ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/102972, October 2016*

Hindi Vowel Classification using QCN-PNCC Features

Shipra[1]* and Mahesh Chandra[2]

[1]Electronics and Communication Engineering Department, BIT Mesra, Near Patna Airport, Patna - 800014, Bihar, India; shipra@bitmesra.ac.in
[2]Electronics and Communication Engineering Department, BIT Mesra, Ranchi - 835215, Jharkhand, India; shrotriya@bitmesra.ac.in

Abstract

This paper present a novel hybridized QCN-PNCC features. These features are obtained by processing Power Normalized Cepstral Coefficients (PNCC) with Quantile based Dynamic Cepstral Normalization Technique (QCN). The robustness of the QCN-PNCC features is compared with PNCC features for the task of Hindi Vowel classification with HMM classifier for Context-Dependent and Context- Independent cases in clean as well as in noisy environment. It is observed that the recognition accuracy of QCN-PNCC features with Hidden Markov Model (HMM) as classifier exhibit an improvement of approximately 8% as compared to PNCC features for Hindi vowel classification task.

Keywords: Power normalized Cepstral Coefficient (PNCC), QCN, QCN-PNCC, Speech Recognition

1. Introduction

Last few decades have seen an explosive growth of technologies in the area of Automatic Speech Recognition (ASR). This remarkable development in the field of Speech recognition has emphasized the ever-present challenge of cancelling background noise in speech. Now a day the applications of ASR have reached to all spheres of day-to-day life including applications such as voice dialing, call routing, interactive voice response, data entry and dictation, voice command and control, appliance control by voice, computer-aided language learning, content-based spoken audio search, and robotics etc. Most of these applications are real-world applications where ASR system is required to work in difficult acoustic environments. In this paper we have evaluated newly proposed PNCC features[1] and we have also studied the effect of Quantile based cepstral dynamic normalization technique (QCN) [2] on the robustness of PNCC features. Power Normalized Cepstral Coefficient (PNCC) is a newly proposed feature extraction technique based on auditory processing. Several implementations of PNCC processing has been introduced[3,4] and evaluated which have proved better for speech recognition compared to other existing algorithms[5,6]; such as zero crossing peak amplitude, RASTA-PLP, Invariant Integration Features (IIF). In our previous work[7], we have processed MFCC features with Quantile based dynamic Cepstral normalization technique to obtain QCN-MFCC features and evaluated them for Hindi vowel classification task and found that QCN-MFCC features provide better recognition accuracy compared to MFCC features.

It has been observed that PNCC features are very much similar to MFCC features in their implementation. Hence we have processed PNCC features with QCN technique to obtain a new feature set that we called QCN-

*Author for correspondence

PNCC features. For evaluation purpose we have chosen the task of Hindi Vowel Classification with HMM classifier[8] in three basic classes; front vowel, mid vowel and back vowel for context-dependent as well as context-independent cases for clean as well as noisy environment. The rest of the paper is organized as follows; section 2 gives the details of the Hindi speech database. Section 3 gives an overview of feature extraction techniques. Section 4 briefly discusses the acoustic-phonetic features of Hindi language. The comparative recognition efficiency of the two feature sets for Hindi vowel classification is presented and discussed in sections 5, 6 and 7.

2. Hindi Speech Database

A Hindi speech database[9], designed at TIFR, Mumbai, India, is used to extract the phones for recognition. The database consists of ten phonetically rich sentences spoken by hundred speakers. Out of the ten sentences spoken by individual speakers, two sentences are common for all the speakers. These sentences cover most of the phonemes of the Hindi language. The database is prepared at CEERI, New Delhi, India. The recording of sentences was done with 16 kHz sampling frequency. Two microphones were used for recording purpose. One of them was a good quality close talking microphone and the other one was an omni-directional desk-mounted microphone which was kept at a distance of one meter from the speaker. The data was stored in the 16-bit PCM-encoded waveform format in mono-mode. In this database, phoneme boundaries are provided for each spoken sentence. Here we have worked with three Hindi vowel classes; front vowel, mid vowel and back vowel. Phonemes were extracted using the labels provided in the database. We have added three types of noises; babble, speech and lynx noise at SNRs of -5dB to 20dB to this database using Noisex-92 database. Thus we obtained a noisy database and a clean database for training and testing of the Hidden Markov Model based phoneme classifier.

