LABORATOIRE INFORMATIQUE D'AVIGNON
RPM2 SUPPORTED BY ANR
ICASSP'11, MAY 22-27

# Speaker diarization of heterogeneous web video files: a preliminary study

Pierre CLÉMENT[1], Thierry BAZILLON[2] and Corinne FREDOUILLE[1]

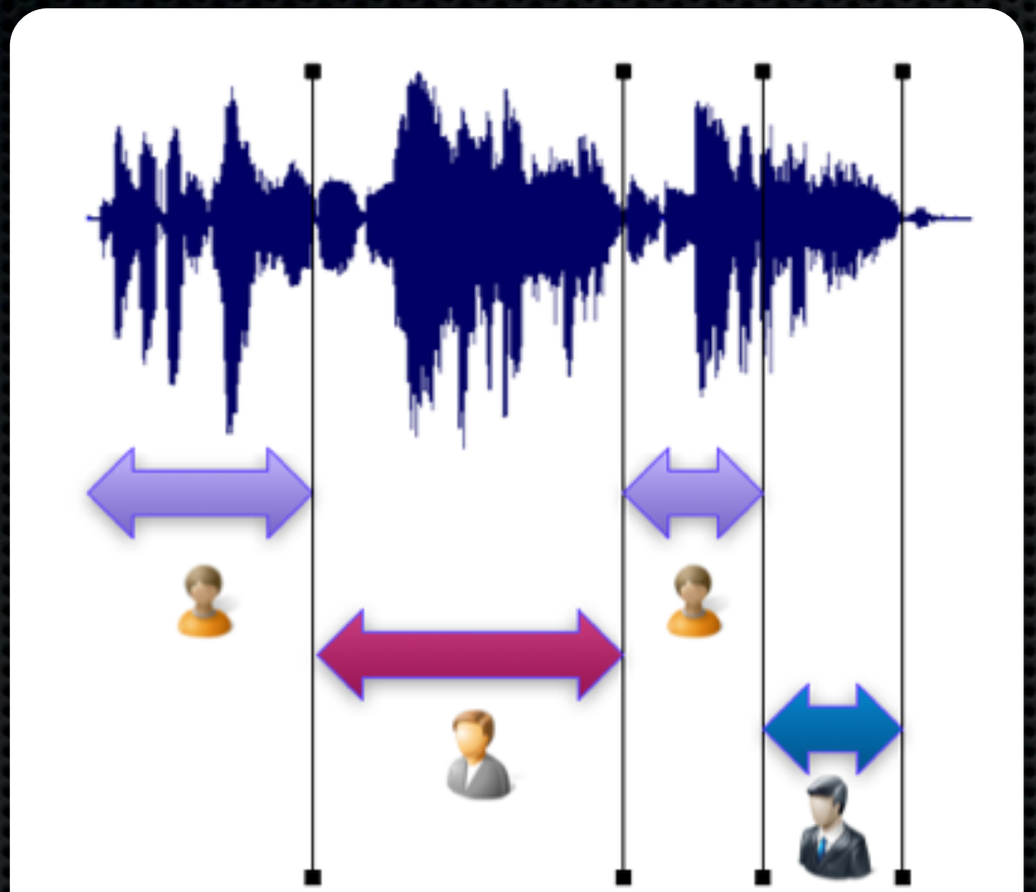
[1]Université d'Avignon - Laboratoire Informatique d'Avignon - CERI/LIA - France

[2]Aix Marseille Université - Laboratoire Informatique Fondamentale - LIF-CNRS - France

# Outline

- Introduction
- LIA speaker diarization system
- Database
- Results
- Conclusion

