Computational Analysis of Bangla Retroflex and Dental Consonants

Hossain. Syed Akhter, Rahman. M Lutfar, and Ahmed. Farruk

***Abstract*—** This paper describes in detail the computational investigation through analysis and synthesis of Bangla Retroflex and Dental consonants /ট/ [ta], /ঠ/ [tha], /ড/ [da], /ঢ/ [dha] and /ত/ [t], /থ/ [th], /দ/ [d] and /ধ/ [dh]. After careful segmentation of speech data, spectral analysis were performed in four long vowel contexts /আ/ [a], /ই/ [i], /ও/ [o], and /উ/ [u] to parameterize the contextual effects. The research was intended for an accurate description of the acoustic characteristics and features of Bangla retroflex and dental consonants for synthesis as well as recognition of consonants in different contexts. The quality of synthetic speech was evaluated through subjective listening along with matching the spectra of synthetic speech with original speech. The synthesized speech segment was fairly clear with good degree of phonetic naturalness.

***Keywords*—** Speech Processing, Synthesis, Phoneme, LPC

I. Introduction

BANGLA is an Indo-Aryan language of more than 300 million people in the world concentrating in the eastern region of Indian subcontinent i.e. Bangladesh, Indian states of West Bengal, Tripura, Assam and around. The history of Bangla begins in the late 18 centuries and before that there was only a family of dialects commonly known as Prakrit [1].

Every language has its own set of basic sounds, which constitute the language. Bangla speech contains a set of eleven vowels with eight being the cardinal vowels and 39 consonants, of which about 30 consonants are of frequent usage. The vowels can be pronounced independently, but the consonants need the support of vowels to be pronounced. These consonants can be conveniently classified according to the manner and place of articulation [1].

The Bangla consonants possess certain special features, which are not so common to other European languages and American English. The major discrepancy results from a difference in the number of places at which the stops can be formed by the tip of the tongue making contact with the front portion of the top of the mouth. Consonants differ from vowels in that they had more energy in the high frequency region compared to the low frequency region. Bangla linguistics classifies consonants on the basis of their place of articulation. In Bangla, there are five ways of distinction exist in the place of articulation for the stop sounds, compared to English where there is only four way contrast [2].

In this work, we investigated spectral properties of Bangla retroflex and dental sounds from 1800 speech corpora containing vowel-consonant-vowel (VCV) and consonant-vowel-consonant (CVC) syllables produced by male and female speakers of different age groups. The revealed spectral properties of the retroflex and dental sounds along with the burst amplitude and duration are used in the synthesis of both the retroflex and dental sounds. The synthetic sound and spectra was evaluated based on the experimental observations.

The rest of the paper is organized as follows: Section II covers literature review along with the analysis procedure and Section III discusses the synthesis procedure with experimental results and discussion in Section IV. Section V concludes the research observations.

II. Analysis Procedure

A. Literature Review

In Bangla linguistics, the vowel and consonant sounds have been further classified into many sub-classes based on place and manner of articulation. Besides these, Bangla sounds are also classified into another two distinct classes according to their mode of excitation: voiced and unvoiced sounds which are subclasses of Plosive sounds [1], [2].

Voiced sounds are produced by forcing air through the glottis with the tension of the vocal chords adjusted so that they vibrate in a relaxation oscillator mode, thereby producing quasiperiodic pulses of air which excite the vocal tract. In Bangla, voiced segments are: /গ/ [ga], /জ/ [ja], /ড/ [da], /দ/ [da], /ব/ [ba] etc. Forming a constriction at some point in the vocal tract, and forcing air through the constriction at a high enough velocity to produce turbulence generates fricative or *unvoiced* sounds e.g. /ক/ [*ka*], /চ/ [*cha*], /ট/ [*ta*], /ত/ [*ta*], /প/ [*pa*] in Bangla. *Plosive* sounds result from making a complete closure in front of the vocal tract, building up pressure behind the closure, and abruptly releasing it [1]-[3].

