Impact Factor: 9.2 ISSN-L: 2544-980X

METHODS OF SPEECH SIGNAL SEGMENTATION FOR MULTIMODAL SPEECH RECOGNITION

L.R.Dalibekov
Fergana branch of TUIT, Fergana, Uzbekistan

Abstract: *Segmentation of a speech signal is one of the most important tasks in the field of computer science and information systems for speech processing and recognition. Segmentation of the speech signal is necessary to isolate the characteristic features of the speaker's voice in certain segments of the speech signal and restore the shape of the vocal tract based on an acoustic feature, which can be used in speech synthesis from the input text and speech recognition.*

Key words: *Zero Cross Rate, Spectral Transition Measure, segmentation, speech signal, speech synthesis, recognition, method, frames, counting.*

INTRODUCTION

In research, manual speech segmentation can be used, but manual speech segmentation slows down the work and is almost impossible to accurately reproduce the results of manual segmentation; it allows for many errors in speech recognition.

In speech recognition information systems for speech signal segmentation, the following is important:

- identifying the main elements (words, syllables, phonemes) of speech recognition;
- segmentation accuracy has a great influence on optimal speech recognition.

There are several main types of automatic speech signal segmentation. One of the types includes speech segmentation, provided that the phoneme sequence of a given phrase is known, but recognition results are often unreliable, and the presence of transcription is possible only at the stage of training lexical models [2]. The other type does not use a priori speech information, and the boundaries of speech segments are determined by the degree of change in the acoustic characteristics of the speech signal. In automatic segmentation, it is desirable to use only the general characteristics of the speech signal, since usually at this stage there is no specific information about the content of the speech.

For simple segmentation of the speech signal into pauses and speeches, there is a "blind" segmentation method. This method is based on the magnitude and rate of change of certain acoustic characteristics - this is the zero crossing rate of the signal level (Zero Cross Rate) and the spectral transition measure (Spectral Transition Measure), but experiments show that these values are not enough for reliable segmentation [3].

The incoming speech signal is recorded as a sequence of reports y_i.

$Y=y_0, y_1, \dots, y_i, \dots;$ where i=0,1,2, … .

The speech signal sequence is divided into frames with a length of 128 samples (respectively (128*1000)/11025~11ms). The size of the frame allows you to accurately determine the boundaries between syllables.

Using the following formula, we find the average energy value in a speech signal frame with a length of 128 samples:

$$E_i = \frac{\sum_{j=i*128}^{i*128+127} y_j^2}{128};$$ where i=0,1,2, … (1.1)

The obtained values according to formula (1.1) are the average energy of a short time over an interval of 11 ms. Let's calculate the average short-time energy value of three neighboring sections using the formula:

$$E_i^* = \frac{E_i+E_{i+1}}{2};$$ where i=0,1,2, … (1.2)

Thus, we calculate the average energy for frames 2*128=256 samples. Frames are taken with overlap and shift of adjacent intervals by 128 samples (Figure 1.1).

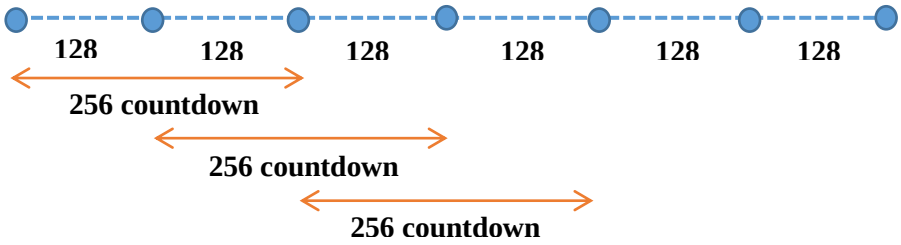

Figure 1. Dividing the speech signal into frames.

The fundamental tone of the Kazakh language is less than 256/11025=0.023 seconds, which corresponds to the fundamental frequency I/0.023=75.5Hz. Therefore, the energy of a 256-count long frame contains the energy of at least one pitch period. Thus, from the speech signal sequence Y=y0, y1, … , yi, …; where i=0,1,2, … we will calculate the sequences of the average energy of sections of 192 counts. $E^* = E_1^*, E_2^*, \dots, E_i^*, \dots$.