3. Feature Extraction

3.1 Power Normalized Cepstral Coefficients (PNCC)

Power Normalized Cepstral Coefficients is a new feature extraction technique proposed by Chanwoo Kim et al. These features have proved to be very useful for real time applications. It is observed by Kim and Stern that PNCC features improve recognition efficiency in the presence of acoustically varying environments without compromising with the performance in clean environments. Many attributes of PNCC processing have been strongly influenced by human auditory processing. If we compare PNCC processing with MFCC processing; which is a very popular feature extraction technique based on human speech perception mechanism; then we may directly conclude that PNCC is very much similar to MFCC in implementation except for certain differences that improves the robustness of PNCC features. For example, in MFCC processing log nonlinearity is performed on Mel-filter bank output but in PNCC processing power-law nonlinearity is performed on Gammatone filter bank output which is chosen to approximate relation between signal intensity and auditory nerve firing rate. This approach improved robustness of features by suppressing small signals. In conventionally proposed PNCC processing medium-time processing with duration of 50-120 ms is used to analyse the parameters characterizing environmental degradation, in combination with the traditional short-time Fourier analysis with frames of 20-30 ms used in conventional speech recognition systems. It is observed that this approach lead to the estimation of environmental degradation more accurately while maintaining the ability to respond to rapidly changing speech signals. In our work we have not performed medium-time analysis. We have simply performed DCT followed by mean-normalization on the samples obtained after application of power law nonlinearity. It has been observed in our previous work that Quantile based Cepstral dynamic normalization technique with MFCC features has improved the recognition efficiency. Therefore, PNCC features are processed with QCN to obtain QCN-PNCC features. The results confirmed that the robustness of any

