Rules for the GENEA Challenge 2023

Only register for this challenge if you actually intend to submit an entry to the challenge and to comply with all the challenge rules.

Goal of the challenge

The GENEA Challenge seeks to advance scientific knowledge relating to automatic gesture generation, by means of open science and a large-scale, joint subjective evaluation. It is therefore a *challenge*, not a competition: the point is not to find *who* does best, but *what* works best. The rules of the challenge have been designed with the intent of best furthering this goal.

Database access

Only registered participants have access to the gesture database during the challenge. Access to the data and participating in the challenge may also require completing and agreeing to a data licence agreement.

Data access (e.g., download passwords/credentials) will be provided after your registration is accepted and you have completed any required licence agreements. Do not share the data or access to it with non-challenge participants.

Materials provided

All participants who have signed any necessary licence agreements will be given access to the following materials:

- 3D full-body motion-capture clips of a speaking and gesticulating person (the *main agent*), in BVH (Biovision Hierarchy) format.
- Aligned 3D full-body motion-capture clips of the interlocutor in the interaction, in BVH format.
- Aligned audio waveforms of the speech associated with the two motion-capture clips, in WAV format.
- Text transcripts of each audio file with word-level timing information, in TSV (tab-separated values) format.
- A metadata file containing information such as labels coding the identity of the persons gesticulating in each recording.
- Code and scripts for training the baseline system or systems to be included in the challenge evaluation (once available).
- A pipeline for visualising system output as videos of a gesticulating avatar, the same as that which will be used to render video stimuli for the challenge evaluation.

The above motion, audio, and transcriptions have been partitioned into an official training set and an official validation set. Please respect this split, and do not train on validation data when developing your system or creating your final submission. The official validation set was created using the same process as the held-out test set for the challenge, and has

similar duration and other characteristics. It is therefore your best guide to what the final, held-out test set will look like.

Approximately one week before the deadline to submit generated motion stimuli, participants will also be given access to:

- Held-out audio waveforms for the main agent, in WAV format, using only speakers present in the training data.
- Audio and aligned 3D full-body motion-capture clips of the interlocutor in the interaction, in WAV and BVH format.
- Text transcripts of each held-out audio file with word-level timing information, in TSV format.
- A metadata file containing information such as labels coding the identity of the persons gesticulating in each recording.

The task of the challenge is to use one's system to generate convincing gesture motion for the main agent given the held-out speech and interlocutor behaviour we provide, and then submit that motion for evaluation. For this reason, we will not provide any target motion data for the main agent in this material. Note that not all of the synthetic motion output submitted to the challenge may be included in the final evaluation.

If, for some reason, you have or gain access to the held-out motion data, we rely on your honesty in not looking at that material or letting it influence your challenge submission.

Limits on participation

Each participating team may only submit one system per team for evaluation. Teams can consist of one or more persons, from zero or more academic institutions and/or commercial entities.

Participants involved in joint projects or consortia who wish to submit multiple systems (e.g., an individual entry and a joint system) should contact the organisers in advance and receive approval first. We will try to accommodate all reasonable requests, provided the evaluation remains manageable. If the number of participating teams is small (e.g., less than five), the organisers may decide to permit multiple entries per team.

Use of external data

- "External data" is defined as data, of any type, that is not part of the provided database. This includes, for example, raw recordings, structured databases, and pre-trained models/systems such as word vectors.
- For this year's challenge, only open external data – data that is available to the public free of charge (possibly after signing a licence agreement) – may be used.
- All external data used in your system must be explicitly acknowledged by providing a citation and/or link in the paper accompanying your submission.
- You are allowed to use external data in any way you wish, subject to any exclusions or limitations given in these rules.

For **data pertaining to text and audio**, any external data may be used, as long as they satisfy the criteria above. There is no limitation on the amount of such data you may use.

For **motion data** (whether 2D, 3D, or video), *only explicitly whitelisted external motion data is permitted.* The reason for this data restriction is that other behaviour-generation challenges have found that system performance often is limited by the amount of training data that can be ingested, which is not an interesting scientific conclusion to replicate.

This year, there is only one permitted source of external motion data, namely the motion from the BEAT dataset that has been retargeted to the challenge skeleton. This data will be made available to challenge participants by the BEAT dataset creators. Note that this data is monadic, not dyadic.

Your system must make use of the main-agent motion data that us organisers provide, but you may exclude parts of that data if you wish. Use of the provided audio or text transcripts is entirely optional and not compulsory, as is the use of any external data.