Each syllable has a syllable peak where the signal energy reaches its greatest value.

Between two syllable peaks there is a point, the corresponding boundary, which separates the syllables [4].

Literature

1. W.P. McNeill, J.G. Kahn, D.L. Hillard, M. Ostendorf. Parse Structure and Segmentation for Improving Speech Recognition // IEEE Spoken Language Technology Workshop. - 2006. –P.90-93.

2. Turgunov, B., Iskandarov, U., Dalibekov, L., & Jurayeva, G. (2024, March). Prospects for using alternative energy sources to generate high power electrostatic fields in the primary processing of raw cotton. In AIP Conference Proceedings (Vol. 3045, No. 1). AIP Publishing.
3. Ergashev, S., Dalibekov, L., Komilov, A., Jo'raeva, G., Xusanova, S., & Komilov, D. (2024, November). Optical electron photo converter. In E3S Web of Conferences (Vol. 508, p. 01002).
4. Dalibekov, L. R. (2023). Innovative applications of apv elements in optoelectronics. International Journal of Advance Scientific Research, 3(10), 286-292.
5. Далибеков, Л. (2023). ИССЛЕДОВАНИЕ АНОМАЛЬНЫХ ФОТО НАПРЯЖЕНИЙ КАК ИНДИКАТОРОВ СЕТЕВЫХ ПРОБЛЕМ. Conference on Digital Innovation : "Modern Problems and Solutions". извлечено от https://fer-teach.uz/index.php/codimpas/article/view/1839
6. Далибеков , Л. (2023). ALOQA TARMOQLARIDA ENERGOBARQAROR TIZIMLARNI TADBIQ ETISH. Conference on Digital Innovation : "Modern Problems and Solutions". извлечено от https://fer-teach.uz/index.php/codimpas/article/view/1846
7. Мадаминов М.Р. и Далибеков Л.Р. (2024). Классификация и режимы работы резервных источников электроснабжения сетей мобильной связи. Лучший журнал инноваций в науке, исследованиях и разработках, 3 (4), 784–789. Получено с https://www.bjisrd.com/index.php/bjisrd/article/view/2137.
8. М.Р.Мадаминов. (2023). Классификация и режимы работы резервных источников электроснабжения сетей мобильной связи. Лучший журнал инноваций в науке, исследованиях и разработках , 2 (12), 32–38. Получено с https://www.bjisrd.com/index.php/bjisrd/article/view/1045.
9. A.X.Abdusamatov. (2023). RADIATION-STIMULATED ANNEALING BY GAMMA QUANTUM OF LEDS BASED ON AlGaAs HETEROSTRUCTURES SUBJECT TO OPERATIONAL FACTORS. Best Journal of Innovation in Science, Research and Development, 2(12), 269–275. Retrieved from https://www.bjisrd.com/index.php/bjisrd/article/view/1111
10. D.R.Komilov. (2023). HYBRID INVERTERS FOR GREEN POWER PLANTS. Best Journal of Innovation in Science, Research and Development, 2(12), 39–42. Retrieved from http://www.bjisrd.com/index.php/bjisrd/article/view/1046
11. M. M. Tillaboev. BESTJOURNAL OF INNOVATION INSCIENCE, RESEARCH ANDDEVELOPMENTISSN: 2835-3579Volume:02Issue:12|2023. In protected areasdangers elimination methods.
12. Халилов, М. М. (2023). СНИЖЕНИЕ ВЕРОЯТНОСТИ ПОТЕРЬ С ПОМОЩЬЮ КОДИРОВАНИЯ СИГНАЛА В ВОЛОКОННО-

ОПТИЧЕСКИХ ЛИНИЯХ СВЯЗИ. European Journal of Interdisciplinary Research and Development, 22, 60-66.

13. MM Khalilov (2023). EFFECTS OF THE INTERACTION OF CHLORINE WITH PbS SURFACE AND THEIR EFFECT KINETIC PARAMETERS OF STRAIN-SENSITIVE FILMS. Best Journal of Innovation in Science, Research and Development, том 2, №2, стр. 229-236.