TRECVID-2004 Story segmentation task: Overview

Wessel Kraaij
TNO TPD
&
Joaquim Arlandis
NIST

Task definition

- Identify the individual news items in a news show
- Introduced at TRECVID 2004, has been studied in speech/IR community (TDT)
- Hope to show the gain of using video features

1. Segmentation task
   - Identify story boundaries in CNN and ABC news shows
   - Ground truth based on TDT 2 annotations
   - Evaluation based on precision & recall, boundaries have to be within +/- 5 seconds interval around ground truth boundaries
2. News classification task: **DISCONTINUED in 2004**
   - § Binary classification proved too easy in 2003
   - § No ground truth available for fine grained categorization
   - § ASR probably dominating feature for multiclass problem.

8 Participating Groups

CLIPS-LSR-LIS (CLIPS+)

IBM Research (IBM)

Imperial College London (ICL)

KDDI R&D Laboratories (KDDI)

National Cheng Kung University ELITE Center (NCKU)

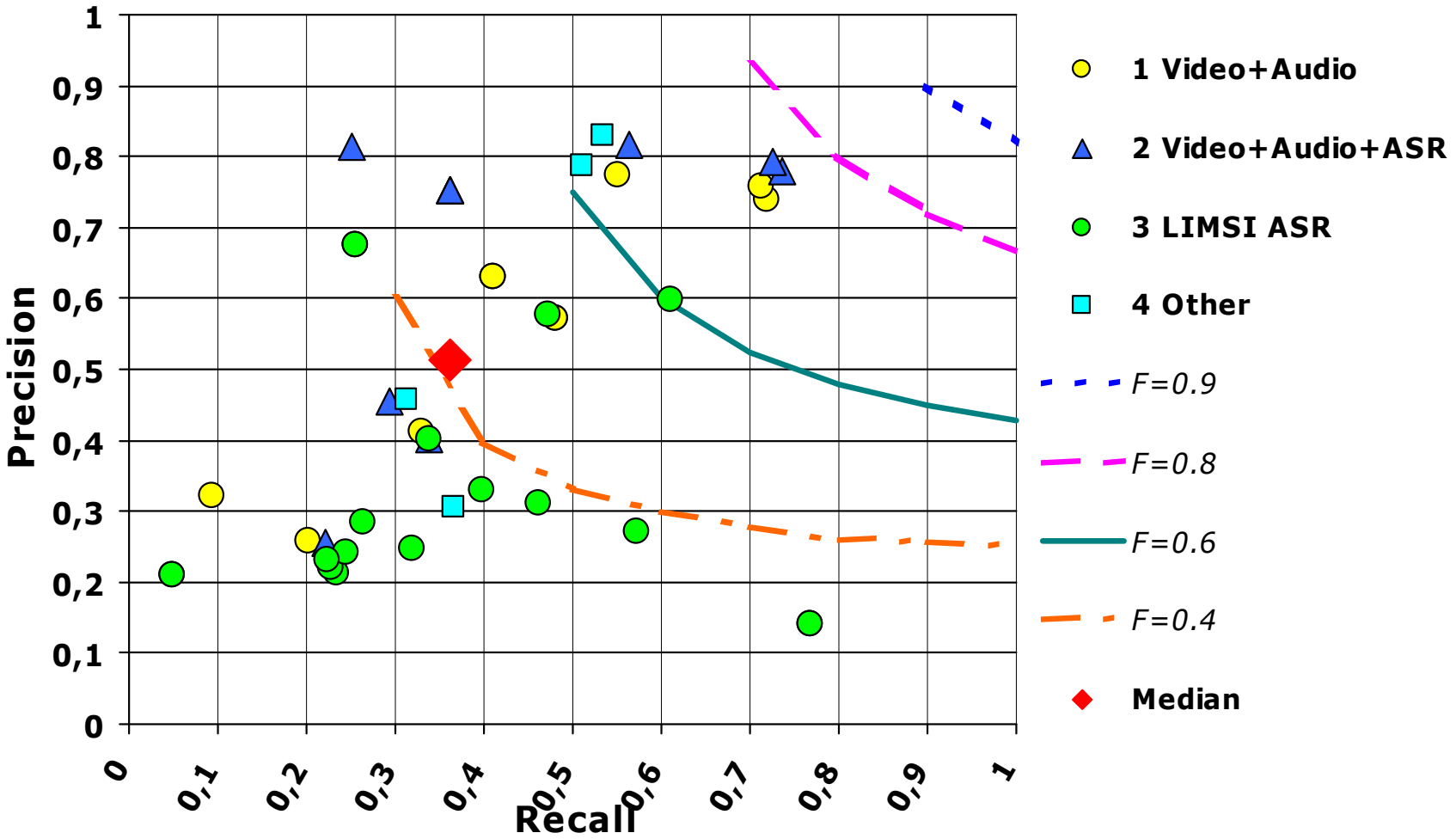
RMIT University (RMIT)

University of Central Florida (UCF)

