Contemporary Engineering Sciences, Vol. 2, 2009, no. 10, 497 - 505

Speech Processing of the Letter 'zha' (ழ) in Tamil Language with LPC

A. Srinivasan[1], K. Srinivasa Rao[2], D. Narasimhan[3] and K. Kannan[4]

[1]Department of Electronics and Communication Engineering,
[2]DST Chair,
[3,4]Department of Mathematics,

Srinivasa Ramanujan Centre, SASTRA University,
Kumbakonam - 612 001 India
E-Mail :[1]asrinivasan78@yahoo.com
[3]dnsastra@rediffmail.com and [4]anbukkannan@rediffmail.com

Abstract : Wideband speech signals of the letter 'zha' (ழ) in Tamil language of 3 males and 3 females were coded using an improved version of Linear Predictive Coding (LPC). The sampling frequency was at 16 kHz and the bit rate was at 15450 bits per second, where the original bit rate was at 128000 bits per second with the help of wave surfer audio tool. The quality of the performance is exhibited through the block diagram voice coder. In the last section, the tradeoffs between the bit rate for a plain LPC vocoder and the bit rate for a voice-excited LPC vocoder with DCT is analyzed.

Keywords : Speech processing, LPC, Tamil Language, Letter 'zha' (ழ), Wavesurfer.

1 Introduction

Tamil is one of the oldest and official languages in India. In Tamilnadu it is the prominent and primary language. It is one of the official languages of the union territories of Pondicherry and Andaman & Nicobar Islands. It is one of 23 nationally recognised languages in the Constitution of India. It has official status in Sri Lanka, Malaysia and Singapore. The art and architecture of the Tamil people encompasses some of the notable contributions of India and South-East Asia to the world of art. With more than 77 million speakers, Tamil is one of the widely spoken languages of the world.

Tamil vowels are classified into short, long (five of each type) and two diphthongs.

Consonants are classified into three categories with six in each category: hard, soft (a.k.a nasal), and medium.

The classification is based on the place of articulation. In total there are 18 consonants. The vowels and consonants combine to form 216 compound characters. Placing dependent vowel markers on either one side or both sides of the consonant forms the compound characters.

There is one more special letter *āytham* used in classical Tamil and rarely found in modern Tamil. In total there are 247 letters in Tamil alphabet. In these 247 letters 'Zha' is the most significant, because of its usage and pronunciation. Many people will not pronounce the letter 'Zha' properly. There are two letters with same sound as 'Zha' (la, lla), so it is necessary to recognize the letter 'Zha'.

2 Vowels and Consonants

2.1 Vowels

There are 12 vowels in Tamil, called uyireluttu (uyir - life, *e<u>l</u>uttu* - letter). These vowels are classified into short (*ku<u>r</u>il*) and long (five of each type) and two diphthongs, /ai/ and /au/, and three "shortened" (*ku<u>rr</u>iyl*) vowels.The long vowels are about twice as long as the short vowels. The diphthongss are usually pronounced about 1.5 times as long as the short vowels.

	SHORT			LONG		
	Front	**Central**	**Back**	**Front**	**Central**	**Back**
Close	i		u	i:		i:
	இ		உ	ஈ		ஊ
Mid	E		o	e:		o:
	எ		ஒ	ஏ		ஓ
Open		A		(ai)	a:	(aw)
		அ		ஐ	ஆ	ஔ

Table 1: Vowels in Tamil

2.2 Consonants

Consonants are known as *meyye<u>l</u>uttu* (mey-body, *e<u>l</u>uttu*-letters) in Tamil. It is classified into three categories with six in each category: *valli<u>n</u>am* (hard), *melli<u>n</u>am* (soft or Nasal) and *i<u>t</u>ayi<u>n</u>am* (medium).Unlike most Indian languages, Tamil does not distinguish aspirated and unaspirated consonants. In addition, the voicing of plosives is governed by strict rules in *centami<u>l</u>* (Pure Tamil). Plosives are unvoiced if they occur word-initially or doubled. Elsewhere they are voiced, with a few becoming fricatives intervocalically. Nasals and approximants are always voiced.

