



The Audio (Voice) and Video Search Challenge:

A paradigm shift to raw data media search

Version 1.0, February 21, 2008

1. Introduction.....	1
2. Evaluation Metric for Voice	3
3. Evaluation Metric for Video.....	5
4. Schedule	6
5. Development & Evaluation Data.....	7
6. Format of results	8
7. Evaluation Condition	9
8. Appendix.....	10
Appendix I. XML format for AT1, AT2 and AT3	10
Appendix II. XML format for VT1, VT2 and VT3	12

1. Introduction

The goal of this document is to define the problems, performance measures and development/test data for the audio (voice) and video search challenge. The objective of this search challenge is to encourage participation from international sites to develop new, interesting and practical search techniques.

The voice search challenge consists of three subtasks, as outlined below:

AT1 Search by IPA

The query is given in International Phonetic Alphabet (IPA); the task is to retrieve all segments that contain the query IPA sequence regardless of its spoken languages;

AT2 Search by example

The query is an utterance spoken by different speakers; the task is to retrieve all segments that contain the query word/phrase/sentence regardless of its spoken languages;

AT3 Search for recurrent voice segments

In the voice archive, certain words/phrases/sentences are repeated more than once by the same or different speakers in the content of the spoken documents. These repetitions are known as recurrent voice segments. The task is to extract all recurrent segments which are *at least 15 seconds* in length. No query is given in this case. The number of *unique* recurrent segments for each document is given.



The video search challenge consists of three subtasks, as outlined below:

VT1 *Search by (Single) Query Image*

The query is provided in the form of a single image. The task is to retrieve all visually similar segments. Note that the similarity is at the perceptual level. That is, the expected results should contain video segments that contain images that *look* similar to the query image, as opposed to the video content being *semantically* similar. There are 20 query images, each depicts an **object or scene**. The general guideline for determining whether an image is similar to the query image is to decide if a similar query object or scene is visibly dominant in the image or not. Note that in some cases (rare, but possible) images may be considered as similar to multiple query images.

VT2 *Search by Video Shot*

The query is a short video shot (<10sec). The task is to retrieve video shots that are *perceptually similar* to the query video clip. Note that compared to **VT1**, there is now additional motion information in the query video shot and the matching criteria should also take into consideration similarity in the motion trajectory. There are 20 query shots, each corresponds to a **scene or event**. The general guideline for determining whether a shot is similar to the query shot is to decide if a similar query scene or event is visibly dominant in the shot. Note that in some cases (rarely but possibly) shots may be considered as similar to multiple query shots.

VT3 *Scene/Event Categorization*

A list of scene/event classes will be defined. For each class, a set of videos depicting the scenes/events will be provided. The participants are expected to develop a model of the class by learning the common visual characteristics from the sample videos of the same class. Then, given a new, unseen test video the task is to categorize the test set into one of the classes. Note that the set of query videos will necessarily be a very large set in the order of 10K. Also, about 10% of these query videos will not belong to any of the predefined scene/event classes, and the desired output result is an 'others' class. There are 20 scene/event classes and one 'others' class. Each query video must be assigned to one of these classes.

2. Evaluation Metric for Voice

a) Mean Average Precision (MAP)

For **AT1** and **AT2**, given a query, the system is supposed to return a ranked list of max 50 returned segments. Each segment is given a relevance score. Each segment has maximum 15 seconds duration. The segments do not overlap each other. As long as the segment contains the search query, it is considered a correct hit. The results of retrieval are checked against the ground truth relevance judgments, and evaluated in terms of the non-interpolated mean average precision (MAP):

$$MAP = \frac{1}{L} \sum_{i=1}^L \left(\frac{1}{R_i} \sum_{j=1}^{R_i} \delta(i, j) \right)$$

where L denotes the total number of queries, R_i the total number of documents relevant to the i^{th} query, and $\delta(i, j)$ is an indicator function which is 1 when there is a hit (i.e. the j^{th} relevant document is in the list output by the retrieval method for query i) and 0 otherwise.