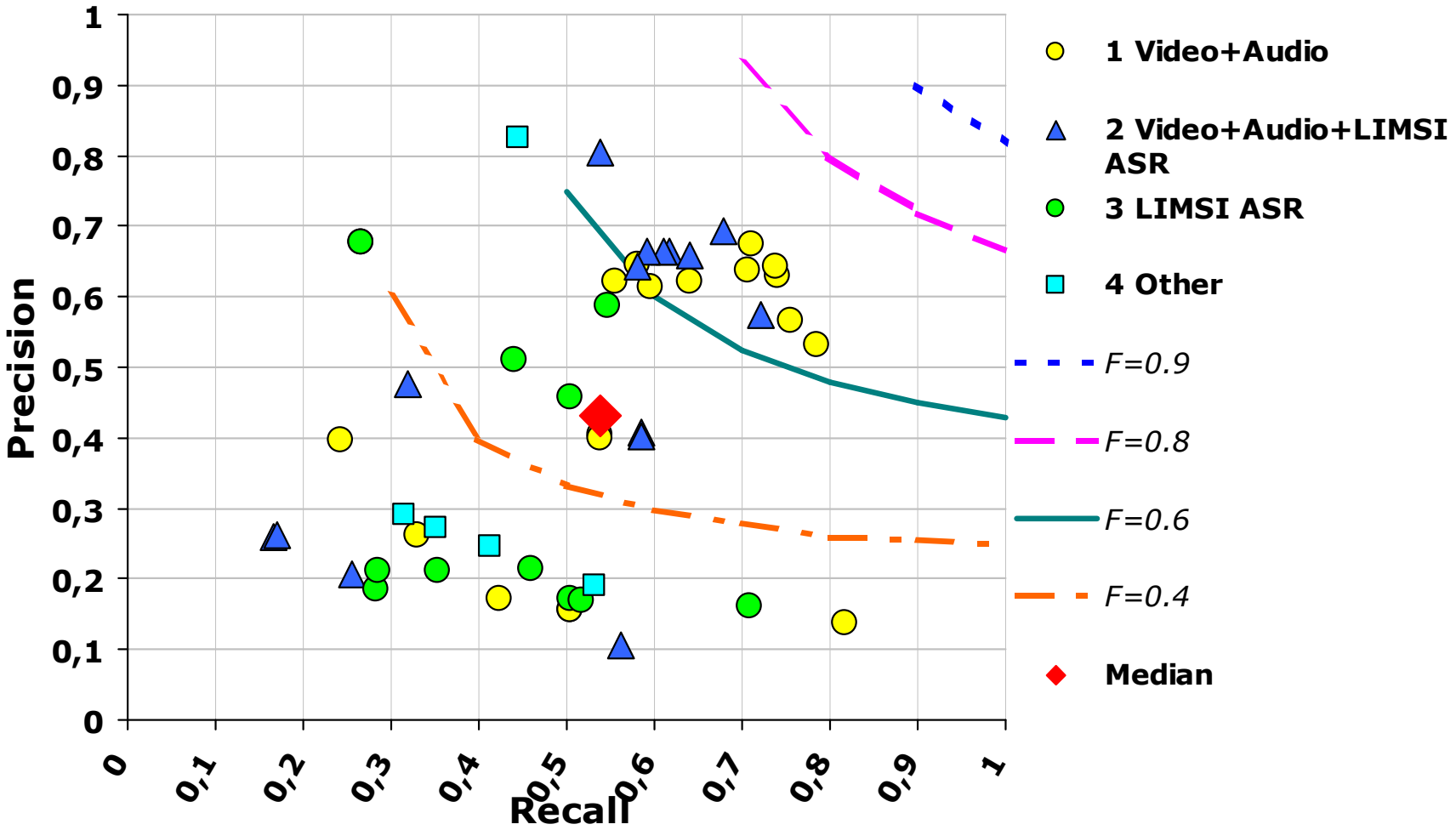
University of Iowa (UIOWA)

(TV2003: 8 groups)

