

Pitch Estimation of Devnagari Vowels using Cepstral and Autocorrelation Techniques for Original Speech Signal

S. S. Nimbhore

Department of CS & IT
Dr.B.A.M.University, Aurangabad
(MS), India.

G. D. Ramteke

School of Computer Sciences
North Maharashtra University,
Jalgaon (MS), India.

R. J. Ramteke

School of Computer Sciences
North Maharashtra University,
Jalgaon (MS), India.

ABSTRACT

This paper deals with pitch estimation of spoken Devnagari vowels from the original speech signals. Devnagari vowels are playing the vital role in pronunciation of any word. Each vowel is classified as starting, middle and end according to the duration of occurrences in the word. The Devnagari script having 12-vowels and 34-consonants are used in some Indian language like Marathi. The Devnagari vowels are categorised into 5-types such as short vowels, long vowels, conjunct vowels, nasal vowel and visarg vowel. The Pitch frequency is estimated from the features of speech signals via pitch detection algorithm through autocorrelation and cepstral methods. These vowels are recorded through PRAAT tools with noisy environment. The pitch estimation of original pitch frequency has been calculated in statistical manner (Mean and standard deviation). The implementation, experiments and result discussions are also existence. These results which have been appropriate match with both techniques.

General Terms

Human Computer Interface, Estimation Techniques, Devnagari Vowels, Pitch Estimation, Cepstral and Autocorrelation Methods.

Keywords

Short vowels, Long vowels, Conjunct vowels, Nasal and Visarg vowel. Feature extraction, Pitch Analysis, HCI, IPA, Speech Synthesis.

1. INTRODUCTION

Pitch extraction is also called fundamental frequency estimation which has been a popular topic in many fields of research such as speech recognition, speech synthesis, speech processing, speaker identification etc. Pitch detection is an essential task in most speech synthesis techniques. Pitch is defined as the apparent period of a signal. The Devnagari vowels are choice for the implementation of such natural text to speech synthesis [1].

The Devnagari vowels cannot pronounce two ways but it can be pronounced only one way e.g. अ . Devnagari 12-vowels are classified with the phonetic transcription structure of phonemes according to organ used in produce the sound.

Devnagari is based on phonetics principles which are considered as Place of articulation (POA) vowels.[2] These Devnagari vowels speech signals are estimated in noisy environment (original signals) for analysis and synthesis. The original speech signals are unbalanced to adjustment of a

pitch interval with help of some feature extraction techniques.[3]

The initial objective is to estimating the pitch of Devnagari vowels with noisy environments speech signals. These original speech signals are extracting the features based on cepstral and autocorrelation techniques.

The paper is summarized as follows. Section-1 contains the Vowels which are used in some Indian languages like Marathi. These 12- Devnagari vowels have been discussed. The Section-2 describes speech signals of 12-Devnagari vowels are deals with IPA structure like short, long, conjunct, nasal and visarg vowels. Devnagari vowels are represented with its modifier. These modifiers are helpful to write a meaningful word in Devnagari script. Section-3 describes the pitch detection and estimation based on two popular methods Cepstral and autocorrelation which are spoken devnagari vowels by different speakers from original speech signals. The section-4 deals with creating standard database through PRAAT tool. The section-5 describes experimental works. The section-6 deals with result discussion. Finally paper is summarized with conclusion and future works.

2. DEVNAGARI VOWELS

The 12-Devnagari vowels are categorised as per IPA (International Phonetics Association) as shown in Table-2. These are used for the speech analysis and synthesis purpose. It describes in different categories such as follows:

2.1 Short vowels (ह्रस्व)

The short vowel is a single vowel (V) in a short word or syllable, that vowel usually makes a short sound. These short vowels usually appear at the beginning of the word or between two consonants.

E.g. the short vowels represent character अ in Marathi अननस and in Hindi अनार. The production of word starting with modifiers as shown in Table-3.

