International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:2, 2007

Bangla Vowel Characterization based on Analysis by Synthesis

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

***Abstract*—**Bangla Vowel characterization determines the spectral properties of Bangla vowels for efficient synthesis as well as recognition of Bangla vowels. In this paper, Bangla vowels in isolated word have been analyzed based on speech production model within the framework of Analysis-by-Synthesis. This has led to the extraction of spectral parameters for the production model in order to produce different Bangla vowel sounds. The real and synthetic spectra are compared and a weighted square error has been computed along with the error in the formant bandwidths for efficient representation of Bangla vowels. The extracted features produced good representation of targeted Bangla vowel. Such a representation also plays essential role in low bit rate speech coding and vocoders.

***Keywords*—**Speech, Vowel, Formant, Synthesis, Spectrum, LPC.

I. Introduction

VOWELS are produced by exciting a fixed vocal tract with quasi-periodic pulses of air caused by vibration of the vocal chords. The resonant frequency of the tract (formants) varies with the cross sectional area and the dependence of cross-sectional area upon distance long the tract, which is known as area function is determined primarily by the position of the tongue, but the positions of the jaw, lips, and, to a small extent, the velum also influences the resulting sound [1].

The chief characteristic of the vowels is the freedom with which the air stream, once out of the glottis, passes through the speech organs. The supra-glottal resonators do not cut off or constrict the air; they cause only resonance, that is to say, the reinforcement of certain frequency ranges [2].

The basic idea of analysis-by-synthesis starts with the representation of the speech signal such as the time-dependent Fourier transform. The basic all pole model of the speech production is considered along with the time-dependent Fourier representation of the model to match that of with the speech signal. Then by varying the parameters of the model in a systematic way, one can attempt to find a set of parameters that cause the model to match the speech signal with minimum error. When such a match is found, the parameters of the model are assumed the parameters of the speech signal [3].

A. Analysis by Synthesis

In general, an analysis-by-synthesis approach attempts to describe the acoustic representation, rx , with some synthetic representation, rs , which is generated from a set of parameters, ry , and a synthesis model, f (), where rs = f (ry) . Although the nature of the modeling has varied, this type of approach has been used previously for performing automatic analyses of the speech signal (Bell et al., 1961; Olive ,1971) and for speech recognition (Blomberg, 1989; Blomberg et al., 1988). The acoustic and synthetic representations have typically been in the spectral domain.

For automatic analysis of speech, it is desirable to find a value for ry which optimizes the match between rx and rs in some way. In the past, recursive procedures have been used to minimize the distance between these representations, although this problem could be cast in a more stochastic framework as well [4].

II. Analysis Procedure

The digital model [2] for voiced speech as shown in the Fig. 1 assumes that a short segment of voiced speech is identical to the same length segment of a periodic sequence

$$x(n) = \sum_{m=-\infty}^{\infty} h_{\upsilon}(n + mN_p) \qquad (1)$$

where $h_v(n)$ represents the convolution of the vocal tract impulse response, v(n), with the glottal pulse, g(n), and the impulse response of the radiation load, r(n).

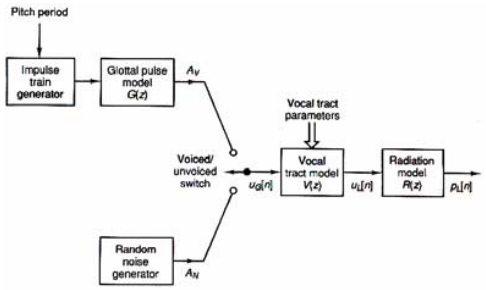

Fig. 2 Discrete Time System Model for Speech Production

That is,

$$h_v(n) = r(n) * v(n) * g(n) \qquad (2)$$

The quantity N_p is the pitch period in samples. The radiation effects, which appear as a differentiation at low frequencies, are adequately modelled for most purposes by a simple first difference, for which the z-transform representation is as follows:

$$R(z) = 1 - z^{-1} \quad (3)$$

Such a representation is required for vowel synthesis and the production model shown in Fig.1 can be simplified to the system of Fig. 2 as shown below.

