

MediaEval 2013 Survey

About MediaEval

MediaEval is a multimedia benchmark that focuses on the human and social aspects of multimedia access and retrieval. It emphasizes the "multi" in multimedia: speech, audio, music, visual content, tags, users, context.

MediaEval attracts participants who are interested in multimodal approaches to multimedia involving, e.g., speech recognition, visual analysis, audio analysis, user-contributed information (tags, tweets), viewer affective response, social networks, user privacy, geo-coordinates and crowdsourcing.

Our goal is to promote the development and evaluation of innovative algorithms that tackle new challenges in multimedia. Participation in MediaEval is open to any group who chooses to sign up. Sign up for MediaEval 2013 will open at the beginning of March 2013 via the MediaEval website <http://www.multimediaeval.org/>

About this survey

This survey collects information from the research community concerning interest in the tasks that have been proposed in response to the [MediaEval 2013 Call for Proposals](#). The most popular tasks will be chosen to run in MediaEval 2013. Also, your comments and suggestions are very valuable to us since they will help to guide the process of task design.

The survey should take you between 5 minutes and 30 minutes depending on the detail with which you choose to respond.

The survey consists of general questions on 12 tasks. After answering the general questions for a task, you may then choose to go to a deeper level and answer more detailed questions on the task.

You may have come to this survey because you have heard about one particular task that you are interested in and wish to give feedback on the more detailed questions. We very much appreciate that you take the time to answer the general questions on all 12 tasks. Tasks in MediaEval benefit from feedback from the wider community and not just from their active participant circles.

Thank you for your interest and input.

Please enter your e-mail address if you wish to receive an e-mail when the MediaEval 2013 call for participation is published. Even if you were on last year's mailing list, please enter your address again here. We create an updated mailing list each year.

Have you or your group participated in one of our benchmarking evaluations in the past?

- [VideoCLEF 2008](#)
- [VideoCLEF 2009](#)
- [MediaEval 2010](#)
- [MediaEval 2011](#)
- [MediaEval 2012](#)
- Yes, but I don't know exactly which.
- No, 2013 would be the first year.

MediaEval 2013

So that you know what to expect next: The survey will ask for your input on the following 12 task proposals, received in response to the [MediaEval 2013 Call for Proposals](#) in this order.

Social Event Detection for Social Multimedia
Search and Hyperlinking of Television Content
Placing: Geo-coordinate Prediction for Social Multimedia
Violent Scenes Detection in Film (Affect Task)
Spoken Web Search: Spoken Term Detection for Low Resource Languages
Question Answering for the Spoken Web
Visual Privacy: Preserving Privacy in Surveillance Videos
Soundtrack Selection for Commercials ("MusiClef")
Similar Segments of Social Speech
Retrieving Diverse Social Images
Emotion in Music (Affect Task)
Crowdsourcing for Social Multimedia

Social Event Detection for Social Multimedia

This task requires participants to discover social events in a collection of social media gathered from the internet (for example from Flickr, Picasa, Instagram or Twitpic). By social events we mean that the events are planned by people, attended by people and that the social media (images/video) are captured by people.

A common dataset will be provided for this task, comprising a large set of images (and videos) together with their associated metadata (including, for instance, time-stamps, tags, geotags, and possibly uploader's ID).

For completing the task, the participants can exploit the actual image/video files, the metadata provided to them as an XML file, as well as information coming from any external resources of the participant's choice (e.g. Wordnet, Wikipedia, or even visual concept detectors trained on external media collections). Ground truth event detection results will be generated by the organizers and will be made available to task participants, following the evaluation of the submissions.

Note that this task involves Social Event Detection, which requires detection of instances of organized events. The goal is not to detect types or classes of activities, as in the case of TRECVID MED (multimedia event detection). In other words, SED would detect a particular soccer match or wedding, whereas MED would detect soccer matches or weddings in general.

What is your opinion about the Social Event Detection Task described above?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Social Event Detection: Additional Questions

These additional questions concern the Social Event Detection Task, which requires researchers to automatically detect events in social multimedia.

Task definition: In 2011 and 2012 we asked participants to return only the media clusters corresponding to the events that we specified in each challenge (e.g. technical events that took place in Germany). In 2013 should we ask participants to organize all images / videos of the collection into clusters, and return this complete clustering to us for evaluation?

- Yes, I am interested in a "complete clustering" challenge.
- No, I am interested in specific challenges that require only relevant clusters to be returned.
- I am interested in both kinds of challenges and would participate in both of them, if such challenges are defined in 2013.

Dataset: In 2011 and 2012 we used a collection of only images. Should the 2013 dataset be more diverse in terms of type of media, i.e., also include videos?

- The dataset should continue to contain only still images.
- The dataset should contain both images and videos.
- Either way is interesting and feasible for me.

If you have comments, add them here:

Dataset: In 2012 we used a collection of ~167.000 Flickr images. In 2013 should the dataset be significantly larger than last year (e.g. in the order of 1 Million images+videos), or would this make it more difficult for you to devote the resources necessary for participating in this task?

- The dataset should remain more or less at the same size (getting larger would make it more difficult for me to participate to the task).
- The dataset should be significantly larger than last year (last year's size does not make the problem interesting enough for me to participate in the task).
- Either way is interesting and feasible for me.

If you have comments, add them here:

Evaluation metrics. We plan to use Precision, Recall, F-Score, and Normalized Mutual Information, for evaluating the accuracy of the results (i.e., the quality of the generated media clusters, each corresponding to a detected event)

- These metrics are appropriate and adequate.
- An additional metric could be used (please explain):

Required runs. A baseline run will be required by all participants. Last year, we specified the required run as a run that “can use any combination of the available image metadata, but no visual information”. This year we should:

- Have a similar required run (using just image metadata; no visual information).
- Have a required run that uses only visual information.
- Have two required runs (i.e., both of the above).
- A completely different required run would be more interesting (please explain):

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Search and Hyperlinking of Television Content (Debuted as Brave New Task in...

