulators*. These provide the finer adjustments to generate speech.

The excitation used to generate speech can be classified into *voiced, unvoiced, mixed, plosive, whisper* and *silence*. Any combination of one or more can be blended to produce a particular type of sound. A *phoneme* describes the linguistic meaning conveyed by a particular speech sound.

The American English language consists of about 42 phonemes, which can be classified into vowels, semivowels, diphthongs and consonants (fricatives, nasals, affricatives and whisper) as shown in Figure 3.

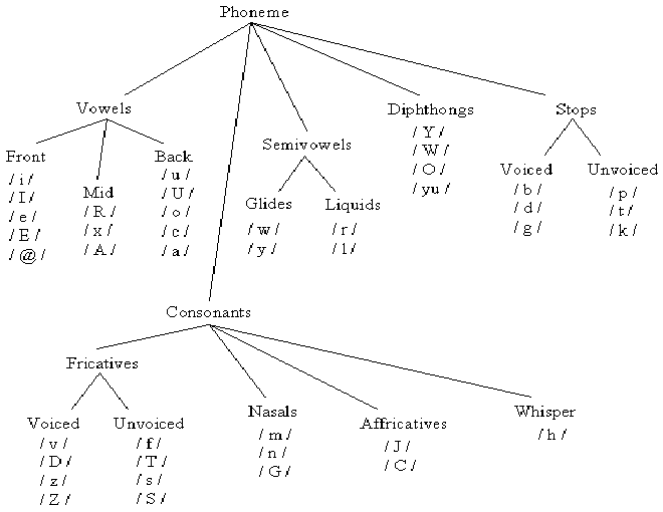


Fig. 3 Phonemes in American English

4.2 Classes of Speech Sounds

Vowels- Vowels (including diapthongs) are voiced, and have usually the largest amplitude among phonemes, and range in duration from 50 to 400 ms in normal speech. Figure 4 shows a brief portion of waveform for a Bangla vowel and its corresponding frequency spectrum. Due to periodicity of the voiced excitation, the frequency spectrum exhibits harmonics with frequency spacing of F0 Hz where F0 is the *fundamental frequency* or the *pitch* of the vocal cord vibrations.

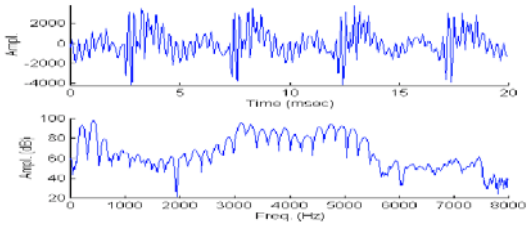


Fig. 4 Time waveform of "aam" / আম / and its corresponding spectrum

Figure 5 shows a brief portion of waveform for a Bangla speech containing vowels with its corresponding spectrogram. A spectrogram is a plot of frequency vs. time. The spectrogram reveals the amount of energy at different frequencies at different times. As seen from the spectrogram, the dark portions in the spectrogram represent the formant frequencies, which are the dominant spectral peaks. The lower bold horizontal line is the first formant frequency (F1) and the upper dark portion represents the second formant frequency (F2). The formants can also be detected by inspection of the spectrum for dominant peaks as seen from Figure 6.

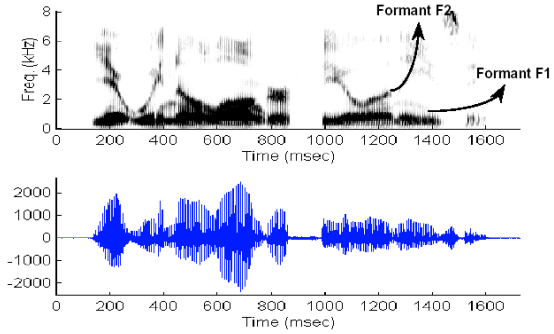


Fig. 5 Spectrogram showing the first two formant frequencies

The dominant peaks in the frequency spectrum can be detected as F1, F2 and F3 formant frequencies. The formant frequencies are normally derived from the Linear Prediction Coding (LPC) plot of the time waveform. Figure 7 shows the time waveform and the corresponding LPC plot with the formant frequencies F1, F2 and F3.

