

Instructions

Welcome to the Agent TTS listening test. You will be asked to evaluate three aspects of synthesized speech: **style consistency**, **audio quality**, and **intelligibility**.

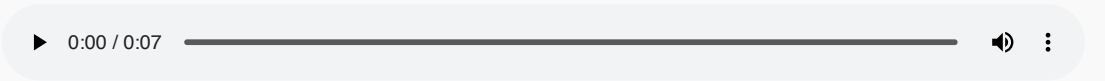
Please carefully read the **style caption** and **transcription** below, then listen to the audio sample.

For each criterion, rate the sample on a scale from 1 (Bad) to 5 (Excellent).

- **Style Consistency:** 1 means the speech does not reflect the caption at all; 3 means partial alignment; 5 means full alignment with all key attributes in the caption.
- **Audio Quality:** 1 indicates very unnatural/robotic audio; 5 means human-like, natural sounding speech.
- **Intelligibility:** 1 means unclear or unintelligible; 5 means perfectly clear and easy to understand.

Notes

When rating intelligibility, please also consider the similarity of the speaker's timbre to the reference voice. Audio that does not resemble the target speaker should receive a lower rating.



FAQ

- Q:** Should I consider other factors when rating a specific aspect?
A: No, please focus only on the current aspect being rated. For example, when rating style consistency,

- ignore intelligibility and audio quality.
Q: Which factor should I consider the transcription?
A: When you rate the intelligibility, please consider whether the speech follows the transcription.

Sample #1



Style Caption: Warm and compassionate voice resonating with empathy.
Transcription: It truly is awe inspiring. The beauty and diversity of marine life never cease to amaze me.

Rate the speech style consistency

☐ 1: Bad

☐ 2: Poor

☐ 3: Fair

☐ 4: Good

☐ 5: Excellent

Rate the audio quality

☐ 1: Bad

☐ 2: Poor

☐ 3: Fair

☐ 4: Good

☐ 5: Excellent

Rate the intelligibility

☐ 1: Bad

☐ 2: Poor

☐ 3: Fair

☐ 4: Good

☐ 5: Excellent

Save and Continue