INTERNATIONAL JOURNAL ON ORANGE TECHNOLOGY

https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071
Volume: 4 Issue: 8 | August 2022

Normalization of the Input Signal During Speech Signal Preprocessing

A.A. Rakhmanov

Chief Specialist of the Territorial Administration of the Samarkand Region of the Ministry for the Development of Information Technologies and Communications of the Republic of Uzbekistan

---***---

Annotation: It is important to pre-process and normalize a voice signal before recognizing and processing it. Since there are a variety of variables that might affect the sound signal's amplitude when recording speech, including the speaker's voice volume, proximity to the microphone, type of recording equipment, etc. In this post, we'll look at the input signal normalization technique and its first processing in preparation for building a model and an algorithm for identifying audio signals.

Keywords: voice signal, amplitude audio signal.

Introduction: Automation of the procedure for managing the monitoring of radio broadcasts is currently one of the essential tasks in Uzbekistan. It is important to develop an algorithm and software program for listening to radio broadcasts' speech in order to automate this operation. Recognizing the sound signal is the first step in solving this issue. It is required to process and normalize the audio signal because when recording speech, a number of elements that cause a wide range in loudness have an impact on the amplitude of the signal. In this post, we'll look at a technique for normalizing the input audio signal so that it can be used for recognition in the future.

When recording speech, the amplitude of the audio signal is affected by a number of factors: the loudness of the speaker's voice, its distance from the microphone, etc.

The volume of the speech signal varies greatly as a result of these factors [1]. When employing heterogeneous sound recording equipment, this phenomena is very obvious. The loudness spread is removed using the amplitude normalizing method. With the use of this method, the signal's amplitude is constrained to the $[\frac{\Delta}{2}, -\frac{\Delta}{2}]$ bounds (Fig. 1). Using the following formula, samples of the normalized signal $\hat{s}\,[n]$can be obtained.:

$$\hat{s}\,[n]=\frac{\Delta}{\max\limits_{m}|\,s[m]\,|}\cdot s[n] \qquad (1)$$

here Δ – width of the normalization band, symmetrical about the x-axis (for example, in Figure 1. Δ=1).

INTERNATIONAL JOURNAL ON ORANGE TECHNOLOGY

https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071

Volume: 4 Issue: 8 | August 2022

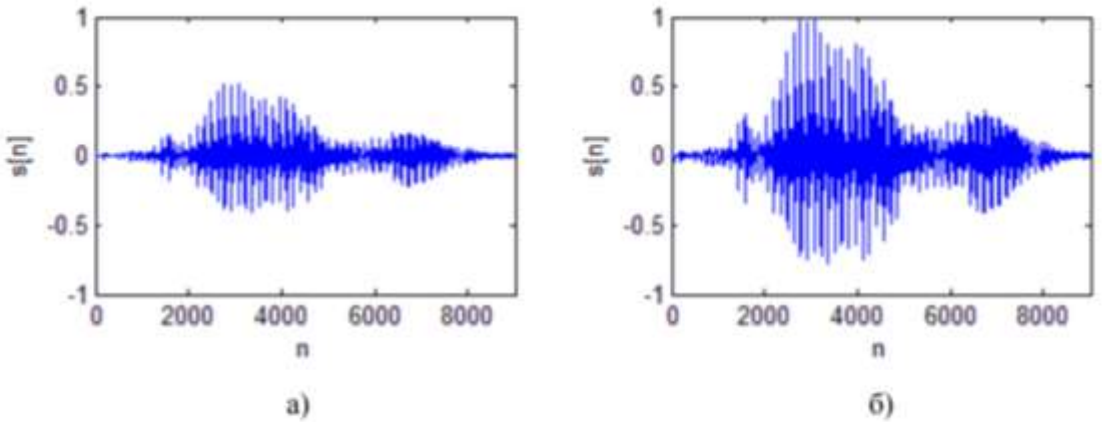

Figure 1. Digitized speech signal before (a) and after (б) normalization

Consider a collection of Q examples of how one or more words are pronounced in order to assess the loudness variability. For each example with a length of N samples, determine the average loudness value $M(q)$ [2] and the average M_Q for the examples:

$$M(q) = \sum_{n=1}^{N} | s[n] | ,q = 1,\dots,Q; \quad (2)$$

$$M_Q = \frac{1}{Q}\sum_{q=1}^{Q} M(q). \quad (3)$$

After that, we calculate the relative deviation $D(q)$ of the loudness of each example from the average:

$$D(q) = \left| 1 - \frac{M(q)}{\mathrm{M_Q}} \right| \quad (4)$$

Formulas (2) and (3) show that the number of samples in the example as well as their absolute value have an impact on the outcome. As a result, it is necessary to assess the loudness variability of a group of examples from the same class that are roughly equal in length as well as the variation in the loudness of the base as a whole. $D(q)$ for Q = 100 recorded instances of the word "three" is shown in Figures 2.a) and 2.b), before and after normalization, respectively. The original samples had a loudness spread of 28.5%, while the normalized ones had a loudness spread of 14.3%. For a speech base of Q = 2000 samples of various pronunciations, graphs 2.c)–d) show $D(q)$. The original base's loudness spread was 25.8%, whereas the new base's was 23.11 percent for the normalized one.

