

Main point: Efficiently compare many video stimuli to each other at the same time



HEMVIP: Human Evaluation of Multiple Videos in Parallel

Patrik Jonell, Youngwoo Yoon, Pieter Wolfert, Taras Kucherenko, and Gustav Eje Henter

Main contributions:

- A proposed framework, called HEMVIP, for parallel and granular evaluation of multiple video stimuli
- A validation study confirming that results obtained using the tool are in close agreement with results of prior studies using conventional multiple pairwise comparisons.

- Code publicly available
<https://github.com/jonepatr/hemvip>

Code available!

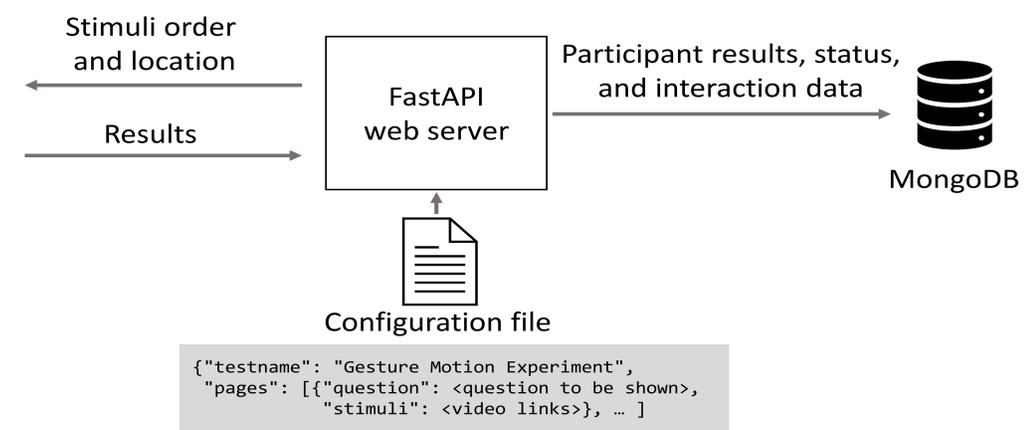
HEMVIP:

- A framework based on MUSHRA
- Users rate multiple videos in parallel
- Scales efficiently, like MOS
- Allows for direct comparisons, like pairwise preference evaluations
- More granular results, which allows using a variety of statistical tests

Evaluation: Similar results to pairwise comparisons, but more efficient!

- We validated HEMVIP against the results of a previous study in the field of gesture generation [13] which used pairwise comparisons
- Results obtained using HEMVIP were largely the same as in the original study (see Table)
- Compared to another study [11] which used binary pairwise comparisons but otherwise was the same, HEMVIP was about 27% more efficient in terms of time spent

System contrast	Full→NoAR				Full→NoPCA				Full→NoFiLM				Full→NoAudio				Full→NoText				Full→NoVel				NoPCA→GT			
Study question	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Significant in [13]?	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
C-P test difference				+								-				-								-				-
t-test difference				+																								



Participants watch videos and rate each on a sliding scale

[11] Can we trust online crowdworkers? Comparing online and offline participants in a preference test of virtual agents. Patrik Jonell, Taras Kucherenko, Ilaria Torre, and Jonas Beskow. Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents. 2020.

[13] Gesticulator: A framework for semantically-aware speech-driven gesture generation. Taras Kucherenko, Patrik Jonell, Sanne van Waveren, Gustav Eje Henter, Simon Alexanderson, Iolanda Leite, Hedvig Kjellström. Proceedings of the ACM International Conference on Multimodal Interaction. 2020.