The search and hyperlinking task involves three sub-tasks: first, the search sub-task where participants have to respond to a known-item query with a ranked list of documents, second, the anchor selection sub-task requiring identifying regions in the video, so-called anchors, for which a user with this query could require additional information, and, third, the link identification sub-tasks that requires generating a ranked list of video segments in the same collection that provide information about these anchors.

The task is a continuation of the MediaEval 2012 Brave New Task. This year, we plan to put more emphasis on the second and third sub-tasks (i.e., the "linking" sub-tasks). We will be using a large set of archival data from a television broadcaster. The data will be in English and we are likely to be able to provide metadata, subtitles and output of visual analysis.

What do you think about the Search and Hyperlinking Task?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Search and Hyperlinking Task: Additional Questions

These additional questions concern the Search and Hyperlinking Task, which requires researchers to automatically generate video-to-video hyperlinks for a collection of television broadcasts.

We currently investigate the use of a dataset of roughly 500 hours of high-quality video. The dataset will be for research only and groups participating in this task will be required to sign a licensing agreement in order to receive the data. Please give us your opinion on the dataset (check all that apply).

- This dataset sounds suitable to me.
- This dataset appears too big for me.
- This dataset appears too small for me.
- I'd like to propose another dataset (Please specify).

In this task we want to promote the use of multiple modalities. We are likely to provide subtitles and the output of several visual analysis tools (face recognition, concept detectors about many concepts). Which of these resources/modalities would you be interested in using (check all that apply):

- Textual modality.
- Audio modality.
- Visual modality.
- My group could provide analysis results (e.g., concept detection results or speech recognition transcripts) for this dataset for use by other participants in the task. Please specify in the box below:

The proposed use scenario for the task assumes two different potential types of users: home users and professionals. What type of queries and search questions might be of more interest for you (check one option):

- I would be more interested in a home search scenario.
- I would be more interested in a professional search scenario (i.e., archivist).
- I have the same interest in both.

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Placing: Geo-coordinate prediction for social multimedia

The Placing Task involves three sub-tasks: First, the "general placing" task involves automatically estimating the geographical location (i.e. the latitude and longitude) of images and/or videos that can be located potentially anywhere in the world using textual, visual and/or aural cues. Second, the "placeability" task involves classifying a media item as being objectively placeable in the world, where something is considered to be placeable when a person with general world knowledge is able to unambiguously place it on the world map (e.g. an image or video of the Eiffel tower is placeable, whereas a close-up image of a flower or a video taken in an arbitrary person's home are not placeable). Third, the "function" task addresses the classification of the shown place in images/videos as being relevant to one or more functions (e.g. landmark, shopping, entertainment, travel), where the estimated location can be used to guide the correct classification. As an example, a video taken outside the Notre Dame church in Paris would be assigned the landmark function, whereas a photo outside of a casino in Las Vegas would be assigned the entertainment function.

In 2010-2012 a dataset of Creative Commons licensed images and videos collected from Flickr was used, where the media items were accompanied by their metadata (e.g. tags, titles, exif, social information). A similar dataset will be used for the 2013 task. For the general placing task, the evaluation is carried out by comparing the estimated geo-coordinates to the actual geo-coordinates associated with the media items in the collection, whereas for the placeability and function tasks standard classification evaluation measures will be used.

Participants are supplied with the video as well as video metadata (e.g., tags, titles, social information). Evaluation is carried out by comparing predicted geo-coordinates to geo-coordinates that have been assigned to the videos by users on Flickr.

What do you think about the Placing Task?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Placing Task: Additional Questions

These additional questions concern the Placing Task, which involves automatically predicting the location of an image or video.

Which of the following tasks would you have an interest in participating in at the Placing Task 2013 (check all that apply)?

- General placing of images (i.e. estimation of the geo-coordinates of an image, as has been done in the past years)
- General placing of videos (i.e. estimation of the geo-coordinates of a video, as has been done in the past years)
- Placeability of images (i.e. classify whether an image can be geo-located)
- Placeability of videos (i.e. classify whether a video can be geo-located)
- Function of images (i.e. determine the function(s) of the place shown in an image)
- Function of videos (i.e. determine the function(s) of the place shown in an image)
- Other (please describe):

If you would participate in the Placing Task 2013, which modalities would you most likely use to address the task (check all that apply)?

- Textual metadata of images/videos
- Image-based visual features
- Video-based visual features that go beyond image features (i.e., exploit the temporal aspects of videos)
- Video-based audio features
- Other (please describe):

It would encourage me and my team to participate in the Placing Task if the organizers would provide us with (check all that apply):

- The source code of a baseline method for each task.
- Extracted visual features for the training and test media item.
- Extracted audio features for the training and test media item.
- Speech transcripts (automatically generated) of the training and test media item.
- Context (automatically derived) of the media items in the training and test sets (e.g. the media item was taken indoor/outdoor)
- Other (please describe):

If you have any additional comments, questions or suggestions about this task, please enter them in the text box below. We are particularly interested in knowing which aspects of previous tasks you would like to see changed, for instance: the size of the dataset, the media items included in the dataset (e.g. too many indoor shots, too many had no tags, too many were taken by the same users), the distribution of items across the training and test sets, the evaluation measures used, etc.