2.2 Long vowels (दीर्घ)

The long vowels a short word or syllable ends with a vowel-consonant (VC). The long vowels usually the combination of e (a-k-e) in Marathi आई. The 'e' at the end of the word is silent. Long vowels when the word or syllable has a single vowel and the vowel appears at the end of the word or syllable, the vowels usually represent makes the long sound in Marathi नाई.

2.3 Conjunct vowels (संयुक्त)

The conjunct vowels are combination of short and long vowels. These phonemes are produced in Marathi e.g. अ + इ = ए as shown in Table-2.

2.4 Nasal vowel (अनुनासिक)

A nasal vowel is produced with a low tune so that air pressure through nose as well as mouth. The term "nasal" is slightly air pressure which does not come exclusively out of the nose in nasal vowels.

2.5 Visarg vowel (विसर्ग)

The Visarg symbol is used rarely in devnagari. The visarg is pronounced as the voiceless ह sound after the vowels. E.g. in Marathi आह [4].

3. PITCH DETECTION AND ESTIMATION

Pitch detection is used to detect the fundamental frequency in a speech signal. The fundamental frequency (F0) is called as pitch. The human pitch lies in the different intervals as shown in Table-I. The pitch frequencies are for men normally around 150Hz, for women 250Hz and children even higher frequencies. Pitch is determined by the frequency pattern of the periodicities in signals and the pitch is needed to construct this part of the speech (POS) signals [5]. The cepstral detector is the Fast Fourier Transform of a pitch signal. It usually has a number of regular peaks that is representing the harmonic spectrum of speech signals. When log magnitude of a spectrum is taken, then peak amplitude values are reduced and it is usable in spectrum [6]. These results are periodic waveform in the frequency domain. Where the period is related to the fundamental frequency of the original speech signal. It means that Fourier transformation of this waveform representing the fundamental frequencies (F0) [7].

3.1 Cepstral Coefficient Method

Cepstral method provides mode for detection of the pitch. In frequency domain, the complexity associated becomes multiplication bond. The cepstral pitch detector is based on the log of the Fourier transformations. Then, using the property of log function;

$$\log AB = \log A + \log B \quad \text{Eq. (1)}$$

Finally, the real cepstrum of a speech signals are carried out through the experiment using mat lab.

$$\left. \begin{aligned} H(w) &= H_g(w)H_f(w) \\ \text{cepstrum} &= \text{IFFT}(\log \text{FFT}(S)) \end{aligned} \right\} \text{Eq. (2)}$$

The cepstrum is a Fourier analysis of the logarithmic amplitude spectrum of the signal. If the log amplitude spectrum contains many regularly spaced harmonics, then the Fourier analysis of the spectrum will show as peak corresponding spacing between the fundamental frequencies (F0). [8]

3.2 Autocorrelation Method

Autocorrelation is used in signal processing for analyzing the series of values, such time domain signals. Periodic waveforms exhibit an interesting autocorrelation characteristic. The autocorrelation function itself is periodic.

The time lag increases to half period of the waveform. The autocorrelation method it indicate a period of the waveform. The pitch detection algorithm is used as short-term analysis in autocorrelation. The autocorrelation is also periodic with the same period. The periodicity in the autocorrelation function indicates periodicity in the signal. The way of estimated the fundamental frequency for various speech waveforms is based on the autocorrelation Eq. (4) which is applied for estimating each pitch frequency. The autocorrelation function of a speech signal is basically transformation of the signal that is useful for displaying structure in the waveform [9, 10].

$$R[m] = \frac{1}{N} \sum_{n=0}^{N-1-m} (w[n]x[n]w[n+m]x[n+m]) \quad \text{Eq.(3)}$$

Where w[n] is window function of length N for the random process in Eq. (3).