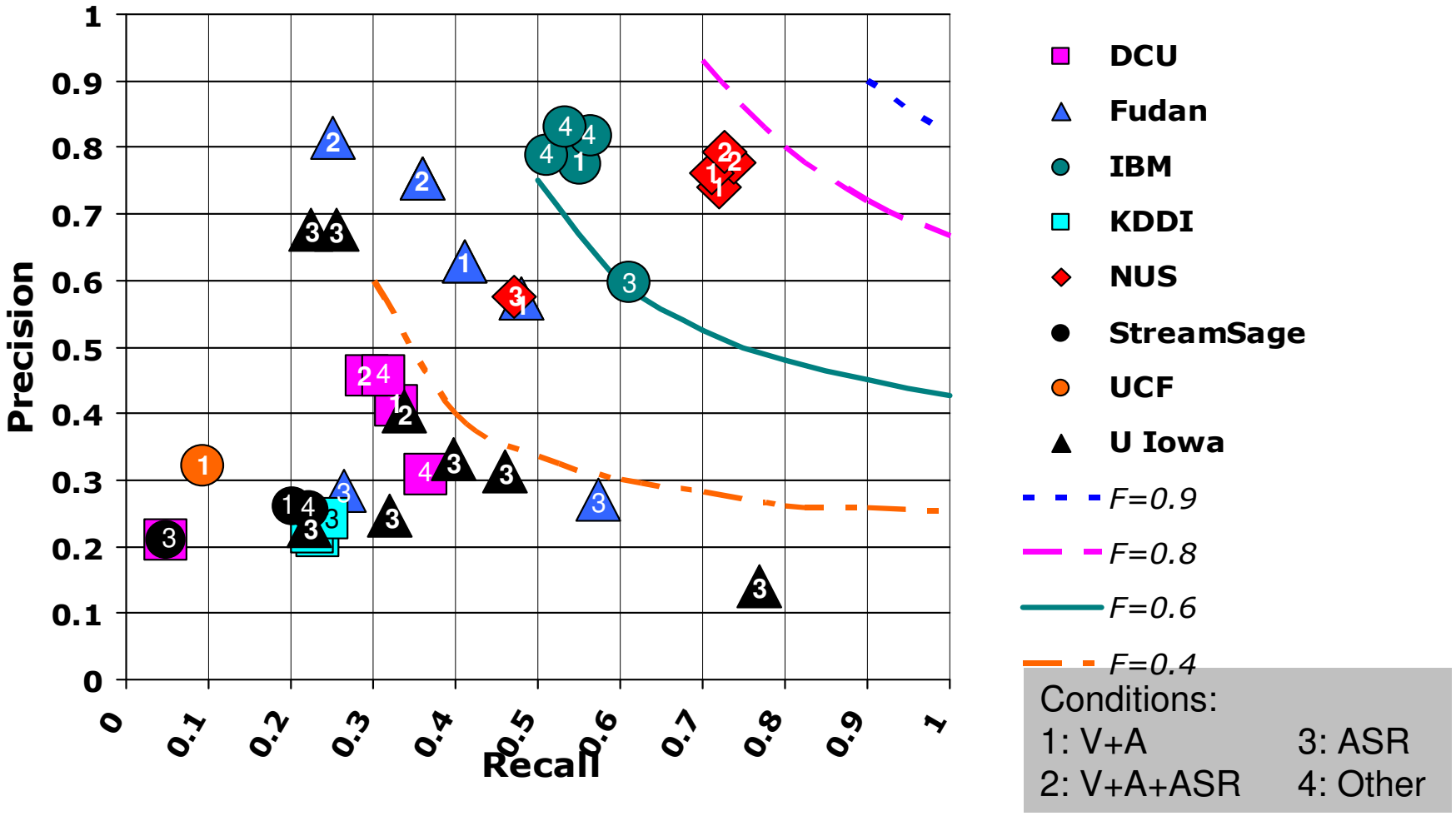
Story segmentation: recall and precision by condition (2003)



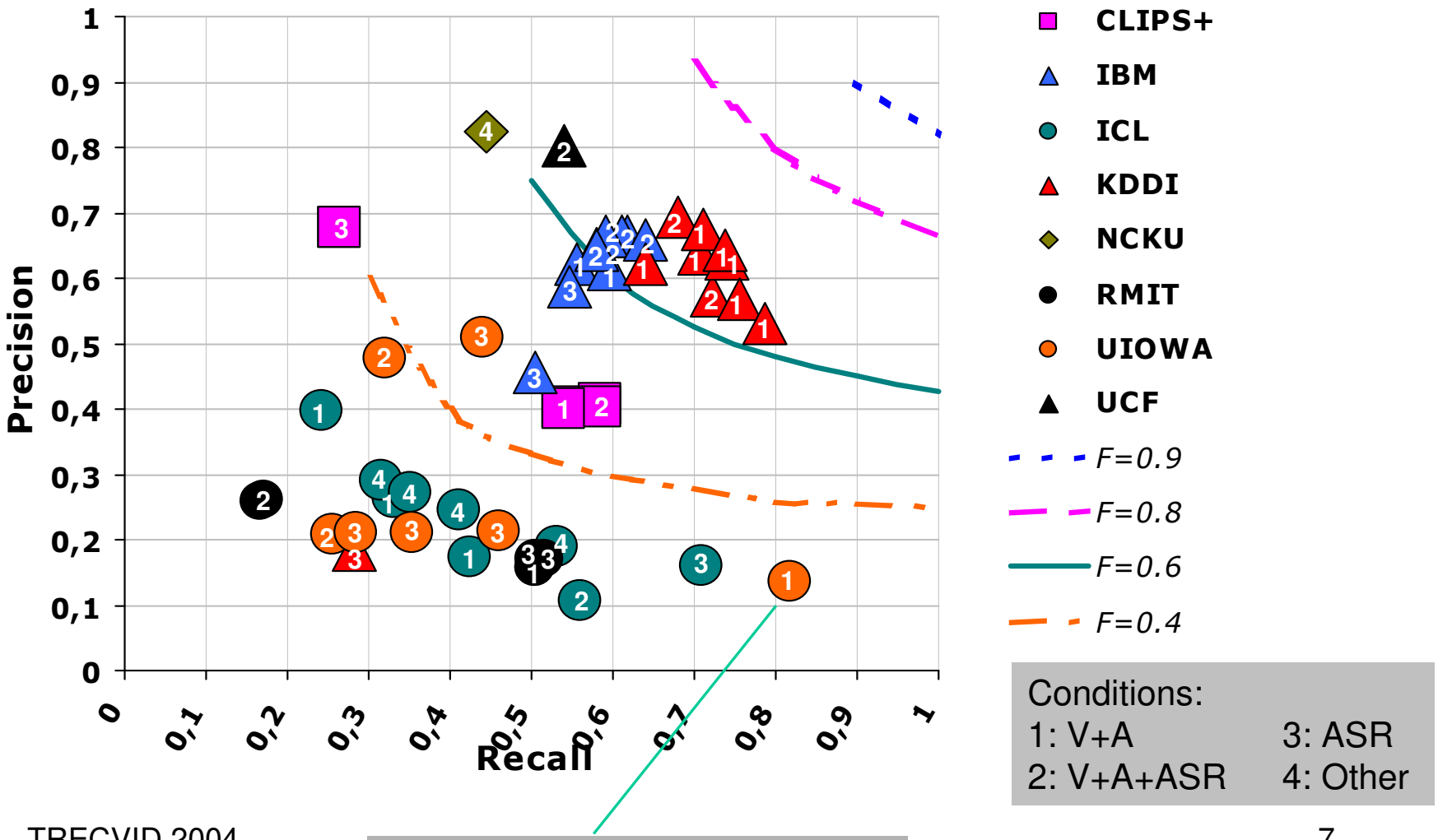
Story segmentation: recall and precision by condition (2004)