Violent Scenes Detection Task (Affect Task)

The Violent Scenes Detection Task consists of detecting violent scenes in movies exploiting only material available on a DVD (video track, audio track and subtitles). This task derives from a use case scenario at Technicolor with the target of helping users of a video on demand portal decide whether a movie is suitable for the family or not. To do so, violent scene detection will be used to generate a preview of the violent portions in a movie thus facilitating the user's choice. The task encourages the use of multiple modalities.

In 2013, the task will essentially be implemented as a follow-up of the MediaEval 2011 and 2012 Affect Task, incorporating new test data. It will also see some changes: two different definitions of the targeted violent events will be proposed. Violence will still be limited to physical violence. A set of 18 annotated movies will be provided for training and development purposes. A minimum of 7 movies will be used for testing. For copyright purposes, participants are required to buy the DVDs online. Organizers will provide the annotations and evaluation software.

Evaluation metrics will be primarily based on retrieval-like distances (e.g., MAP) but false alarms and missed detection will also be considered.

What do you think about the Violent Scenes Detection Task?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Violent Scenes Detection Task (Affect Task): Additional Questions

These additional questions pertain to the Violent Scenes Detection Task (i.e., Affect Task), which involves detecting violent portions or segments in movies.

In 2012, the Violent Scenes Detection task was run with a set of 18 movies (including 3 test movies), which participants had to order via Amazon for copyright reasons. Please indicate your opinion on the number of movies in the data set (check all that apply):

- We need a much larger data set (training and test), even if that means to order more movies.
- We need a considerably enlarged test set.
- I would be able to volunteer some time to help annotate, if that meant having a larger data set.

Please include any other comments or questions below. If you are volunteering to help annotate, please include your name in the box so that the organizers can contact you.

The Affect Task in 2013 will involve detecting violent portions or segments in movies. Beyond this basic task definition, we would like to know your interests and preferences with regard to the details of task design. Which of the following reflect your needs, interests or preferences (check all that apply)?

- I am interested in having an (automatic) shot segmentation provided (thus considering the task as a shot classification one).
- I could provide automatic shot segmentation to all participants.
- I would be interested in submitting a system that does not use a predefined shot segmentation (i.e., finding the boundaries is part of the task).
- I would be interested in a version of the task in which it is required to use a specific combination of modalities, e.g., audio and video, in order to better support comparison of performance between teams
- I would be interested in a version of the task in which it is required to use a particular combination of modalities, e.g., one given modality combined with a single free-choice additional modality, in order to better support comparison of performance between teams.

Please list additional aspects of violent scenes detection that are interesting to you, but are not included above. If you have comments about why certain aspects are interesting, please include them here, as well:

Which of the different annotations do you think are of interest for the task (check all that apply):

- high level video concepts (presence of blood, presence of fire, presence of firearms, etc.)
- high level audio concepts (gunshots, screams, etc.)
- high level text-related concepts

If high-level concepts were provided, would you use them:

- Yes, I would.
- No, I don't think so.

Please indicate your opinion on evaluation metrics by selecting one or more of the options below:

- The output of the system should be a list of violent segments.
- A weighted sum of false alarm and miss detection is a good way to evaluate violent scene detection.
- Evaluation should not be performed at a specific operating point but rather on the whole range of false alarm/miss detection points (e.g., area under the curve).
- Evaluation should be performed at different operating points (e.g., false alarm at several predetermined miss rates).
- If shot segmentation is provided, the output of the system should be a ranked list of shots from most to least violent.

In 2012, the definition of violence was limited to: "physical violence resulting in pain or injury", therefore excluding any example of psychological or verbal violence.

For 2013, two different definitions of violence (but still restricted to physical violence) will be provided, leading to two different subtasks. These two subtasks will differ only in the violent segments to detect: same systems may therefore answer the two subtasks provided that the learning step is re-done on a different training set. Will you participate in the two subtasks or do you think it is too much work for one task in a benchmark?

- Yes, I would participate in both subtasks
- No, even if it only means re-training the same system, it is too much work. I would only target one of the subtasks.

For 2013, would you be interested in the following definitions being added (check all that apply or none if physical violence is enough)?

- Events associated with violence (gunshots, screaming, fighting, car chase)
- Objects associated with violence (guns, knives, blood)
- Acts of violence against humans
- Acts of violence against animals
- Acts of violence against objects
- Segments in which the consequences of violence are evident, although not the violence itself
- Segments in which the intention to act violently is portrayed (someone threatens to shoot, but does not)
- Segments containing verbal violence
- Segments containing psychological violence

For 2013, participants will be asked to provide a video summary of the most violent extracted segments. Such a summary will not be part of the evaluation process. Do you think that such feature could be valuable for the future of the task?

- Yes, definitely.
- Yes, definitely and it should be included in the evaluation process in the future.
- No, the notion of video summary is still too vague to serve as evaluation.

For 2013, would you be interested in having the possibility to submit a run in which all kinds of external data sources can be used (EXCEPT for information on actual annotation, i.e., frame- or shot-wise annotations themselves)?

- Yes
- No

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Spoken Web Search Task

The Spoken Web Search Task involves searching FOR audio content WITHIN audio content USING an audio content query. This task is particularly interesting for speech researchers in the area of spoken term detection.

The task specifically addresses the challenge of multiple, resource-limited languages. It requires researchers to build a language-independent audio search system so that given an audio query, it should be able to find the appropriate occurrence(s) in the test data. This requires identification of the audio file and the location of query term(s) within that audio file.

The language-independent speech retrieval system is to be built given almost no training data. Researchers will be provided with a development and test data set consisting of recordings in multiple languages (labels will be available for the development portion), from different channels/ sound qualities, and using different speaking styles.

