Video Hyperlinking (LNK) TRECVid 2016

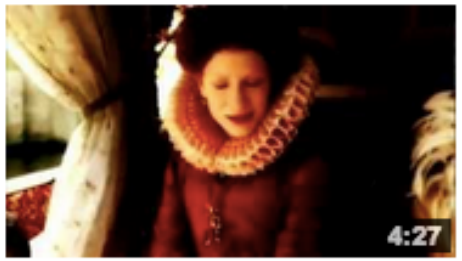
Maria Eskevich, Martha Larson (Radboud University)
Robin Aly, Roeland Ordelman (University of Twente)
Gareth J. F. Jones (Dublin City University)
Benoit Huet (EURECOM)

Task intuition

Fish and Chips Filet Featuring Jamie Oliver
by **EpicMealTime** • 1 month ago • 1,625,276 views
LIKE/FAVORITE this video!!! EpicMealTime featuring a
HD

FISH 'N' CHIPS - VIDEO RECIPE
by **robjnixon** • 1 year ago • 112,781 view
http://tinyurl.com/NickosFacebook http://tir
http://tinyurl.com
HD

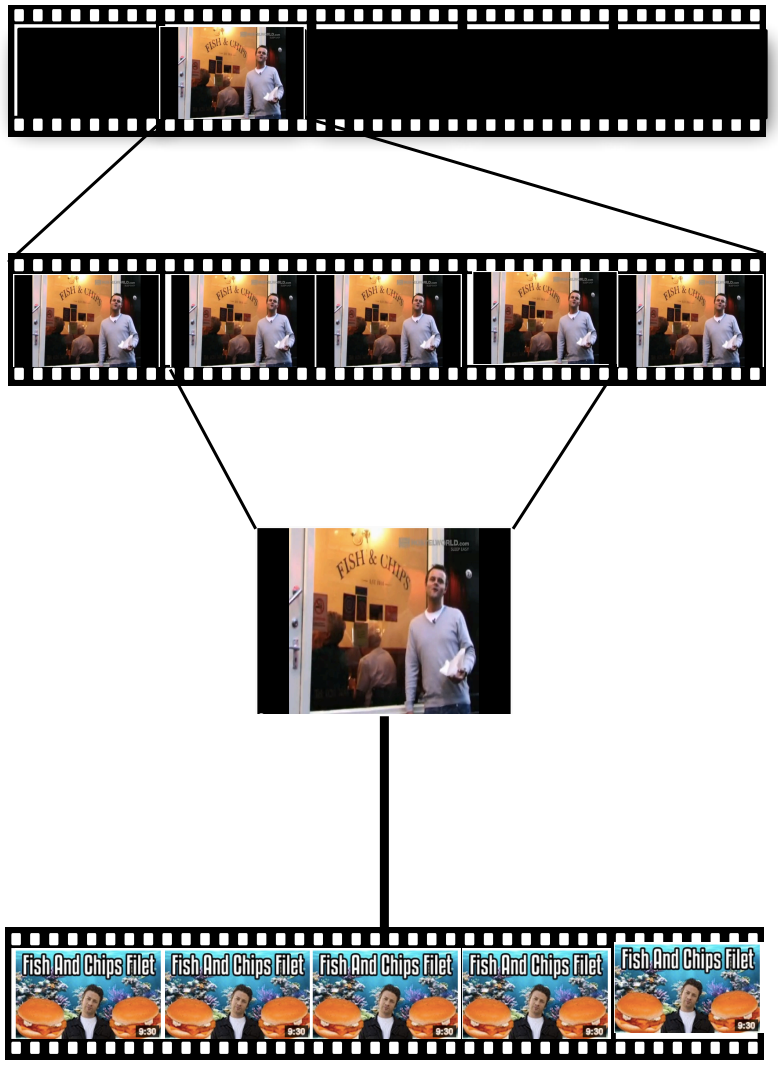
Task intuition



Queen Elizabeth I // I
by Animalcrossing95 • 1
Just a tribute to Cate Blan
did
NEW HD

Queen Elizabeth's Historic Move,
Discrimination
by ABCNews • 1 week ago • 3,850 views
Queen takes bold step in defending wome
HD

Terminology



- **Video** (e.g., 2 hours)
- **Video clip** (e.g., 10 min)
- **Anchor**: segment (unconstraint) for which a user requests a link (e.g., 1 min) “I want to know more about this”
- **Hyperlink**
- **Target**: relevant segment for given anchor (No segmentation enforced)

Collection details

- Blip10000 dataset : 14,838 semi-professionally created videos
- ASR transcripts: LIMSI(2012, 2016) and LIUM (2012)
- Shot segmentation (2012)
- Visual features (AlexNet) : 1000 visual concepts

Anchor creation

Observation from 2015

- Too much ambiguity in anchors
- Previously created anchors were not always multimodal

Strategy for 2016

- True multimodality of anchors reduce ambiguity
- We assume more reliable relevance assessment

TV'16: focus on multimodality

Look for anchors containing a combination of **verbal-visual** information:

- Verbal linguistic cues: `can see', `seeing here', `this looks', `looks like', `showing', `want to show'
- Visual cues: actions and objects crucial for this video are not explicitly named or mentioned

Anchor creation: practical details

- **Dev set:** 28 anchors collected for Search and Hyperlinking 2012 MediaEval task
- **Test set:** 94 anchors defined by 2 human annotators (media professionals)
- Anchor Verification stage on Amazon MTurk:
  - Feedback collection from 3 workers for each anchor (26 and 93 respectively from development and test sets)
  - Examples in the HIT were taken from both development and test sets (2 + 1)
  - 3 anchors from the test set were discarded as confirmed not to be properly verbal-visual
  - **90 test set anchors verified for task evaluation usage**

Anchor Verification: HIT Layout

Task: Watch the video segment and describe why would someone share it.

Imagine that you are watching videos on a site like YouTube. We know that people upload those videos for certain reasons.

We ask you to think about what the person who made the video was trying to communicate to viewers during this short piece.

In other words, what are viewers supposed to understand by watching this particular short piece of the video.

Here you can see a video description example.

1) How would you describe the content you see to another person?