As common place in languages of India, Tamil is characterised by its use of more than one type of coronal consonants. Retroflex consonants include the retroflex approximant /‘zha’ (ழ)/ (H) (example Tamil), which among the Dravidian languages is also found in Malayalam (example Kozhikode), disappeared from Kannada in pronunciation at around 1000 AD (the dedicated letter is still found in Unicode), and was never present in Telugu. Dental and alveolar consonants also contrast with each other, a typically Dravidian trait not found in the neighboring Indo-Aryan languages. In spoken Tamil, however, this contrast has been largely lost, and even in literary Tamil, e and d may be seen as allophonic.

A chart of the Tamil consonant phonemes in the International Phonetic Alphabet follows. Phonemes in brackets are voiced equivalents. Both voiceless and voiced forms are represented by the same character in Tamil, and voicing is determined by context. The sounds /f/ and /§/ are peripheral to the phonology of Tamil, being found only in loanwords and frequently replaced by native sounds. There are well-defined rules for elision in Tamil, categorised into different classes based on the phoneme which undergoes elision.

	Labial	Dental	Alveolar	Retroflex	Palatal	Velar
Plosives	p (b)	t̪ (d̪)		ʈ (ɖ)	tʃ (dʒ)	k (g)
	ப	த		ட	ச	க
Nasals	M	n	N	ɳ	ɲ	Ŋ
	ம	ந	ன	ண	ஞ	ங
Tap		ɾ̪				
		ர				
Trill			R			
			ற			
Central Approximants	ʋ			ɻ	j	
	வ			ழ	ய	
Lateral Approximants		l̪		ɭ		
		ல		ள		

Table 2: Consonants in Tamil

3 Special letter- $\bar{A}$ytam

Classical Tamil also had a phoneme called the $\bar{A}$ytam. Tamil grammarians of the time classified it as a dependent phoneme (or restricted phoneme) (*cārpeḻuttu*), but it is very rare in modern Tamil. The rules of pronunciation given in the *Tolkāppiyam*, a text on the grammar of Classical Tamil, suggest that the *āytam* could have glottalised the sounds it was combined with. It has also been suggested that the *āytam* was used to represent the voiced implosive (or closing part or the first half) of geminated voiced plosives inside a word.The *Āytam*, in modern Tamil, is also used to convert pa to fa (not the retroflex zha [ℓ]) when writing English words using the Tamil script.

4 LPC Vocoder

The detailed LPC speech coding technique is viewed the specific modifications, additions are given to improve this algorithm. However, before jumping into the detailed methodology of our solution, it will be helpful to give a brief overview of speech production. Sounds of speech are produced when velum is lowered to make it acoustically coupled with the vocal tract. Nasal sounds of speech are produced in this way. Speech signals consist of several sequences of sounds. Each sound can be thought of a unique information. Generally the speech sounds are classified into two types namely, voiced and unvoiced. The fundamental difference between these two types of speech sounds comes from the way they are produced. The vibrations of the vocal cords produce voiced sounds. The rate at which the vocal cords vibrate dictates the pitch of the sound. On the other hand, unvoiced sounds do not rely on the vibration of the vocal cords. The unvoiced sounds are created by the constriction of the vocal tract. The vocal cords remain open and the constrictions of the vocal tract force air out to produce the unvoiced sounds.

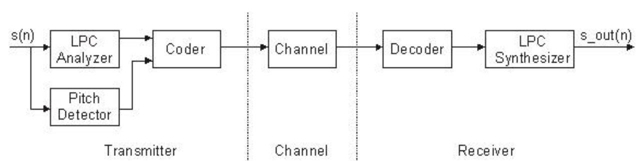

Figure 1: LPC Vocoder

LPC technique will be utilized in order to analyze and synthesize speech signals. This technique is used to estimate the basic speech parameters like pitch, formants and spectra. A block diagram of an LPC vocoder can be seen in Fig.1. The principle behind the use of LPC is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. This could be used to obtain a unique set of predictor coefficients. These predictor coefficients are normally estimated in every frame, which is normally 20 ms long. The predictor coefficients are represented by a_k. The synthesis filter takes the error signal as an input and it is filtered and the output is the speech signal. The transfer function of the time-varying digital filter is

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

where, G is gain. For LPC-10 algorithm p is 10 and p is 18 for the improved algorithm. The two most commonly used methods to compute the coefficients are, the covariance method and the auto-correlation formulation. For our implementation, we will be using the auto-correlation formulation. However, if the frame is unvoiced, then white noise is used to represent it and a pitch period of T=0 is transmitted. Therefore, either white noise or impulse train becomes the excitation of the LPC synthesis filter. It is important to re-emphasize that the pitch, gain and coefficient parameters will be varying with time from one frame to another.