In the Eq. (3) input speech signal is quasi-waveforms which has maxima of pitch period and its waveforms.

$$\left. \begin{aligned} x[n] &= \cos(\omega_0 n + \Phi) \\ R[m] &= E\{x^*[n]x[n+m]\} \\ R[m] &= \frac{1}{2} \cos(\omega_0 n) \end{aligned} \right\} \text{Eq. (4)}$$

In Equation (4) x [n] - the input speech signal with n=0 to N. N - Number of sample. ω_0 - Fundamental angle frequency.

4. DATABASE ACQUISITION

In this paper speech database acquisition of Marathi vowels have been written in Devnagari script. Which are spoken by various genders (3-Male and 2-Female), these speakers' age is between 19 to 23 years and sampling speech frequency is 20 KHz (22,050 Hz). The speech database of Devnagari vowels from the original speech signals is recorded through standard PRAAT tool. The recorded Devnagari vowels speech database size is 60.

5. EXPERIMENTAL WORK

In this experiment, the numbers of utterances with original speech in noisy environment are used. The experimental work is done by using mat lab tool. Pitch estimation of spoken Devnagari vowels from the original speech signals is based on autocorrelation and cepstral methods. Both methods are used for detecting of pitch which is found pitch listing and number of undefined signal. Undefined signal is closest to zero hertz. Those pitch reading compared with standard Praat tool pitch listing. The input speech samples or spoken Devnagari vowels are delivered to the pitch frequency estimation from original speech signals which are accepted through pitch detection algorithm. These original speech frequency transitions using a hamming window apply to each frame. Autocorrelation method and cepstral method generating the smoothed pitch track of devnagari vowels.

It shows the one type of vowel in Fig.1 [A] shows original speech signal waveforms of short vowel अ . Fig.1 [B] shows the estimated pitch track frequency waveform of original speech signal using autocorrelation method. Fig.2 [A] shows original speech signal pitch waveform using cepstral method of short vowel अ . Fig.2 [B] shows the estimated pitch track frequency waveform of original speech signal using cepstral method of short vowel अ is varied with

triangular waveform. Pitch estimation of spoken Devnagari vowels from the original speech signals through pitch detection algorithm is applied to all spoken utterances.

6. RESULT AND DISCUSSION

Results are discussed with one devnagari vowel result e.g. short vowel अ standard PRAAT tool reading of Male-1 is 151.90. Short vowel autocorrelation reading of Male-1 is 148.66. Short vowel cepstral reading of Male-1 is 167.99. It means that experimental results are approximately matched with PRAAT tool. Autocorrelation reading is almost reaching of standard PRAAT tool but cepstral reading is sometime near of standard PRAAT tool. It is not easy way to find out undefined speech signal in cepstral method. But undefined speech signal can easily find out in autocorrelation method.

Table-4.(a) to Table-4.(e) results are showing the calculated mean and standard deviation of estimated pitch frequency from original speech signals of Devnagari vowels such as Short, Long, Conjunct, Nasal and Visarg respectively, based on autocorrelation and cepstral Method. It observed that mean reading of short vowel अ in PRAAT tool is 210.06 and standard deviation reading is 66.58. The mean reading of short vowel अ is 204.26 in autocorrelation method and standard deviation is 62.90. The mean reading of short vowel अ is 210.04 in cepstral method and standard deviation is 54.43. The experimental results are comparatively shown in various tables.

7. CONCLUSIONS AND FUTURE WORKS

In autocorrelation methods the original pitch frequency values are approximately matched with PRAAT tools. This experimental work throughout applied in all speech databases. Standard statistical methodologies are applied in all original speech signals of Devnagari vowels. Estimating the pitch frequencies which have been calculated statistical manner through Mean and Standard Deviation. In this throughout experimental works the PRAAT tool pitch frequency results are compared with autocorrelation and cepstral techniques. These results which have been appropriate match with both techniques.

In future work an attempt will be made to estimate the speech frequency using LPC and MFCC of Devnagari consonants, spoken by the speakers with noisy environments.

