On Crowdsourcing-Design with Comparison Category Rating for Evaluating Speech Enhancement algorithms
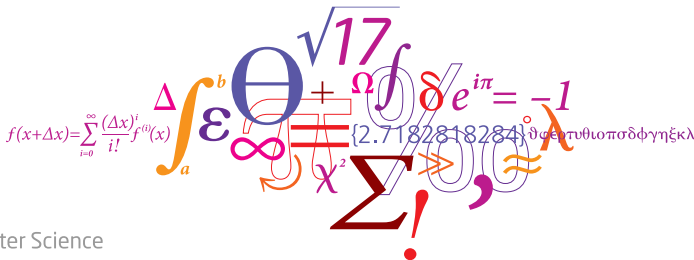
Angélica S. Z. Suárez [1], Clément Laroche [2], Line H. Clemmensen [1], Sneha Das [1]

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark

[2] GN Audio Research, Ballerup, Denmark

Objectives and Contributions

Objectives:

❶ Assess the suitability of CCR for speech enhancement

Contributions:

❶ User ratings are bimodal $\rightarrow$ We explore models that account for the bimodality.

❷ We assess the suitability of the CCR methodology to evaluate speech enhancement algorithms and investigate CCR ratings against objective measures.

❸ Design choices in terms of the conditions, number of clips and speaker types and their influence on the crowd responses are discussed.

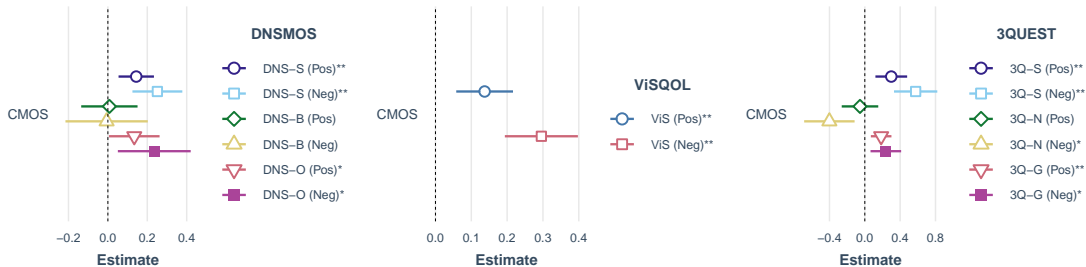❹ Dataset generated in this work is open-sourced.

Dataset design

- Clean speech: VCTK
- Background noise: 2017 DCASE dataset.
- Transient noise: Freesound
- RTC algorithms: Teams, Zoom, Amazon, Chime
- RTC(noisy-signal): Enhanced signal.

RESULTS

1. Biomodality in CMOS
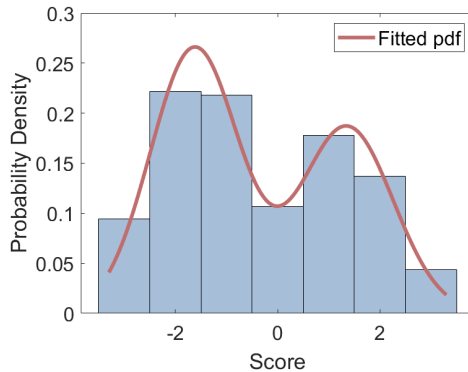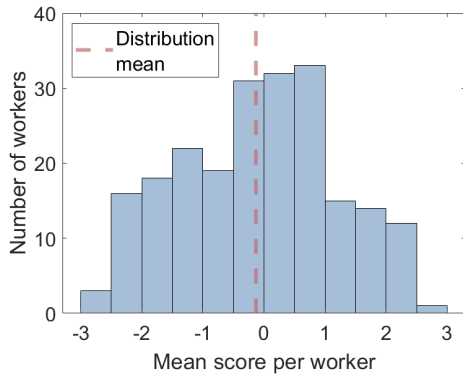
2. CCR-CMOS versus objective metric

① Higher coefficients for scores from negative mode.

② Highest (significant) correspondence of CMOS to 3QUEST.

3. Crowdsourcing-design based on statistical evidence

1. Influence of condition parameters:

$$C_{MOS_{P-N}} = \alpha + \beta SNR + \gamma S_L + \delta N_E + \eta N_T + \zeta F_{alg} + \epsilon. \tag{1}$$

- SNR shows large coefficients for both modes $\implies$ higher influence on the crowd responses.
- Speech-level seems to effect the negative mode
- Noise-type, specifically park shows higher predictability for the positive mode.

3. Crowdsourcing-design based on statistical evidence

2. Limits on crowd-size
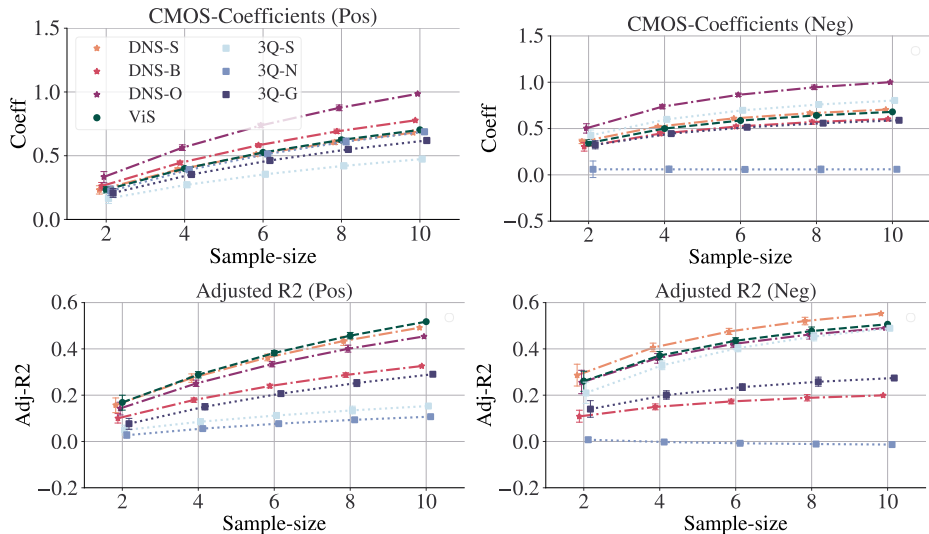
Conclusions

❶ Analyse positive-negative modes of responses separately or incorporate positive-negative affiliation within analysis model.

❷ Higher and significant correlation $\rightarrow$ positive workers and noise type.

❸ Correlation better between negative workers and speech level.

❹ SNR showed significant effect over both positive and negative groups of CMOS responses.