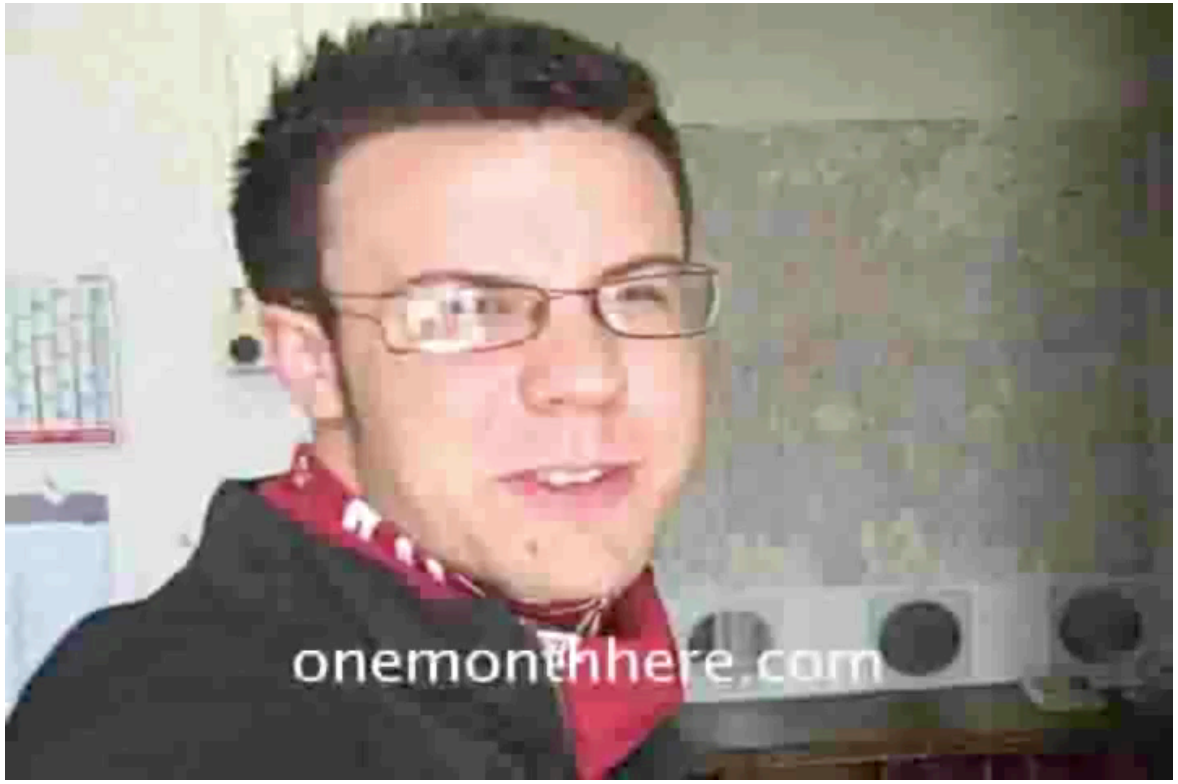
Please write a sentence in the box below to describe the current video segment.

Please keep in mind that this description that you write will be shown to someone else (working on Amazon Mechanical Turk), who will be asked to find other videos that are related to this video segment.

General feedback (optional): If you encounter any problems with this HIT or have any other comments, please report them here. You can also use this textbox to tell us if you liked this HIT or have any suggestions.

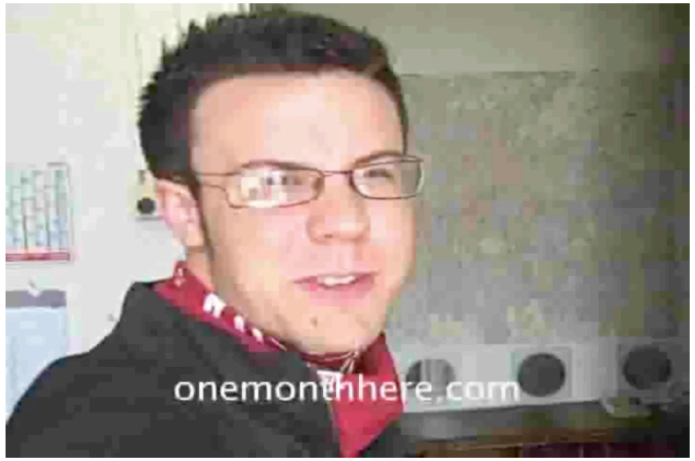
Anchor example

- Textual description:
  - “We are looking for videos of people explaining where and how they live.”



Relevance assessment

Target Vetting: Forced choice



1) Please make sure that you read all of the descriptions below and then choose the description that you find would be the best match with this video. Note: You might feel your best match is a good or even an excellent match. It is also possible that this question is difficult to answer because none of the choices is a particularly good match. If there is no particular good match, please make a choice the best you can.

- *${}{DescriptionOption1}*
- *${}{DescriptionOption2}*
- *${}{DescriptionOption3}*
- *${}{DescriptionOption4}*
- *${}{DescriptionOption5}*

2) This question allows us to gather feedback on whether the question was difficult to answer.

- I felt the video I chose was a relatively good match. The question was easy to answer.
- I felt that the video I chose was the best possible in the list. It wasn't a particularly good match, and for this reason the question was not easy to answer.

3) Please write 2-3 sentences in the box below to explain your decisions.

Case ID	Choice of target description	Feedback on decision making process	Relevance decision
1	Correct	Positive	Relevant
2	Correct	Negative	Manual Check
3	Other	Positive	Non-Relevant
4	Other	Negative	Non-Relevant

Participants

- Number of registrations : 16
  - **INF** **Beijing University of Posts and Telecommunication University Autonoma de Madrid Shandong University Xian Jiao Tong University Singapore**
  - VIREO City University of Hong Kong
  - **IRISA** **CNRS, IRISA, INSA, Universite de Rennes 1**
  - **EURECOM** **EURECOM**
  - **FXPAL** **FX PALO ALTO LABORATORY, INC**
  - NII_Hitachi_UIT National Institute of Informatics, Japan (NII); Hitachi, Ltd; University of Information Technology, VNU-HCM, Vietnam (HCM-UIT)
  - ORAND ORAND S.A. Chile
  - PKU_ICST Peking University
  - VideoVision PES University
  - **EURECOM_POLITO** **Politecnico di Torino Eurecom**
  - REGIMVID REGIM, Research Group on Intelligent Machines National School of Engineers, University of Sfax, Tunisia
  - Trimps The Third Research Institute of the Ministry of Public Security, P.R. China
  - HAWKEYE Tsinghua University
  - TUZ TUBITAK UZAY
  - Waseda Waseda University
  - IIP_WHU Whuhan University