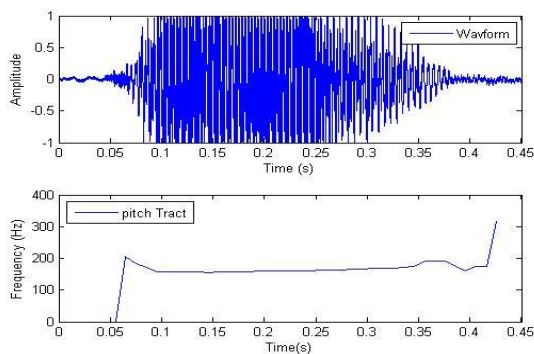


Fig. 1: A) Original speech signal waveform using autocorrelation method of short vowel अ
B) Original speech signal pitch tracking using autocorrelation method of short vowel अ

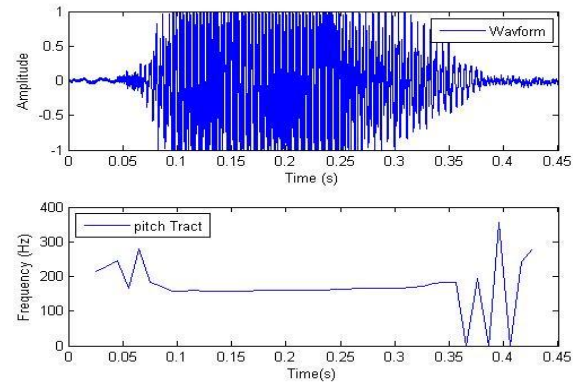


Fig. 2: A) Original speech signal pitch waveform using cepstral method of short vowel अ
B) Original speech signal pitch track using cepstral method of short vowel अ

8. ACKNOWLEDGMENTS

The authors are grateful to University Grants Commission (UGC) for funding to Major Research Project (Ref. No.: 39-121/2010(SR)). We would like to thank our Prof. Dr. S. C. Mehrotra (UGC-BSR Fellow), Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS-India) for numerous discussions and Encouragement.

9. REFERENCES

- [1] X. Huang and A. Acero, H. Hon, "Spoken Language Processing: A Guide to Theory, Algorithm, and System Development", published by Prentice Hall, 2001.
- [2] Alina Nica, Alexandra Caruntu et. al., "Analysis and Synthesis of vowels Using Mat lab", IEEE proceeding, Vol. 2 PP. 371-374, May-2006.
- [3] M. Habibullah Pagarkar, Lakshmi Gopalakrishnan, et.al. "Language Independent Speech Compression using Devnagari Phonetics", 2002.
- [4] R.K. Bansal, and J.B. Harrison, "Spoken English for India, A Manual of Speech and Phonetics", published by Orient Longman, 1972.
- [5] Rashmi Makhijani, Urmila Shrawankar, Dr. V. M. Thakare, "Speech Enhancement using Pitch Detection Approach for Noisy Environment", International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 2 PP. 1764-1769, Feb-2011.
- [6] Hiroshi Shimodaira and Mitsuru Nakai, "Robust Pitch Detection By Narrow Band Spectrum Analysis", Second International Conference on Spoken Language Processing (ICSLP'92) PP. 1597-1600 Banff, Alberta, Canada: Oct-1992.
- [7] A. D. Cheveigne and H. Kawahara, "A fundamental frequency estimator for speech and music" Journal of the Acoustical Society of America, PP. 1917-1930, April-2002.
- [8] Mehdi S. Javadi, "Pitch Period of Speech Signals Introduction", Determinations and Changing".
- [9] Li Tan and Montri Karnjanadecha, "Pitch Detection Algorithm: Autocorrelation Method and AMDF",

Intelligent Signal Processing and Communication Systems (ISPACS), PP.453 – 456 King Mongkut's Univ. of Technol. Thonburi, Bangkok, Thailand: Jan-2009.