5 Analysis of 'Zha' using WaveSurfer

WaveSurfer is a simple but powerful interface. The sound can be visualized and analyzed in several ways with the help of this tool. In addition, a spectrum window can be opened using Popup Spectrum Section for analyze Spectrum section plot (Magnitude Vs Frequency). Further the special control windows are available for Waveforms and Spectrograms, which allow the user to make quick modifications such as sound edit, noise elimination etc. The basic document we work with is sound files of 3 male and 3 female speakers with letter 'zha'. The standard speech analysis of the letter 'zha' such as Waveform, Spectrogram, Pitch, and Power panes are analyzed and the samples are shown in following figures.

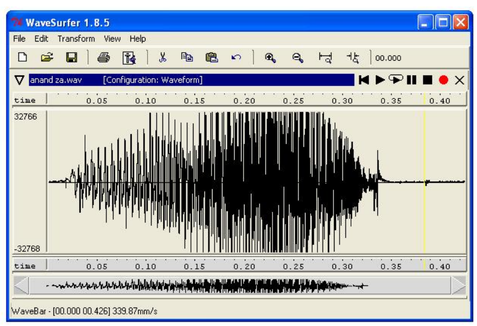

Figure 2: Waveform of letter 'zha' (ழ)

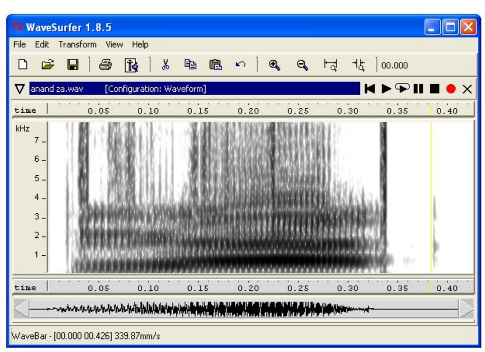

Figure 3: Spectrogram of letter 'zha' (ழ)

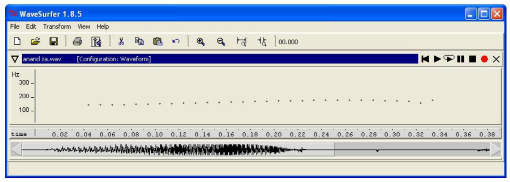

Figure 4: Pitch panes of letter 'zha' (ழ)

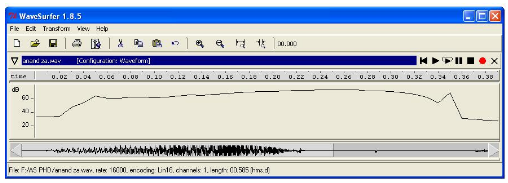

Figure 5: Power panes of letter 'zha' (ழ)

6 Experimental Results

The variation of the Spectrum section plot in LPC is measures for the letter 'zha' is analyzed with the following parameters and the sample of magnitude Vs frequency plot is shown in figure 6 and figure 7.

- Analysis type: LPC
- Analysis order: 20
- Speech signal bandwidth B = 8 kHz
- Sampling rate Fs = 16000 Hz (or samples/sec.)
- Channel: All
- Window type: Hamming
- Window length (frame): 512 points (20ms)
- Number of predictor coefficients of the LPC model = 18

6.1 Spectrum plot

Sample spectrum section plot of letter 'zha' (ழ)

