Applied Soft Computing 13 (2013) 2281–2291

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

An ANN based approach to recognize initial phonemes of spoken words of Assamese language

Mousmita Sarma, Kandarpa Kumar Sarma*

Department of Electronics and Communication Technology, Gauhati University, Guwahati 781014, Assam, India

ARTICLE INFO

Article history:
Received 6 April 2012
Received in revised form 25 August 2012
Accepted 2 January 2013
Available online 1 February 2013

Keywords:
Formant
Phoneme
LPC
RNN
SOM
PNN

ABSTRACT

Initial phoneme is used in spoken word recognition models. These are used to activate words starting with that phoneme in spoken word recognition models. Such investigations are critical for classification of initial phoneme into a phonetic group. A work is described in this paper using an artificial neural network (ANN) based approach to recognize initial consonant phonemes of Assamese words. A self organizing map (SOM) based algorithm is developed to segment the initial phonemes from its word counterpart. Using a combination of three types of ANN structures, namely recurrent neural network (RNN), SOM and probabilistic neural network (PNN), the proposed algorithm proves its superiority over the conventional discrete wavelet transform (DWT) based phoneme segmentation. The algorithm is exclusively designed on the basis of Assamese phonemical structure which consists of certain unique features and are grouped into six distinct phoneme families. Before applying the segmentation approach using SOM, an RNN is used to take some localized decision to classify the words into six phoneme families. Next the SOM segmented phonemes are classified into individual phonemes. A two-class PNN classification is performed with clean Assamese phonemes, to recognize the segmented phonemes. The validation of recognized phonemes is checked by matching the first formant frequency of the phoneme. Formant frequency of Assamese phonemes, estimated using the pole or formant location determination from the linear prediction model of vocal tract, is used effectively as a priori knowledge in the proposed algorithm.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Spoken word can be represented phonetically by a finite set of symbols called the phonemes of the language. Identification of these phonemes is an important part of any phoneme based or language specific speech recognition system [10]. The phonemical structure of every language provides various phonemical groups for both vowel and consonant phonemes each having distinctive features. This work provides an approach to recognize initial phoneme from some consonant–vowel–consonant (CVC) spoken words by taking advantages of such phonemical groups of Assamese language. Here all words of the recognition vocabulary are initially classified into six distinct phoneme families and then the initial phoneme is identified by certain combination of segmentation and classification technique which shall be described in the subsequent sections.

Recognition of individual phoneme from a word requires some efficient segmentation algorithm which can extract the differentiating characteristics of that particular phoneme. In most approaches, the speech signals are segmented using constant-time segmentation, for example into 25 ms blocks [1]. Constant segmentation may loose information about the phoneme, because phonemes are naturally of varying lengths. Different sounds may be merged into single blocks and as a result individual phonemes are lost completely. Discrete wavelet transform (DWT) is another frequently used technique for such kind of phoneme segmentation [2]. DWTs approach to formulating segmentation boundaries is based on a filter bank which suffers from sampling/aliasing effects and fails to generate near perfect reconstruction [9]. It leads to distortion for which success-rates suffer. DWTs limited spectral response within the recorded frequency range of the speech signals prevents formulation of segmentation boundaries at low frequency range where voiced sounds occur. Further DWT has the problem that at each level, signal has to be reconstructed. It limits the development of a one pass algorithm, where the segmentation and recognition take place consequently. Here we present a new algorithm to segment and recognize the initial phoneme from certain two alphabet Assamese words, where a self organizing map (SOM) segments the word into its constituent phonemes in an unsupervised manner whereas two supervised artificial neural network (ANN) blocks namely, recurrent neural network (RNN) and probabilistic neural network (PNN), respectively play the role of recognizing the phoneme family to which the initial

* Corresponding author. Fax: +91 361 2700311.
E-mail addresses: go4mou@gmail.com (M. Sarma), kandarpaks@gmail.com (K.K. Sarma).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.01.004