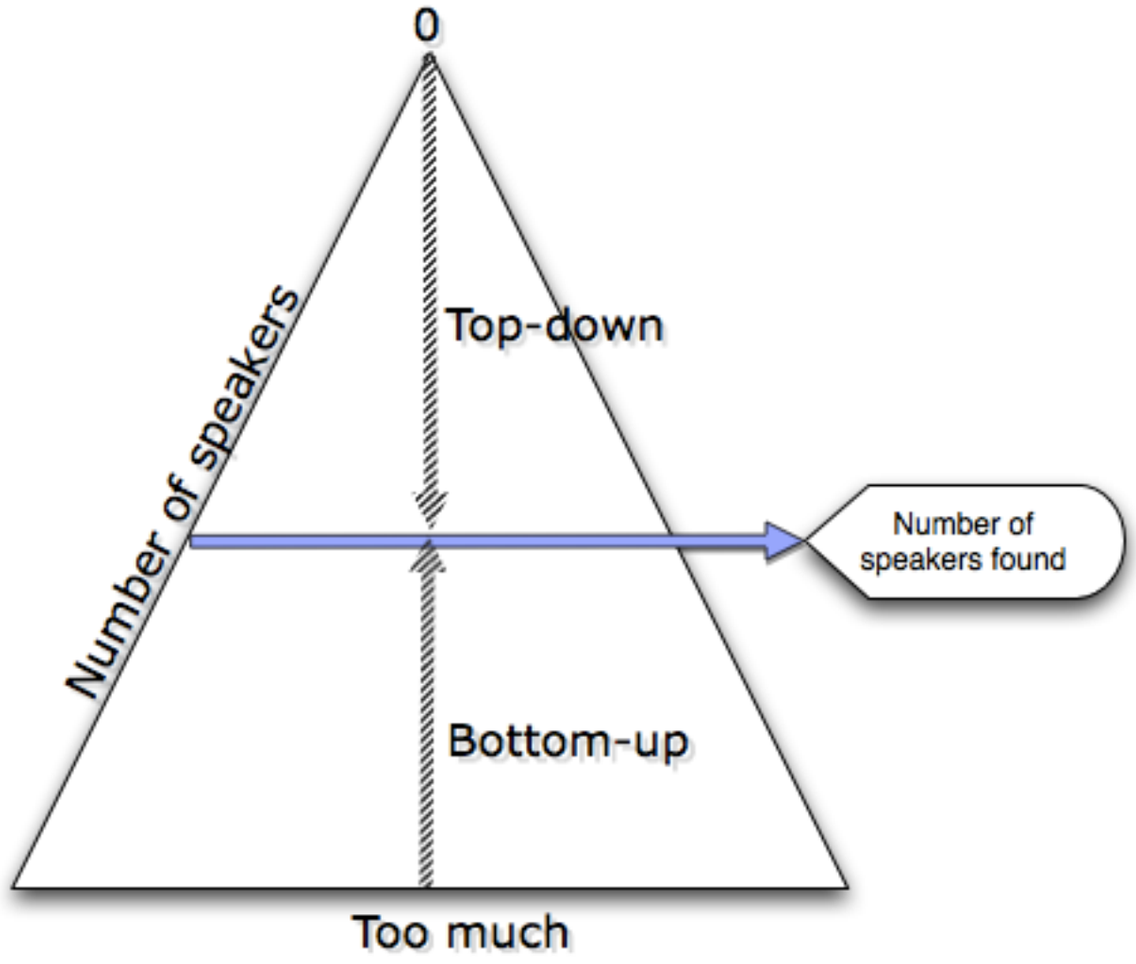
# Speaker Diarization in 3 words



- Who spoke when?
  - No prior information
  - No speaker identification

# Diarization systems

2 approaches: top-down (e.g. LIA system) & bottom-up (e.g. LIUM system)