- **Submitting participants : 5 groups (20 official runs)**

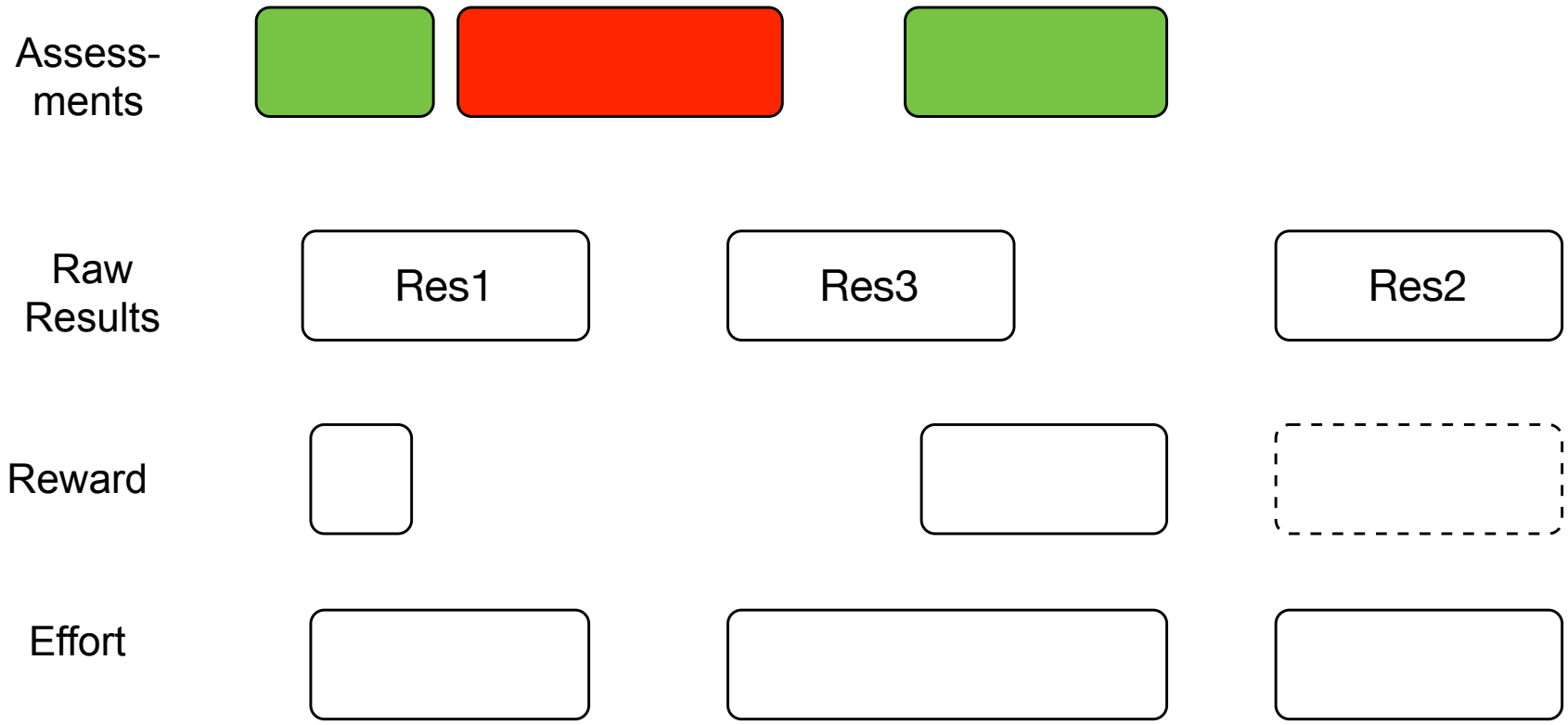
Target vetting: details

- Variability in 20 runs submissions:
  - **Top 5 ranks : 7 216 diverse anchor-target pairs - NEW!**
  - Top 10 ranks : 14 271 diverse anchor-target pair
- Target vetting : HIT processing
  - 3 judgments for each anchor-target pair from the top 5 (21 648 HIT submissions)
  - Batch target vetting assessment: forced logic
  - Final relevance judgment assignment for each anchor-target pair is based on majority decision over 3 crowd workers decisions

