

The VoiceMOS Challenge 2022

Wen-Chin Huang^{1,a)} Erica Cooper² Yu Tsao³
Hsin-Min Wang³ Tomoki Toda¹ Junichi Yamagishi²

Abstract: We present the VoiceMOS Challenge, a scientific event that aims to encourage research in automatic prediction of Mean Opinion Scores (MOS) for synthesized speech. While human listening tests have been the gold standard for evaluating synthesized speech, the high cost motivates the development of automatic objective measures. However, recent popular data-driven neural network-based models often have a low correlation with human ratings, and the generalization abilities of current data-driven quality prediction systems suffer significantly from domain mismatch. The focus of the VoiceMOS challenge is on understanding and comparing current MOS prediction techniques using a standardized dataset. It has been accepted as a special session at INTERSPEECH 2022 and has attracted attention from research teams worldwide.

This talk introduces details of the VoiceMOS challenge, including the datasets, task designs, baseline systems, and the analyses of results. We also discuss future directions for the next version of the challenge.

Keywords: VoiceMOS, MOS prediction, speech quality assessment

1 Nagoya University, Japan

2 National Institute of Informatics, Japan

3 Academia Sinica, Taiwan

a) wen.chinhuang@g.sp.m.is.nagoya-u.ac.jp