phoneme belongs and identifying the initial phoneme from the SOM generated segments. SOM with its ability to reorganize to different topological states captures contextual variations better with changing number of epochs. Therefore, the proposed ANN based phoneme segmentation method shows distinct advantage in terms of recognition success rate than the conventional discrete wavelet based speech segmentation method. A few two class PNNs are trained to learn the patterns of all the consonant phonemes of Assamese language. The phoneme segments obtained by training SOM with various iterations are then matched with the PNN patterns. Before the global decision is taken by the PNN, an RNN takes some local decision about the incoming word and classifies them into six phoneme families of Assamese language. The segmentation and recognition are performed separately at each of the RNN decided family. Since every natural language provides a distinct set of phonemes and some distinct phonetic groups, therefore it will be a novel approach to apply pattern classification techniques to classify spoken words according to the specific phonemical structure of the language which naturally exists. In that sense the proposed method is a new proposal toward language specific speech recognition. The method is experimented in case of Assamese language, which is spoken by a population of around 40 million in North-East India (NEI). As per our knowledge, no such method has been reported for speech recognition in Assamese and other Indian languages. An important aspect of the classification algorithm designed using RNN and PNN is that it uses the prior their knowledge of first formant frequency (F1) of the Assamese vowel phonemes while taking decisions. This is due to the fact that phonemes are distinguished by own unique pattern as well as in terms of their formant frequencies. Here, the concept of pole or formant location determination related to the linear prediction model (LPC) model of vocal tract is used while estimating F1 [3]. Assamese CVC words are recorded from five girls and five boys, so that the classification algorithm can remove the speaker dependance limitation. Assamese, which is a widely spoken language in NEI, has its own distinctive phonological features and provides a sound area for speech researchers. The work shows almost 98–100% success rate, which is well above the previous reported results. The proposed work provides a new phoneme segmentation technique which can be considered as a new contribution toward ANN based speech analysis. The description included here is organized as below. Section 2 provides briefly the phonemical details of Assamese language. The paper then provides a brief account on how linear prediction models the formant and antiformant frequencies of the speech signal in Section 3. The proposed phoneme segmentation and recognition algorithm are described in Section 4. The results and the related discussion are included in Section 5. Section 6 concludes the description.

2. Certain phonemical features of Assamese language

Assamese is an Indo-Aryan language originated from the Vedic dialacts, and therefore, a sister of all the northern Indian languages. Although the exact nature of the origin and growth of the language are yet to be clear, it is supposed that like other Aryan languages, Assamese also developed from *Apabhraṁśa* dialects developed from *Mãgadhi* Prakrit of the eastern group of Sanskritic languages [6]. Retaining certain features of its parent Indo-European family it has got many unique phonological characteristics. Some of those may be cited as below:

- A unique feature of the Assamese language is a total absence of any retroflex sounds. Instead the language has a whole series of alveolar sounds, which include oral and nasal stops, fricatives, laterals, approximants, flaps and trills, unlike other Indo-Aryan and Dravidian languages [4].

Consonants

	Bilabial		Alveolar		Palatal	Velar		Glottal
	Vl.	Vd.	Vl.	Vd.	Vd.	Vl.	Vd.	Vd.
Unaspirated	p	b	t	d		k	G	
Aspirated	ph	bh	th	dh		kh	gh	
Spirant			s	z		x		fi
Nasals		m		n			ŋ	
Lateral				l				
Trill				r				
Frictionless Continuant		w			j			

Fig. 1. Table showing Assamese consonant phonemes, Vl., voiceless; Vd., voiced.

- Another striking phonological feature of the Assamese language is the extensive use of velar nasal / η/. In other New Indo Aryan languages this / η/ is always attached to a homorganic sound like /g/. In contrast it is always used singly in Assamese.
- The voiceless velar fricative / x / is a distinct characteristic of Assamese language which is not to be found in any language in the entire country. It is similar to the velar sound in German of Europe. It may be an Indo- European feature, which has been preserved by 'Axomiya'. It is an important phoneme in the language [4].

There are other phonological uniqueness of Assamese pronunciation which show minor variations when spoken by people of different regions of the state. This makes Assamese speech unique and hence requires a study exclusively directly to develop a speech recognition / synthesis system in Assamese [4].

There are twenty-three consonants and eight vowel phoneme in the standard colloquial Assamese. The consonants may be grouped into two broad divisions: the stops and the continuants. For the stops there are contrast in three points of articulation: the lips, the alveola, and the velum; and four-way contrasts in every point as to the presence or otherwise of voice and aspiration. Therefore, a stop may be voiced or voiceless, aspirated or unaspirated. There are continuants – two frictionless, viz, the semivowels /w, j/, four spirants /s z x h/, one lateral /l/, one trill /r/, and three nasals /m, n, η/ which are stops as well as continuants both at once [6]. The Assamese phoneme tables obtained from [6] are shown in Figs. 1 and 2.

3. Use of linear prediction coding to model speech signal

The speech signal is produced by the action of the vocal tract over the excitation coming from the glottis. Different conformations of the vocal tract produce different resonances that amplify frequency components of the excitation, resulting the different

Vowels

	Front	Central	Back
High	i̯		u
Higher-mid	e		o
Lower-mid	ɛ		ɔ
Low		a	ɒ

Fig. 2. Table showing Assamese vowel phonemes.

Download English Version:

https://daneshyari.com/en/article/495926

Download Persian Version:

https://daneshyari.com/article/495926

Daneshyari.com