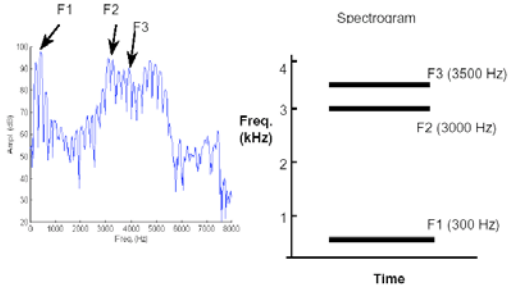


Fig. 6 Spectrum showing the first three formant frequencies and corresponding spectrogram

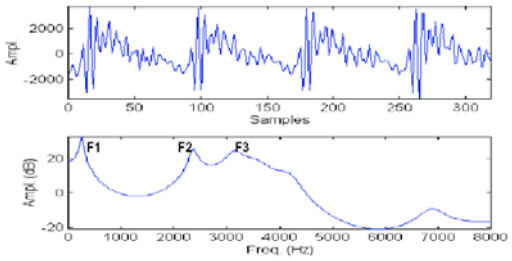


Fig. 7 LPC Spectrum of a vowel segment containing "aa" / আ /

5 Spectral Characteristics

5.1 Bangla Vowels and Consonants

Vowels

Vowels are associated with well-defined formant frequencies, which have provided the dominant approach to acoustic characterization of these vowels. The Peterson and Barney's study helped to relate the vowel formant frequencies to vowel articulation. It was shown that F1 varies mostly as the tongue height and F2 varies mostly with the tongue advancement.

According to Bangla Linguistics, there are eight classified cardinal vowels grouped into categories of frontal and back vowels and one central or neutral vowel "aa" /আ/. The frontal vowels are "e" /ই/, "a" /এ/, "ae" /এই/ and back vowels are "ao" /অ/,"o" /ও/,"ou" /ঔ/ and "u" /উ/ respectively [1].

The following Figure 8 shows the time waveform, gray scale spectrogram and formant tracking chart for a male voice utterance of the Bangla word "aam" /আম/ containing the neutral vowel.

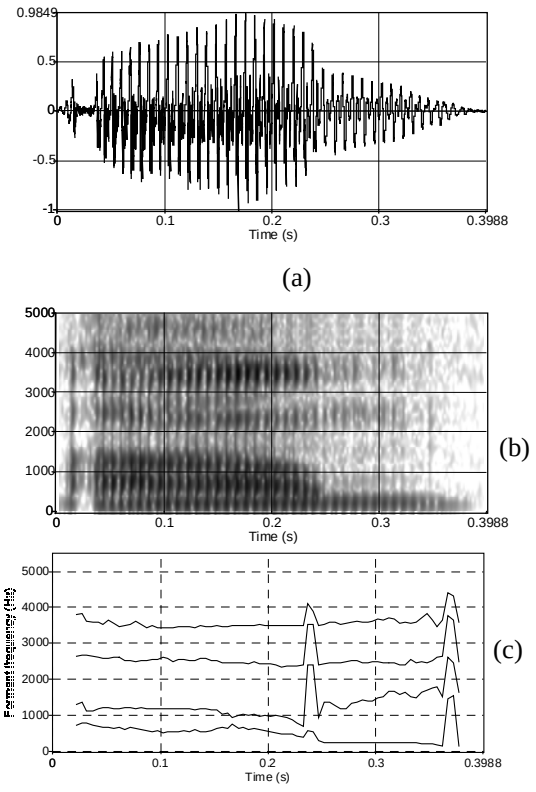


Fig. 8 Bangla vowel "aa" / আ/ in word "aam" / আম/ (a) Timewaveform (b) Spectogram (c) Formant track

It is clearly visible in the above spectrum of the vowel "aa" /আ/ that it is made up of a large number of harmonics, with those harmonics occurring at frequencies close to the resonant frequencies of the tract (formants) having the greatest amplitude.