# Main idea

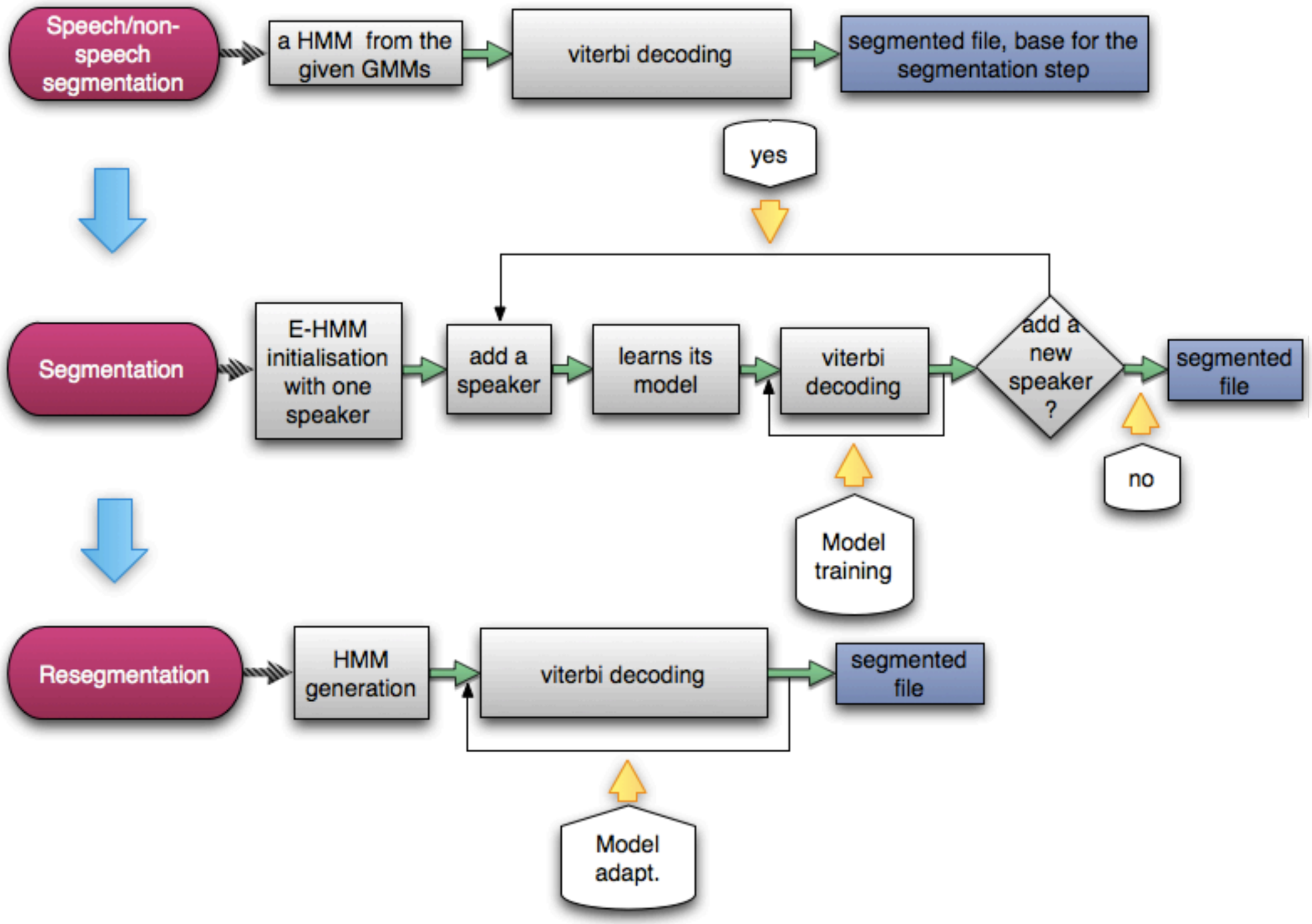
- To test the LIA Speaker Diarization system
- Analyze its behavior

| | Type | Evaluation |
|---|---|---|
| ✔ | Phone | |
| ✔ | Broadcast news | ESTER'08 |
| ✔ | Meeting | RT'09 |
| ? | Web videos | |

# Outline

- Introduction
- LIA speaker diarization system
- Database
- Results
- Conclusion

# LIA speaker diarization system



# Outline

- Introduction
- LIA speaker diarization system
- Database
- Results
- Conclusion

# New context: web videos

- Uncontrolled content:
  - multigenre
  - multisource
- LIA GERARD limited to 7 categories with similar sources

# LIA GERARD: global presentation

- 856 videos in 7 categories
  - documentary, movie trailer, cartoon, commercial, news, sport and music video
- 5 categories handled for now
  - documentary, movie trailer, cartoon, commercial and news
- audio annotations on 129 files (about 10h30)