In 2013, the Spoken Web Search would like to introduce an innovation intended to lower the entrance barrier for new groups wishing to work on the task. A basic speech recognition set up will be provided to the participants in the form of a virtual machine, following the [Speech Recognition Virtual Kitchen](#) framework.

What is your opinion about the Spoken Web Search Task described above?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

If a baseline spoken term detection system with all data available and ready to run the task would be made available as a [virtual kitchen](#), would you consider participating in the evaluation?

- Yes
- No

Please give us an explanation of why you chose this answer:

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Spoken Web Search Task: Additional Questions

These additional questions are on the Spoken Web Search Task, which requires researchers to build a language-independent search system to locate occurrence of terms in audio data.

The following are suitable metrics for this evaluation (tick all that you find suitable):

- NIST AWTV (as in spoken term detection)
- Precision/ Recall
- F-Measure
- Other

Do you think that search speed should be used as a metric (maybe secondary metric, after one from those above) in order to measure how fast a system can match a query in the reference?

- Yes.
- No.

Please give us an explanation of why you chose this answer:

In addition to different languages (some low-resourced, some not), do you think it would also be interesting to measure how robust might systems be regarding when tested on different acoustic/recording conditions and heavily accented speech?

- Yes
- No

Please explain why you think so:

Are around 4 hours of reference speech and 100 queries per test set/language sufficient in your opinion?

- Yes, it is enough.
- No, more are necessary.

If you have additional comments on how the datasets should be designed/created, or other questions on the datasets, please add them here.

What sort of approaches would you consider applying to the task (check all that apply)?

- Search over phonetic lattices
- Search directly over the audio
- all of the above
- other approaches (please specify)

If you have additional comments or questions, please add them here.

Do you know of similar spoken audio data that could/ should be used to carry out the Spoken Web Search Task? We would be very interested in including it for this year or next year's evaluation. (Please describe it in the box):

What types of query terms would you be interested in working with (check all that apply)?

- full matches, single word (all words spoken in the query are a match)
- full matches, multiple words with random amount of silence in between (all words spoken in the query are a match)
- partial matches (not all the spoken words in the query are a match)
- multiple matching query terms with filler words (not matching) in-between
- Other (please specify)

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Question Answering for the Spoken Web (New task in 2013 based on Spoken Web...)

This task builds on the low-resource spoken term detection carried out in the Spoken Web Search task and takes a step closer to the vision of a Spoken Web search engine.

Participants are required to match questions spoken in Gujarati to answers spoken in Gujarati. The query set will consist of 100 questions and the target collection of 4000 answers. The data will be provided by IBM Research, India.

Graded relevance judgments will be generated by human judges. It is planned to carry out in parallel at MediaEval and also at [FIRE \(Forum for Information Retrieval Evaluation\)](#).

What do you think about the Question Answering for the Spoken Web Task?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Question Answering for the Spoken Web Task: Additional Questions

These additional questions concern the Question Answering for the Spoken Web Task, which requires participants to match questions and answers for a low-resource language (Gujarati).

Are you interested in participating in just the MediaEval task, which involves actual spoken questions and actual spoken content, or might you also participate in the FIRE search task, which involves (the same) transcribed questions and (the same) transcribed content?

- MediaEval only
- Possibly also FIRE
- I don't know at this point

Are you able to process the audio, or would you need some pre-processed searchable form derived from that audio (e.g., a phone lattice or phone n-grams) in order to be able to participate?

- I am able to process the audio
- I would like a phone-based transcription
- If no phone-based transcription is provided, I won't be able to participate

If you are able to process the audio, would you be willing and able to share some pre-processed searchable form derived from the audio (e.g., a phone lattice or phone n-grams) for both the questions and the corpus with teams who are interested in searching representations of speech but who are not (yet) able to generate such representations? If yes, could you generate those representations two months earlier than results are due?

We plan to report many of the standard IR effectiveness measures computed by trec_eval (e.g., MAP, NDCG, and R-Precision), but we need to designate one primary evaluation measure for which teams can optimize. We think that Mean Reciprocal Rank would be a good choice because it models the case in which only a single answer is desired. Would you prefer a different primary evaluation measure? If so, which one?

- No, I think Mean Reciprocal Rank is good choice
- MAP
- NDCG
- R-Precision
- Other (please specify)

We are planning to provide some sample questions (and the complete content) along with some (sparse) relevance judgments for training. Would 20 questions be enough?

- Yes
- No, more questions are needed (please specify how many)

We will be transcribing all of the content for FIRE. Do you need some of the content transcribed for ASR training? If so, we would need to hold that training content out from the search task in both MediaEval and FIRE. How much training data (in words) would be a practical minimum for you to participate?

All of the content is in Gujarati, and at present we do not plan to translate it into any other language. Will that cause you any problems?

- No that should be fine
- Yes it can cause problems for me (please specify)

We plan to use 100 rather long questions (>30 seconds) for evaluation because large numbers of questions are required to achieve statistically informative MRR evaluation results (indeed, we would like more, but we have a limited number of long questions). Would you have any problem processing that many long questions.

- No
- Yes (please specify)

What research questions are you most interested in exploring?

Who else should we contact to inform them of the opportunity to comment on this proposed MediaEval task (either because of their expertise or their potential interest, or hopefully both!)?

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Visual Privacy Task

The Visual Identity Protection Task requires participants to automatically detect and obfuscate face regions in digital imagery in a reversible manner, such that the original image or frame can be recovered from the result of the obfuscation. Critically, the obfuscation technique must take into account the attractiveness or acceptability of the resulting obfuscated or scrambled region to human viewers.

This task builds on the 2012 Privacy Task and aims at finding ways to protect the privacy of people in videos. This year, the task is considering extending face obfuscation to also obfuscation of other items that reveal personal identity. Object detection output will be made available to support the participants of this task.

