Using of Beylkin Wavelet for Speech Recognition

N A Leontiev

1Institute of Physics and Technologies, North-Eastern Federal University, Belinskogo 58, Yakutsk, 677013, Russia

E-mail: leonza@mail.ru

Abstract. This paper describes the application of the Beylkin wavelet for speech segmentation. The problem of speech segmentation in the Yakut language is that there are segmentation difficulties due to the peculiarities of the language. The use of long vowels and double consonants in the Yakut language complicates the correct segmentation of oral speech. For the analysis, the window method of analyzing the energy of the wavelet signal is used. The experience of using different wavelet functions has shown that it is not always possible to accurately find the segment boundaries in some cases. The Scilab package has a large library of wavelets that allows extensive research into their applications in speech recognition. The results of the study show that there are difficulties due to various reasons, one of which is the presence of double sonorant consonants. The graphs of the analysis of doubled sonorant consonants are given.

1. Introduction

Currently, speech recognition systems are widespread, there are many software products, online systems, hardware devices. There is a problem with the representation of languages, the most common are English, European languages and other major languages of the world. First of all, the languages of peoples with a small number do not have software systems for speech recognition. After all, it takes tens of years and a lot of work to study a language, including the collection of material.

There are works on the segmentation of speech in Tibetan [1], in Chinese there is a problem of word segmentation [2], in Arabic and English using a dynamic threshold [3], segmentation of syllables in Chinese [4]. A lot of research is going on in the recognition of the Arabic language [5-6], Indian languages, in particular Hindi [7].

Wavelets have many types. The most popular wavelets are Daubechies, Haar, Meyer, Koifman. In addition to them, there is such a wavelet as the Belkin wavelet. It is part of the wavelet library for Scilab.

The development of software with computer libraries of mathematical methods for signal processing makes it possible to use a variety of types of wavelets. A survey of wavelets is given in [8].

Wavelets are also used to compress and encrypt speech signals [9]. The wavelet transform shows a promising solution for a non-stationary signal. Noise suppression of a noisy speech signal transmitted using the wavelet thresholding technique [10]. Speech enhancement aims to improve the quality and intelligibility of speech using a variety of methods and algorithms, which is very important for hearing aids [11]. Automatic recognition of a continuous speech signal in the Gujarati language [12].

ASR is designed for continuous Kannada speech recognition. Acoustic and language models were created using the Kaldi toolkit. Speech database created using male and female Kannada speakers.
80% of the collected speech data is used to train acoustic models, and 20% of the speech database is used to test the system [13].

Recognizing speech emotion is one of the challenging research tasks in a knowledge-based system, and various methods have been recommended to achieve high classification capabilities. To achieve high efficiency of classification in recognition of speech emotions, a nonlinear multilevel model of feature generation using a cryptographic structure is presented [14].

Vowel formants provide information about how vowels are pronounced. Formant frequencies are important in human speech processing applications. However, such implementations are mostly done with non-Spanish speakers. Thus, the characterization of Spanish vowels is for further study. This paper presents a formant extraction method based on a discrete wavelet transform. The work is aimed at Hispanic residents of Antioquia, Colombia. Wavelet analysis parameters are tuned to set the appropriate vowel response in the frequency formant space. The results show that vowel-specific wavelet analysis yields well-defined clusters in the formant space. [15]

This article presents the analysis of a speech signal using multilevel wavelet transform and signal decomposition. This research work is carried out by recording the voices of 40 different speakers. In this work, the speakers pronounced the same set of words, these 4 words - "Bhavana", "How", "Is" and "You" [16].

In this study, we proposed sensor matrices based on 1D Discrete Wavelet Transform (DWT) for speech compression. This study explores the performance analysis of various DWT-based sensor arrays such as: Daubechies, Coiflets, Simlets, Beylkin, and Vaidyanathan of the wavelet family [17]. There are segmentation methods for languages that do not have word boundaries in writing, such as Burmese [18].

The author carried out research on the automated processing of the Yakut language, speech segmentation was carried out in [19-20]. In these works, the MathCad package and the Daubechies wavelet were used, interesting results were obtained that show the possibility of segmentation of oral speech using a computer. Also, the properties of speech were identified, which must be paid attention to in order to obtain a good result. It is necessary to select the type of wavelet that can be used to perform the correct segmentation of continuous speech.

2. Results
To study the segmentation of speech in the Yakut language, audio recordings of speech from native speakers were collected. The recording format is WAVE without compression, with a sampling rate of 44100 Hertz, with a bit depth of 16 bits.

Due to the peculiarity of the Yakut speech, there is a softening of double sonorant consonants. This does not allow for clear segmentation of speech. It is necessary to clearly capture the presence of a gap between the segments.