Story segmentation: recall and precision by system and condition (1-4) (2003)



Story segmentation: recall and precision by system and condition (1-4) (2004)

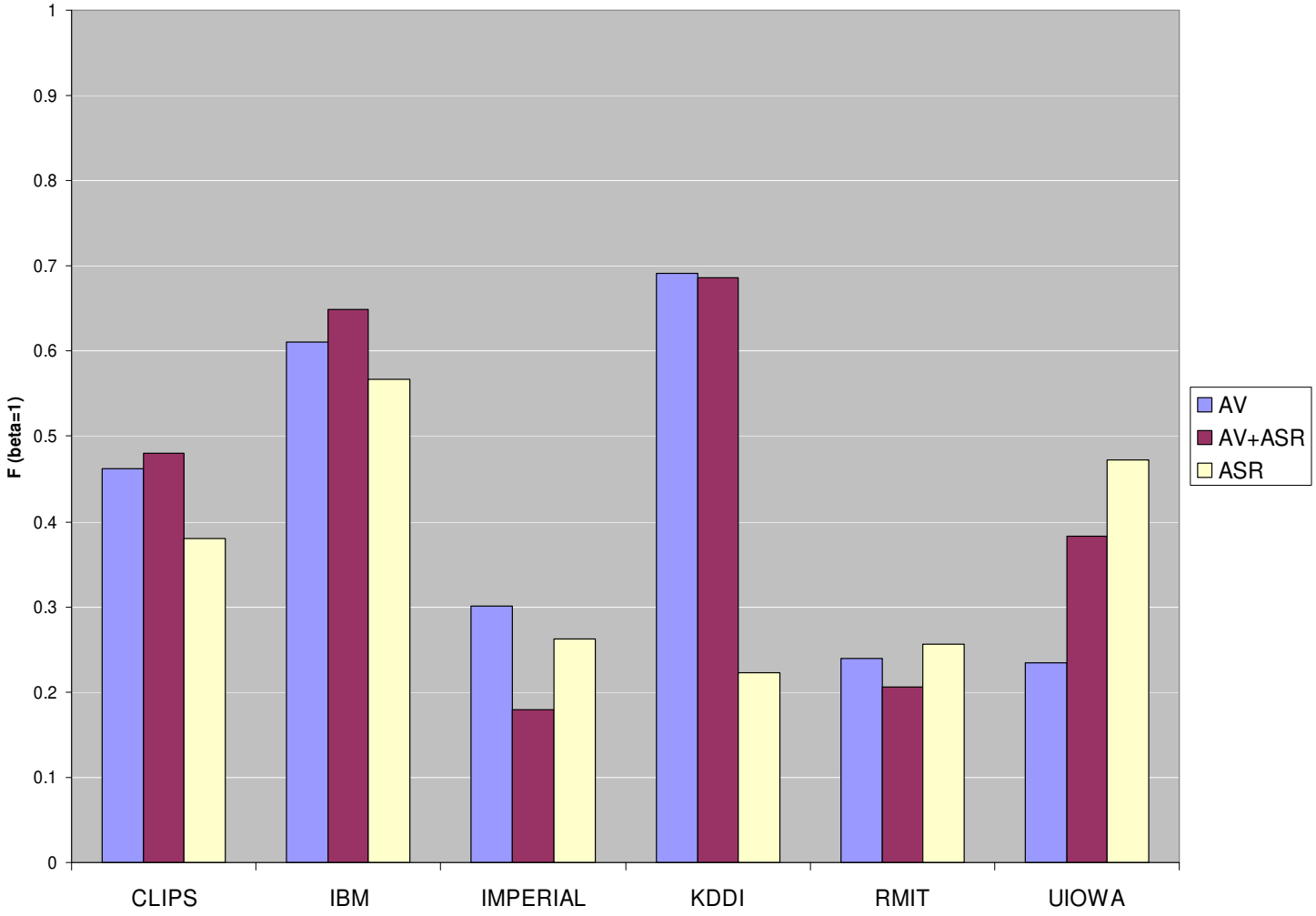


Conditions:

1: V+A	3: ASR
2: V+A+ASR	4: Other

Shot boundary = story boundary

Segmentation, within system (F)
Best run each team



CLIPS-LSR-LIS: headlines

- Approach
  - logical combination of single feature detectors
  - candidates (AV) : shot boundaries plus long pauses
- Features
  - audio change: peak detection on BIC curve
  - Speaker segmentation (LIMSI)
  - example bases jingle detection
  - ASR: cue phrases
- Results
  - audio features boost precision of SB baseline at the cost of some recall
  - adding ASR improves bothe recall and precision with 0.05
  - Performance close to median
  - Feature combination method might be improved

IBM Research: headlines

- Approach
  - Visual clue cluster construction (information bottleneck principle [Tishby, Pereira, Bialek99])
- Features
  - Text
  - Video
  - Rich prosody
- Results
  - Well above median
  - Best system: AV+ASR
  - Best overall ASR only system
- *PRESENTATION FOLLOWS*

Imperial College London: headlines

- Approach
  - AV: Anchor shot starts new story and finishes previous story
  - ASR: merge segments when similarity exceeds threshold. Window size: 5 tokens, score determined by number of similar words, weighted by type and distance
- Features
  - Anchor detection : K-NN classifier trained on 2003 data
- Results
  - Runs score below median

KDDI R&D Laboratories: headlines

- Approach
  - Feature combination using SVM
  - Section-specialized segmentation (top stories, headl. sports)
  - Anchor shot segmentation (pauses)
- Features
  - Shot segmenter, shot duration, shot density
  - Audio: average RMS, avg RMS first n frames, freq of audio class (silence, speech, music, noise)
  - Motion: H,V, total, intensity
  - Color layout and distance of various frames
- Results
  - Best for TV2004
  - Section-sp. segmentation: RecallÆ , PrecisionÆ
  - Anchor shot segmentation: RecallÆ
  - Post filtering: helps and hurts
- *PRESENTATION FOLLOWS*

National Cheng Kung University ELITE Center: headlines

- Approach
  - unknown (condition 4 run)
- Results
  - well above median (best precision result)

RMIT University: headlines

- Approach
  - Condition 1: story boundaries of anchor shots or speech pauses
  - Condition 2: segment ASR using shot boundaries. Story boundaries are defined by detecting minima of a weighted vocabulary overlap score in a window consisting of several shots.
  - Condition 3: similar, but candidate boundaries are based on speech pauses only.
- Features
  - Anchor detection, pause detection
  - Vocabulary overlap using a search engine
- Results
  - Runs score below median
  - Problems with “evidence combination”

University of Central Florida: headlines

- Approach
  - News show grammar: story:= <anchor><shot>+
  - Special treatment of weather and sports stories (no anchors)
  - Merge adjacent similar stories (to reduce FA rate)
- Features
  - Anchor and sub-anchor detection
  - Color similarity
  - Normalized text similarity
- Results
  - Only a single condition 2 run was submitted, performing well above median
- *PRESENTATION FOLLOWS*

University of Iowa: headlines

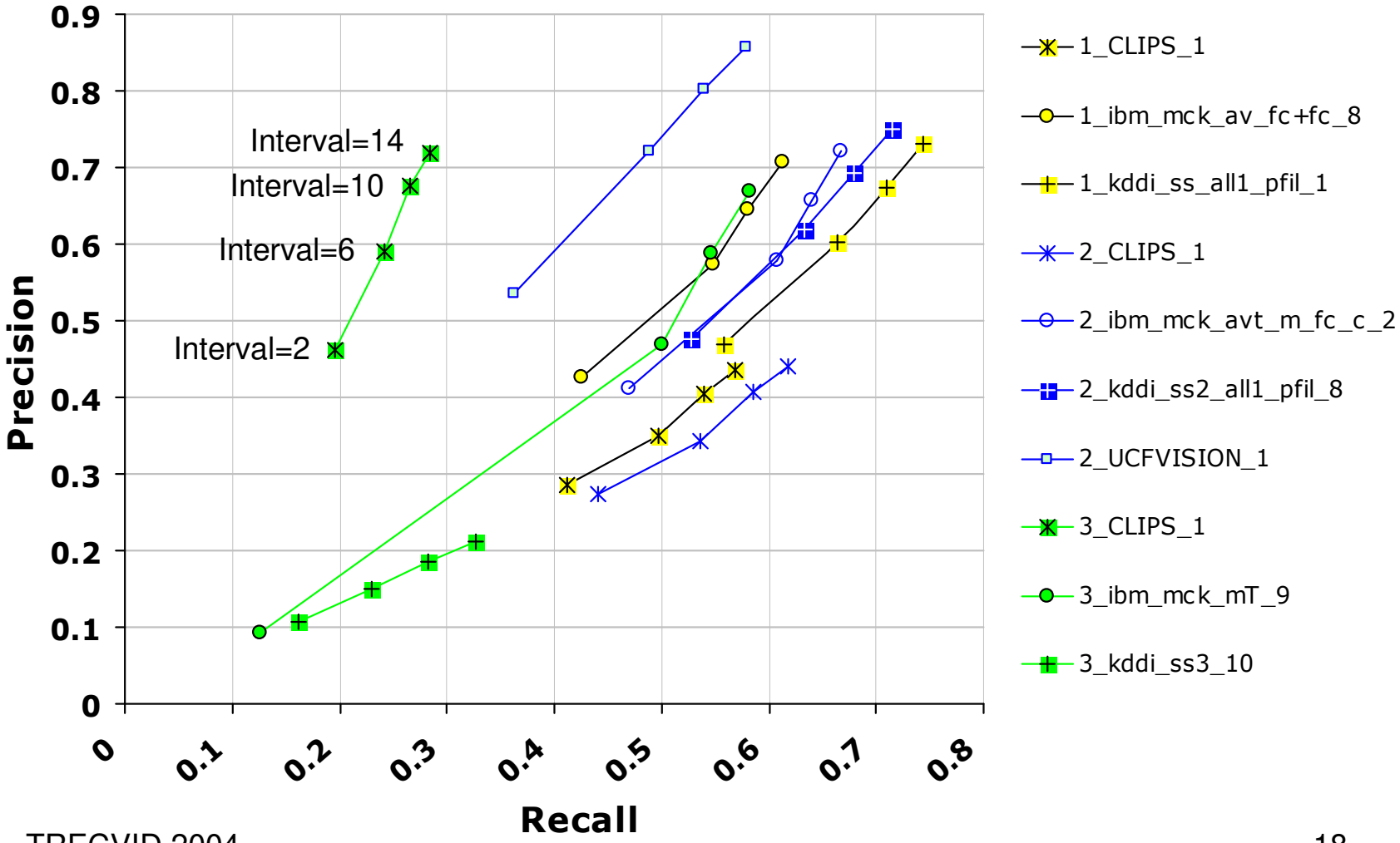
- Approach
  - Condition1: story boundary=shot boundary
  - Condition 3: extended set of trigger phrases
- Features
  - Color histogram, pixel similarity, edge similarity
  - Trigger phrases
  - Speech pauses
- Results
  - 1 << 2 << 3
  - Best run (3) is close to median