The data set will consist of about 60 high resolution video files of an average length of 30s each. The scenes will be varied with a mixture of indoors and outdoors scenarios, as well as some night-time videos. The people featured in the videos will be performing various actions, such as exchanging objects, talking, fighting or simply walking by. The ground truth will be created by a team of human annotators.

What is your opinion about the Visual Privacy Task described above?

- Very interested. I'd like to participate in one or both.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on these tasks?

- Yes. Please take me to the more detailed questions on these tasks.
- No, thank you. I'd like to skip the detailed questions for these tasks.

Visual Privacy Tasks: Additional Questions

These additional questions concern the visual privacy tasks, which involve detecting humans in video in a privacy-preserving manner. Detected humans are obfuscated with a minimum impact on the acceptability/appeal of the video to human viewers.

To perform high quality evaluation it might be essential for participants to provide binary files of their methods. Please indicate if you would be willing to submit the results in the form of the binary file or other forms (check one or more)?

- Binary executable file.
- Source code.
- Resulting video

What is your opinion on the way privacy should be defined for the purposes of evaluation in this task?

- I would prefer to have a set of clearly defined objectives that can be evaluated individually (e.g. hiding accessories, hiding ethnicity).
- I would prefer to have a set of clearly defined objectives but have an evaluation procedure that would still reward me for finding alternative methods and/or objectives
- Other (please specify)

What elements would you be interested in obscuring from the video footage (check all that apply)?

- Faces
- Gender
- Ethnicity
- Accessories
- Proximal objects
- Gait
- Clothing
- Other (please specify)

How would you prefer the Visual Identity Protection Task submission to be evaluated (choose one)?

- Conduct subjective tests of the privacy protection techniques, i.e., based on the results of questionnaire from human subjects to evaluate the social/human aspects of the proposed blurring approach.
- Based on the reviews by experts in privacy and ethics together with experts in video analytics.
- Based on the scores from a non-reference/reference 'privacy metric', which would measure how 'recognizable' the encrypted/scrambled faces are compared to the original images.
- Based on the hybrid approach: a combination of scores from the 'privacy metric' and the review from the experts.

If you have additional comments or questions on the evaluation of this task, please add them here.

Please give any other comments or feedback you might have on the Virtual Privacy task:

Soundtrack Selection for Commercials ("MusiClef")

This task focuses on the topic of music usage in TV commercials. When selecting a song to accompany the advertisement of a brand or a product, music consultants base their selections on both context- and content-based information.

With this topic, we move beyond the MusiClef 2012 music auto-tagging brave new task, towards an application with a strong multimedia component and a clear need for multimodal approaches.

Our task concretely would be to, given a TV commercial, predict the most suitable song for this commercial from a list of candidate songs (which does not include the original song in the commercial). We will require at least one run to include more than one modality.

The task has a strong user component, and as such, the ground truth will be strongly based upon human input (ratings and/or annotations). We plan to adopt crowdsourcing methods to acquire this input.

What do you think about the Soundtrack Selection Task?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Soundtrack Selection Task: Additional Questions

These additional questions are on the Soundtrack Selection Tasks, which requires researchers to match soundtracks with video commercials.

Which of these feature types would you be interested in using (check all that apply):

- low-level audio features (timbre, rhythm, ...)
- high-level music features (genre, sound-alike, ...)
- social tags on music (gathered from Last.fm and other online services)
- web pages describing the brand/product
- web pages describing the artist
- images features (album covers, brand logo, advertised product, ...)
- video features of the commercial
- social tags on commercial (gathered from YouTube)
- microblogs on brand/product/artist/song
- Other (please describe):

Do you prefer to work on audio features extracted from the actual commercials or from the music files of the respective songs? (check all that apply)

- commercials
- songs

If you have comments, add them here.

Due to copyright restrictions, we may not be able to distribute 'raw' data (audio and video signals, copyrighted text). In that case, we would have to resort to distributing pre-computed features. Are there any specific features you would really like to be pre-computed for this task?

If your group could provide analysis results (e.g., concept detection results or speech recognition transcripts) for this dataset to be use by other participants in the task, please specify them in the box below:

In terms of the size of the music collection, what is more important to you?

- The collection should be large-scale, even if this means having less metadata available for long-tail items.
- The data establishing the collection should be comprehensive in terms of metadata and contextual data (such as web pages, social media data, and collaborative tags).

Evaluation is a challenge, and if our task gets sufficient interest, we would like to discuss

this with our core participants. Possible evaluation criteria under current consideration include:

1 Does the music relate to the video?

2 Does the music fit the user's perception of the product and the company?

This has impact for the way ground truth and evaluation will be set up. Please rate the suitability of the following evaluation scenarios.

1 (Not
suitable)

10 (Highly
suitable)

Ask users to assess the fit between a music track and the commercial video, and compare this to the original video-soundtrack combination.

Ask users to describe the perceived 'image' of a product/company, and compare this to descriptions obtained for the original video-soundtrack combination.

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Similar Segments of Social Speech (New in 2013)

This task requires that participants take regions of interest in audio/video (ca. 10 seconds each) and return an ordered list of all similar regions, where similarity is based on the perceptions of human searchers.

Use Scenario: A new member has joined an organization or social group that has a small archive of conversations among its members. The person starts to listen, looking for any information that can help understanding, participating in, enjoying, finding friends in, and succeeding in this group. The person marks segments of interest in the audio/video and the system returns a list of similar segments.

What is your opinion on the Similar Segments of Social Speech Task described above?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Similar Segments of Social Speech: Additional Questions

These additional questions concern the Similar Segments of Social Speech Task, which requires participants to identify portions of audio recordings of conversations that people perceive to be similar.

