

VisualSpeaker: Visually-Guided 3D Avatar Lip Synthesis

Supplementary Material

A. Supplementary Material

A.1. AutoAVSR Feature Alignment

The confusion matrix in Figure 1 shows how cosine similarity scores are strongest along the diagonal, with non-matching videos elsewhere scoring far lower. This indicates that the features of the 3D Gaussian Splatting (3DGS) render closely match those of the input frames.

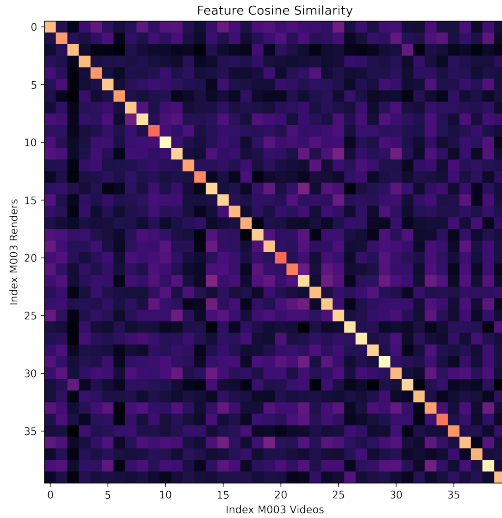


Figure 1. Cosine similarity confusion matrix.

A.2. User Study Details

The user study consisted of 51 participants, each evaluating 20 pairs of videos. They were recruited via a departmental mailing list. We circulated 5 variants of the study, each with a different set of 20 pairs, and each video was rated by at least 4 participants. The participants were asked to rate which video they preferred based on the realism of the lip movements, with the options shown in Figure 2. The following instructions were given to the participants:

In this task, you will be presented with pairs of short video animations, shown side-by-side. Each video features an animated 3D character speaking a short sentence.

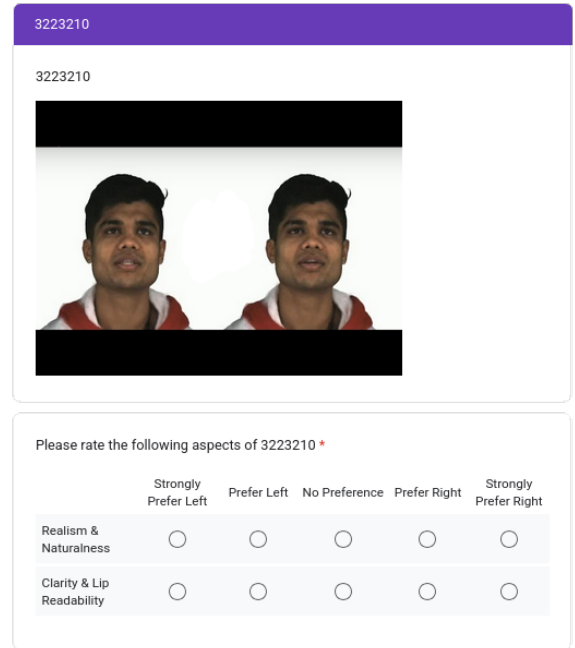
Each form contains 20 randomized samples. Your goal is to carefully evaluate and compare them based on two main criteria:

- **Realism and Naturalness:** How believable, human-like, and natural the lip movements appear.
- **Clarity and Lip Readability:** How clear the lip movements are in representing the spoken words, and how easy it would be to understand what is being said by **only** watching the lips.

Audio is provided with the videos. If possible, we kindly request that you use headphones for this task to ensure you can hear clearly.

Please take your time to consider each pair carefully. There are no right or wrong answers.

The rankings were then converted to a $\{-2, -1, 0, 1, 2\}$ to calculate the preference percentages.



	Strongly Prefer Left	Prefer Left	No Preference	Prefer Right	Strongly Prefer Right
Realism & Naturalness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clarity & Lip Readability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 2. Sample User Study Interface.