- [10] Jong-soon Jung, Jeong-jin Kim and Myung-jin Bae, "Pitch Alternation technique in speech synthesis system", IEEE Transaction on Acoustics, Speech and Signal Processing, PP. 163-167 Vol. 47 No.1, Feb 2001.

Table 1. Range of Human Speech

Gender	Fundamental frequency (F0) Min Hz	Fundamental frequency (F0) Max Hz
Male	80	200
Female	150	350

Table 2.12-Devnagari vowels classified into five types

Types of Devnagari Vowels	1	2	3	4
Short	अ	इ	उ	-
Long	आ	ई	ऊ	-
Conjunct	अ+इ=ए	अ+ई=ऐ	अ+उ=ओ	अ+उ=औ
Nasal	अं	-	-	-
Visarg	अः	-	-	-

Table 3. Devnagari vowels represents with modifiers

Vowels	आ	इ	ई	उ	ऊ	ऋ	ॠ	ए	ऐ	ओ	औ
Modifiers	ा	ि	ी	ु	ू	ृ	ँ	े	ै	ॉ	ो

Short Vowels

अ	Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	151.90	148.66	167.99	
Male -2 (20)	180.86	177.80	177.83	
Male-3 (20)	155.18	151.87	167.30	
Female-1 (19)	294.58	282.76	282.24	
Female-2 (19)	267.58	260.20	254.83	
Mean	210.06	204.26	210.04	
SD	66.58	62.90	54.43	

इ	Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	161.88	160.03	161.19	
Male -2 (20)	174.40	172.10	192.04	
Male-3 (20)	160.04	159.88	176.15	
Female-1 (19)	278.50	272.78	270.29	
Female-2 (19)	269.90	255.25	261.82	
Mean	208.94	204.01	212.30	
SD	59.90	55.35	50.36	

उ	Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	160.56	161.88	163.68	
Male -2 (20)	182.90	156.25	217.65	
Male-3 (20)	143.24	138.90	180.15	
Female-1 (19)	289.93	274.16	271.57	
Female-2 (19)	272.85	230.83	226.45	
Mean	209.90	192.40	211.90	
SD	67.03	57.58	42.24	

Long Vowels

आ	Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	127.71	154.08	166.67	
Male -2 (20)	166.36	169.63	185.40	
Male-3 (20)	192.81	171.41	186.27	
Female-1 (19)	279.42	267.45	268.46	
Female-2 (19)	237.58	234.46	218.65	
Mean	200.77	199.41	205.09	
SD	59.43	48.95	40.06	

ई	Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	172.31	170.60	179.62	
Male -2 (20)	173.24	170.61	173.12	
Male-3 (20)	168.07	164.10	176.47	
Female-1 (19)	273.44	268.26	249.87	
Female-2 (19)	259.55	239.52	236.23	
Mean	209.32	202.62	203.06	
SD	52.46	47.97	36.89	

ऊ			
Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	159.33	156.68	173.56
Male -2 (20)	148.11	147.99	180.11
Male-3 (20)	143.77	143.14	162.51
Female-1 (19)	273.35	268.24	235.76
Female-2 (19)	250.21	252.21	256.46
Mean	194.95	193.65	201.68
SD	61.81	61.23	41.69

Conjunct Vowels

अ + इ = ए			
Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	184.97	187.33	178.39
Male -2 (20)	169.52	175.78	171.52
Male-3 (20)	173.92	170.40	184.62
Female-1 (19)	282.61	275.08	262.99
Female-2 (19)	291.61	245.08	236.51
Mean	220.52	210.74	206.81
SD	61.12	46.68	40.57

अ + ई = ऐ			
Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	165.63	168.93	183.64
Male -2 (20)	191.63	171.52	183.78
Male-3 (20)	176.84	166.39	194.17
Female-1 (19)	274.68	268.13	254.82
Female-2 (19)	285.73	242.81	244.51
Mean	218.90	203.56	212.18
SD	56.85	48.26	34.67