Observations

- *Wide range of approaches and of system complexity*
- *AV runs have improved recall w.r.t. 2003*
- Combination of AV and ASR gives a small gain for segmentation
- Most approaches are generic
- Most systems use just “local” features

- Are the combination methods optimal?
- Are the ASR segmentation runs state of the art?
- Still room for improvement or are we overfitting the LDC annotation standard?

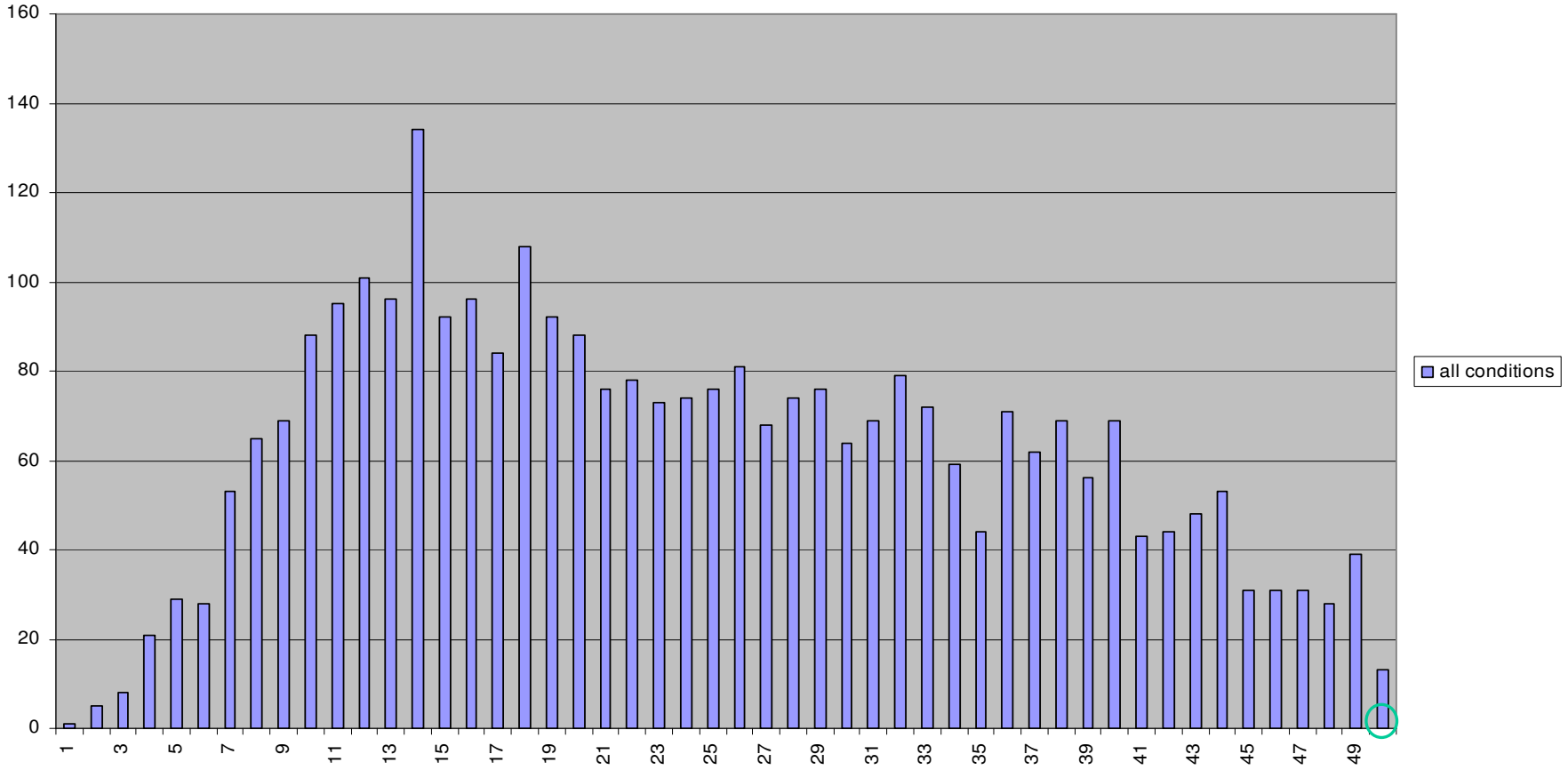
Story segmentation: system degradation as a function of the evaluation interval