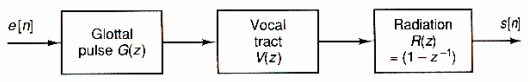

Fig. 2 Simplified Model for Synthesizing Voiced Sound

The excitation signal *e[n]* is a quasi-periodic impulse train and the glottal pulse model could be either the exponential or the Rosenburg pulse. The vocal tract is characterized by a transfer function of the form:

$$V(z) = \frac{A}{\prod_{k=1}^{M}(1 - 2e^{-\sigma_k T}\cos(2\pi F_k T)z^{-1} + e^{-2\sigma_k T}z^{-2})} \quad (4)$$

where the number of poles included depends upon the sampling frequency of the input data. Finally the glottal pulse is of finite duration, implying that the z-transform of g(n) is a polynomial in z of the form:

$$G(z) = \sum_{n=0}^{N_g} g(n)z^{-n} = B\prod_{n=1}^{N_g}(1 - z_n z^{-1}) \quad (5)$$

where N_g is less than N_p. From Eq. (2) it is seen that the z-transform of $h_v(n)$ is

$$H_v(z) = R(z) \,.\, V(z) \,.\, G(z) \quad (6)$$

and the corresponding Fourier transform would be

$$H_v(e^{j\omega}) = R(e^{j\omega})\, V(e^{j\omega})G(e^{j\omega}) \quad (7)$$

The Fourier transform of the periodic signal x(n) will consist of very sharp spectral lines at multiples of the fundamental frequency [5-8].

The basic vowel sounds are recorded for both the adult male and female speakers based on selective isolated Bangla word for each of the vowel sound. Vowels are then visually labeled and segmented from the isolated Bangla words like the vowel আ from the spoken word আম [9].

After the pre-emphasis, speech sample is down sampled to 10kHz and the Linear Predictive Coding (LPC) technique is applied on the selected vowel segment based on auto-correlation method with a prediction order of 10 and an analysis window of 20msec with an overlap segment size of 5msec.

In order to determine the five formants which is, in most of the speech application, cover wide resolution in the resonance spectrum of the vocal chord. From the LPC co-efficient vector and the corresponding gain, the formant tracking and peak picking is applied to each analysis frame. The following Fig 3(a) and 3(b) shows the findings on LPC spectrum and formant tracks and the Table I shows the different formants and their respective bandwidths.

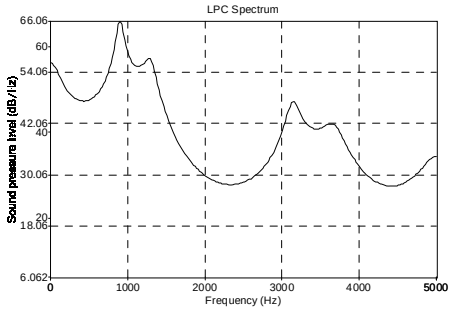

Fig. 3 (a) LPC Spectrum of a Bangla Vowel Segment

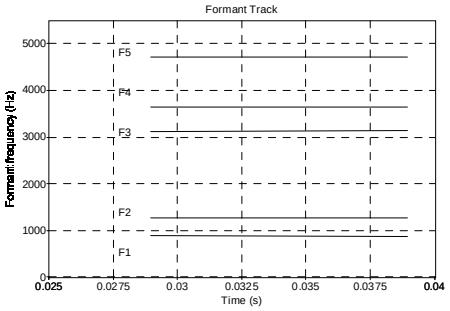

Fig. 3 (b) Formant track of আ in the utterance আম (female)

TABLE I
FORMANT FREQUENCIES

Formant	Freq(Hz)	Bandwidth (dB)
F1	897.28	85
F2	1272.56	604
F3	3146.06	144
F4	3672.08	805
F5	4716.46	635

LPC coefficients extracted from the vowel segment are used for the spectral estimates of the vowel and the same coefficients are used in the synthesis of the vowel thorough inverse filtering operation as shown in the vowel production model in the Fig. 1. The synthesized vowel is shown in the Fig. 4(a).