b) F-measure

For **AT3**, the results of retrieval are checked against the ground truth and evaluated in terms of the F-measure, defined as follows:

$$F_{\alpha} = \frac{1}{D} \sum_{d=1}^D \frac{1}{|U(d)|} \sum_{u \in U(d)} \frac{(1 + \alpha) P_d^u R_d^u}{\alpha P_d^u + R_d^u}$$

where D is the total number of test documents and $U(d)$ is the set of unique recurrent segments in document d . P_d^u and R_d^u denote the precision and recall respectively of the segment u in document d . α can be adjusted to change the weights of precision and recall in the measure. In this challenge, $\alpha = 1.0$ is used. The precision and recall are computed as:

$$P_d^u = \frac{\sum_{t=1}^{T_d} \delta(l_h(t), u) \delta(l_r(t), u)}{\sum_{t=1}^{T_d} \delta(l_h(t), u)}$$

$$R_d^u = \frac{\sum_{t=1}^{T_d} \delta(l_h(t), u) \delta(l_r(t), u)}{\sum_{t=1}^{T_d} \delta(l_r(t), u)}$$

where

$$\begin{aligned}l_r(t) &= r \text{ where } r_s \leq t < r_e \\l_h(t) &= h \text{ where } h_s \leq t < h_e \\ \delta(a,b) &= \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}\end{aligned}$$

$l_r(t)$ and $l_h(t)$ denote the reference and hypothesis labels respectively at time t . r_s and r_e are the start and end times of the reference segment, r . Likewise, h_s and h_e are the start and end times of the hypothesis segment, h . $\alpha(a,b)$ is an indicator function.

c) Time constraints between the release of queries and the submission of results

- i) 48 hours for 1st, 3rd knockout rounds, qualifying race;
- ii) 12 hours for Grand Final;

3. Evaluation Metric for Video

a) (Non-interpolated) Average Precision (AP) and Mean Average Precision (MAP)

For both **VT1** and **VT2**, given a query (a single image for **VT1** and a short video shot for **VT2**), a participating system should retrieve a ranked list of max **50** video shots. Each shot is to be given a relevance score. For a particular retrieved shot, as long as any of the ground truth images is contained in the shot, the retrieved shot result is considered a correct hit. The results of retrieval are checked against the ground truth relevance judgments, and evaluated in terms of the non-interpolated **Average Precision** (AP):

$$AP = \frac{1}{R} \sum_{r=1}^N (P(r) * rel(r))$$

where R is the total number of ground-truth relevant shots, N is the total number of retrieved shots (here equals to 50), $P(r)$ refers to the precision at the r -th retrieved document, given by:

$$P(r) = \frac{1}{r} \sum_{i=1}^r rel(i)$$

$rel(r)$ is the relevance indicator of the r -th shot (ie, $rel(r)=1$ if the r -th shot is relevant, and 0 if otherwise)

To measure the overall retrieval performance of a method, the **Mean Average Precision** (MAP) simply averages over all the APs for all test queries.

b) Classification

For **VT3** (Scene/Event Classification), the classification results are checked against the ground truth and evaluated in terms of classification accuracy, that is, the number of correct classification divide by the total number of test runs. There is no penalty against a wrong classification.

c) Time constraints between the release of queries and the submission of results

- i) 48 hours for 2nd, 3rd knockout rounds, qualifying race;
- ii) 12 hours for grand final;

4. Schedule

Table 1 outlines the schedule for the challenge: This challenge will be conducted in several rounds. For each round, every registered team will receive the database (video with sound track in MPEG-1 format) from the organizer via express courier delivery. Each team is given 4 days to process the database. The participants will need to obtain the queries from the main website on the dates specified in the table and to submit the results to the main website within 24 hours. There are in total 5 rounds:

- 1. 1st knockout round (Voice search only):**
 - Up to 25 hours monolingual database for **AT1** and **AT2** tasks only
- 2. 2nd knockout round (Video search only):**
 - Up to 25 hours monolingual database for **VT1** and **VT2** tasks only
- 3. 3rd knockout round (Voice+Video search):**
 - Up to 25 hours multilingual database for **AT1 & AT2, and VT1 & VT2** tasks only
- 4. Qualifying Race (Voice+Video search):**
 - Up to 50 hours multilingual database for **AT1, AT2 & AT3, and VT1, VT2 & VT3** tasks
 - A team must participate in at least two out of three AT tasks, and two out of three VT tasks. If a team chooses to participate in more tasks, then the two AT tasks of best results and the two VT tasks with best results will be used for ranking.
 - Top 5 teams will qualify for the grand final
- 5. Grand Final (Voice+Video search):** - Up to 50 hours multilingual database for **AT1, AT2 & AT3, and VT1, VT2 & VT3** tasks
 - A team must participate in at least two out of three AT tasks, and two out of three VT tasks. If a team chooses to participate in more tasks, then the two AT tasks of best results and the two VT tasks of best results will be used for ranking.
 - On-site evaluation



5. Development & Evaluation data

The organizer will release a set of development data, sample queries and answers, recurrent voice segment annotation and evaluation tools to the registered participants on April 15, 2008. The participants can use the development data to fine-tune and test their algorithms.

The evaluation data will be made available by the organizer at the appropriate time. The participants will be informed about the date(s) and method of delivery in advance. The test queries will be announced at the website <http://www.thestarchallenge.sg/>. The participants will submit their results electronically through the same website before the respective submission deadlines.

Registration deadline	March 31, 2008
1st knockout round*: - release of test queries - result submission deadline - announcement of qualifiers	May 4, 2008 May 5, 2008 May 15, 2008
2nd knockout round*: - release of test queries - result submission deadline - announcement of qualifiers	June 4, 2008 June 5, 2008 June 20, 2008
3rd knockout round*: - release of test queries - result submission deadline - announcement of qualifiers	July 4, 2008 July 5, 2008 July 20, 2008
Qualifying race: - release of test queries (race starts) - announcement of finalists	August 20, 2008 August 23, 2008
Grand final: - release of test queries (race starts) - announcement of winner	October 23, 2008 October 23, 2008

Table 1: Schedule overview

* Please note that the total number of competitive knockout rounds is at the discretion of the organizers.



6. Format of results

For each task, the results are to be submitted in one single XML (Extensible Markup Language) document. The file format specifications are given, in terms of the Document Type Definition (DTD), in Appendix I and Appendix II, respectively.

The DTD files will be made available for download at <http://www.thestarchallenge.sg/>. All the participating teams are advised to validate their XML result files against these DTD files before submission. A tool to validate the output file will also be provided at <http://www.thestarchallenge.sg/>.



7. Evaluation Condition

The collection of audio (voice) and video documents for this challenge are multilingual MPEG-1 documents consisting of hundreds of hours of audio track from TV broadcasting, including documentaries, news, interviews, commercials, MTV etc.

8. Appendix

Appendix I. XML format for AT1, AT2 and AT3

For the **AT1** and **AT2** submissions, the results should be put into a single XML file for each subtask. The XML file should conform to the Document Type Definition (DTD) shown in Table 2;

<pre> <!ELEMENT results (query)> <!ELEMENT query (nbest)> <!ELEMENT nbest EMPTY> <!ATTLIST results team CDATA REQUIRED> <!ATTLIST query id ID REQUIRED> <!ATTLIST nbest rank CDATA REQUIRED> <!ATTLIST nbest docid CDATA REQUIRED> <!ATTLIST nbest score CDATA REQUIRED> </pre>

Table 2: DTD for the AT1 and AT2 submissions

A sample output file is given in Table 3.

<pre> <?xml version="1.0"?> <!DOCTYPE results SYSTEM "t1t2.dtd"> <results team="team name"> <query id="query id1"> <nbest rank="1" docid="id1" score="s1"/> <nbest rank="2" docid="id2" score="s2"/> <nbest rank="3" docid="id3" score="s3"/> ... <nbest rank="50" score="s50"/> </query> <query id="query id2"> <nbest rank="1" docid="id1" score="s1"/> <nbest rank="2" docid="id2" score="s2"/> <nbest rank="3" docid="id3" score="s3"/> ... <nbest rank="50" score="s50"/> </query> ... </results> </pre>