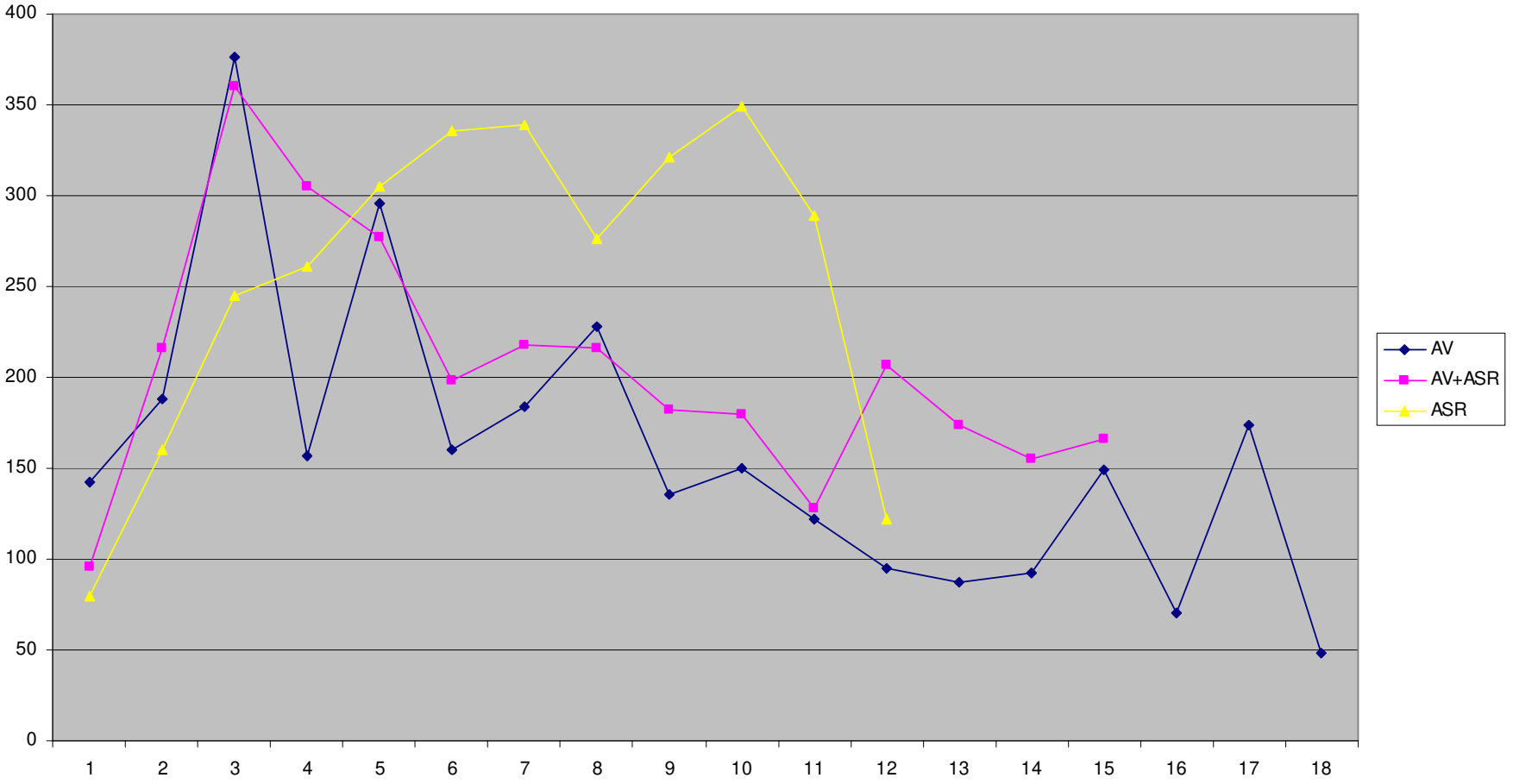
Preliminary error analysis across systems

- Count FN (FP, TP , TN) across systems
- Select FN that occur for most/all systems
- Watch corresponding clips
- Categorize problems across conditions

#boundaries missed by the same #systems



#boundaries missed by the same #systems (per condition)



