2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Performance Enhancement of a Pathological Voice Quality Evaluation System Using a Self-Attention Model

Ching-Ju Hsiao, Ji-Yan Han, Wei-Zhong Zheng, Guan-Min Ho, Chia-Yuan Chang, Ying-Hui Lai*

Abstract— **Auditory-perceptual evaluation is a common assessment used for evaluating the voice quality of patients; however, it suffers from inter- and intra-rater reproducibility problems. To this end, we propose an objective evaluation metric utilizing multi-input self-attention technology to evaluate voice quality using the grade (G), roughness (R), and breathiness (B) scale. The results showed that our proposed system has better performance accuracies of 69.25% for G, 77.5% for R, and 82.25% for B, as compared to that reported in previous studies. The findings also suggests that our system serves as a potential approach for voice quality evaluation applications in the future.**

I. Introduction

The GRBAS scale is a common auditory-perceptual approach that uses five categorical traits, namely grade (G), roughness (R), breathiness (B), asthenia (A), and strain (S), to evaluate a patient's voice quality using four scales (0 to 3). Although the GRBAS scale is widely used in clinical practice, it suffers from a critical problem of subjectivity; hence, inter- and intra-rater issues [1] often appear in real application conditions when using this scale. We believe that an objective evaluation metric can alleviate this issue. Following this concept, we proposed a voice quality evaluation system using deep learning. We aim to provide patients with voice therapy a highly accurate and objective system to quantify the progress of their therapy based on voice quality evaluation.

II. Method

Fig. 1 shows the flow chart of our proposed system. As can be seen, we have used the multi-input self-attention model in our system [2]. Specifically, the multi-input self-attention model considers the relationship between pitch, vowel, and time sequence of sustained phonations. Two critical units, namely long short-term memory (LSTM) and self-attention model, are used in our system. The Saarbrücken voice database [3] was used to train and test the proposed system; 80% of the data used for training and 20% for testing. Furthermore, because of the traits A and S were unreliable, a simpler GRB scale was used in this database. Finally, the confusion matrix and accuracy were compared to the results of a previous study to evaluate the advantages of the proposed system [3].

III. Results

Based on the results (Fig. 2), our proposed system exhibited a higher accuracy of 69.25% on G, 77.5% on R, and 82.25% on B as compared to the baseline system [2], which showed accuracies of G, R, and B on 60.61%, 55.29%, and 60.75%, respectively. This implies that our method can evaluate the quality of voice from two aspects, namely pitch and vowel, which is beneficial for addressing inter- and intra-rater issues, as mentioned earlier.

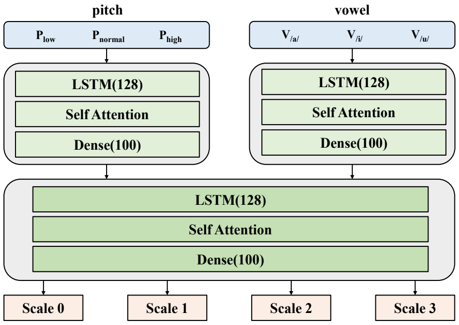

Fig. 1. Proposed system

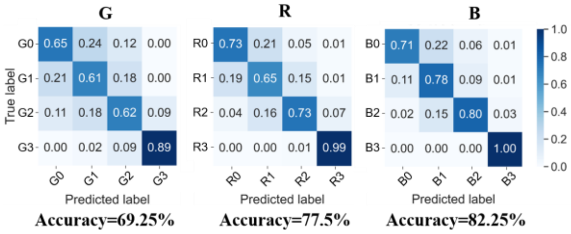

Fig. 2. Confusion matrix and the accuracy of G, R and B.

IV. Conclusion

This study proposed an objective evaluation metric for a voice quality evaluation system with a multi-input self-attention structure. The results showed that the proposed system performs better than the classical baseline system. The results also suggest that the proposed system with its multi-input self-attention structure is a potential approach for further improving the accuracy of voice quality evaluation tasks.

Acknowledgment

This study was supported by the Ministry of Science of Technology of Taiwan under the 110-2218-E-A49A-501 project and industry–academia cooperation project with APrevent Medical Inc. (109J052).

References

[1] M. S. D. Bodt, F. L. Wuyts, P. H. V. D. Heyning, and C. Croux, "Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality," vol. 11, no. 1, pp. 74-80, 1997.

[2] Ashish Vaswani et al., "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[3] J. D. A. Londoño, J. A. G. García, and J. I. G. Llorente, "Multimodal and multi-output deep learning architectures for the automatic assessment of voice quality using the grb scale," vol. 14, no. 2, pp. 413-422, 2019.

Ching-Ju Hsiao, Ji-Yan Han and Wei-Zhong Zheng are with the Department of Biomedical Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan.

Guan-Min Ho, Chia-Yuan Chang are with the APrevent Medical Inc.

*Ying-Hui Lai is with the Department of Biomedical Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan. *corresponding author, phone: +886-2-28267021; e-mail: yh.lai@nycu.edu.tw

978-1-7281-1178-0/21/$31.00 ©2021 IEEE