A formant in the Figure 8 as a dark band on the spectrogram, which corresponds to a vocal tract resonance. Technically, it represents a set of adjacent harmonics, which are boosted by a resonance in some part of the vocal tract. Thus, different vocal tract shapes will produce different formant patterns, regardless of what the source is doing in source filter model of speech production. The spectrogram of Figure 8 represents the presence of Bangla neutral vowel "aa" /আ/. It is noticeable in the formant trajectory that the first formant is very much steady during the resonance. The first formant correlates (inversely) roughly to the height (or directly to openness) of the vocal tract. The next formant, F2 corresponds to backness and/or rounding since it is also steady indicating the nature of the neutral vowel.

A full account of the acoustic cues for vowel perception would seem to require consideration of each of the following factors: formant frequencies, vowel duration, fundamental frequency and formant bandwidth. The shape of the vowel spectrum provides extra information regarding the perception of vowels. Spectral tilt in the spectrum of the vowel does not have a significant effect on the perception of the vowels. But a pronounced effect in vowel perception is observed if there is a shift in the relative position of spectral peaks. Hence the location of peaks and their movement due to addition of noise or any other reason may contribute to change in the perception of vowels [12,14].

Vowel duration helps distinguish spectrally similar vowels whereas the fundamental frequency of the vowels may help distinguish the speaker. Formant bandwidth and amplitude can help perceive the naturalness of the spoken vowel. Yet another factor that may affect vowel identification is spectral contrast. Spectral contrast for a vowel is defined as the ratio of the maximum amplitude in the spectrum of the vowel to the minimum amplitude.

Consonants

Consonants differ from vowels in that they had more energy in the high frequency region compared to the low frequency region. Stop consonants can be divided in three classes viz. labials, alveolars and velars, each having a distinct release burst spectrum shape.

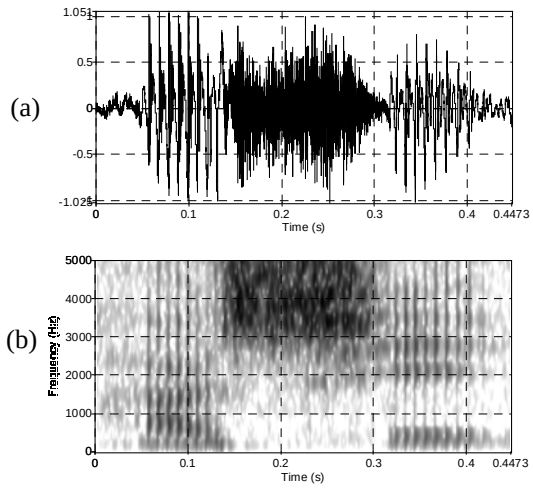
For the stop consonants the peak in burst frequency reveals important information regarding the place of articulation. A peak in the spectrum of the burst at low frequencies was associated with the labials such as /b/ and /p/, where as a peak at higher frequencies was found for alveolars such as /t/ and /d/ for English. Velars such as /g/ and /k/ were found to have a peak in middle of the spectrum as shown by Steven and Blumstein (1978) [13,17].

Bangla linguistics also classifies consonants based on the manner of articulation. The different classes are as follows [1]:

- Glottal or Laryngeal: "haa" হ
- Velar: "kaa" /ক/ "kha" /খ/ "gaa" /গ/ "gha" /ঘ/ "umo" /ঙ/
- Dorso Alveolar: "chaa" /চ/ "chhaa" /ছ/ "zaa" /জ/ "zhaa" /ঝ/
- Post Alveolar: "shaa" /শ/
- Alveolar-Retroflex: "taa" /ট/ "thaa" /ঠ/ "daa" /ড/ "dhaa" /ঢ/ "raa" /ঢ়/ "rraa" /ড়/
- Alveolar: "raa" /র/ "laa" /ল/ "shaa" /শ/ "shaa" /ষ/ "saa" /স/ "zaa" /য/ "naa" /ন/
- Dental: "taa" /ত/ "thaa" /ঠ/ "daa" /দ/ "dhaa" /ধ/
- Labial: "paa" /প/ "phaa" /ফ/ "baa" /ব/ "bhaa" /ভ/ "maa" /ম/
- Labio-Dental: "faa" /ফ/ "bhaa" /ভ/