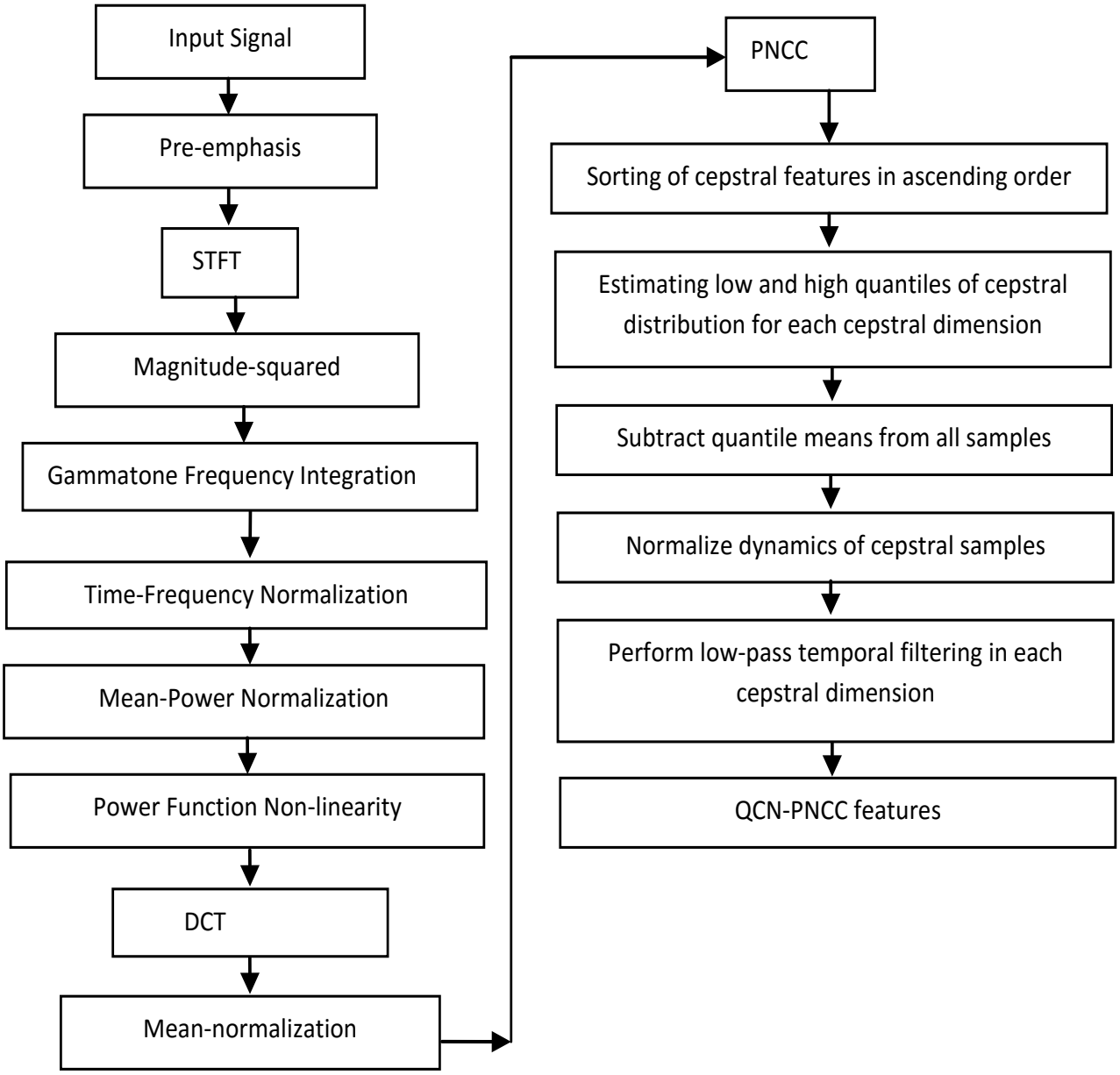

Figure 1. EEG Block diagram of PNCC and QCN-PNCC feature extraction technique.

Table 1. Hindi language Acoustic classes and their Phoneme members

Vowels	Front	इ	ई	ए	ऐ
	Middle	अ	आ		
	Back	उ	ऊ	ओ	औ
Consonants	Velar	क	ख	ग	घ
	Affricate	च	छ	ज	झ
	Retroflex	ट	ठ	ड	ढ
	Dental	त	थ	द	ध
	Bilabial	प	फ	ब	भ
	Nasal	ञ	ण	न	म
	Glides	य	व		
	Liquids	ल	र		
	Fricatives	श	ष	स	ह
Silence		.	।	?	

Cepstral based features can be improved by processing it with QCN technique. The steps of PNCC processing and QCN-PNCC processing is represented in Figure 1.

4. Acoustic-phonetic Feature of Hindi

The acoustic-phonetic features of Hindi are very different from any European language. There are 10 vowels, 4 semivowels, 4 fricatives and 25 stop consonants in Hindi alphabet. The 10 vowels of Hindi alphabet include 2 diphthongs. The classification of Hindi phonemes as given by Samudravijaya et al.[9] is given in Table 1. The Table 1 has three sections consisting of vowels, consonants, semivowels and fricatives respectively. Comparative classification accuracy has evaluated of Hindi vowels in three classes: Front vowel, Mid Vowel and, Back Vowel.

5. Experimental Setup

In this work the Hindi vowel classification is achieved. The experimental setup is given by Figure 2. The Hindi phonemes are obtained from Hindi speech database[9], designed at TIFR, Mumbai, India. From this database 50 speakers were taken for phoneme extraction. Out of these 50 speakers, 33 speakers were male and 17 speak-

ers were female. Here, the phonemes are extracted from the database using the transcription file given with the database[10–11]. After extracting the phonemes, two types of feature extraction techniques are used for obtaining features of phonemes. In first technique 13 PNCC features were obtained for each phoneme by applying 40 channels Gammatone filter bank on the preprocessed speech signal. Pre-processing includes framing with 10ms frame period and windowing with Hamming window of 25.6 ms. Lower 13 Cepstral coefficients are taken as features. In second set Cepstral vectors are read and sorted in ascending order. Then the low and high quantiles of the cepstral dimension is estimated. The quantile means are subtracted from all the samples. The dynamics of the cepstral coefficients are normalized. Low- Pass temporal filtering is performed in each cepstral dimension.

One HMM model is prepared for front vowel, mid vowel and back vowel each having 3 emitting states and 4 Gaussian Mixture Components, with spherical covariance. The accuracy of classification is calculated by the following equation:

$$efficiency = \frac{(total\ \ test\ \ samples - error)x100}{total\ \ test\ \ samples}$$