Evaluation metrics

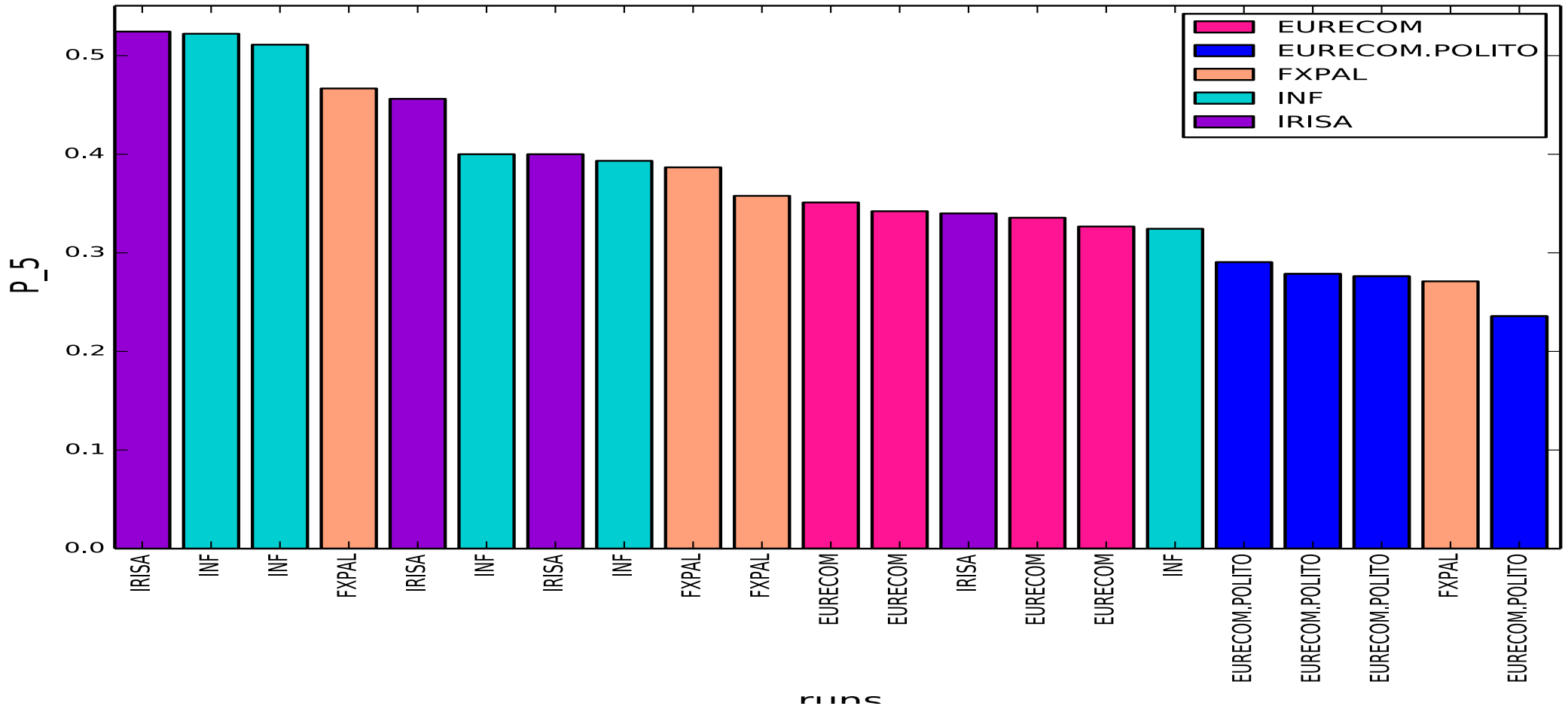
Official metrics

- Precision at rank 5
- MAiSP

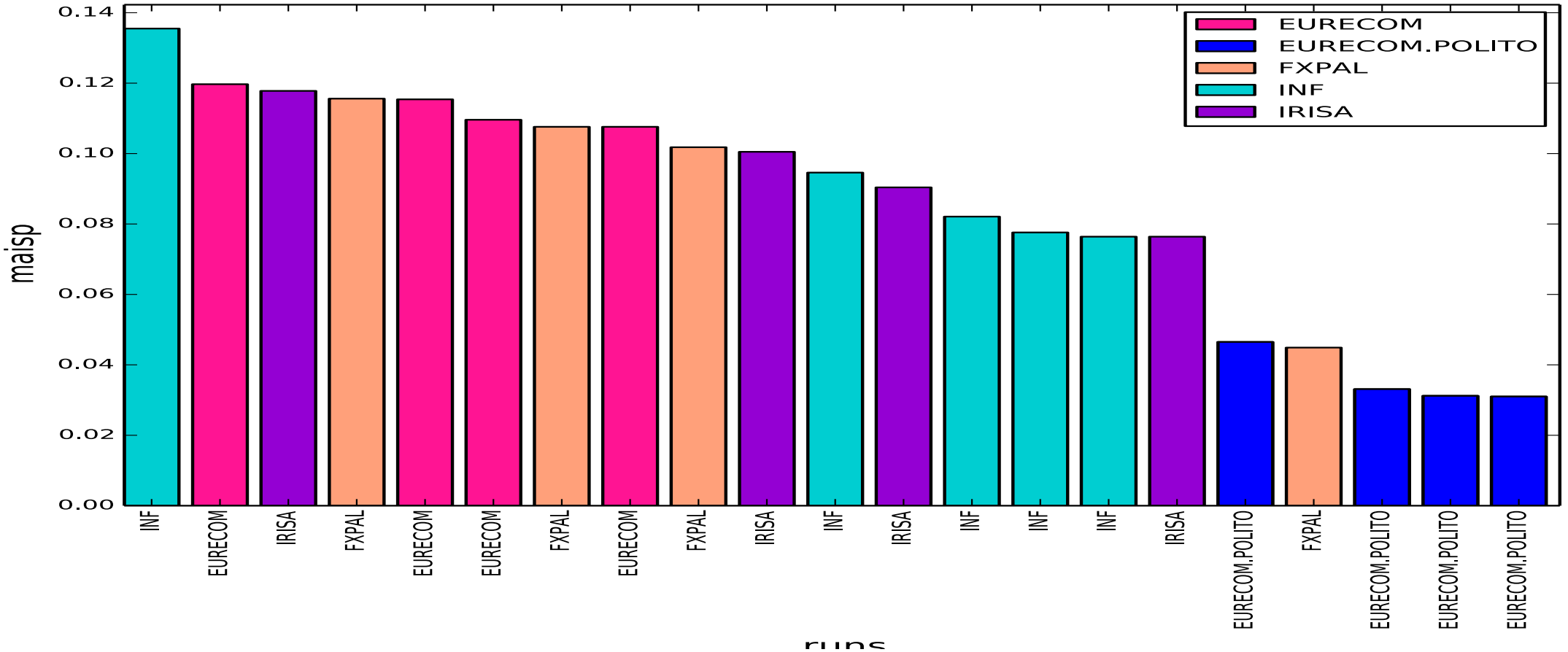


Results

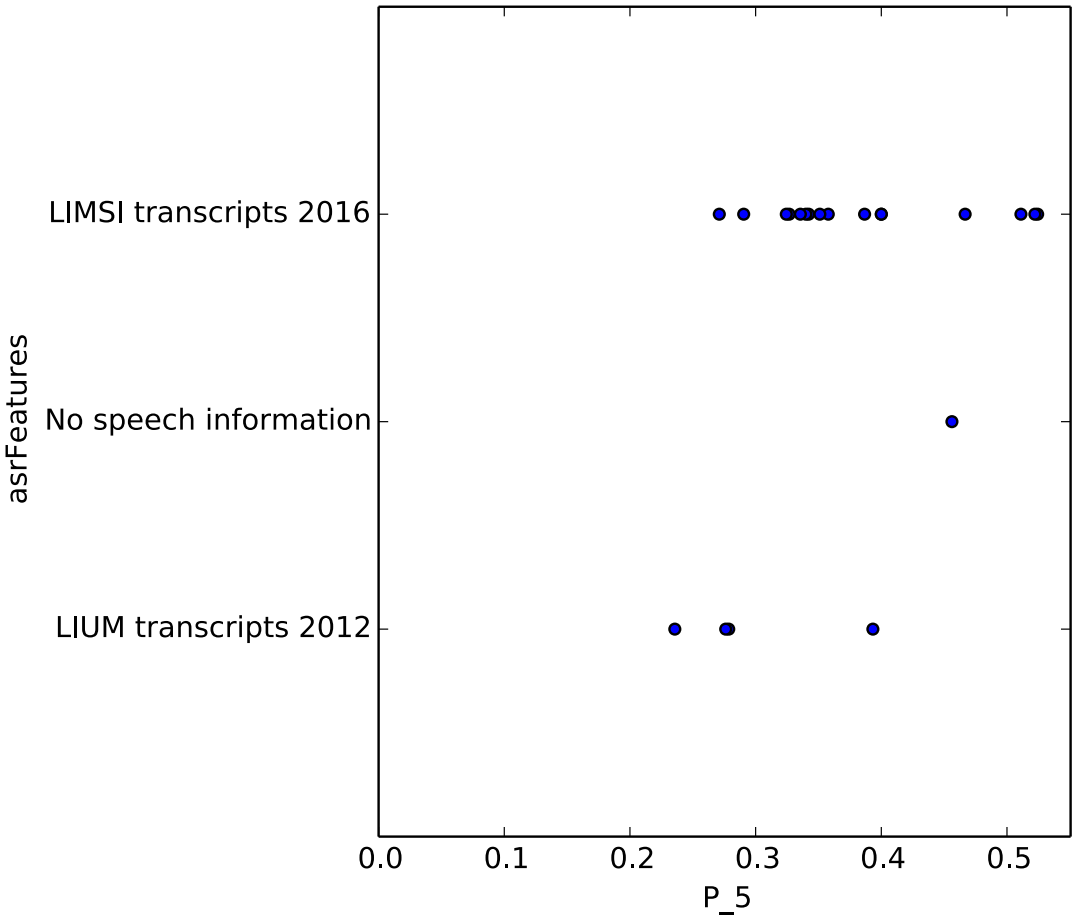
TV'16: Teams (P@5)

