ARIA
APPLIED RESEARCH IN ACTION

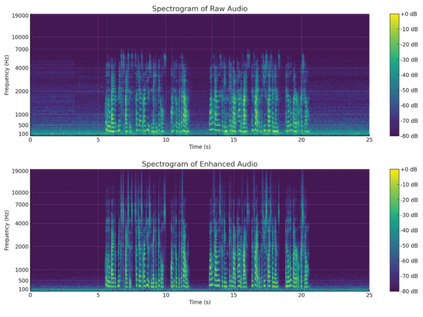
Towards Broadcast-Quality Audio: Enhancement of Conference Microphones and Exploration of Objective Evaluation Techniques

Merging Deep Learning and Established Metrics for Accurate Audio Quality Assessment

Wenzhe Xu

Gerald Penn
ACADEMIC SUPERVISOR

Paul Ferry
INDUSTRY SUPERVISOR



	PESQ[1]	STOI[2]	SRMR[3]	Composite[4]	NISQA[5]	DNSMOS[6]
SIG	0.19	0.07	0.19	0.48	0.79	0.84
BAK	0.07	0.05	0.03	0.07	0.58	0.35
OVRL	0.19	0.01	0.22	0.46	0.77	0.80

Table 1: Estimated correlation coefficients for six objective measures

PROJECT SUMMARY

Modern audiences demand high-quality audio in speech content to ensure listener engagement and satisfaction. Yet, recordings from consumer-grade equipment often exhibit quality degradations, including noise, reverberation, and equalization distortion. While traditional enhancement methods are adept at addressing specific issues, they often fall short in achieving optimal speech quality.

This study investigates the combination of various speech enhancement techniques, aiming to emulate professional broadcast microphone output using audio from conference microphone systems. In addition, emphasis is placed on the exploration of objective metrics for assessing the enhanced audio quality. Many benchmarked quality measures, although designed for specific domains, are often employed beyond their intended scope. This mismatch can result in unreliable quality estimations. Our research contrasts traditional metrics frequently referenced in academia with more recent deep-learning-based methods. By consolidating these objective measures through linear and nonlinear regression analysis, we aim to derive a metric with improved correlation that better aligns with human-perceived audio quality. This represents a pivotal step towards an automated evaluation process that parallels Mean Opinion Scores from human listening panels.

REFERENCES

[1] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of Speech Quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221). doi:10.1109/icassp.2001.941023

[2] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, 2010. doi:10.1109/icassp.2010.5495701

[3] T. H. Falk, C. Zheng, and W.-Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 7, pp. 1766–1774, 2010. doi:10.1109/tasl.2010.2052247

[4] Y. Hu and P. C. Loizou, "Evaluation of objective measures for speech enhancement," Interspeech 2006, 2006. doi:10.21437/interspeech.2006-84

[5] G. Mittag, B. Naderi, A. Chehadi, and S. Möller, "Nisqa: A deep CNN-self-attention model for multidimensional speech quality prediction with crowdsourced datasets," Interspeech 2021, 2021. doi:10.21437/interspeech.2021-299

[6] C. K. Reddy, V. Gopal, and R. Cutler, "Dnsmos: A non-intrusive perceptual objective speech quality metric to evaluate noise suppressors," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021. doi:10.1109/icassp39728.2021.9414878

nureva

Computer Science
UNIVERSITY OF TORONTO
Master of Science in Applied Computing