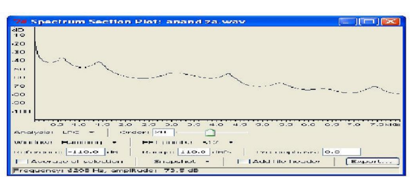

Figure 6: Sample 1

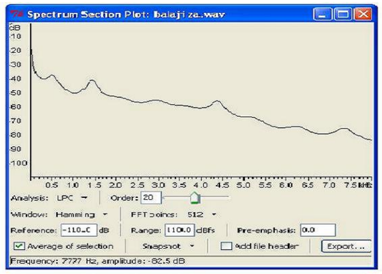

Figure 7: Sample 2

6.2 Magnitude and frequency comparison of 3 male and 3 female speakers

Sl.No.	Frequency (Hz)	†F1 (dB)	†F2 (dB)	†F3 (dB)	††M1 (dB)	††M2 (dB)	††M3 (dB)
1	15.625	-20.47	-20.98	-19.67	-21.02	-21.86	-20.71
2	140.625	-32.68	-32.73	-32.01	-32.94	-33.13	-32.75
3	390.625	-36.63	-36.74	-36.13	-36.94	-37.14	-37.01
4	640.625	-41.36	-34.99	-40.97	-41.48	-41.92	-40.99
5	1015.625	-51.20	-47.91	-51.00	-51.90	-52.65	-50.90
6	2046.875	-51.68	-53.12	-51.02	-52.02	-52.98	-53.05
7	3140.625	-56.56	-58.12	-56.10	-57.00	-57.60	-56.28
8	4203.125	-61.15	-68.51	-61.12	-61.66	-61.98	-63.22
9	5234.375	-73.12	-70.22	-72.97	-73.92	-74.48	-73.62
10	5703.125	-72.42	-68.77	-72.27	-72.95	-73.25	-72.99
11	6140.625	-76.70	-71.34	-76.61	-77.13	-77.89	-77.69
12	7171.875	-73.45	-72.08	-73.39	-73.99	-74.19	-73.16
13	7953.125	-84.67	-84.46	-84.51	-85.23	-86.00	-85.92
14	7984.375	-84.74	-84.52	-84.57	-85.56	-86.02	-85.83

Table 3: Magnitude and frequency comparison of 3 male and 3 female speakers. † Female, †† Male

6.3 Bit rates

The bit rate for a plain LPC vocoder and the bit rate for a voice-excited LPC vocoder with DCT is calculated and shown in Table 4 and Table 5.

Void	Number of bits per frame
Predictor coefficients	18 * 8 = 144
Gain	5
Pitch period	6
Voiced/unvoiced switch	1
Total	156
Overall bit rate	50 * 156 = 7800 bits / second

Table 4: Bit rate for plain LPC vocoder

Void	Number of bits per frame
Predictor coefficients	18 * 8 = 144
Gain	5
DCT coefficients	40 * 4 = 160
Total	309
Overall bit rate	50 * 309 = 15450 bits / second

Table 5: Bit rate for voice-excited LPC vocoder with DCT

7 Conclusion

It is observed from voice excited LPC with Wavesurfer tool, there is a variation in magnitude of the letter 'Zha' among different people. To strengthen the results, more samples could collected from TamilNadu, Srilanka and Malasiya infuture. For this analysis, synthesis and numerous simulations are needed. The synthesis is based on Hidden Markov Model(HMM). Further research may be carried out by using HMM.

References

[1] B. S. Atal, M. R. Schroeder, and V. Stover, "Voice-Excited Predictive Coding Systetm for Low Bit-Rate Transmission of Speech", *Proc. ICC*, pp.30–37 to 30–40, 1975.

[2] Daniel Jurafsky, James H. Martin, "Speech and Language Processing", *Pearson education*, (ISBN 8178085941), 2002.

[3] Harold F. Schiffman, "A Reference Grammar of Spoken Tamil", *Cambridge University Press* (ISBN-10: 0521027527), 2006.

[4] B. H. Juang; L. R. Rabiner, "Hidden Markov Models for Speech Recognition", *Technometrics*, Vol. 33, No. 3. (Aug., 1991), pp. 251-272.

[5] L. R. Rabiner and R. W. Schafer, "Digital Processing of Speech Signals", *Prentice- Hall*, Englewood Cliffs, NJ, 1978.

[6] J. Srinonchat, "New Technique to Reduce Bit Rate of LPC- 10 Speech Coder", *IEEE Transaction on Audio and Speech Language processing*, Sweden, Sep 2006.

[7] G.F. Sudha and S. Karthik, "Improved LPC Vocoder using Instantaneous Pitch Estimation Method", *International Journal of Wireless Networks and communications*, Vol. 1, No. 1 (2009), pp. 43-54.

[8] R. Thangarajan, A.M. Natarajan, M. Selvam, "Word and Triphone Based Approaches in Continuous Speech Recognition for Tamil Language", *Wseas Transactions on Signal Processing*, Vol. 4, No. 3 March 2008.

[9] C. J. Weinstein, "A Linear Predictive Vocoder with Voice Excitation", *Proc. Eascon*, September 1975.

[10] Wavesurfer manual: http://www.speech.kth.se/wavesurfer/man.html

[11] Wavesurfer Tool: http://mac.softpedia.com/get/Audio/WaveSurfer.shtml

Received: August, 2009