Are you interested in audio only, or would you also be interested in video modality.

- Audio recordings only.
- It would also be interesting to have the video.
- I would definitely use the video if it were available.

Would you mind if the dataset includes dialogs involving noticeably non-native speakers?

- I prefer to concentrate on native speakers.
- Non-native speakers are ok, but not particularly interesting for me.
- Non-native speech is of particular interest to me.

We'll probably supply 2 hours of data and 70 similarity-sets in the training set; will this be adequate for you?

- Yes. This is the right size.
- It's a bit small, but fine for the first year of this task.
- I would strongly prefer a larger data set. Please comment on your reasons:

The evaluation metric will probably evaluate the ability of the system to output 'jump-in points' at or closely preceding the start of a relevant region of speech, that is, timepoints where a human searcher would like to start viewing/listening. In particular we are considering mGAP.

- That sounds reasonable.
- That would be inconvenient for my team since our methods assume a predefined segmentation.
- mGAP is not really the best metric for this (please comment)

Do you have interest in using this data to support different tasks, and if so what annotations or metadata would you like to have?

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Retrieving Diverse Social Images (New in 2013)

The task on Diversity in Social Photo Search involves automatically refining photo search results from common social media platforms with the objective of retrieving not only representative, but also diversified photos depicting the query in a comprehensive and complete way.

Participants are supplied with a ranked list of photos with monuments - ca. 25k photos depicting around 100 monument locations - retrieved from platforms such as Picasa, Panoramio or Flickr using textual and GPS tags. These results are typically very noisy and redundant. Participants will have to improve the results by selecting only a small sub-set of photos that are equally representative (accurate matches of the query), diverse (diverse representations of the query) and compact (very few, e.g., 10-15 photos for each monument). To serve as reference information, each location is accompanied by a representative photo and a monument description from Wikipedia.

The specificity of the task compared to other diversity related tasks such as the ImageCLEF 2009 Photo Retrieval is in addressing the social dimension with explicit focus on users that is reflected both in its nature (variable quality of photos and of metadata) and in the methods devised to retrieve it.

Target communities involve both machine and human media analysis such as image retrieval, re-ranking, relevance feedback, crowd-sourcing and automatic geo-tagging. To solve the challenge, participants are free to choose from any approaches, from human oriented, machine-based to hybrid machine-human; as well as to take advantage of any additional data sources, e.g., the Internet. To encourage participation of groups from different research areas, additional resources such as state-of-the-art visual descriptors and textual location models will be provided for the entire collection. Evaluation of performance is going to be carried out by comparison with human ground truth.

What is your opinion on the Retrieving Diverse Social Images Task described above?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Retrieving Diverse Social Images: Additional Questions

These additional questions concern the task on Diversity in Social Photo Search that requires participants to automatically select a small set of equally representative and diverse photos from typical noisy results provided by common social media search engines.

Representativeness. For this task, a representative image of a monument has been defined as “a canonical image of a monument depicted from outside”.

1 Strongly disagree 2 3 4 5 6 7 Strongly agree

Please rate from 1 to 7 how much you agree with this definition:

How would you complete/change the above definition to better suit your opinion of a representative photo (please specify):

Diversity. For this task, a set of pictures is considered to be diverse if the set contains pictures that depict, partially or entirely, the monument under different conditions such as light, perspective and background.

1 Strongly disagree 2 3 4 5 6 7 Strongly agree

Please rate from 1 to 7 how much you agree with this definition.

In your opinion, when is a set of photos a diverse representation of a monument (please specify):

Question on the approach. Which of the following reflect your needs, interests or preferences (check all that apply)?

- I am interested in working mainly with machine analysis techniques (e.g., machine learning).
- I am interested in working mainly with human-based approaches (e.g., crowd-sourcing).
- I am interested in working mainly with hybrid machine-human approaches.
- I am interested in considering the use of external data sources from the Internet.
- Other (please specify)

Question on data sources. Which of the existing social photo search platforms you find the most interesting and appropriate for this task (check all that apply)?

- Flickr.
- Picasa.
- Panoramio.
- Other (please specify)

Question on the size of the data corpus. What would you consider to be the appropriate size of the photo corpus for this task?

- Thousands.
- Tens of thousands.
- Hundreds of thousands.
- Millions of photos.

If none of the above, please specify:

Question on evaluation metrics. We plan to use average precision (measure of how many photos are relevant), completeness (computed as the total number of clusters - i.e. the groups of photos depicting the monument in a very similar way - that can be obtained from the true positive photos), cluster precision (measure of how many of the clusters in the final refinement are relevant), cluster recall (measure of how many of the existing clusters are represented in the final refinement) and F-measures (with different weighting):

- These metrics are sufficient.
- An additional metric could be used (please explain):

Please enter a few keywords describing your area or areas of expertise.

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Emotion in Music (New type of affect task proposed for 2013)

This task is a new task on emotional characterization of music. This task comprises two subtasks. The first task, the per song emotion characterization task, requires participants to deploy multimodal features to automatically detect arousal and valence on 9 points scale for each song. In the second task, the continuous emotion characterization task, the emotional dimensions, arousal and valence, should be determined for a given song continuously in time, the quantization scale will be per frame (e.g., 20ms). We will also take advantage of metadata crawlable from the internet, e.g., last.fm, twitter, etc.

These affective features can be used in recommendation and retrieval platforms. There are already examples of mood based or emotion based online radios, e.g, Streamood (<http://www.stereomood.com>), however, these systems do not really take advantage of classical definitions of emotion and only rely on user generated tags.