For the research, the open source software package Scilab was used. There are many computational libraries, including digital signal processing libraries. The authors used the Wavelet Toolbox wavelet library. This library contains discrete transform of wavelets, including one-dimensional, 2D and 3D. There is a large selection of wavelet types including Daubechies, Meyer, MHAT and Belkin wavelets. The window size selection is 512 samples or 11.6 milliseconds. The windows overlap each other by half.

Figures 1 and 2 shows the correlation graph of the wavelet windows. For close frequencies, the correlation coefficient is close to 1. This shows that the windows refer to one frequency, that is, to one sound unit. When analyzing the word "oonnyuur" - Daubechies wavelet shows the best result when separating double sonar consonants.

The boundaries between phonemes are due to the change in the correlation coefficient.
3. Conclusions
The use of discrete Beylkin wavelet allows for better segmentation of dual sonar consonants.

4. References
[1] Wang L, Yang H 2019 Tibetan Word Segmentation Method Based on BiLSTM-CRF Model Proceedings of the 2018 International Conference on Asian Language Processing IALP 2018 8629257 pp 297-302
[2] Zhao Y, Li H, Yin S, Sun Y 2018 A new Chinese word segmentation method based on maximum matching Journal of Information Hiding and Multimedia Signal Processing 9(6) pp 1528-1535
[3] Jazyah Y H 2018 Speech segmentation using dynamic windows and thresholds for Arabic and English languages (Open Access) Journal of Computer Science 14(4) pp 485-490
[4] He S, Zhao H 2017 Automatic syllable segmentation algorithm of Chinese speech based on MF-
DFA Speech Communication 92 pp 42-51

[5] Jazyah Y H 2018 Speech segmentation using dynamic windows and thresholds for Arabic and English languages Journal of Computer Science 14(4) pp 485-490

[6] Ahcene A, Aissa A, Abdelkader D, Khadidja B, Ghania D-H 2016 Automatic segmentation of Arabic speech signals by HMM and ANN Proceedings of 2016 International Conference on Electrical Sciences and Technologies in Maghreb CISTEM 2016 8066776

[7] Khan R A, Chitode J S Hindi syllable segmentation using ZCR and dual band energy ratio International Journal on Communications Antenna and Propagation 7(7) pp 581-585

[8] Rashid O, Amin A, Lone M R 2020 Performance analysis of DWT families Proceedings of the 3rd International Conference on Intelligent Sustainable Systems ICISS 2020 pp 1457 – 1463

[9] Hameed A S 2021 Speech compression and encryption based on discrete wavelet transform and chaotic signals Multimedia Tools and Applications vol 80 9 pp 13663-13676

[10] Chiluveru S R & Tripathy M 2021 Speech Enhancement using a Variable Level Decomposition DWT National Academy Science Letters vol 44 3 pp 239-242

[11] Brahim M 2021 Denoising and Enhancement Speech Signal Using Wavelet Journal of Information Systems and Telecommunication vol 9 33 pp 37-44

[12] Pandit P, Makwana P & Bhatt S 2021 Automatic Speech Recognition of Continuous Speech Signal of Gujarati Language Using Machine Learning Mathematical Modeling, Computational Intelligence Techniques and Renewable vol 1287 (Springer, Singapore)

[13] Mahadevaswamy & Ravi D J 2021 Robust Perceptual Wavelet Packet Features for Recognition of Continuous Kannada Speech, Wireless Personal Communications, https://doi.org/10.1007/s11277-021-08736-1

[14] Tuncer T, Dogan S & Acharya U R 2021 Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighborhood component analysis techniques Knowledge-Based Systems vol 211

[15] Orellana S & Ugarte J P 2021 Vowel characterization of Spanish speakers from Antioquia–Colombia using a specific-parameterized discrete wavelet transform analysis Applied Acoustics vol 172

[16] Returi K D, Radhika Y, Mohan V M & Srujan Raju K 2021 A method of speech signal analysis using multi-level wavelet transform Vol 1171 pp 711 – 721

[17] Parkale Y V, Nalbalwar S L 2016 Application of 1-D discrete wavelet transform based compressed sensing matrices for speech compression SpringerPlus Open Access Vol 5 Issue 11 2048

[18] Ma C, Yang J 2018 Burmese Word Segmentation Method and Implementation Based on CRF Proceedings of the 2018 International Conference on Asian Language Processing IALP 8629163 pp 340-343

[19] Leontiev N A, Nyurova A G 2019 The Use of Discrete Meyer Wavelet for Speech Segmentation Article number 8934292019 International Multi-Conference on Industrial Engineering and Modern Technologies FarEastCon

[20] Leontiev N A 2018 The use of wavelets for the Yakut speech recognition problems AIP Conference Proceedings 2041 050005 doi: 10.1063/1.5079374