# LIA GERARD in numbers

| Category | Nb of files | File length (av.) | Part of speech (in %) | Spks by file (av.) | Spk turns by file (av.) | Spk turns length (av. in s.) |
|---|---|---|---|---|---|---|
| Documentary | 29 | 0:06:51 | 72 | 8 | 84 | 3.51 |
| Movie trailer | 30 | 0:02:14 | 54 | **9** | 34 | **2.06** |
| Cartoon | 30 | 0:08:23 | 64 | **11** | **113** | 2.87 |
| Commercial | 10 | 0:01:34 | 68 | 5 | 25 | 2.56 |
| News | 30 | 0:03:05 | 88 | **5** | 26 | **6.31** |

# Outline

- Introduction
- LIA speaker diarization system
- Database
- Results
- Conclusion

# Experimental context

- LIA system compared with the LIUM bottom-up system
- Different sets
  - RT'09: NIST speaker diarization eval. campaign, meeting
  - ESTER'08: French evaluation campaign ESTER II, broadcast news
  - LIA GERARD subset with automatic and manual speech/non-speech segmentation (speech activity detection, SAD), heterogeneous

# Preliminary results

| | RT'09 eval | | ESTER'08 dev | | ESTER'08 eval | | LIA GERARD (auto SAD) | | LIA GERARD (man SAD) | |
|---|---|---|---|---|---|---|---|---|---|---|
| System | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM |
| DER | 18.9 | NA | 14.6 | 8.8 | 15.5 | 8.2 | 73.2 | 55.6 | 38.7 | 34.3 |
| $E_{missed}$ | 0.5 | NA | 1.8 | 0.5 | 1.3 | 0.2 | **9.5** | **13.9** | 0 | 0 |
| $E_{fa}$ | 2.9 | NA | 0 | 2.5 | 1.7 | 1.6 | **27.2** | **13** | 0 | 0 |
| $E_{spk}$ | 15.5 | NA | 12.8 | 5.8 | 12.5 | 6.4 | **36.5** | **28.7** | **38.7** | **34.3** |

All results are in percent

# Preliminary results

| | RT'09 eval | | ESTER'08 dev | | ESTER'08 eval | | LIA GERARD (auto SAD) | | LIA GERARD (man SAD) | |
|---|---|---|---|---|---|---|---|---|---|---|
| System | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM |
| DER | 18.9 | NA | 14.6 | 8.8 | 15.5 | 8.2 | 73.2 | 55.6 | 38.7 | 34.3 |
| $E_{missed}$ | 0.5 | NA | 1.8 | 0.5 | 1.3 | 0.2 | **9.5** | **13.9** | 0 | 0 |
| $E_{fa}$ | 2.9 | NA | 0 | 2.5 | 1.7 | 1.6 | **27.2** | **13** | 0 | 0 |
| $E_{spk}$ | 15.5 | NA | 12.8 | 5.8 | 12.5 | 6.4 | **36.5** | **28.7** | **38.7** | **34.3** |

All results are in percent

# Preliminary results

| | RT'09 eval | | ESTER'08 dev | | ESTER'08 eval | | LIA GERARD (auto SAD) | | LIA GERARD (man SAD) | |
|---|---|---|---|---|---|---|---|---|---|---|
| System | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM |
| DER | 18.9 | NA | 14.6 | 8.8 | 15.5 | 8.2 | 73.2 | 55.6 | 38.7 | 34.3 |
| $E_{missed}$ | 0.5 | NA | 1.8 | 0.5 | 1.3 | 0.2 | **9.5** | **13.9** | 0 | 0 |
| $E_{fa}$ | 2.9 | NA | 0 | 2.5 | 1.7 | 1.6 | **27.2** | **13** | 0 | 0 |
| $E_{spk}$ | 15.5 | NA | 12.8 | 5.8 | 12.5 | 6.4 | **36.5** | **28.7** | **38.7** | **34.3** |