We are still deciding on the dataset to be used but a good candidate for the continuous subtask is: <http://music.ece.drexel.edu/research/emotion/moodswingsturk>

We are considering open music archives to be able to release the original tracks, however, that might need extra selection efforts. If you decide to fill in the additional questions, we will ask you to comment on this choice.

What is your opinion on the Emotion in Music Task described above?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Emotion in Music: Additional Questions

These questions concern the Emotion in Music Task, which involves predicting the mood or emotion of music.

For this new task, two different subtasks on tagging and continuous representation are being offered. These two subtasks will differ on their dataset, and their machine learning methods.

Will you participate in the two subtasks or do you think it is too much work for one benchmark year?

- Yes, I will participate in the two subtasks
- No, even if it only means developing two systems, it is too much work. I will only target one of the subtasks.

In a hypothetical case that the organizers provides features extracted by, MIRToolbox, Marsyas, Rastamat and Echonest. How important is it for you to be able to extract your own features?

- Not really important, those features will be enough for me.
- Very important, I would like to extract my own set of features.
- Don't know.

Should we use dimensional representation (arousal-valence) or categorical?

- Categorical.
- Dimensional.
- Fine with both.
- We should use other than arousal and valence (please specify):

Would you or members of your team be able to help us with annotations? If yes, how much capacity do you have?

- We can participate in the task, but don't have time to help with annotations.
- We can help with annotations (please add your name so that the organizers can contact you):

How many annotators per song do you think is adequate for annotation?

- at least 10.
- at least 5.
- I don't have an opinion.
- Another number, please indicate and possibly provide your reasons

Do you think it is important to release the raw data of annotation (i.e., without averaging across different annotators)?

- Yes.
- No.
- I don't have an opinion.

If the annotation is only made for a short segment (e.g., 10 sec or 30 sec) of a song, do you think it makes sense to extract features from the whole song or only from the segment?

- It is important to extract features from the whole song.
- Only the features from the segment are necessary.
- I don't have an opinion.

Do you think it important/interesting to have lyrics?

- Yes, lyrics are useful
- No, we prefer instrumental music
- No, we do not use lyrics
- I don't have an opinion.

Should we annotate valence and arousal at the same time (using a 2-D interface) or annotate valence and arousal separately?

- At the same time, since they are correlated.
- Separately.
- I don't have an opinion.

Do you think it important to use songs that the annotators are unfamiliar with, or in a foreign language?

- No, we should not use them since the semantics of the lyrics will be lost
- Yes, since we are interested in prosodic features and not the lyrics
- I don't have an opinion.

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

Crowdsourcing for Social Multimedia (New in 2013)

The ultimate goal of this task is to create an automated approach to create ground truth (clean labels) from noisy crowd annotations.

In this task a dataset of 4810 Creative Common images with focus on fashion images is provided. Each image has a tag provided by the person who published the picture. Furthermore, each image is annotated by three Amazon Mechanical Turk (AMT) workers. The annotation process consists of validating the given tag to an image and verifying whether the image is fashion-related or not. Participants can use the annotations of AMT workers as well as any other image content and context analysis methods to create clean labels for each image.

The dataset contains actual images, annotation of images by AMT workers and metadata. The metadata includes contextual information (such as title, description, geo-tags) and social information (such as user profile, comments).

The ground truth is provided by a separate set of annotations which are done by trusted annotators. We used accuracy as evaluation metric. The accuracy is calculated as the number of correct labels predictions divided by the total number of images.

What is your opinion on the Crowdsourcing for Social Multimedia Task described above?

- Very interested. I'd like to participate.
- Sounds quite interesting. Don't know for sure if I'd participate.
- If I had more information, I might consider it.
- This task is not for me.

If you have comments, add them here. For example, Why (or why not) would you be interested in this task? Do you have suggestions for the task? Is the task innovative? Already solved? Too novel to yet be of interest? Just right to be of value to the research community? Do you know of an existing benchmark, data set or project related to this task (please specify)?

Would you like to answer more detailed survey questions on this task?

- Yes. Please take me to the more detailed questions on this task.
- No, thank you. I'd like to skip the detailed questions for this task.

Crowdsourcing for Social Multimedia: Additional Questions

These questions concern the Crowdsourcing for Social Multimedia, which involves creating clean labels from noisy crowdsourcing annotations.

Which source of information do you plan to use to create clean labels? (check all that applies)

- Workers annotations
- Image content features
- Image contextual features
- Social features
- Other (please specify)

What do you think about the number of AMT workers per image?

- Three workers per image is sufficient
- There should be more workers per image (please specify)

The workers are asked to confirm image labels with one of three possible options ("Yes": the label match with image, "No": the label does not match with image, and "Not sure"). Do you prefer the third option ("Not sure") to be considered as a final label for an image or you think the image labels should be ultimately a yes/no label?

- I prefer to consider "Not sure" option as an ultimate label for an image
- I prefer only the "Yes/No" options to be considered as ultimate labels

Please explain why you think so:

Evaluation Metric: Do you think the evaluation metric for this task (accuracy) is sufficient?

- Evaluation metric is sufficient
- Additional evaluation metric can be used (Please specify)

Do you know any dataset similar to the dataset used for this task?

- No
- Yes (please specify)

Do you wish to do the same task on any other domains (like video, speech, etc.)?

- No
- Yes (please specify)

If any additional comments, questions or suggestions about this task occurred to you while you were answering the detailed questions, it would be helpful if you could enter them here:

General MediaEval questions

Your answers to questions on the MediaEval 2013 workshop and other general MediaEval questions will help us to make good choices for organization in 2013 and moving forward.

What is your estimate of the probability that you would attend the MediaEval 2013 workshop (to be held in Barcelona most probably just before <http://acmmm13.org/>)?