The burst spectrum for alveolars had a diffuse- rising pattern wherein the peaks were evenly spaced (diffuse) and\or the peaks at the higher frequencies had higher energy than those at lower frequencies. The burst spectrum of velars exhibited a compact spectrum which had high number of peaks were concentrated in the mid-frequency region than the low and high frequencies.

The following Figure 9 shows the time waveform, gray scale spectrogram and formant tracking chart for a male voice utterance of the Bangla word "hashi" / হাসি / containing only laryngeal or glottal stop "haa" / হ /.



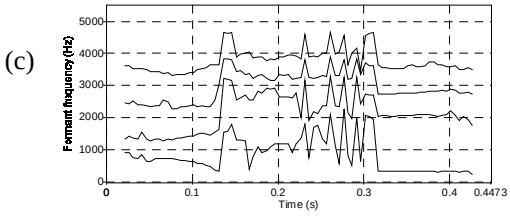
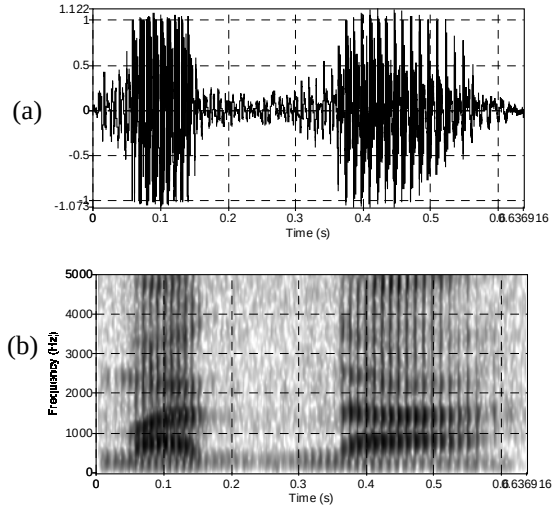


Fig. 9 Bangla glottal stop "haa" / হ / in word "hashi" / হাসি /
(a) Time-waveform (b) Spectrogram
(c) Formant track

As seen in the spectrogram of the Figure 9, there is no voicing during the initial closure of the stop "haa" /হ/. Then suddenly, there is a burst of energy and the voicing begins, goes for a couple of milliseconds or so, followed by an abrupt loss of energy in the upper frequencies, followed by another burst of energy, and some noise. The first burst of energy is the release of the initial stop. It has been observed that the formants moving into the vowel, where they sort of hold steady for a while and then move again into the final stop. The little blob of energy at the bottom is voicing, only transmitted through flesh rather than resonating in the vocal tract. The final burst is the release of the final stop, and the last bit of noise is basically just residual stuff echoing around the vocal tract. In brief, the major spectral characteristics of the stop consonants, important for identification, are the release burst frequencies, the shape of the burst spectrum and the formant transitions.

The following Figure 10 shows the time waveform, gray scale spectrogram and formant tracking chart for a male voice utterance of the Bangla word "magna" /মাগনা / containing Bangla nasal consonant "naa" /ন/.



