LITHUANIAN COMPUTER SOCIETY
VILNIUS UNIVERSITY
INSTITUTE OF DATA SCIENCE AND DIGITAL TECHNOLOGIES
LITHUANIAN ACADEMY OF SCIENCES

11th International Workshop on

DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS

Druskininkai, Lithuania, Hotel "Europa Royale"
http://www.mii.lt/DAMSS

November 28–30, 2019

VILNIUS UNIVERSITY PRESS
Vilnius, 2019

Co-Chairmen:

Dr. Saulius Maskeliūnas (Lithuanian Computer Society)
Prof. Gintautas Dzemyda (Vilnius University, Lithuanian Academy of Sciences)

Programme Committee:

Prof. Juris Borzovs (Latvia)
Prof. Albertas Čaplinskas (Lithuania)
Prof. Robertas Damaševičius (Lithuania)
Prof. Janis Grundspenkis (Latvia)
Prof. Janusz Kacprzyk (Poland)
Prof. Ignacy Kaliszewski (Poland)
Prof. Yuriy Kharin (Belarus)
Prof. Tomas Krilavičius (Lithuania)
Prof. Julius Žilinskas (Lithuania)

Organizing Committee:

Dr. Jolita Bernatavičienė
Prof. Olga Kurasova
Dr. Viktor Medvedev
Laima Paliulionienė
Dr. Martynas Sabaliauskas

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt
Tel. +370 5 2109 315

https://doi.org/10.15388/Proceedings.2019.8

ISBN 978-609-07-0325-0 (digital PDF)

The Impact of Data Augmentation Based on White Noise for Noise-Robust CNN-Based Speech Recognition

Gintautas Tamulevičius, Gražina Korvel,
Jolita Bernatavičienė, Povilas Treigys

Institute of Data Science and Digital Technologies
Vilnius University
gintautas.tamulevicius@mif.vu.lt

The languages of small nations or dialects are considered as low-resource languages. It is a big problem if we want to explore data-driven approaches like Deep Neural Networks (DNNs). Deep learning requires a large training set to achieve a correct assessment of model performance. In order to provide a sufficient amount of data, the augmentation procedure should be done.

In the deep learning approach, images are most commonly used as network input data. In the case of speech, the recordings can be converted to the two-dimensional feature space and exported as grayscale images. The traditional image augmentation methods are not suitable for speech signals as they can lead to loss or distortion of key characteristics (for example, time-scale properties). In signal processing domain, researchers apply deformations directly to the acoustic signal before converting it into images. The literature review shows that the main focus of most of the scientific papers covering data augmentation is to increase network performance. In this work, we look at the data augmentation from another perspective. We are seeking for augmentation that can also be used for improving noise robustness of the speech recognition. In this research, a thorough analysis of speech signals in the presence of noise is done. For this purpose, we added white noise to the speech signals before converting it to a two-dimensional feature space. The following signal-to-noise ratio (SNR) levels were considered: 30 dB, 25 dB, 20 dB, 15 dB, 10 dB, 5 dB, and 0 dB. Also, the experiment was extended by testing clear recordings with an SNR value of more than 100 dB. These signals were obtained by applying the Wiener filter.

An experiment was performed on the Lithuanian speech recordings using 111-word utterances by 36 females and 26 males. Different tests were conducted on the same recordings, but in the presence of noise. For the experiment, we chose the time-frequency domain representation called spectrograms.