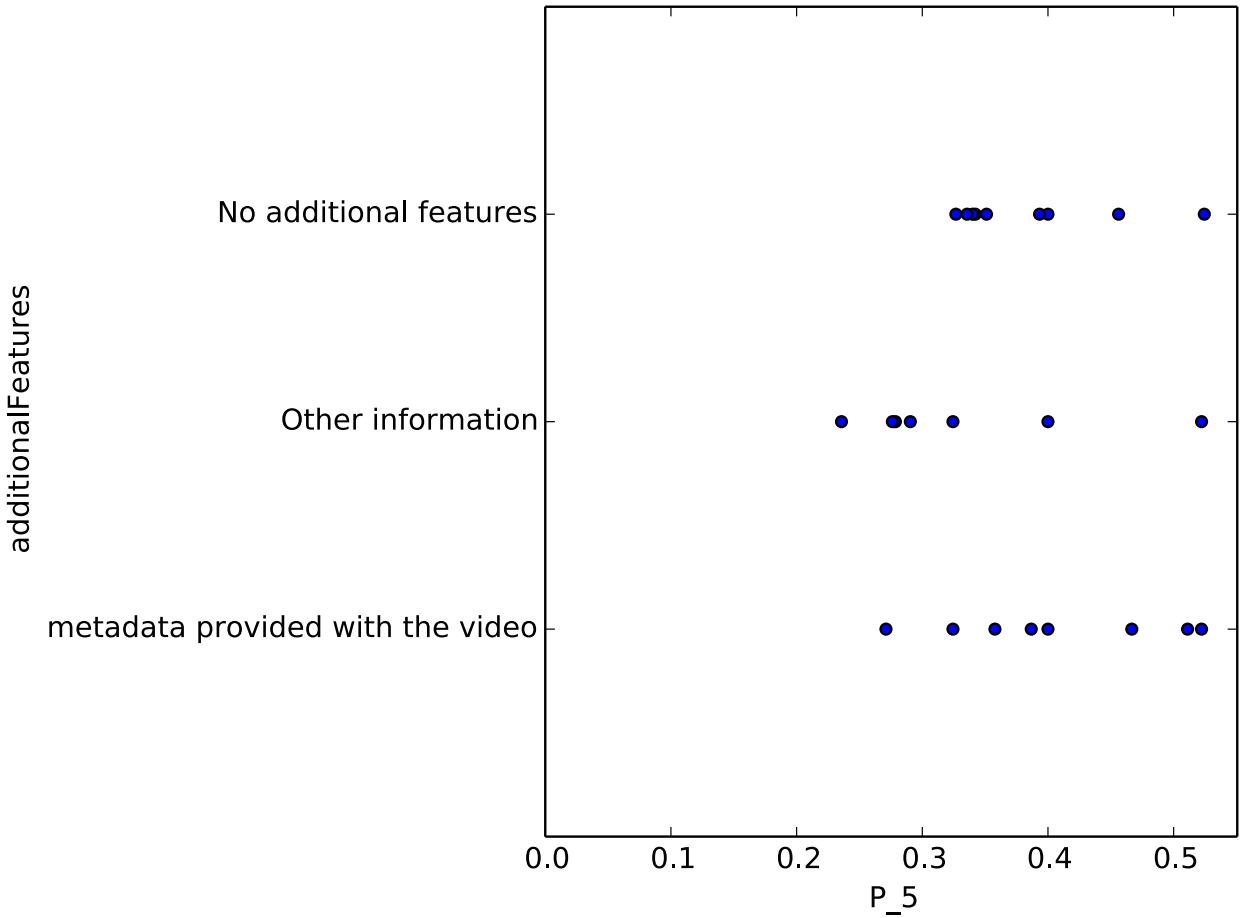


TV'16: Teams (MAiSP)

