Suppressing Noise Disparity in Training Data for Automatic Pathological Speech Detection

Mahdi Amiri

Problem Statement

- **Assumption**: recordings from both healthy and pathological speakers are acquired in identical noise-free environment

- [1] demonstrate this assumption is not valid for commonly used datasets

- DL-based approaches learn noise-discriminant cues instead of actual pathological-discriminant ones

- Propose an approach to suppress noise disparity

[1] G. Schu et al., “On using the UA-Speech and TORGO databases to validate automatic dysarthric speech classification approaches”, ICASSP 2022

Proposed Method

$$y_h = s_h + n_h, \qquad y_p = s_p + n_p$$

$$\hat{y}_h = \underbrace{s_h + n_h}_{y_h} + \alpha_h \hat{n}_p, \quad \hat{y}_p = \underbrace{s_p + n_p}_{y_p} + \alpha_p \hat{n}_h,$$

- Conditions for having same noise characteristics:

$$\left.\begin{aligned} \mathrm{SNR}_{\hat{y}_h}^{n_h} &= \mathrm{SNR}_{\hat{y}_p}^{\alpha_p \hat{n}_h} \\ \mathrm{SNR}_{\hat{y}_h}^{\alpha_h \hat{n}_p} &= \mathrm{SNR}_{\hat{y}_p}^{n_p} \\ \mathrm{SNR}_{\hat{y}_h}^{n_h + \alpha_h \hat{n}_p} &= \mathrm{SNR}_{\hat{y}_p}^{n_p + \alpha_p \hat{n}_h} \end{aligned}\right\} \Rightarrow \alpha_h = \sqrt{\frac{P_{s_h}}{P_{s_p}}}, \quad \alpha_p = \sqrt{\frac{P_{s_p}}{P_{s_h}}}$$

Experimental Results

- Models learn noise-discriminant cues in the standard setting
- Oracle setting confirms the validity of the proposed method
- Performance gap between oracle and practical settings

Test	CNN-based			wav2vec2-based		
	Standard	Oracle	Practical	Standard	Oracle	Practical
Noisy	99.5 ± 0.5	71.5 ± 6.5	76.7 ± 3.8	95.5 ± 2.8	78.5 ± 2.0	86.2 ± 3.0
Clean	51.3 ± 1.9	70.3 ± 1.9	68.2 ± 1.7	53.6 ± 5.9	80.0 ± 2.1	77.6 ± 5.0

Thank you!