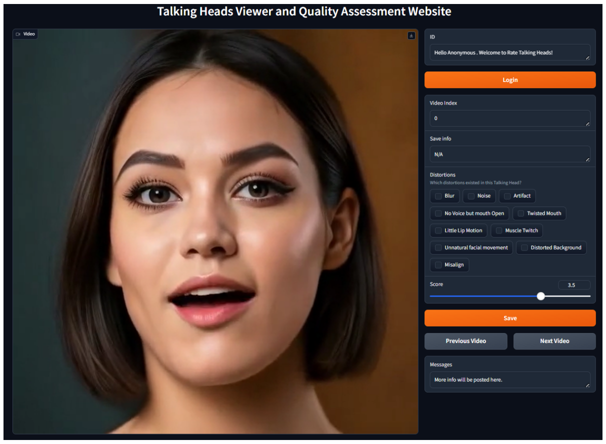
Who is a Better Talker: Subjective and Objective Quality Assessment for AI-Generated Talking Heads

Supplementary Material

1. Subjective Assessment Interface & Details

To facilitate the subjective evaluation of AI-Generated Talking Heads (AGTHs) within the THQA-10K dataset and the subsequent data collection, we develop a subjective scoring interface by integrating *Python* with *Gradio* [1], as illustrated in Fig. 1. The interface consists of two main sections: the left panel, labeled "Video," displays the AGTHs, while the right panel serves as the subjective scoring area. This area is divided into two dimensions: distortion type and overall quality score. Specifically, the distortion type is provided as a multiple-choice option to allow for a more comprehensive categorization of the distortions present in the AGTHs. On the first day of the subjective experiment, all participants underwent a training session lasting approximately 30 minutes. During this session, participants were instructed on how to evaluate the quality of each AGTH using the Absolute Category Rating (ACR) method, as well as how to identify the types of distortion present. Following the training, participants were introduced to the assessment interface and were given adequate time to familiarize themselves with the scoring procedure.



ID: Hello Anonymous . Welcome to Rate Talking Heads!

Video Index: 0

Save info: N/A

Distortions
Which distortions existed in this Talking Head?
☐ Blur ☐ Noise ☐ Artifact
☐ No Voice but mouth Open ☐ Twisted Mouth
☐ Little Lip Motion ☐ Muscle Twitch
☐ Unnatural facial movement ☐ Distorted Background
☐ Misalign

Score: 3.5

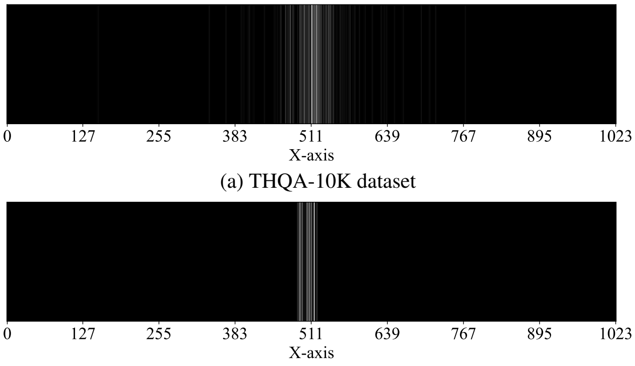
Messages: More info will be posted here.

Figure 1. The Gradio-based quality assessment website, designed for multi-distortion categorization and quality rating.

2. Limitations & Solutions

Although the proposed FSCD method achieves state-of-the-art (SOTA) performance on THQA-10K, THQA [2], and THQA-3D [3] datasets, it does exhibit certain limitations: L1) FSCD depends on the AI-Generated portrait images (AGPIs) for detecting the center point of the mouth, imposing specific requirements on both AGPIs and the face keypoint detection algorithm. In extreme cases, such as when the quality of the AGPI is too poor for accurate face detection, FSCD cannot be applied directly; L2) While slicing the mouth centroids against AGTHs is a straightforward implementation, it may not represent the most optimal approach.

To address these limitations, we propose optimizations and future directions for FSCD: S1) For instances where face detection fails in AGPIs, we propose fixing the x-coordinate of the mouth centroid at half the width of the AGPI. To validate this approach, we conduct a statistical analysis of the mouth centroids in all AGPIs where face detection is successful, with the results presented in Fig. 2. The figure clearly demonstrates that the vast majority of AGPIs in THQA-10K and THQA [2] datasets feature the face near the center. Furthermore, in THQA-3D [3], the viewpoint is intentionally centered on the head model when rendering the 3D talking head as a video. Thus, statistically, it is reasonable to assume that the mouth centroid is most likely located at half the image width in cases where face detection fails; S2) The selection of tangent locations and their corresponding forms warrants further investigation. While the current method uses a Y-T slice through the mouth centroid, alternative adaptive methods may offer richer quality features and lead to more effective results.



(a) THQA-10K dataset

(b) THQA dataset

Figure 2. Distribution of mouth centroid x-coordinates across different AGPIs, with brighter regions indicating higher frequency of occurrences. All AGPIs are rescaled to 1,024 × 1,024 resolution.

References

[1] Abubakar Abid et al. Gradio: Hassle-free sharing and testing of ml models in the wild. *arXiv preprint arXiv:1906.02569*, 2019. 1

[2] Yingjie Zhou et al. Thqa: A perceptual quality assessment database for talking heads. In *ICIP*, pages 15–21, 2024. 1

[3] Yingjie Zhou et al. Subjective and objective quality-of-experience assessment for 3d talking heads. In *ACM MM*, pages 6033–6042, 2024. 1