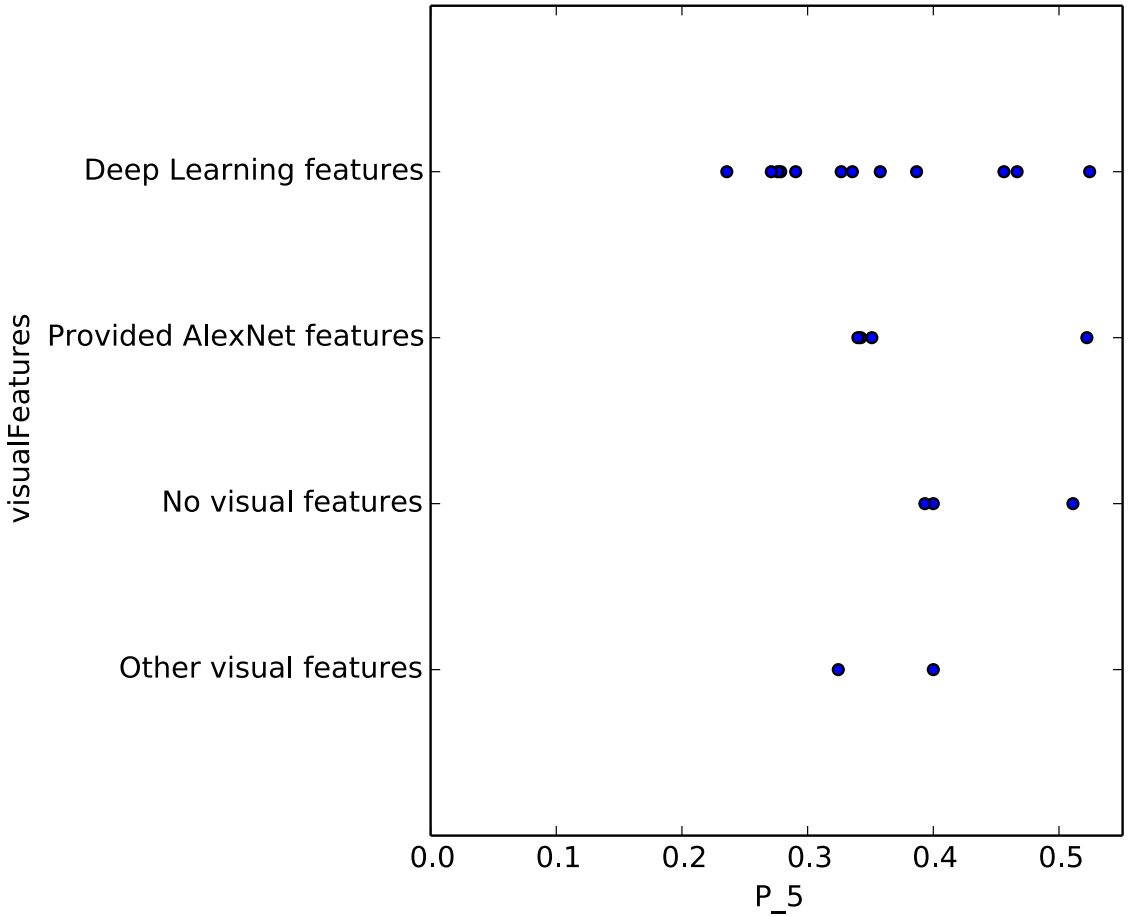


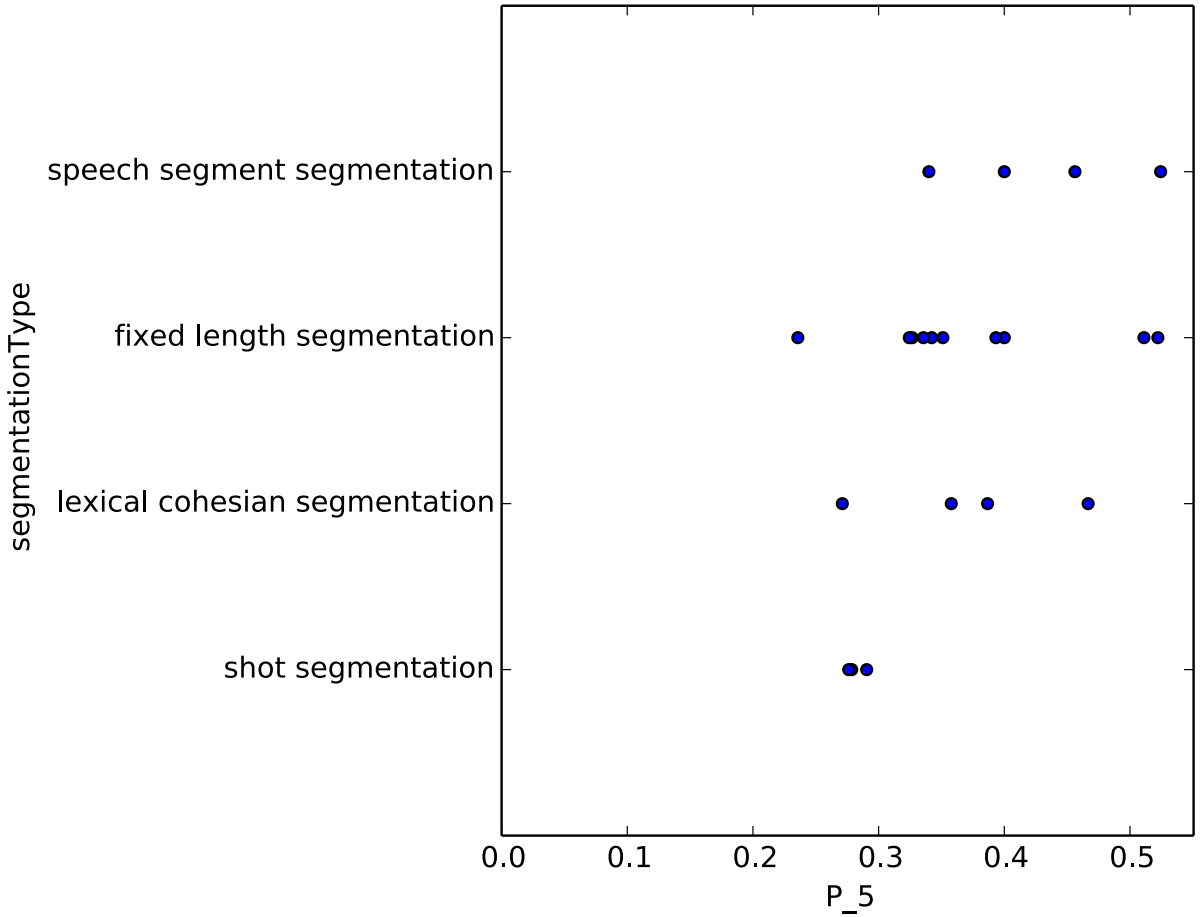
TV'16: Textual features (P@5)



