Investigating MIKNN for Objective Speech Intelligibility Assessment of Dysarthric Speech

Kruthika Reddy Kowkuntla | Supervisors: Jorge Martinez Castaneda, Dimme de Groot

k.r.kowkuntla-1@student.tudelft.nl

EEMCS, Delft University of Technology, The Netherlands

TUDelft

Background

- Dysarthria is a motor speech disorder that is a common symptom of neurological diseases and it results from disturbances in muscular contraction necessary for speech.
- Dysarthria significantly affects the phonation, prosody, resonance and articulation for the speech production leading to reduced intelligibility.
- Traditional subjective assessments of speech intelligibility are resource-intensive, time-consuming, and prone to bias.
- MIKNN (Mutual Information with K-Nearest Neighbours) is an objective speech intelligibility measure in the speech enhancement domain.
- MIKNN uses information theory to evaluate intelligibility by quantifying mutual dependencies in speech signals.
- Research gap: there is very limited research on how the objective intelligibility metrics can be used for assessment of dysarthria speech.
- Previous studies have explored Short-Time Objective Intelligibility measure (STOI) and ESTOI (Extended STOI) as the metric to assess pathological speech [1]. This research aims to use MIKNN as the metric to do the same.
- STOI and ESTOI, rely on short-term temporal resolutions to assess intelligibility, while MIKNN tries to address both linear and non-linear dependencies.

Research Question

To investigate the applicability of MIKNN as an objective intelligibility measure for dysarthric speech.

Methodology

Signal preprocessing: speech signals were converted to time-frequency representations using Short-Time Fourier Transform (STFT) and were mapped onto 1/3 octave band scales.

Reference Signal Construction: created utterance-specific reference templates by averaging aligned representations from multiple healthy speakers.

Dynamic Time Warping (DTW): applied DTW to align the pathological speech signals with clean reference templates, accounting for temporal variability.

MIKNN-Based Intelligibility Assessment: used MIKNN to compute mutual information between the temporal envelopes of aligned pathological and reference speech signals. Calculated normalized intelligibility scores based on mutual information.

Experimental Setup and Evaluation

Dataset: UA-Speech; 10 spastic dysarthric speakers were considered and 13 healthy speakers as references.

Mapping function: a logistic mapping function applied to align objective scores (MIKNN) with subjective intelligibility ratings.

State-of-the-art metrics: P-STOI, P-ESTOI, K_{LP}, %V, σ_Δ, Δ_{f0}, LHMR[2]

Comparative Analysis: Compared MIKNN's performance against state-of-the-art metrics using Pearson (R) and Spearman (R_s) correlation coefficients and their p-values.

Measures	R	p	R_s	P
MIKNN	**0.66**	0.0389	0.62	0.0537
P-STOI	**0.90**	5E-04	**0.82**	7E-03
P-ESTOI	**0.95**	4.3E-5	**0.91**	2E-04
K_{LP}	0.41	0.23	0.42	0.23
σ_Δ	0.45	0.2	0.51	0.13
%V	-0.40	0.25	-0.58	0.08
Δ_{f0}	**-0.70**	0.02	-0.61	0.06
LHMR	-0.55	0.09	-0.54	0.10

Table 1: Performance of MIKNN and other state-of-the-art measures.
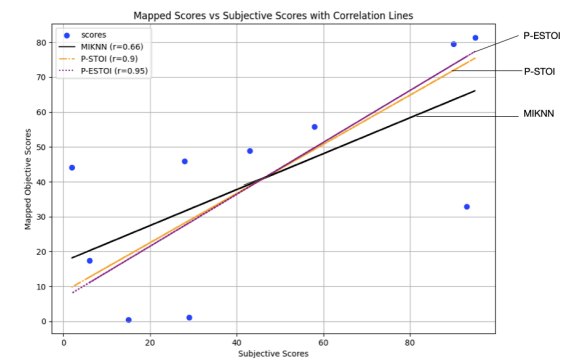
Results

Figure 1: Scatter plot of the intelligibility scores. The vertical axis is the 'mapped objective intelligibility scores' and horizontal axis is the 'subjective intelligibility scores'. Each line represents the linear fit corresponding to the respective measure, with the slope and alignment of the lines indicating the strength of correlation between the subjective and objective scores.

Conclusions

- Results suggest that MIKNN performs better than metrics like K_{LP}, %V, σ_Δ, Δ_{f0} and LHMR in terms of correlation with subjective scores.
- MIKNN shows a significantly moderate correlation with subjective intelligibility scores. **R = 0.66** with **p < 0.05**
- However, it is outperformed by P-STOI and P-ESTOI.
- Correlation results (p < 0.05) indicate MIKNN as a potential complementary tool for objective dysarthric speech intelligibility assessment.

References

[1] P. Janbakhshi, I. Kodrasi and H. Bourlard, "Pathological Speech Intelligibility Assessment Based on the Short-time Objective Intelligibility Measure," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 6405-6409, doi: 10.1109/ICASSP.2019.8683741.

[2] T. H. Falk, W.-Y. Chan, and F. Shein, "Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility," *Speech Communication*, vol. 54, no. 5, pp. 622–631, 2012, doi: 10.1016/j.specom.2011.03.007.