6. Results

The classification task is carried out with 13 PNCC features and 13 QCN-PNCC features for each Hindi phoneme segmented from the database. At first experiment is performed for the Context-Independent (CI) phoneme classification case, and then same experiments is performed for Context-Dependent (CD) phoneme classification case. The results obtained are shown in Table 2 and Figure 3. The results show that the CD phoneme classification results are better than the CI phoneme classification for clean as well as noisy data. It is also observed that as signal to noise ratio decreases the recognition efficiency decreases for all cases. For clean data an overall improvement of 8% is observed with QCN-PNCC features over PNCC features for context independent cases while for context-dependent cases the improvement is 5%. For noisy database an improvement of 3% to 5% is observed with QCN-PNCC features over simple PNCC features for CI as well as CD cases. The reason for this improvement lies in the fact that though in the development of PNCC features care is taken for suppression of noise by incorporating power law nonlinearity to closely approximate the relationship between incoming signal

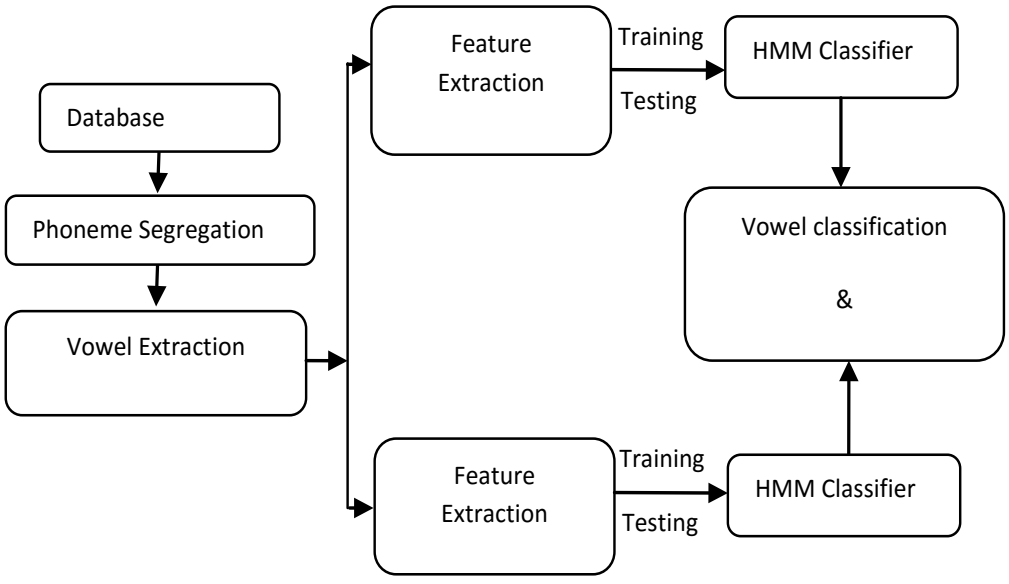

Figure 2. Block diagram of Experimental setup.

Table 2. Comparative % recognition efficiency of PNCC and QCN-PNCC features for Hindi vowel classification

Dataset	Features Front		Phoneme			
			Mid	Back	Average	
clean	CI	PNCC	82.87	81.01	83.2	82.02
		QCN-PNCC	91.03	87.6	90.14	89.59
	CD	PNCC	92.6	89.3	89.0	90.3
		QCN-PNCC	94.5	99.5	91.5	95.17
10dB	CI	PNCC	63.4	60.45	61.7	61.85
		QCN-PNCC	65.01	65.4	63.5	64.63
	CD	PNCC	68.87	65.69	64.01	66.19
		QCN-PNCC	72.34	71.98	68.63	70.98
5dB	CI	PNCC	55.89	53.7	54.66	54.75
		QCN-PNCC	56.90	60.85	56.03	57.92
	CD	PNCC	58.05	59.98	56.97	58.33
		QCN-PNCC	60.53	67.1	58.01	61.88
0dB	CI	PNCC	44.01	42.87	43.76	43.55
		QCN-PNCC	44.89	48.73	44.81	46.14
	CD	PNCC	47.1	44.66	44.5	45.42
		QCN-PNCC	47.78	48.98	47.3	48.02

amplitude in a given frequency channel and the corresponding response of the processing model but nothing is done to counter the involuntary adjustment of vocal parameters by speaker in presence of noise (Lombard Effect). QCN-PNCC features have the characteristics of PNCC features processed with QCN technique.

7. Conclusions

In this paper a novel hybridized feature set QCN-PNCC is proposed. The speech recognition accuracy of QCN-PNCC features are evaluated for the task of Hindi vowel classification. It is observed that for clean data QCN-PNCC features provide 8% improved accuracy over PNCC features for CD case and an improvement of 5% is observed for CI case. For noisy data an improvement of 3% to 5% is observed. The results obtained are tabulated in Table 2 and the same results are graphically represented in Figure 3.