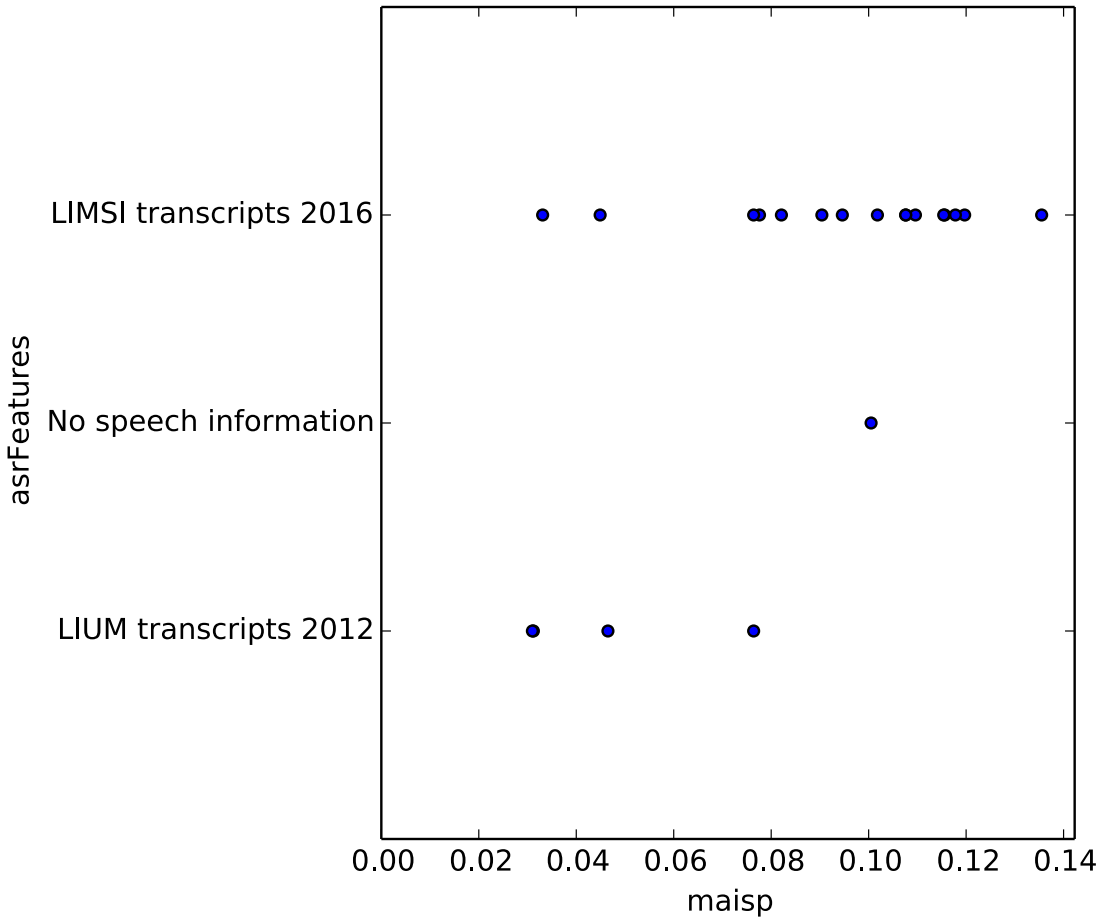


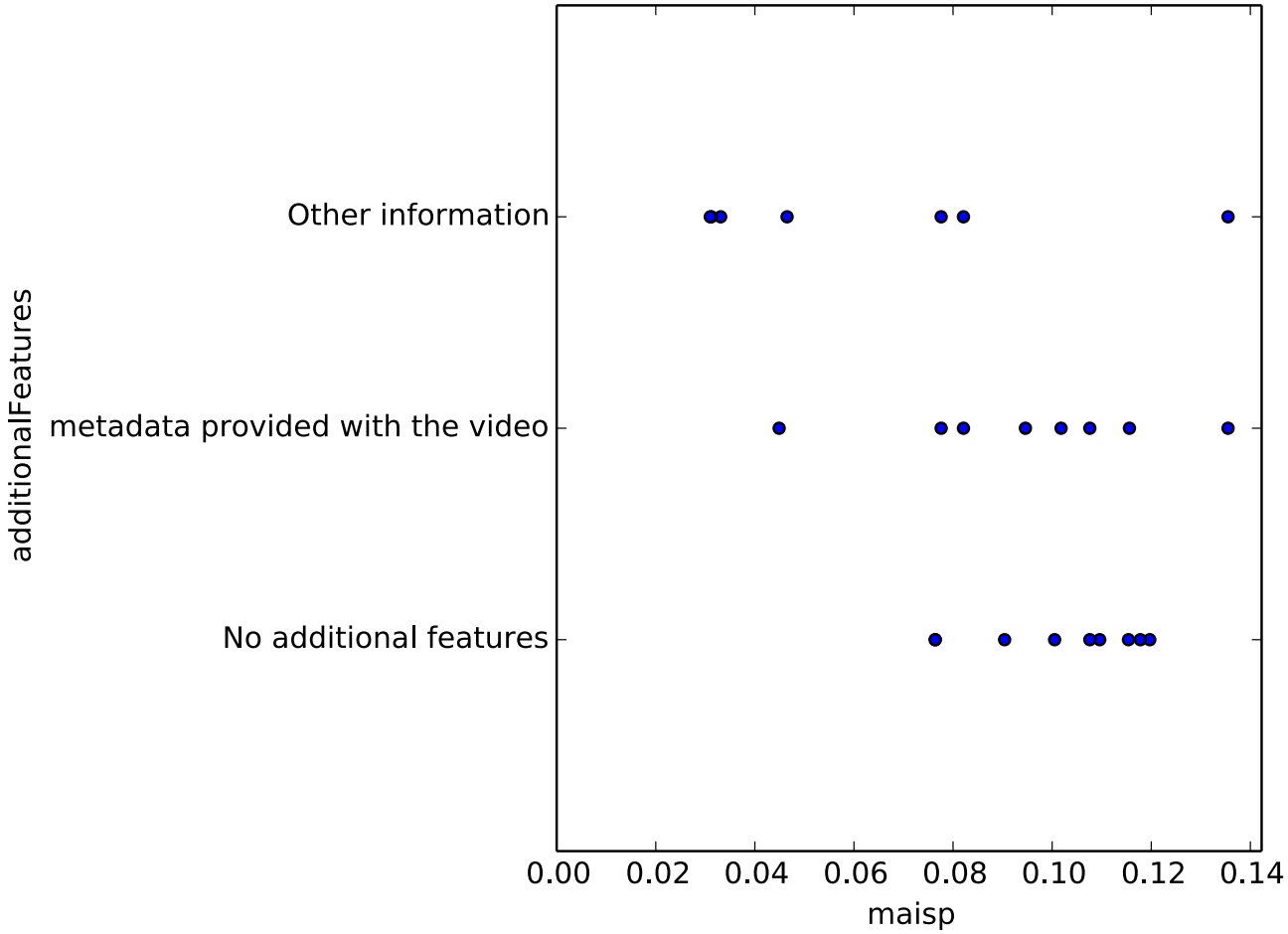
TV'16: Visual features and segmentation (P@5)



