Masterthesis

Evaluation of Speech Quality Using Synthetic Data and Transfer Learning

Description:

Based on the principle of ultrasound beam focusing, Magnetic Resonance-guided Focused Ultrasound (MRgFUS) enables the precise and non-invasive ablation of brain tissue through localized heat induction. This technique offers significant advantages in functional neurosurgery, as real-time Magnetic Resonance Imaging (MRI) monitoring allows for accurate control and assessment of the treatment process [1]. Currently, MRgFUS is being utilized in the treatment of movement disorders, such as Parkinson's disease-related tremor and essential tremor, demonstrating its potential as a precise and minimally invasive therapeutic approach [2]. Since these disorders are often associated with speech impairments, a comprehensive assessment of speech quality throughout the treatment process is essential. This allows for the identification of potential side effects and ensures a holistic evaluation of the therapy's impact [3,4].

A previous study evaluated speech data across five categories using a survey and trained a Convolutional Neural Network (CNN) with the resulting scores as labels to automatically assess speech quality based on Mel-Frequency Cepstral Coefficients (MFCCs). However, when applied to newly evaluated data, the model exhibited a decrease in prediction performance, suggesting overfitting to the training set and limited suitability for clinical application [5].

Building upon the entire labeled dataset from the prior study, the present work aims to develop a new model that integrates synthetic data generation and transfer learning to address these limitations and enhance speech assessment. Following an extensive literature review, generative AI methods will be employed to produce synthetic data, which, as predicted by [6], can help to cover edge cases within the long-tail data distribution and thus augment small datasets in a targeted manner.

Furthermore, various pre-trained audio feature extraction networks, trained on millions of data samples, will be utilized to derive high-level audio representations. These feature vectors will then serve as input to a custom-trained Multi-Layer Perceptron (MLP) for the final speech quality evaluation. Model performance will be assessed on a held-out, stratified test set to ensure robustness and generalizability.

Ideally, the final model will be integrated into an existing C/C++-based signal-processing pipeline that supports automated audio acquisition and multiple configurable feature-extraction pathways. Embedding the trained network into this framework would enable a seamless end-to-end workflow—from real-time speech recording through feature computation to direct inference of speech quality scores. Such an integration would allow newly captured patient data to be assessed immediately and consistently, providing a complete and efficient processing chain while still leaving room for future adaptations or extensions.

Tasks:

- Speech Analysis and Data Preprocessing
- Synthetic Data Generation
- Transfer Learning

Requirements:

- Interest in medical topics
- Good knowledge of Python and C/C++ programming
- Prior knowledge of deep learning and deep learning frameworks like PyTorch (ideally for audio applications)
- Structured and independent way of working

Contact / Supervision:

Prof. Dr.-Ing. Gerhard Schmidt	DSS, Kiel University E-Mail: gus@tf.uni-kiel.de Phone: +49-431-880-6125	Karolin Krüger, M. Sc.	DSS, Kiel University E-Mail: kkru@tf.uni-kiel.de Phone: +49-431-880-6141
Prof. Dr. Sören Pirk	VCAI, Kiel University E-Mail: sp@informatik.uni-kiel.de Phone: +49-431-880-7064	Marten J. Finck, M. Sc.	VCAI, Kiel University E-Mail: mafi@informatik.uni-kiel.de Phone: +49-431-880-7096
Dr. med. Steffen Paschen	Medical Faculty, UKSH / Kiel University E-Mail: steffen.paschen@uksh.de Phone: +49-431-500- 23980		

References:

[1] Trumm, C. G., Napoli, A., Peller, M., Clevert, D. A., Stahl, R., Reiser, M., & Matzko, M. (2013). MR-gesteuerter fokussierter Ultraschall. Der Radiologe, 53(3), 200-208.

[2] Lev-Tov, L., Barbosa, D. A., Ghanouni, P., Halpern, C. H., & Buch, V. P. (2022). Focused ultrasound for functional neurosurgery. Journal of Neuro-Oncology, 156(1), 17-22.

[3] Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and Hearing Disorders, 43(1), 47-57.

[4] Hägglund, P., Sandström, L., Blomstedt, P., & Karlsson, F. (2016). Voice tremor in patients with essential tremor: effects of deep brain stimulation of caudal zona incerta. Journal of Voice, 30(2), 228-233.

[5] Finck, M., Krüger, K., Paschen, S., Helmers, A. K., & Schmidt, G. (2024). Analyzing Impaired Speech in Context of Magnetic Resonance-guided Focused Ultrasound Using Convolutional Neural Networks. DEGA eV (Hrsg.), Fortschritte der Akustik–DAGA 2024, 2024.

[6] Nolte, L., Finck, M. J., Pirk, S., & Tomforde, S. (2025). Feedback-guided Dataset Shaping for Automated Downstream Task Optimization. In International Conference on Architecture of Computing Systems (pp. 477-481). Cham: Springer Nature Switzerland.