According to the physiological aspect, consonants can be defined as, “A sound (in normal speech), in forming which the air issues out of the lungs through the pharynx and mouth or the different articulators, there being facing obstruction and

Hossain. Syed Akhter is with the Computer Science and Engineering Department, Daffodil International University, Dhaka, Bangladesh, (corresponding author’s phone: +8801817382645; fax: +88029131947; e-mail: aktarhossain@daffodilvarsity.edu.bd).

Rahman. M Lutfar is with Daffodil International University as the Vice Chancellor, Dhaka, Bangladesh (e-mail: lrahman@daffodilvarsity.edu.bd).

Ahmed. Farruk is with the School of Engineering and Computer Science, Independent University Bangladesh, Dhaka, Bangladesh (e-mail: farruk@iub.secs.edu.bd).

narrowing such as would cause audible friction" [3]-[4].

Kewley-Port [5] explored the spectral change over time by focusing on the time-varying properties of stop consonants from the release burst into the vowel portions of Consonant-Vowel (CV) syllables. One of the metric adopted was spectral tilt of the burst. The study done by Kewley-Port mainly focused on observing the absolute spectral tilt in different stop consonants in the release burst spectrum. But the investigator did not take into account the relative changes in the spectral energy, which occurred, from burst to the formant transitions. It was then proposed by Lahiri *et al* [6] that the invariant properties for the labial and dental/alveolar place of articulation could be contained in relative changes in the distribution of energy at high and low frequencies when measured at the release burst and the beginning of formant transitions.

A study done by Dorman and Loizou [7] verified that the relative spectral change were indeed effective in classifying the labials and alveolar stops in English. It was also verified by Dorman and Loizou that an attribute of the signal, which allows accurate sorting of the signals by algorithm, fails to exert a large influence on perception when other cues are present.

One of the earliest acoustic study on Bangla speech was conducted by Paramanik [8, 9] during the doctoral research in Japan. The study was limited to Bangla vowels for a small set of Bangla vowel phonemes and extracted formants of isolated Bangla vowels and drawn a comparison with those of Japanese vowels and vowel like sounds.

B. Analysis Procedure

All the consonants were recorded in the form of CVC and VCV syllables spoken by male and female native Bangla speakers in all four vowel contexts /অ/ [a], /ই/ [i], /ও/ [o], and /উ/ [u] at the sampling rate of 16 kHz. These CVC and VCV syllables were then broken into CV and VC syllables for the entire analysis through spectral techniques. The starting and ending times of the vocalic nuclei was measured by hand from high-resolution digital spectrograms. Only the first 10 ms of the release burst or the whole burst was considered.

Prior to the analysis, the Bangla vowel data set was manually and visually segmented to consonant vowel consonant segments. The starting and ending times of each vocalic nuclei were measured by hand from high-resolution digital spectrograms. In order to avoid the effect of formant transitions due to the role of consonants, acoustic measurements were made starting form 20% of the vowel duration to 80% of the vowel duration. The vocalic segments were segmented into frames of 10-20 ms and acoustic analyses were done on these frames [10]-[12].

For Bangla retroflex and stop consonants, the release burst were analyzed over a 10 ms interval or the total burst duration, whichever was found smaller, starting from the burst.

A half-Hamming window of 20 ms was used that is (only the latter half of the Hamming window was multiplied with the 10 ms of the release burst samples). The LPC spectrum was obtained using a 512-point FFT. The frequency of the maximum amplitude of spectral peak was extracted from LPC spectrum using a global peak picking algorithm.

C. Synthesis Procedure

Articulatory synthesis was used as computational models of the articulators(e.g. jaw, tongue, lips etc.) and their movements to synthesize speech. In this model speech is synthesized through well defined rules [13]-[17].

Same formant frequencies of burst (and transitions) were used for consonants belonging to same place of articulation. For example same values of first four formants were used for all four consonants /ত/ [t], /থ/ [th], /দ/ [d] and /ধ/ [dh] in the dental group for a given vowel context.

