



# DAMSS

DATA ANALYSIS  
METHODS FOR SOFTWARE  
SYSTEMS



**15th Conference on**

# **DATA ANALYSIS METHODS for Software Systems**

**November 28–30, 2024**

Druskininkai, Lithuania, Hotel “Europa Royale”

<https://www.mii.lt/DAMSS>

**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. **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)

Assoc. Prof. **Gintautas Tamulevičius** (Lithuania)

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*

Tel. (+370 5) 2109 315

Prof. Olga Kurasova

*olga.kurasova@mif.vu.lt*

Copyright © 2024 Authors. Published by Vilnius University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.15388/DAMSS.15.2024>

ISBN 978-609-07-1112-5 (digital PDF)

© Vilnius University, 2024

# Noise Dataset for Deep Learning-Based Speech Enhancement

Faustas Žilijaėvas, Justina Ramonaitė, Daniel Zakševski,  
Gražina Korvel, Gintautas Tamulevičius

Institute of Data Science and Digital Technologies  
Vilnius University

*grazina.korvel@mif.vu.lt*

In the real world, ambient noise often impeded human speech, such as passing cars, background chatter, or other environmental sounds. While humans have adapted to filter out most of this noise, computers struggle to do the same. Even with advanced noise cancellation algorithms, unpredictable and ever-changing noise remains a significant challenge. Researchers use many types of real-world noise to improve these algorithms to train artificial neural networks. This requires diverse and regularly updated noise datasets. The goal of this work is to create a noise dataset that will help advance this field. This work presents a comprehensive noise dataset recorded in Vilnius, Lithuania, in 2024. The recordings capture high-quality single-channel audio at a sampling rate of 44.1 kHz, stored in WAV format, and quantized at 24 bits. The dataset was produced using a Shure SM78b dynamic microphone and a MiX-Prie 3 audio interface. The dataset consists of a variety of indoor and outdoor sounds. Long-duration sounds, each lasting 30 minutes, include construction noise, a city street, a bus stop, neighborhood sounds, forest ambiance, rainfall, and the sound of a vaporizer. In addition to these longer recordings, the dataset includes shorter audio samples of specific sounds, such as a garbage truck, a kettle, a coffee maker, a car wash, and a microwave oven. The variety of noise recordings and multichannel format of the dataset make it an optimal choice for tasks such as noise classification, speech enhancement, and noisy speech recognition.

**Acknowledgements:** This project has received funding from the Research Council of Lithuania (LMTLT), agreement No S-MIP-24-118.