Spectral Subspace Analysis for Automatic Assessment of Pathological Speech Intelligibility

Parvaneh Janbakhshi[1,2], Ina Kodrasi[1], Hervé Bourlard[1,2]

[1]Idiap Research Institute, Speech and Audio Processing Group, Martigny, Switzerland

[2]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

{parvaneh.janbakhshi, ina.kodrasi, herve.bourlard}@idiap.ch

Motivation

- Objective pathological speech intelligibility assessment is crucial for the management of speech disorders
- P-ESTOI (our previously proposed technique) was shown to be a successful objective intelligibility measure for pathological speech
- P-ESTOI requires time-alignment using Dynamic Time Warping (DTW)
  - High computational complexity
  - Inapplicable to text-independent scenarios

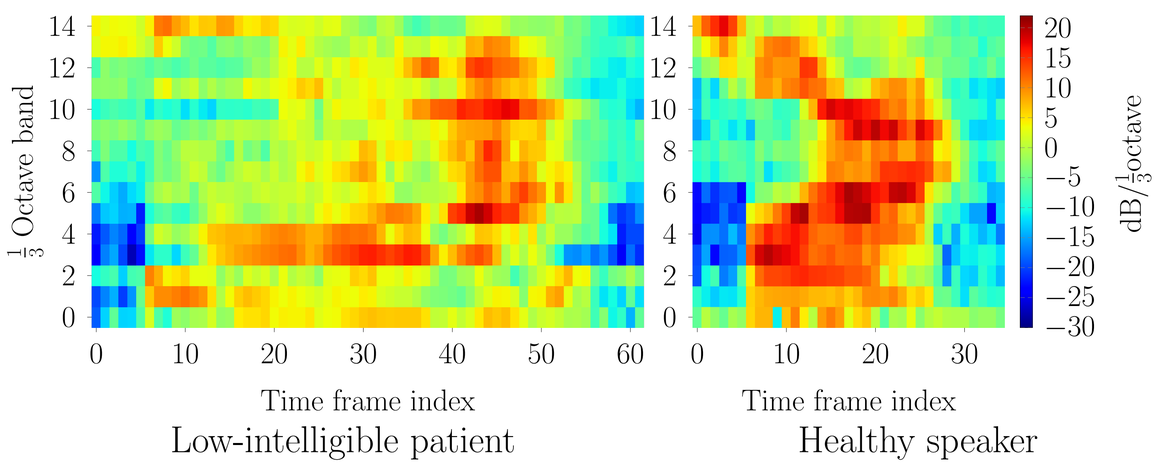
Aim

Automatic pathological speech intelligibility measure, which

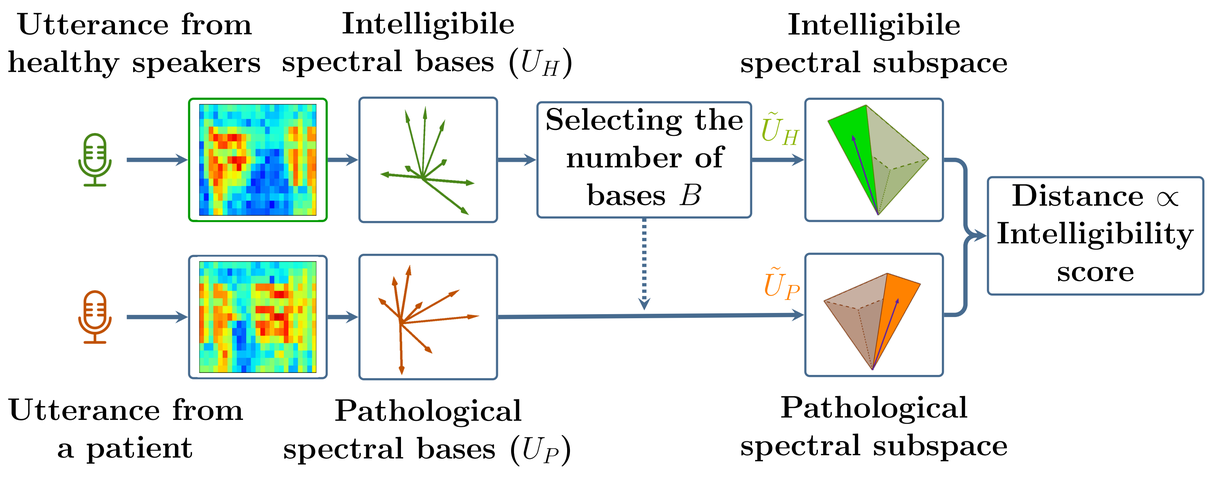
- is reliable and efficient
- does not require time-alignment or large amounts of training data
- is applicable to text-independent scenarios

Hypothesis

- Difference in spectral patterns of intelligible and pathological speech ⇒ Intelligibility



Proposed method



1. Computing intelligible spectral bases from healthy speakers' utterances
2. Computing pathological spectral bases from pathological test utterance
3. Automatic selection of the number of dominant spectral bases
4. Computing a distance measure between the spectral subspaces

Computing spectral bases

- Principal Component Analysis (PCA)
- Approximate Joint Diagonalization (AJD)

Automatic selection of the number of spectral bases B

- L-curve method

Computing a distance measure between spectral bases

- Grassman distance measure

$$d_{CF} = 2\sqrt{\sum_{i=1}^{B} sin^2(\theta_i/2)}$$

Evaluation

Database

- 15 English-speaking Cerebral Palsy (CP) patients and 13 healthy speakers

Considered scenarios

i) Word-level analysis
ii) Text-level analysis
iii) Text-independent analysis with several common words across healthy and pathological speakers
iv) Text-independent analysis without any common words across healthy and pathological speakers

Criteria

- Pearson (R) and Spearman rank (R_s) correlation coefficients between estimated scores and the subjective intelligibility scores (along with p-values)

Comparison

- State-of-the-art reference-based measures, i.e., P- ESTOI and iVector-based regression method

Results

Measures	R	p	R_S	p
i) Word-level				
P-ESTOI	0.94	2.5e−7	0.94	9.3e−7
iVector	0.74	–	–	–
AJD	−0.82	1.7e − 4	−0.88	1.8e−5
PCA	−0.83	1e−4	−0.88	1.8e−5
ii) Text-level				
P-ESTOI	0.93	9e−7	0.95	3e−7
AJD	−0.80	3.9e−4	−0.78	5.7e−4
PCA	−0.81	2.4e−4	−0.83	1.2e−4
iii) Text-level with several common words				
AJD	−0.79	6e−4	−0.78	9.5e−4
PCA	−0.78	0.008	-0.76	0.019
iv) Text-level without any common words				
AJD	−0.78	9.5e−4	−0.77	0.001
PCA	−0.7	0.009	−0.65	0.047

- P-ESTOI gives the highest correlations values while being computationally more expensive than the proposed measure and cannot be used in text-independent scenarios
- Proposed subspace-based intelligibility measure (PCA or AJD) achieves high and significant correlations in all considered scenarios.

Conclusion

- Proposed measure can be used as a reliable and efficient objective intelligibility measure for pathological speech without imposing any constraints on the speech material