The synthesis flow diagram used in this study is shown on the flow diagram of Fig. 1. As shown in the Fig. 1, the synthesis model process frames of segment of speech and determine necessary coefficients and gain from the LPC based analysis of the segment. The high frequency noise is removed and required contextual modification is achieved through the prosodic measurements.

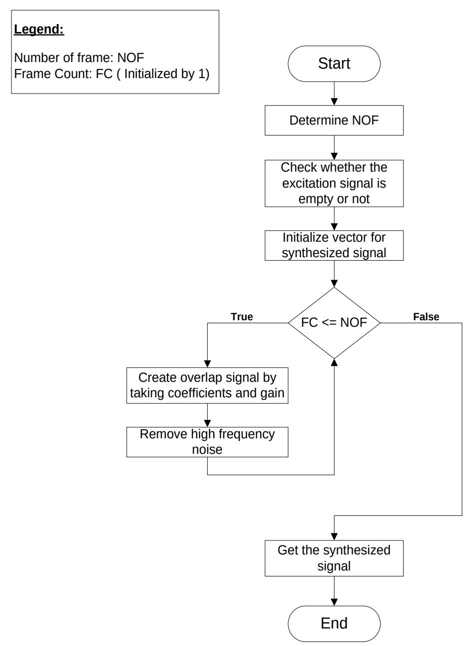

Fig. 1 Flow diagram of Bangla consonant synthesis

III. Experimental Results and Discussion

Prior to any analysis, the complete consonant data set was manually segmented to Vowel-Consonant-Vowel segment. The starting and ending times of the vocalic nuclei was measured by hand from high-resolution digital spectrograms. Only the first 10 ms of the release burst or the whole burst was considered.

As can be seen from the Figure 2, the speech segment /আত/ [aat] is displayed on the top frame and the segment length is about 65msec. The FFT and LPC smoothed spectrum is captured in another frame whereas frame wise formant contour and log Mel power spectrum of the same speech segment is captured in another two frames.

The formant contour clearly demonstrates the transition of formant frequency due to consonant segment /ত/ [ta]. The first formant F1 tend to increase due to the closure of the oral tract. The FFT and LPC smoothed portion of the spectrum shown is the last frame of the speech segment under analysis.

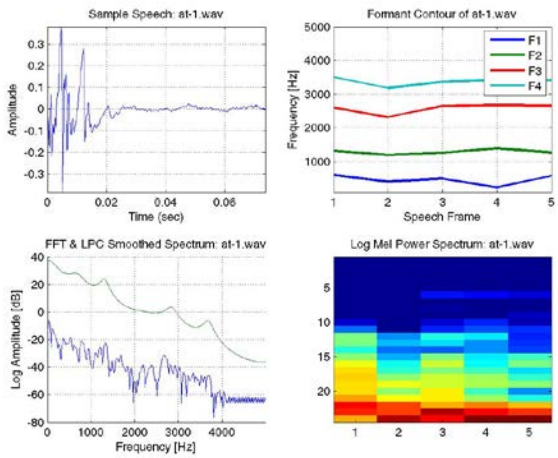

Fig. 2. Analysis of Bangla speech segment /আত/ [aat] from the Bangla corpus আতা [ata] produced by a male speaker

Fig. 3 shows the analysis of the speech segment আত [aat] and the speech segment তা [ta] same as the male speaker. The segment length is about 85 msec and 95 msec respectively. The FFT and LPC smoothed spectra along with the formant contour and log Mel power spectrum of the same speech segments are captured in another two frames. The formant contour clearly demonstrates the transition of formant frequencies due to vowel-consonant-vowel phonation. The first formant F1 tends to increase due to the closure of the oral tract and decreases again due to the opening.

The experimental results are shown in Table I and Table II and as can be seen, the place of articulation and the vowel context determine the burst frequencies of retroflex and dental sounds. The burst and formant transition of the target vowel varies across the stop consonants. The same stop consonant shows variation in the formant frequency transition in different vowel context.

