16th Conference on

DATA ANALYSIS METHODS
for Software Systems

November 27–29, 2025

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

Co-Chairs:

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

Programme Committee:

Dr. **Jolita Bernatavičienė** (Lithuania)
Prof. **Juris Borzovs** (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
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt
Tel. (+370 5) 2109 315

Initial Study of Lithuanian Emotional Speech Synthesis

**Gediminas Navickas, Gražina Korvel,
Gintautas Tamulevičius**

Institute of Data Science and Digital Technologies
Vilnius University

gediminas.navickas@mif.vu.lt

Emotional speech synthesis is a complex area of research that aims to generate speech that sounds natural and conveys human emotions. Despite the rapid progress of neural text-to-speech (TTS) methods, synthesis of emotional expression poses significant challenges in all languages, even high-resourced ones. The main challenges are related to still clearly undefined acoustic features of emotions, different levels and types of emotions (e.g., cold anger and hot anger), mixed emotions, limited interpretability and control of the emotional speech synthesis process. Lastly, the absence of an emotional speech corpus also restricts the capabilities of modern models. For low-resource languages (such as Lithuanian), these challenges and tasks become even more complex.

Recent literature has identified a shift from traditional rule-based and statistical parametric methods to deep generative approaches. Emotional TTS systems are based on deep neural networks (DNNs), variational autoencoders (VAEs), generative adversarial networks (GANs), transformers, and diffusion models. These models lead to the following emotional speech synthesis strategies:

- Explicit training of models using an emotionally labelled speech corpus.
- Transfer learning of emotional speech, thus avoiding the need for large amounts of data.
- Semi-supervised training methods, based on learning from both unlabelled and labelled data.

To achieve synthesis control and interpretability, another paradigm should not be dismissed: modifying neutral synthesized speech to provide the desired emotional content. This paradigm would require

a detailed analysis of emotional speech, a large corpus of emotional speech data, and additional models for the speech transformation.

Initiating the study of emotional Lithuanian speech synthesis, we began by assessing the State-of-the-Art methods in speech synthesis, the availability of emotional speech corpora, and model transferability. This report summarizes the main trends, methods, and challenges identified in recent studies, outlining how these insights can be used in the development of emotional speech synthesis for the Lithuanian language.