अ + उ = औ			
Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	146.08	149.00	189.53
Male -2 (20)	128.39	128.46	164.40
Male-3 (20)	137.25	135.53	167.57
Female-1 (19)	252.97	248.01	241.50
Female-2 (19)	243.09	237.74	227.51
Mean	181.56	179.75	198.10
SD	61.11	58.21	34.96

अ + उ = औ			
Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	171.45	169.72	171.79
Male -2 (20)	191.89	156.67	190.48
Male-3 (20)	165.59	163.21	190.39
Female-1 (19)	273.51	267.90	268.23
Female-2 (19)	265.16	232.94	231.53
Mean	213.52	198.09	210.48
SD	51.96	49.56	38.98

Nasal Vowel

अं			
Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	131.19	135.51	171.97
Male -2 (20)	146.52	145.14	160.56
Male-3 (20)	138.97	140.78	175.76
Female-1 (19)	280.89	274.96	266.89
Female-2 (19)	240.87	236.85	230.69
Mean	187.69	186.65	201.17
SD	68.51	64.74	45.66

Visarg Vowel

अः			
Speakers with Age	Praat Tool in Hz	Auto Correlation Method in Hz	Cepstral Method in Hz
Male -1 (21)	136.62	138.71	173.31
Male -2 (20)	126.39	127.38	181.54
Male-3 (20)	137.19	136.48	166.27
Female-1 (19)	277.74	266.76	245.73
Female-2 (19)	222.58	208.77	215.41
Mean	180.10	175.62	196.45
SD	67.00	60.47	33.38

Table-4. (a) to (e): which are showing the calculated mean and standard deviation of estimated pitch frequency from spoken Devnagari vowels such as Short, Long, Conjunct, Nasal and Visarg respectively, based on autocorrelation and cepstral Method.

Table 4 (a). Short Vowels Pitch Frequency

Short Vowels Methods	अ		इ		उ	
	Mean	SD	Mean	SD	Mean	SD
Praat Tool in Hz	210.06	66.58	208.94	59.90	209.90	67.03
Auto Correlation Method in Hz	204.26	62.90	204.01	55.35	192.40	57.58
Cepstral Method in Hz	210.04	54.43	212.30	50.36	211.91	42.24

Table 4 (b). Long Vowels Pitch Frequency

Long Vowels Methods	आ		ऋ		ऌ	
	Mean	SD	Mean	SD	Mean	SD
Praat Tool in Hz	200.77	59.43	209.32	52.45	194.95	61.81
Auto Correlation Method in Hz	199.41	48.95	202.62	47.97	193.65	61.23
Cepstral Method in Hz	205.09	40.06	203.06	36.89	201.68	41.69

Table 4 (c). Conjunct Vowels Pitch Frequency

Conjunct Vowels Methods	अ+इ=ए		अ+ई=ऐ		अ+उ=ओ		अ+र=औ	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Praat Tool in Hz	220.52	61.12	218.90	56.85	181.56	61.11	213.52	51.96
Auto Correlation Method in Hz	210.74	46.68	203.56	48.26	179.75	58.21	198.09	49.56
Cepstral Method in Hz	206.81	40.57	212.18	34.67	198.10	34.96	210.48	38.98

Table 4 (d). Nasal Vowel Pitch Frequency

Nasal Vowel Methods	अं	
	Mean	SD
Praat Tool in Hz	220.52	61.12
Auto Correlation Method in Hz	210.74	46.68
Cepstral Method in Hz	206.81	40.57

Table 4 (e). Visarg Vowel Pitch Frequency

Visarg Vowel Methods	अः	
	Mean	SD
Praat Tool in Hz	213.52	51.96
Auto Correlation Method in Hz	198.09	49.56
Cepstral Method in Hz	210.48	38.98