Please keep in mind that the point of the challenge is to gain better insight into the synthesis and perception of motion and gestures, not to see who has the best data and resources. Consequently, participants are strongly encouraged to share processed material they are using in their entries with other participants and with the organisers. Example data that may be valuable to share include: improved transcripts and alignments; denoised and reconstructed motion data; sub-selected data; bug fixes to baseline systems; etc.

If you are in any doubt about how to apply these rules, please contact the organisers for clarification!

Synthesising test motion

Synthetic gesture motion must be submitted in the same format as that used by the challenge gesture database (BVH, same skeleton, frame rate, etc.). The organisers take no responsibility for any effects that may occur when processing motion that was not submitted in the correct format.

Since the idea is to evaluate how systems would perform in an unattended setting, manually tweaking the test inputs or the generated output motion is not allowed. For systems with stochastic output, it is similarly not permitted to manually cherry pick which random samples are submitted for evaluation. Only automated processing is permitted at synthesis time.

Retention and distribution of submitted stimuli

Any stimuli that you submit for evaluation will be retained by the organisers for future use. The evaluated stimuli and any associated user ratings and comments will also be made publicly available for non-commercial purposes, labelled by the corresponding anonymised system label.

Evaluation

The GENEA Challenge focuses on subjective human perception, not objective metrics. A large-scale formal evaluation by means of several user studies will be conducted to jointly evaluate and compare the submitted co-speech gestures. These user studies will be carried out online using crowdsourced raters who speak and comprehend the language featured in the database.

The evaluation of the submitted gesture motion will likely consider aspects such as

- its perceived human-likeness, without accounting for the speech or the interlocutor;
- its appropriateness for the associated held-out main agent speech, in terms of, e.g., timing, semantic content, or both; and
- its appropriateness for the speech and/or motion of the interlocutor.

Aside from stimuli based on motion submitted by challenge participants, the evaluation will also incorporate stimuli based on natural speech and motion; motion generated from one or more baseline approaches based on code available to challenge participants; and checks on raters' attention.

The results of the evaluation, including a statistical analysis, will be made public, albeit with the identity of participating systems anonymised. Participating teams will be informed of the results and which system is theirs, so that they can draw conclusions and describe what they learnt in papers describing their submissions.

Paper

The ultimate goal of the GENEA Challenge is to advance scientific knowledge, which is disseminated through papers submitted to and presented at the conference and workshop associated with the challenge. Therefore, *d*o not enter the challenge if you are unable to comply with any of the requirements below!

- Each participant **must** submit a paper (using the template specified) describing their challenge entry for double-blind review. This submission and its reviews will be permanently available on OpenReview.
    - Although submitted systems will be anonymised in the results of the challenge as published by the organisers, participants are encouraged to report which anonymised label is associated with their system in their paper and in any other publications based on their challenge submission, since this increases the scientific value of the challenge as a whole.
- Each participant is also required to complete a form giving the general technical specification of their system, to facilitate easy cross-system comparisons.
    - This form may include questions such as "Is your motion generated based on playback such as motion graphs, or on continuous motion generation, or a hybrid approach?" "Is the output deterministic or stochastic?" "Does your system use text input, audio input, or both?", etc.
- One of the **authors** of each accepted paper **must** register and present their work at the conference workshop associated with the challenge, if their paper is accepted, or if the organisers otherwise ask them to do so.

- Paper submission, copyright, publishing, and presentation is subject to ACM rules and the rules of the parent conference at which the GENEA Challenge is hosted. You must accept and comply with their rules.
- There is no penalty for dropping out of the challenge prior to the start of the evaluation. However, teams whose stimuli are included in the evaluation are required to submit a paper on their system and present their work at the workshop.
    - Failure to submit a paper and give a presentation at the workshop where required may result in being blacklisted from participating in future challenges.

Use of results

This gesture-generation challenge is a scientific exercise. You may use the results only for the purpose of scientific research. Specifically, you may **not** use the results (e.g., your team's ranking in the evaluations) for any commercial purposes, including but not limited to advertising products or services.

How are these rules enforced?

This is a challenge designed to advance scientific knowledge, and not a competition. The point is not to find *who* does best, but *what* works best. Therefore, we depend on your honesty in preparing your entry.

Blatant rules violations and behaviour that goes against the spirit of the challenge – for example participating in the evaluation but not submitting a paper to describe your entry – may result in being blacklisted from participating in future challenges.

Please contact the organisers at genea-challenge@googlegroups.com if you have any questions about these rules.