- More than 90%
- More than 75%
- More than 50%
- Less than 50%
- Impossible to estimate at this moment.

The probability that I will take part in MediaEval 2013 (i.e., carry out a task, write a working notes paper, attend the workshop) is high enough that I would like to take the time to answer a set of detailed questions concerning organizational aspects of the benchmark.

- Yes, please take me to the more detailed questions on MediaEval.
- No, thank you. I'd like to skip additional detailed questions concerning MediaEval.

Venue of the MediaEval 2013 Workshop

This question concerns the venue for the MediaEval 2013 workshop. It should only be answered by people who are intending to attend the workshop.

There are four considerations used to choose a venue for the MediaEval workshop:

1. The workshop should be co-located with an established conference that is related to the "multi" in multimedia. (Past experience shows that co-location helps to fortify international participation and also introduces new people to the MediaEval community.)
2. The workshop should take place in late September, October, or early November. (We need to hold the whole summer free so that everyone has enough time to work on their run submissions.)
3. The workshop should take place at a venue that has something to do with medieval history.
4. The workshop should be as convenient as possible for people to reach by both ground and air travel.

On the basis of these considerations, we have decided to hold the MediaEval 2013 workshop in Barcelona around the time of <http://acmmm13.org/>. We have several possible scenarios and in the questions below we would like to ask your feedback on each.

ACM Multimedia 2013 (<http://acmmm13.org/>) will be held Monday 21 October - Friday 25 October in Barcelona. We would like to hold the MediaEval workshop around the same time. Currently, the preferred dates for the workshop are Friday 18 and Saturday 19 October 2013. There is a chance that the logistics of the workshop location will more or less dictate the dates. However, we would like to collect the opinion of the community about which dates would be most popular and take that into account if at all possible. Please rank the following options according to your preference (1 is your first choice and 5 is your least favorite).

<input type="checkbox"/>	Thursday 17 and Friday 18 October 2013
<input type="checkbox"/>	Friday 18 and Saturday 19 October 2013
<input type="checkbox"/>	Saturday 19 and Sunday 20 October 2013
<input type="checkbox"/>	Saturday 26 and Sunday 27 October 2013
<input type="checkbox"/>	Sunday 27 and Monday 28 October 2013

We are attempting to get a venue in Barcelona. However, there is a chance that this won't work out. There are some very nice medieval castles in Catalonia that we have also been exploring. Also, we are a bit worried (independently of castles) that Barcelona will be more expensive than Pisa. Please check the following statements with which you agree.

- I like the castle in Catalonia and I don't mind spending a couple of hours on the train to get there from Barcelona.
- No adventurous trip to a mediaeval castle for me! I would prefer to stay in Barcelona.
- Either would work for me and it's nice that we can maintain the tradition of a medieval venue (i.e., hold the workshop in a building built in the Middle Ages).
- Let's drop the idea of a mediaeval venue and just hold it in a modern venue.
- I agree with keeping the price of the workshop as low as possible, but I don't mind if it goes up from what it was last year (e.g., two-day workshop registration would go from 190 Euros in 2012 to ca. 250 Euros in 2013)
- Hey! Don't raise the price. I can only attend if the registration stays below 200 Euros for the two-day workshop.
- Wherever we are it would be nice to have a internet connection.
- Wherever we are, I need to be able to get there with public transportation (e.g., train, bus, metro).
- I have an idea for where the workshop should be or have another comment:

Detailed questions on the general organization of MediaEval

Please indicate how important each of the following points concerning MediaEval tasks is to you:

	Very important	Important	Not really important	Don't care
MediaEval tasks are based on real-world use cases.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MediaEval tasks emphasize social media.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MediaEval tasks focus on multimedia access and retrieval (as opposed to, e.g., multimedia sensor information processing)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MediaEval tasks emphasize human aspects of multimedia (i.e., aspects related to human perception, interpretation or affective reaction).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MediaEval tasks encourage multimodal approaches.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The tasks offered by MediaEval match the topics on which I am currently working.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If you have any general comments or suggestions that you would like to communicate to the MediaEval organizers, please add them here:

Please indicate how important each of the following additional points about MediaEval is to you:

	Very important	Important	Not really important	Don't care
The MediaEval workshop takes place in Europe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My group receives a prize or certificate (e.g., 1st, 2nd, 3rd place) that reflects our quantitative performance relative to other participants.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My group is involved in the organization of MediaEval tasks.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The working notes for the MediaEval workshop are published online (i.e., MediaEval 2012 published at http://ceur-ws.org/Vol-927/)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The working notes for the MediaEval workshop are short papers -- intended to be extended later for submission elsewhere.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MediaEval focuses on collaboration more than competition.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If you have any general comments or suggestions that you would like to communicate to the MediaEval organizers, please add them here:

In 2011, MediaEval doubled in size. Please give us your opinion on the following points related to the growth of MediaEval (Tick all boxes that apply).

- MediaEval should restrict the number of tasks to 5-6.
- MediaEval should not extend its focus to other areas of multimedia less directly related to search and retrieval.
- The decision of the number of tasks should be based upon the number of participants interested in the task. In other words, any task that has enough participants to run should run if it "fits" MediaEval (i.e., emphasizes the 'multi' in multimedia and has a human/social aspect). It doesn't matter if there are 6 or more tasks.
- The workshop size should be limited to 60 (i.e., the size of the 2011 and 2012 workshops)
- The workshop should be allowed to grow larger.

If you have any general comments or suggestions about the growth of MediaEval, please add them here:

Thank you for your responses. For more information on MediaEval please refer to the website <http://www.multimediaeval.org> or contact Martha Larson (m.a.larson@tudelft.nl)