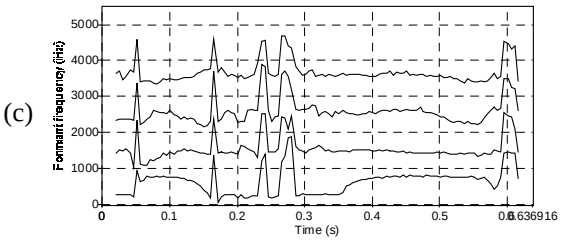


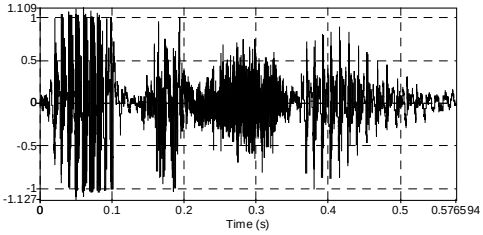
Fig. 10 Bangla nasal “naa” /ন/ in word “magna” / মাগনা / (a) Time-waveform (b) Spectogram (c) Formant track

Nasals have some formant stucture as shown in the Figure 10, but are better identified by the relative 'zeroes' or areas of little or no spectral energy. In spectrogram shown for the nasal “naa” /ন/ in word “magna” /মাগনা/, the final nasals have identifiable formants that are lesser in amplitude than in the vowel, and the regions between them are blank. Nasality on vowels can result in broadening of the formant bandwidths, and the introduction of zeroes in the vowel filter function.

The real trick to recognizing nasals stops is a) formant structure, but b) relatively lower-than-vowel amplitude. Place of articulation can be determined by looking at the formant transitions, and sometimes, based on the voice knowledge, and the formant/zero structure itself.

Looking at the spectrogram in Figure 10, it can be can be seen that the nasal “naa” /ন/ in word “magna” / মাগনা / has an F2/F3 'pinch'--the high F2 of “naa” /b/ moves up and seems to merge with the F3. In the nasal itself, the pole (nasal formant) is up in the neutral F3 region.

The following Figure 11 shows the time waveform, gray scale spectrogram and formant tracking chart for a male voice utterance of the Bangla word “habshi” /হাবশী/ containing Bangla fricative consonant “baa + shaa” /ব + শ/.



(a)

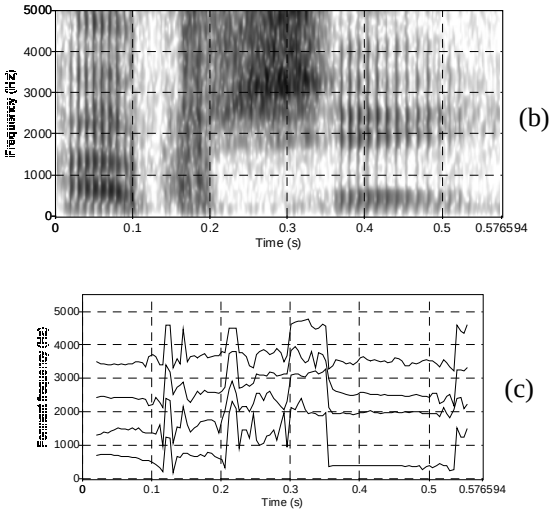


Fig 11: Bangla fricatives “baa + shaa” /ব + শ/ in word “habshi” /হাবশী / (a) Time-waveform (b) Spectrogram (c) Formant track

Fricatives, by definition, involve an occlusion or obstruction in the vocal tract great enough to produce noise (frication). Frication noise is generated in two ways, either by blowing air against an object (obstacle frication) or moving air through a narrow channel into a relatively more open space (channel frication). In both cases, turbulence is created, but in the second case, it's turbulence caused by sudden 'freedom' to move sideways.

The spectrogram of Bangla fricatives “baa+shaa” /ব+শ/ in word “habshi” /হাবশী/ is shown in the Figure 11. The sound “shaa” /শ/ is by far the loudest fricatives. The darkest part of “shaa” /শ/ noise is off the top of the spectrograms, even though these spectrograms have a greater frequency range than the others. “shaa" /শ/ is centered (darkest) and has most of its energy concentrated in the F3-F4 range.

6 Conclusion

This paper discussed various issues of speech production and perception. The role of various articulators in the classification of sound is discussed. The acoustic and articulatory features are observed both for the vowels and consonants. The linguistic classification of Bangla phoneme along with the English phoneme is also discussed with the allophones and phonetic similarity features. The computational model for the production of speech is discussed with the characterization of phoneme based on spectral properties. This is apparent that linguistic classification which is based on position and manner of articulation

does not provide sufficient spectral characteristics needed for the synthesis and recognition of speech due to the nature of the phonemes as well as of the speech. The result of applying the speech processing on the selected Bangla words containing different vowels and consonants shows more pragmatic features than their linguistic counterpart.