Although QCN-PNCC features show improved recognition efficiency over PNCC features, but in noisy conditions, where SNR is low, even QCN-PNCC features do not provide impressive recognition efficiency.

8. References

1. Harvilla MJ, Stern RM. Histogram-based sub band power warping and spectral averaging for robust speech recognition under matched and multistyle training. IEEE International Conference on Acoustics, Speech Signal Processing; 2012 May.
2. Boˇril H. Robust speech recognition: Analysis and equalization of lombard effect in czech corpora, Ph.D. Thesis, Czech Technical University in Prague, Czech Republic; 2008.
3. Kim C, Stern RM. Feature extraction for robust speech recognition using a power-law nonlinearity and power-bias subtraction. INTERSPEECH-2009; 2009 Sep; p. 28–31.
4. Kim C, Stern RM. Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring. IEEE International Conference on Acoustics, Speech, and Signal Processing; 2010 Mar. p. 4574–7.
5. Kelly F, Harte N. A comparison of auditory features for robust speech recognition. EUSIPCO-2010; 2010 Aug. p. 1968–72.
6. Kelly F, Harte N. Auditory features revisited for robust speech recognition. International Conference on Pattern Recognition. 2010 Aug; p. 4456– 9.

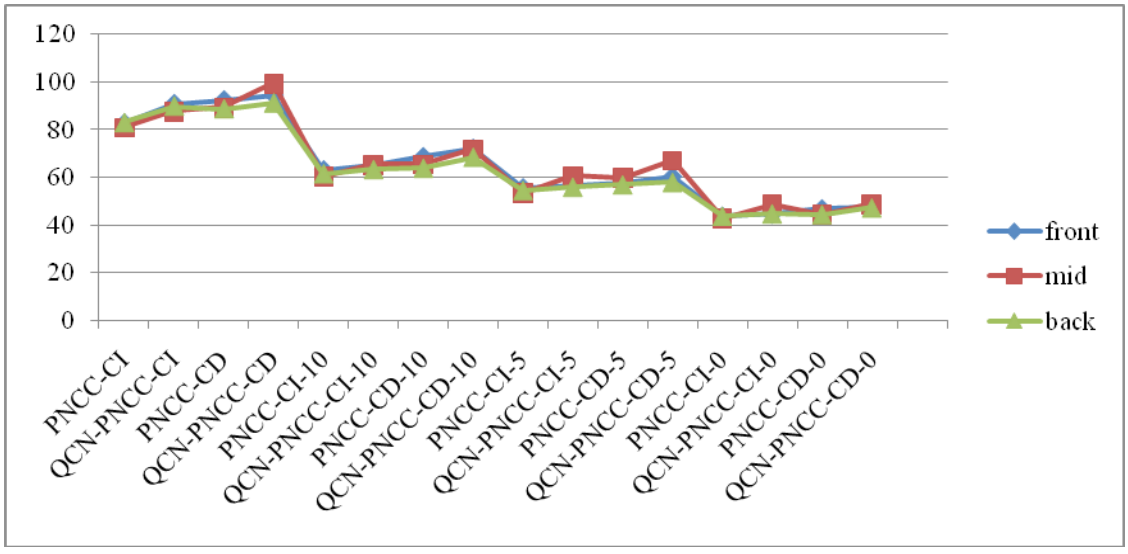

Figure 3. Comparative % recognition efficiency of PNCC and QCN-PNCC features for Hindi vowel classification.

7. Shipra, Chandra M. Hindi vowel classification using QCN-MFCC features. Perspectives in Science. 2016 Sep; 8:28–31. DOI: dx.doi.org/10.1016/j.pisc.2016.01.010.
8. Rabiner LR. A Tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989; 77(2):257–85.
9. Samudravijaya K, Rao PVS, Agrawal SS. Hindi speech database. International Conference on Spoken Language Processing (ICSLP00). Beijing; 2002. p. 456–9.
10. Biswas A, Sahu P, Chandra M. Admissible wavelet packet features based on human inner ear frequency response for Hindi consonant recognition. Computers and Electrical Engineering. 2014; 40(4):1111–22.
11. Biswas A, Sahu P, Bhowmick A, Chandra M. Feature extraction technique using ERB like wavelet sub-band periodic and aperiodic decomposition for TIMIT phoneme recognition. International Journal of Speech Technology. 2014; 17:389–99.