All results are in percent

# Preliminary results

| | RT'09 eval | | ESTER'08 dev | | ESTER'08 eval | | LIA GERARD (auto SAD) | | LIA GERARD (man SAD) | |
|---|---|---|---|---|---|---|---|---|---|---|
| System | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM |
| DER | 18.9 | NA | 14.6 | 8.8 | 15.5 | 8.2 | 73.2 | 55.6 | 38.7 | 34.3 |
| $E_{missed}$ | 0.5 | NA | 1.8 | 0.5 | 1.3 | 0.2 | **9.5** | **13.9** | 0 | 0 |
| $E_{fa}$ | 2.9 | NA | 0 | 2.5 | 1.7 | 1.6 | **27.2** | **13** | 0 | 0 |
| $E_{spk}$ | 15.5 | NA | 12.8 | 5.8 | 12.5 | 6.4 | **36.5** | **28.7** | **38.7** | **34.3** |

All results are in percent

# System and data genre influences

| | Av. nb of spks found | | DER (in %) | | Min DER (in %) | | Max DER (in %) | |
|---|---|---|---|---|---|---|---|---|
| **System** | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM |
| News | 2.9 | 4.6 | 25.7 | 12.8 | 0.1 | 0 | 66.4 | 61.6 |
| Documentary | 4 | 7 | 26.4 | 22 | 4.2 | 0 | 65.6 | 73.8 |
| Movie trailer | 1.4 | 2.8 | 54.3 | 51.1 | 18.9 | 17.7 | 79.2 | 79.2 |
| Cartoon | 4.3 | 10.2 | 49.7 | 53.1 | 22.1 | 31.6 | 72.2 | 73.4 |
| Commercial | 1.5 | 1.9 | 33.6 | 27.7 | 0 | 0 | 62 | 45.9 |

Results on LIA GERARD with manual SAD

# System and data genre influences

| | Av. nb of spks found | | DER (in %) | | Min DER (in %) | | Max DER (in %) | |
|---|---|---|---|---|---|---|---|---|
| **System** | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM |
| News | 2.9 | 4.6 | 25.7 | 12.8 | 0.1 | 0 | 66.4 | 61.6 |
| Documentary | 4 | 7 | 26.4 | 22 | 4.2 | 0 | 65.6 | 73.8 |
| Movie trailer | 1.4 | 2.8 | 54.3 | 51.1 | 18.9 | 17.7 | 79.2 | 79.2 |
| Cartoon | 4.3 | 10.2 | 49.7 | 53.1 | 22.1 | 31.6 | 72.2 | 73.4 |
| Commercial | 1.5 | 1.9 | 33.6 | 27.7 | 0 | 0 | 62 | 45.9 |

Results on LIA GERARD with manual SAD

# System and data genre influences

| | Av. nb of spks found | | DER (in %) | | Min DER (in %) | | Max DER (in %) | |
|---|---|---|---|---|---|---|---|---|
| **System** | LIA | LIUM | LIA | LIUM | LIA | LIUM | LIA | LIUM |
| News | 2.9 | 4.6 | 25.7 | 12.8 | 0.1 | 0 | 66.4 | 61.6 |
| Documentary | 4 | 7 | 26.4 | 22 | 4.2 | 0 | 65.6 | 73.8 |
| Movie trailer | 1.4 | 2.8 | 54.3 | 51.1 | 18.9 | 17.7 | 79.2 | 79.2 |
| Cartoon | 4.3 | 10.2 | 49.7 | 53.1 | 22.1 | 31.6 | 72.2 | 73.4 |
| Commercial | 1.5 | 1.9 | 33.6 | 27.7 | 0 | 0 | 62 | 45.9 |

Results on LIA GERARD with manual SAD

# Outline

- Introduction
- LIA speaker diarization system
- Database
- Results
- Conclusion

# Study outlines

- Difficulties for diarization systems on heterogeneous web content
- Difficult database:
  - lot of variability between categories
  - high interactivity (number and duration of speaker turns)
  - lot of background noise

# Perspectives

- Deal only with the categories where systems are least worst (news and documentaries)
- Use high-level information from the video stream to help the decision

# Questions

## DER by set and system