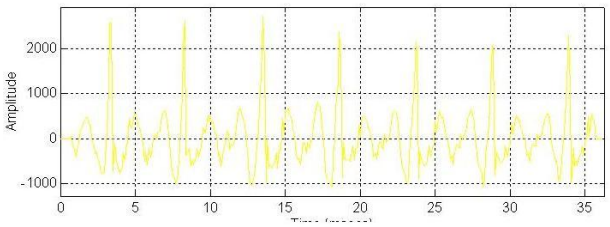

Fig 4 (a) Synthesized vowel আ generated by LPC coefficients

The synthesized vowel আ is now applied in the same procedure as the glottal source waveform and the LPC coefficients are calculated along with the formant tracking of the synthetic vowel [10-13].

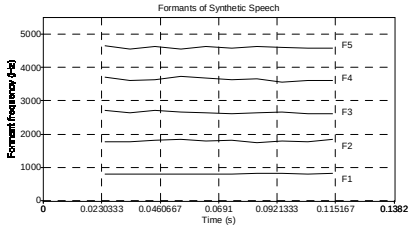

Fig. 4 (b) Formant Track of synthetic vowel আ (female)

The Fig 4(b) shows the formant tracks of the synthetic vowel and in Table 2, the formant frequencies and their respective bandwidths.

TABLE I1
FORMANT FREQUENCIES (SYNTHETIC VOWELS)

Formant	Freq(Hz)	Bandwidth (dB)
F1	819.85	231
F2	1697.15	602
F3	2640.94	547
F4	3690.52	423
F5	4684.88	323

Similar analysis by synthesis is carried out with all the basic vowel sounds in Bangla language.

III. RESULTS

Analysis-by-Synthesis is applied for Bangla vowels for both male and female speakers and the following tables show the respective results:

TABLE III (A)
FORMANT FREQUENCY AND BANDWIDTH (FEMALE SPEAKER)

A. Original Vowel			
Vowel	F1/B1	F2/B2	F3/B3
অ	690/65	1068/76	3173/202
আ	897/85	1272/604	3146/144
ই/ঈ	660/110	2893/140	3504/131
উ/ঊ	270/327	3279/341	4278/184
এ	447/93	2447/140	3016/936
ও	369/440	1054/1139	3144/161

TABLE III (B)
FORMANT FREQUENCY AND BANDWIDTH (FEMALE SPEAKER)

B. Synthetic Vowel			
Vowel	F1/B1	F2/B2	F3/B3
অ	865/216	1819/572	2727/606
আ	819/231	1697/602	2640/547
ই/ঈ	687/169	1531/434	2750/376
উ/ঊ	2423/534	3471/378	4539/858
এ	477/361	1936/496	2733/792
ও	473/780	1563/829	2561/548

TABLE IV (A)
FORMANT FREQUENCY AND BANDWIDTH (MALE SPEAKER)

C. Original Vowel			
Vowel	F1/B1	F2/B2	F3/B3
অ	624/98	2542/115	3332/185
আ	743/130	1258/182	2585/297
ই/ঈ	546/157	2205/99	2850/1016
উ/ঊ	349/50	2461/105	3188/267
এ	394/45	1865/744	3432/433
ও	459/94	2497/139	3368/151

TABLE IV (B)
FORMANT FREQUENCY AND BANDWIDTH (MALE SPEAKER)

D. Synthetic Vowel			
Vowel	F1/B1	F2/B2	F3/B3
অ	780/245	1760/103	2721/561
আ	807/345	1765/285	2404/288
ই/ঈ	710/890	1687/195	2566/377
উ/ঊ	455/77	1589/746	2579/790
এ	344/444	1746/604	2534/137
ও	633/507	2125/413	3102/123

IV. CONCLUSION

In this paper, a sequence of experiment has been performed to explore the analysis-by-synthesis techniques. Several factors make this approach attractive. The speech analysis to speech synthesis is still in need of acoustical data, which is phonetically well distributed. We investigated different parameters specially predictor order, threshold and the analysis frame size with overlap segment size in order to get accurate estimation of the speech parameters for the synthesis.

We have done synthesis of Bangla vowel which seems work well but requires more probabilistic analysis procedure to include alternative synthesis model in the future work for the Bangla vowel classification, synthesis and recognition.