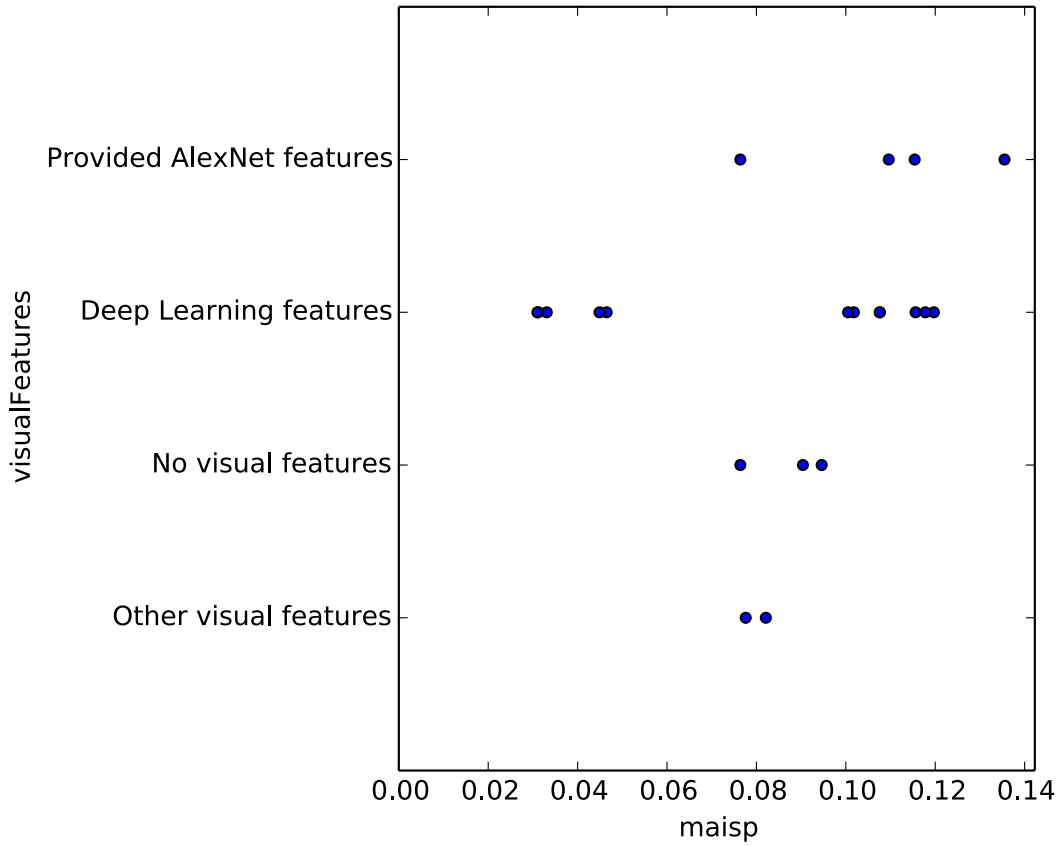


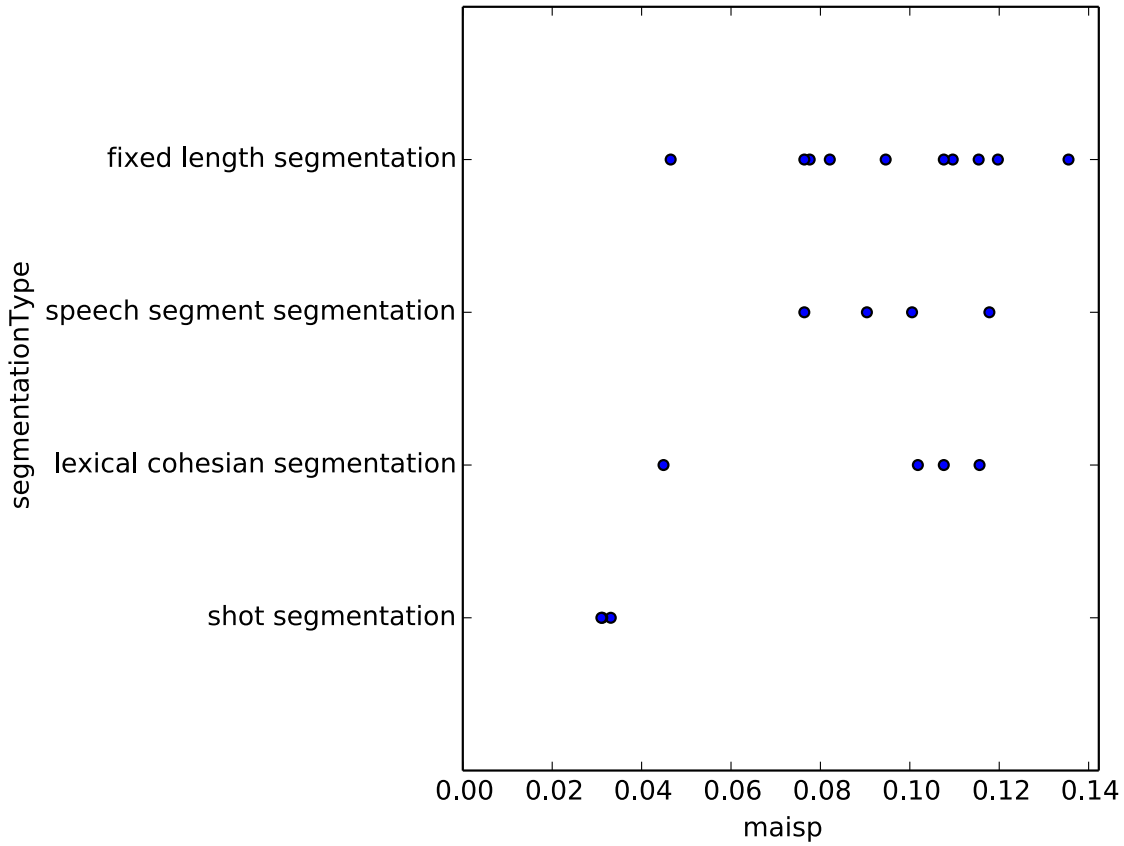
TV’16: Textual features (MAiSP)





TV'16: Visual features and segmentation (MAiSP)





Summary and observations so far

- Disruptive data issue in 2016: no direct comparison with 2015
- However: the task happened!
  - Collection with state-of-the-art features
  - Improved anchors creation process that leads to higher relevance assessment reliability
- Post-workshop analysis is in progress

Practical plans for 2017

- Continuation of the task
- Default: BlipTV collection
  - 2016 multimodal anchors
- Potential addition: BBC 2015 collection
  - No multimodal development anchors
  - What is the added value?
  - Different use scenario
  - Different anchor/target density in professional content
- Task focus:
  - Multimodality in anchors
  - Features of interest?