The vowel context /ই/ [i] shows significant difference in formant transition between first formant F1 and second formant F2 for all the retroflex sounds of Bangla. The formant transition between the second and third formant frequency F2 and F3 is consistent throughout different dental sounds. Besides, the context /উ/ [u] and /ও/ [o] shows persistent formant transitions for different retroflex sounds as can be seen from the Table I and Table II.

TABLE I
COMPARISONS OF BURST FREQUENCIES OF DENTAL SOUNDS

(a) /আ/ [a] context

Formant	/ত/ [t]	/থ/[th]	/দ/ [d]	/ধ/ [dh]
F1 (Hz)	600	950	380	750
F2 (Hz)	1450	2000	1700	1900
F3 (Hz)	2700	2900	2700	3000
F4 (Hz)	3400	3900	3350	3850

(b) /ই/ [i] context

Formant	/ত/ [t]	/থ/[th]	/দ/ [d]	/ধ/ [dh]
F1 (Hz)	400	450	350	450
F2 (Hz)	2150	1850	2150	2150
F3 (Hz)	2550	2950	2900	2800
F4 (Hz)	3800	4150	3250	3750

(c) /উ/ [u] context

Formant	/ত/ [t]	/থ/[th]	/দ/ [d]	/ধ/ [dh]
F1 (Hz)	550	450	375	350
F2 (Hz)	1600	1750	1650	1600
F3 (Hz)	2750	3150	2750	3000
F4 (Hz)	3850	3550	3450	3950

(d) /ও/ [o] context

Formant	/ত/ [t]	/থ/[th]	/দ/ [d]	/ধ/ [dh]
F1 (Hz)	550	500	250	450
F2 (Hz)	1500	1900	1725	1450
F3 (Hz)	2850	2550	2550	2850
F4 (Hz)	3550	3950	3950	3900

The consonant /ট/ [t] shows higher relative burst amplitude in higher formant frequencies in all vowel contexts. In case of consonant /ঠ/ [th] the relative burst amplitude is high in higher formant frequencies except for the vowel context /আ/ [aa]. The consonant /ড/ [d] also shows higher relative burst amplitude in higher formant frequencies except for the vowel contexts /আ/ [aa] and /উ/ [u]. In case of consonant /ঢ/ [dh] the relative burst amplitude is higher in higher formant frequencies for all vowel contexts under the present study.

Burst frequencies depend on the place of articulation of the stop as well as the vowel context. The spectral difference among the stop consonants across the places of articulation are primarily reflected in the spectra of the burst and formant transitions of the target vowel. The burst frequencies of the same consonants do show slight variation in different vowel contexts.

Formant frequency measures were taken one-third into the vowel (i.e., at the point determined by adding one-third of the total duration to vowel onset). Formant values were both automatically computed by spectral analysis using Matlab and visually verified using the spectrographic display; these methods almost always converged.

TABLE II
COMPARISONS OF BURST FREQUENCIES OF RETROFLEX SOUNDS

(a) /আ/ [a] context

Formant	/ট/ [t]	/ঠ/ [th]	/ড/ [d]	/ঢ/[dh]
F1 (Hz)	950	600	350	550
F2 (Hz)	1900	1800	1850	1750
F3 (Hz)	3100	2750	2850	2650
F4 (Hz)	3450	3500	3600	3650

(b) /ই/ [i] context

Formant	/ট/ [t]	/ঠ/ [th]	/ড/ [d]	/ঢ/[dh]
F1 (Hz)	1050	350	300	350
F2 (Hz)	1750	2050	2100	2100
F3 (Hz)	3150	2850	2950	2950
F4 (Hz)	3350	3850	3900	3950

(c) /উ/ [u] context

Formant	/ট/ [t]	/ঠ/ [th]	/ড/ [d]	/ঢ/[dh]
F1 (Hz)	400	400	250	300
F2 (Hz)	1650	1950	1500	1625
F3 (Hz)	2850	3050	2650	3250
F4 (Hz)	3450	3500	3700	4150