Table 3: Example result file for the AT1 and AT2 submissions

For the **AT3** submission, the results should be put into a single XML file for each subtask. The XML file should conform to the Document Type Definition (DTD) shown in Table 4;

```

<!ELEMENT results (query)>
<!ELEMENT query (nbest)>
<!ELEMENT recurrence (segment)>
<!ELEMENT segment EMPTY>
<!ATTLIST results team CDATA REQUIRED>
<!ATTLIST query id ID REQUIRED>
<!ATTLIST recurrence id ID REQUIRED>
<!ATTLIST segment start CDATA REQUIRED>
<!ATTLIST segment end CDATA REQUIRED>

```

Table 4: DTD for the AT3 submission

A sample output file is given in Table 5.

```

<?xml version="1.0"?>
<!DOCTYPE results SYSTEM "t3.dtd">
<results team="team name">
  <query id="query id1">
    <recurrence id="r1">
      <segment id="1" start="s1" end="e1"/>
      <segment id="2" start="s1" end="e1"/>
      ...
      <segment id="N" start="s1" end="e1"/>
    </recurrence>
    <recurrence id="r2">
      <segment id="1" start="s1" end="e1"/>
      <segment id="2" start="s1" end="e1"/>
      ...
      <segment id="N" start="s1" end="e1"/>
    </recurrence>
    ...
    <recurrence id="rN">
      <segment id="1" start="s1" end="e1"/>
      <segment id="2" start="s1" end="e1"/>
      ...
      <segment id="N" start="s1" end="e1"/>
    </recurrence>
  </query>
  ...
</results>

```

Table 5: Example result file for the AT3 submission



Appendix II. XML format for VT1, VT2 and VT3

For the **VT1** and **VT2** submissions, the results should be put into a single XML file for each subtask. The XML file should conform to the Document Type Definition (DTD) shown in Table 6;

```
<!ELEMENT results (query)>
<!ELEMENT query (nbest)>
<!ELEMENT nbest EMPTY>
<!ATTLIST results team CDATA REQUIRED>
<!ATTLIST query id ID REQUIRED>
<!ATTLIST nbest rank CDATA REQUIRED>
<!ATTLIST nbest docid CDATA REQUIRED>
<!ATTLIST nbest score CDATA REQUIRED>
```

Table 6: DTD for the VT1 and VT2 submissions

A sample output file is given in Table 7.

```
<?xml version="1.0"?>
<!DOCTYPE results SYSTEM "t1t2.dtd">
<results team="team name">
  <query id="query id1">
    <nbest rank="1" docid="id1" score="s1"/>
    <nbest rank="2" docid="id2" score="s2"/>
    <nbest rank="3" docid="id3" score="s3"/>
    ...
    <nbest rank="50" score="s50"/>
  </query>
  <query id="query id2">
    <nbest rank="1" docid="id1" score="s1"/>
    <nbest rank="2" docid="id2" score="s2"/>
    <nbest rank="3" docid="id3" score="s3"/>
    ...
    <nbest rank="50" score="s50"/>
  </query>
  ...
</results>
```

Table 7: Example result file for the VT1 and VT2 submissions

For the **VT3** submission, the results should be put into a single XML file for each subtask. The XML file should conform to the Document Type Definition (DTD) shown in Table 8;

```
<!ELEMENT results (query)>
<!ELEMENT query (class)>
<!ELEMENT class EMPTY>
<!ATTLIST results team CDATA REQUIRED>
<!ATTLIST query id ID REQUIRED>
<!ATTLIST class id CDATA REQUIRED>
<!ATTLIST class score CDATA REQUIRED>
```

Table 8: DTD for the VT3 submission

A sample output file is given in Table 9.

```
<?xml version="1.0"?>
<!DOCTYPE results SYSTEM "t3.dtd">
<results team="team name">
  <query id="query id1">
    <class id="c1" score="s1">
  </query>
  ...
  <query id="query idN">
    <class id="cN" score="sN">
  </query>
</results>
```

Table 9: Example result file for the VT3 submission

Notes:

1. Please note that this white paper may be updated to clarify queries as they arise.
2. If you have any queries regarding the whitepaper, please contact us via the website at <http://www.thestarchallenge.sg/contact.do> or email to helpdesk@thestarchallenge.sg