INTERNATIONAL JOURNAL ON ORANGE TECHNOLOGY

https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071

Volume: 4 Issue: 8 | August 2022

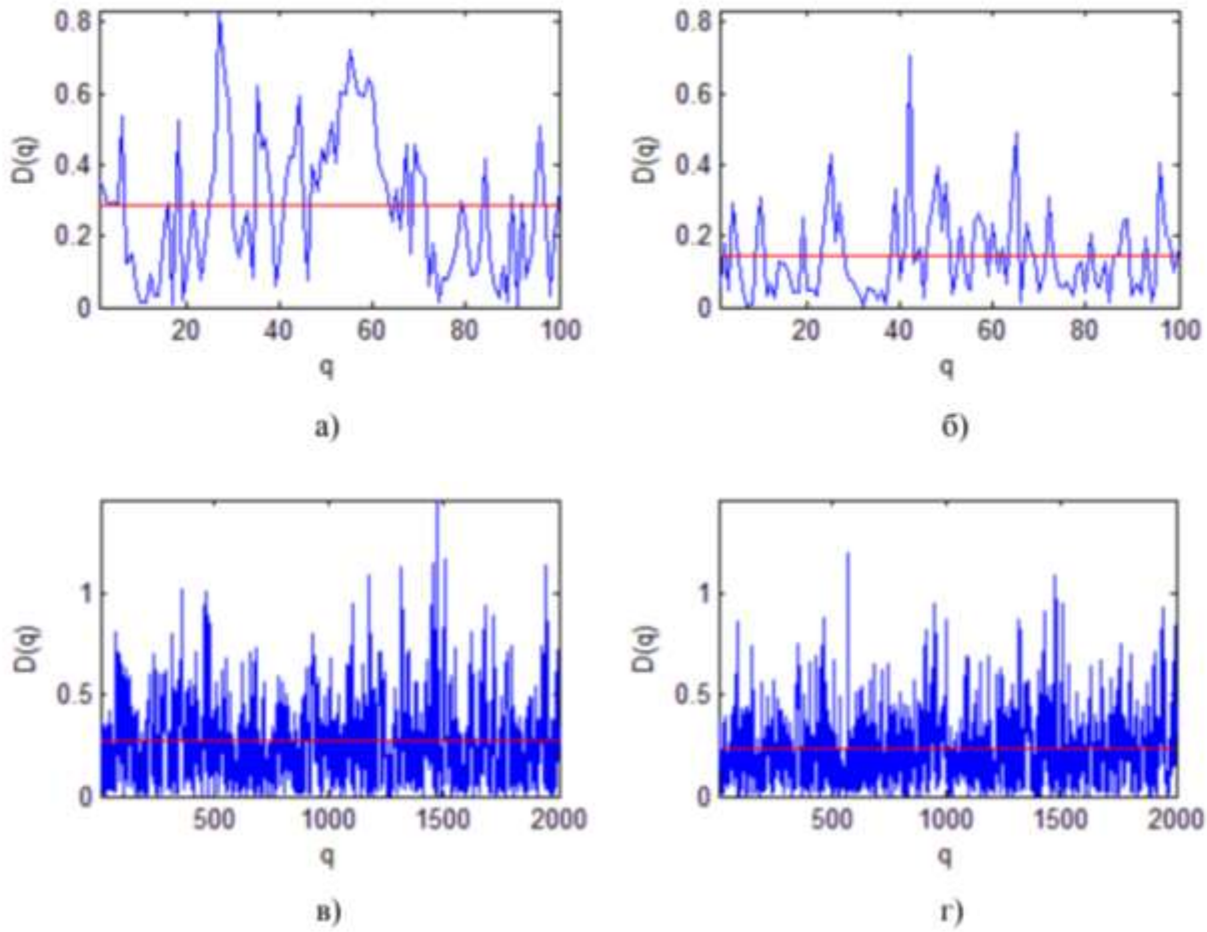

Figure 2. Deviation of the example energy from the average energy value for (а, в) initial examples, (б, г) normalized ones (а, в)

The study shows that using normalization always makes the variation in loudness for various utterances smaller. These results were achieved for a voice base that was recorded under the same settings and the only equipment available, therefore normalization in general had little impact. However, normalization is required when a speech signal is received under varied circumstances in a real-world system.

Conclusion:

This article examines a technique for normalizing and preparing voice signals.

The following are the work's principal findings:

1. Receiving readings of the normalized signal;
2. determining the average loudness value;
3. calculating the relative loudness departure of each example from the average.
4. The use of normalizing was the subject of research.

https://journals.researchparks.org/index.php/IJOT e-ISSN: 2615-8140 | p-ISSN: 2615-7071
Volume: 4 Issue: 8 | August 2022

References

1. Огнев И. В., Парамонов П.А. Предварительная обработка речевого сигнала для построения базы произношений одиночных слов // Информационные средства и технологии: труды Международной научно-технической конференции (20 – 22 октября 2012 г.): в 3 т. – М.: МЭИ, 2012. – 1 т. – с. 53-58.
2. Л. Рабинер, Р. Шафер, Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – 496 с.