Equal Speech : Exploring Non-Native Accent Bias in OIMs

David Dobin

Understanding speech is an important part of our daily routine , and sometimes it can even save lives. Non-Native Speakers generally struggle more to be understood by their peers, especially under high background noise [1].

SI (Speech Intelligibility) is often analyzed using SIMs (Subjective Intelligibility Metrics), which consist of survey results, or OIMs (Objective Intelligibility Metrics) , which are based on mathematical score estimates. OIMs often get trained or tested on SIM results sourced from Native Speaker data. [2]

This research investigates potential OIM estimation biases towards Non-Native-accented speech , and explores the potential influence of sound distortions and the environment on OIM prediction accuracy.

Research Questions

To what extent do modern OIM solutions generalize to non-native speech?

- To what extent does Speech-Shaped-Noise distortion affect OIM performance on Non-Native speech samples?

- How well does OIM estimate Non-Native SI?

- To what extent can Non-Native speech be used to estimate the average intelligibility in a certain environment?

Hypothesis

OIMs will perform less accurately on L2 (Non-Native) speech samples than on L1 (Native) ones in all areas.

Methodology

- **Dataset:** SpeechBox ALLSSTAR[3] Dataset with Native + Non-Native (Hebrew, Mandarin, Cantonese, Turkish) Speech Samples
- **OIMs:** SIIB[2], STOI[4], SRMR[5]
- **Reference Metrics:** SNR (Signal To Noise Ratio), SIM[3] Scores, STIPA[6] (Environment Intelligibility Measure)

Bias Calculation

1.Pearson's Rho

$$\rho = \frac{\sum_i (s_i - \mu_s)(f(d_i) - \mu_{f(d)})}{\sqrt{\sum_i (s_i - \mu_s)^2 \sum_i (f(d_i) - \mu_{f(d)})^2}}$$

2.RMSE

$$\sigma = \sqrt{\frac{1}{S}\sum_i (s_i - f(d_i))^2}$$

***Bias = Difference in L1-L2 Correlations**

Linearization Function for Pearson's Rho:

$$\frac{1}{1 + e^{a\left(\frac{d}{x}\right)+b}}$$

Experiments:

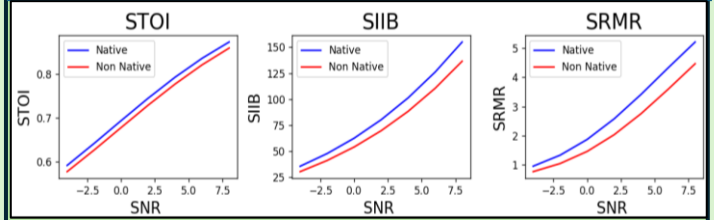
1. The Influence of Speech-Shaped-Noise Distortion on L1/L2 OIM Performance

2. OIM Accuracy of SI estimation (relative to SIM results) for L1/L2 speech.

3. OIM Accuracy of average SI estimation for an environment with L1/L2 Speech

Results

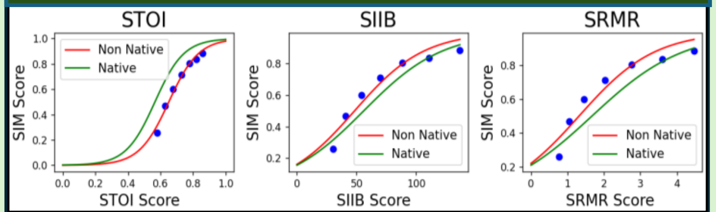
1.Influence of Distortion on OIM Evaluation
(SNR-OIM Correlation)



Pearson Rho (ρ) Diff.			RMSE (σ) Diff,		
STOI	SIIB	SRMR	STOI	SIIB	SRMR
-0.0003	0.0007	**0.0065**	0.0017	0.0046	**-0.034**

All Metrics contain a slight bias against L2 speech for **all SNRs**
Highest bias found in **SRMR** (ρ=0.0065, σ=0.034)

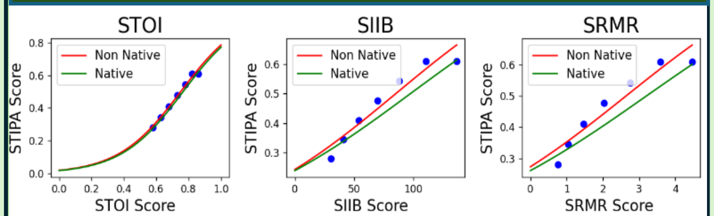
2.Accuracy of L2 SI Estimation
(SIM-OIM Correlation)



Pearson Rho (ρ) Diff.			RMSE (σ) Diff,		
STOI	SIIB	SRMR	STOI	SIIB	SRMR
0.007	0.024	**0.034**	0.164	**-0.164**	-0.149

SIIB + SRMR contain similar bias against L2 speech
STOI contains bias against L1 speech
Highest bias found in **SIIB** (ρ=0.024, σ=0.164)

3.Accuracy of Average SI Estimation in an Environment (STIPA-OIM Correlation)



Pearson Rho (ρ) Diff.			RMSE (σ) Diff,		
STOI	SIIB	SRMR	STOI	SIIB	SRMR
-0.001	-0.004	**-0.016**	**-0.014**	0.007	-0.001

All Metrics contain a slight bias against L2 speech
All perform with approximately the same efficiency

Future Considerations:

-Using different datasets, other than ALLSSTAR
-Exploring different OIMs (PESQ, SII)

[1] P. Adank, B. G. Evans, J. Stuart-Smith, and S. K. Scott, "Comprehension of familiar and unfamiliar native accents under adverse listening conditions," Journal of Experimental Psychology: Human Perception and Performance",2009
[2] S. Van Kuyk, W. B. Kleijn, and R. C. Hendriks, "An instrumental intelligibility metric based on information theory," IEEE Signal Processing Letters, vol. 25, no. 1,pp. 115–119, 2018.
[3] A. R. Bradlow, "ALLSSTAR: Archive of L1 and L2 Scripted and Spontaneous Transcripts And Recordings," retrieved from SpeechBox database.
[4] R. H. J. J. C.H. Taal, R.C. Hendriks, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech, "IEEE Trans. Audio Speech Lang. Process., p. 2125–2136,2011
[5] C. Z. T. H. Falk and W. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech, " IEEE Trans. Audio, Speech, Lang. Process., 2010.
[6] Speech Intelligibility Measurement with the XL2 Analyzer, NTI, 2020'