(d) /ও/ [o] context

Formant	/ট/ [t]	/ঠ/ [th]	/ড/ [d]	/ঢ/[dh]
F1 (Hz)	500	550	450	300
F2 (Hz)	1850	2050	1550	1600
F3 (Hz)	2650	2770	2550	2850
F4 (Hz)	3500	3350	3450	3800

It has also been observed that the bandwidths of the formants are narrower (i.e. the formant peaks are better defined) in the case of voiced aspirated sounds /দ/ [d], /ধ/ [dh], /ড/ [d], and /ঢ/ [dh] as compared to unvoiced aspirated sounds /ত/ [t], /থ/ [th], /ট/ [t], and /ঠ/ [th]. The formant transitions from burst frequency to the target vowel frequency are part of the aspiration segment. Aspiration duration is about 90-110 ms in case of retroflex sounds, while the duration is 80-100 ms for dental sounds.

Burst frequencies depend on the place of articulation of the stop as well as the vowel context. The spectral difference among the dental consonants across the places of articulation is primarily reflected in the spectra of the burst and formant transitions of the target vowel. The burst frequencies of the same consonants do show slight variation in different vowel contexts.

IV. CONCLUSION

We identified the contextual parameters and spectral differences of Bangla retroflex and dental consonants based on the burst amplitude as well as formant transitions. These burst amplitude and formant transition can be used for natural synthesis of consonant segments within the vowel context. We also conclude that the burst position at third and fourth formant contains most of the energy in case of dental sounds while fourth formant contains most of the energy in case of retroflex sounds. Despite the lowering of the third and fourth formant in case of dentals, retroflex sounds have a very strong burst release as compared to the corresponding dental sounds. Also there is general lowering of the third and fourth formants in case of retroflex as compared to dentals. Voice bar frequency is about 10 to 15% lower in case of retroflex as compared to dentals. These spectral parameters in different vowel context may be used for identification of dental and retroflex consonants in speech recognition beside speech synthesis.

REFERENCES

[1] M.Abdul 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] Md. Abdul Hai, *Dhvani Vljnan 0 Bangla Dhvani-Tattwa (Bangla)*, Mullick Brothers, 2000.

[4] 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.

[5] D. Kewley-Port, "Time-varying features as correlates of place of articulation in stop consonants," *J. Acoust. Soc. Amer*. 73, 322-335, 1983.

[6] L. Gerwirth Lahiri and S. Blumstein, "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants. Evidence from a cross-language study," *J. Acoust. Soc. Am.*, vol. 76, pp. 391-404, 1984.

[7] 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.

[8] Abdul Kader Paramanik, *Acoustical Study on the Vowel Structure of Bengali Language*, Doctoral Dissertation, Tohoku University, Japan, 1977.

[9] M Paramanik and K Kido, "Bengali Speech:Formant Structure of Single Vowels and intial vowels of words," *IEEE International Conference on ICASSP*, Japan, 1976.

[10] A K Datta, R Sengupta, N Dey, B M Banerjee & Dipali Nag, "Perception of Nasality in Bangali Vowels: Role of Harmoinics between F0, F1," *Proceedings of Internation Conferences on Computational Linguistics, Speech and Document Processing*, India, 1998.

[11] A K Datta, R Sengupta, N Dey, B M Banerjee & Dipali Nag, "Spectral Cues of Nasality in Inherently Nasal Bengali Vowels," *Proceedings of Internation Conferences on Computational Linguistics, Speech and Document Processing*, India, 1998.

[12] A K Datta, R Sengupta, N Dey, B M Banerjee & Dipali Nag, "Speaker Identification using pitch in Bengali," *Proceedings of Internation Conferences on Computational Linguistics, Speech and Document Processing*, India, 1998.

