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

14th Conference on

DATA ANALYSIS METHODS for Software Systems

November 30 – December 2, 2023

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

VILNIUS UNIVERSITY PRESS
Vilnius, 2023

Co-Chairmen:

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

Programme Committee:

Dr. Jolita Bernatavičienė (Lithuania)
Prof. Juris Borzovs (Latvia)
Prof. Robertas Damaševičius (Lithuania)
Prof. Janis Grundspenkis (Latvia)
Prof. Janusz Kacprzyk (Poland)
Prof. Ignacy Kaliszewski (Poland)
Prof. Bożena Kostek (Poland)
Prof. Tomas Krilavičius (Lithuania)
Prof. Olga Kurasova (Lithuania)
Assoc. Prof. Tatiana Tchemisova (Portugal)
Prof. Julius Žilinskas (Lithuania)

Organizing Committee:

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

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt
Tel. +370 5 2109315

https://doi.org/10.15388/DAMSS.14.2023
ISBN 978-609-07-0985-6 (digital PDF)

Investigation of Speech Signal Processing Parameters in Wave-U-Net Source Separation

Justina Ramonaitė, Pooja Gore, Gražina Korvel, Gintautas Tamulevičius

Institute of Data Science and Digital Technologies
Vilnius University

justina.ramonaite@mif.stud.vu.lt

Separating clean speech from noise in order to enhance speech quality and intelligibility is a challenging task known as speech denoising, where the input is noisy speech. The challenge is mainly caused by non-stationary noise and low signal-to-noise ratio (SNR). Deep learning models are being increasingly used to solve this task due to their superior performance in non-stationary noisy environments compared to conventional approaches. Deep learning methods model the nonlinear relationship between clean speech and noisy speech signals without prior knowledge of noise statistics. In this study, we focus on predicting end-to-end audio source separation. We use a waveform-based method, namely Wave-U-Net, which is an adaptation of the U-Net architecture to a one-dimensional time domain. The goal of this research is to investigate the correlation between Wave-U-Net performance and speech processing parameters, including speech sampling frequency and frame length. For this purpose, the speech signal is downsampled using sampling frequencies of 8 kHz, 11.025 kHz, 16 kHz, and 48 kHz, and divided into short-time intervals ranging from 10 ms to 50 ms in 10 ms increments. The experimental results reveal the significant influence of these parameters on denoised speech quality.