References:

[1] M.A. Hai, Bengali Language Handbook, Center for Applied Linguistics, Washington D.C., 1966

[2] R. Islam, An Introduction to Colloquial Bengali, Chapter 1, Central Board for Development of Bengali, Dhaka, 1970

[3] M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Procs*., pp. 208-211, Apr. 1979.

[4] S.Blumstein and K.Stevens, "Acoustic invariance in speech production," *J. Acoust. Soc. Am.*, vol. 66, pp. 1001-1017, 1979.

[5] S. Blumstein and K. Stevens, "Perceptual invariance and onset spectra for stop consonants in different vowel environments," *J. Acoust. Soc. Am.*, vol. 67, pp. 648-662, 1980.

[6] S. Blumstein, E Issac and J. Mertus, "The role of gross spectral shape as a perceptual cue to place of articulation", *J. Acoust. Soc. Am.*, vol. 72, pp. 43-50, 1982.

[7] S.Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol.27, pp. 113-120, Apr. 1979.

[8] Cheng and D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence," *ICASSP*, vol.2, pp. 961-964, Apr. 1991.

[9] F. Cooper, P. Delattre, A. Liberman, J. Borst and L. Gerstman, "Some experiment on perception of synthetic speech sounds", *J. Acoust. Soc. Am.*, vol. 24, pp. 597-606, 1952.

[10] J. Deller Jr, J. Proakis and J. Hansen, *Discrete-time processing of speech signals*, Macmillan, 1993.

[11] M. Dorman, M. Studdert-Kennedy and L. Ralphael, "Stop consonant recognition: Release bursts and formant transitions as functionally equivalent context-dependent cues," Percept. Psychophys., vol. 22, pp. 109-122, 1977.

[12] M. Dorman and P. Loizou, "Relative spectral change and formant transitions as cues to labial an alveolar place of articulation," *J. Acoust. Soc. Am.*, vol. 100, pp. 3825-3830, 1996.

[13] G.Fant, *Acoustic Theory of Speech Production*, 's-Gravenhage, The Netherlands: Mounton and Co., 1960.

[14] J. Flanagan, "A difference limens for vowel formant frequency," *J. Acoust. Soc. Am.*, vol. 27, pp. 288-291, 1955.

[15] B. Gold and N. Morgan, *Speech and audio signal processing*, Wiley, 2000.

[16] J. Hawks, "Difference limens for formant patterns of vowel sounds," *J. Acoust. Soc .Am.*, vol. 95, no. 2, pp. 1074-1084, 1994.

[17] J. Hillenbrand and R. Gayvert, "Identification of steady-state vowels synthesized from the Peterson and Barney measurements," *J. Acoust. Soc. Am.*, vol. 94, pp. 668-674, 1993.

[18] Syed Akhter Hossain, M Lutfar Rahman , Farruk Ahmed "Vowel Space Identification of Bangla Speech", Dhaka University Journal of Science, 51(1): 31-38 2003(January)

[19] Syed Akhter Hossain, Farruk Ahmed, Mozammel Huq Azad Khan, M A Sobhan, and M Lutfar Rahman, "Analysis by Synthesis of Bangla Vowels", 5th International Conference on Computer and Iinformation Technology Proceeding, 2002, pp. 272-276

[20] Syed Akhter Hossain, M A Sobhan, Mozammel Huq Azad Khan, "Acoustic Vowel Space of Bangla Speech", International Conference on Computer and Iinformation Technology 2001 Proceeding, pp. 312-316

[21] Syed Akhter Hossain and M Abdus Sobhan, "Fundamental Frequency Tracking of Bangla Voiced Speech" –Proceedings of the 1st National Conference on Computer and Information System Proceeding 1997, pp. 302-306