[13] Sarwar-E-Alam, *Bengali Voice Synthesis*, MSc Thesis, Department of Computer Science and Engineering, University of Dhaka, Dhaka, Bangladesh, 1995.

[14] Examination Roll:841, *A Text-To-Speech Synthesis System for Bangla Language Using Concatenative Technique*, MS Thesis, Department of Applied Physics and Electronics, Rajshahi University, Rajshahi, Bangladesh, 1997.

[15] Syed Khairuzzaman Tanbeer, Md Sarwar-E-Alam and Md Abdul Mottalib, "Study of Phonemes for Bangla Voice Synthesis," *Proceedings of International Conferences on Computer and Information Technology*, Dhaka, Bangladesh, 2000.

[16] Soumen Chowdhury, Moushumi Goswami and Anupam Basu, "Partneme as the Speech Inventory for Bangla Text to Speech Synthesis," *Proceedings of International Conferences on Computer Processing of Bangla*, Dhaka, Bangladesh, 2005.

[17] Firoz Alam, S. M. Murtoza Habib, Mumit Khan, "Acoustic Analysis of Bangla Consonants," *Spoken Language Technologies for Under-resourced language* (SLTU'08), Hanoi, Vietnam, May 5 - 7, 2008

Syed Akhter Hossain is currently working as Professor and Head of the Department of Computer Science and Engineering at Daffodil International University. Professor Hossain obtained MSc in Applied Physics and

Syed Akhter Hossain is currently working as Professor and Head of the Department of Computer Science and Engineering at Daffodil International University. Professor Hossain obtained MSc in Applied Physics and Electronics from Rajshahi University and PhD in Computer Science and Engineering from University of Dhaka. He attended Erasmus Mundus post doctoral fellowship in the area of Informatics and Industrial Engineering with University Lumiere Lyon 2 in France. He has more than 19 years of working experience in industry, education, research and training. He is actively involved in research guidance/ research projects/ research collaborations with Institutes/ Industries and has more than 50 publications/ presentations and his work is listed in DBLP, IEEE Explore and other research databases. He received International awards for teaching including other national awards for scholastic works. He is also member of several national and international professional bodies. More details about him can be found at http://syedakhterhossain.blogspot.com/

M. Lutfar Rahman is Vice-Chancellor of Daffodil International University, Dhaka, Bangladesh. He was founding Vice-Chancellor of Begum Rokeya University, Rangpur, Bangladesh and founding Chairman of the Department of Computer Science and Engineering, University of Dhaka. Professor Rahman obtained MSc in Physics in Bangladesh and MSc and PhD in Electronic and Electrical Engineering in England. He has teaching and research experience in Microprocessor Applications, Computer Networks, Information Security and Bangla Language Processing. He has over 100 research papers to his credit and he authored/co-authored sixteen books on Electronics, Communications and Computer Engineering.

Professor Rahman has contributions for expansion of Computer Science education and research in Bangladesh. He initiated a number of national and international conferences and edited a number of research journals. He received awards from several institutions and societies for contributions to education and research. He is member of a number of professional societies and is currently President of Bangladesh Electronics Society.

Farruk Ahmed is working, at present, as a Professor in the School of Engineering and Computer Science, Independent University, Bangladesh (IUB), Dhaka. Previously he has worked as the Professor and Chairman of the Department of Applied Physics, Electronics and Communication Engineering, University of Dhaka and also in the Department of Electrical Engineering and Computer Science in North South University, Bangladesh. He has taught wide range of subjects which include Digital Signal Processing, Communication Engineering, Microcomputer and Microprocessor based systems, Feedback Control Systems, Electronic Instrumentation etc and his research activities include all of them. He has supervised a big number of B.Sc, M.Sc projects / theses, and PhD theses. He has published nearly 180 papers and research articles in various national and international journals and conference proceedings of repute. He holds a PhD degree in Electronic and Electrical Engineering from the University of Salford in Manchester, England. He has served on various University and Governmental Committees as member or chairman/convener. He has been the President of Bangladesh Electronics Society for about 10 years.