REFERENCES

[1] Rabiner L. R., Schafer R. W., "Digital Processing of Speech Signals", Trentice-Hall Inc, Englewood Cliffs, 1978.
[2] John R Deller, John G Proakis, John H L Hansen, "Discrete-Time Processing of Speech Signals", Macmillan Publishing Company, 1993
[3] G.Fant, *Acoustic Theory of Speech Production*, 's-Gravenhage, The Netherlands: Mounton and Co., 1960.
[4] 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.
[5] S.Blumstein and K.Stevens, "Acoustic invariance in speech production," *J. Acoust. Soc. Am.,* vol. 66, pp. 1001-1017, 1979.
[6] 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.
[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] B. Gold and N. Morgan, *Speech and audio signal processing*, Wiley, 2000.
[9] Muhammad Abdul Hai, *Dhvani Vijnan O Bangla Dhvani-Tattwa*, Mullick Brothers, 2000.
[10] S Akhter Hossain, M Lutfar Rahman , Farruk Ahmed "Vowel Space Identification of Bangla Speech", Dhaka University Journal of Science, 51(1): 31-38 2003(January).
[11] S Akhter Hossain, Md Faruk Ahmed, Mozammel Huq Azad Khan, M A Sobhan, and Md Lutfar Rahman, "Analysis by Synthesis of Bangla Vowels", 5th International Conference on Computer and Iinformation Technology Proceeding, 2002, pp. 272-276.
[12] S 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.
[13] S Akhter Hossain & M Abdus Sobhan, "Fundamental Frequency Tracking of Bangla Voiced Speech" –1st National Conference on Computer and Information System Proceeding 1997, pp. 302-306.

Syed Akhter Hossain was born in Khulna, Bangladesh in 1965. He received B.Sc.(Hons) and M.Sc. in Applied Physics and Electronics from Rajshahi University in 1992. He started his teaching career in Institute Technology Pekerja Pekerja, Malaysia as Lecturer from 1993 to 1996. He served also as Lecturer in the Department of Computer Science and Engineering at Rajshahi University. He joined East West University, Dhaka as the Assistant Professor in the Department of Computer Science and Engineering in 1999. He has served both the software industry and academia at home and abroad during his thirteen years of professional career. He is member of ACM and IEEE. His main research interests include signal and speech processing, neural networks, software engineering, and bio-informatics.

Professor M. Lutfar Rahman obtained B.Sc. (Hons) and M.Sc. in Physics from University of Dhaka and MS and Ph.D. in Electronic and Electrical Engineering from University of Sheffield, UK. He has over 300 research papers and scientific and technical articles. He authored/co-authored eighteen books on Electronics and Computer Science and was awarded Halima Sharfuddin Science Writer Prize by Bangla Academy, Dhaka. Dr Rahman is the founder chairman of the Department of Computer Science and Engineering in University of Dhaka. He worked as the director of the Computer Center of the University of Dhaka. Before joining the Department of Computer Science and Engineering in 1992, Dr Rahman worked in the Department of Applied Physics and Electronics, University of Dhaka, Computer Department of the Higher Institute of Electronics initially located in Malta and later shifted to Libya and Atomic Energy Center, Dhaka. He has served as the expert member in different national committees.

Dr. Rahman is guiding several MS, PhD students in the department. He is also serving as the Dean, Faculty of Science s and Engineering at Daffodil International University of Dhaka.

Professor Farruk Ahmed obtained B.Sc. (Hons) and M.Sc. in Physics from University of Dhaka in 1966 and D.T.S (Electronics) from University of Manchester and Ph.D. in Electrical Engineering from Salford, UK in 1979. He started his teaching career in 1967 at University of Dhaka and became Professor and Chairman of the Department of Applied Physics and Electronics in 1993. He is engaged in teaching and research with various branches of Electronic Engineering and Computer Technology. He has guided and supervised more than 60 projects including M.Sc., M.Phil and Ph.D. theses and innumerable number of undergraduate projects. He has published nearly 150 papers in different journals of repute in the respective fields. He has served as the expert in different national committees. He worked as the founder and Honorary Advisor and Dean of Computer Science and Engineering department of Darul Ihsan University.

Dr. Farruk is now Professor in the Department of Computer Science and Engineering at North South University.