Total # boundaries= 3105

Overlap between missed boundaries of different conditions

- 14 boundaries are not found by all runs
- No boundary found by all runs
- Analysis of boundaries missed by all runs of a condition:

	total	C1	C2	C3	All
C1	48	13	21	0	14
C2	166	21	96	35	14
C3	122	0	35	73	14

Analysis of False Negatives

- Method:
  - Determine story boundaries missed by all systems, specific for a certain condition.
- Condition 1 (13)
  - A: 7/13 two adjacent stories of same type (e.g. sports: hockey follows basketball)
  - B: 3/13 substories on same issue (boundary is questionable)
  - C: 4/13 special cases: intro, trailing anchor shot, short anchor shots, a story without voice
  - D: 0/13 stories are not related
  - Some of these missed story boundaries do not coincide with a shot boundary (4/13)

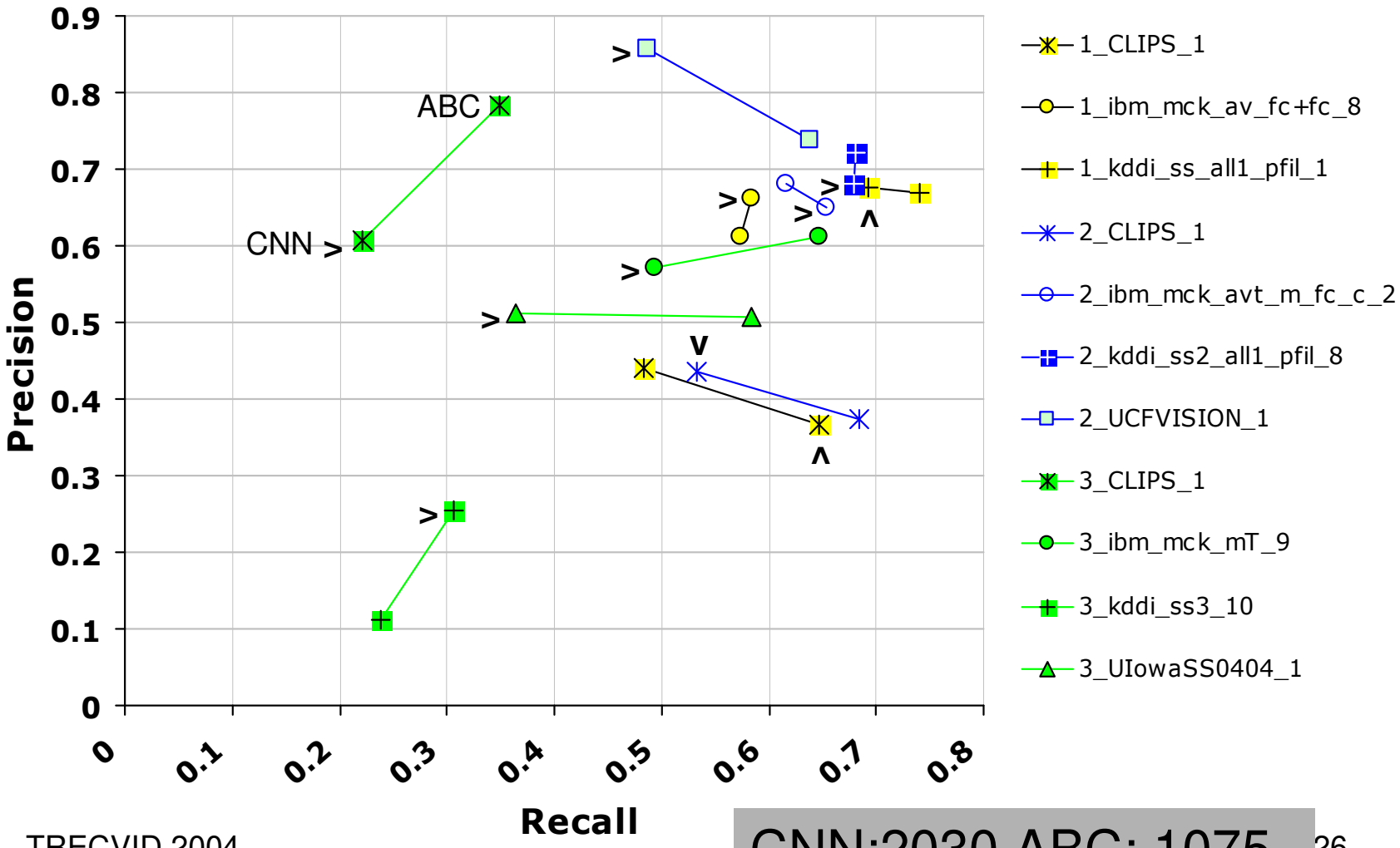
Analysis of False Negatives (continued)

- Condition 2 (96)
  - A: 11/32 two adjacent stories of same type (e.g. sports: hockey follows basketball)
  - B: 9/32 substories on same issue (boundary is questionable)
  - C: 5/32 special cases: start, end or break of program special anchor “bridge”
  - D: 4/32 stories are not related
  - Some of these missed story boundaries do not coincide with a shot boundary (12/32)
  - Some of these missed story boundaries do not coincide with a speech pause

Analysis of False Negatives (continued)

- Condition 3 (73)
  - A: 11/25 two adjacent stories of same type (e.g. sports: hockey follows basketball)
  - B: 9/25 substories on same issue (boundary is questionable)
  - C: 5/25 special cases: just breaks
  - D: 4/25 stories are not related (but do use some equivalent words e.g. money)
  - E: 10/25 CNN "top stories"
  - Most of these missed story boundaries do not coincide with a shot boundary (20/25), in about 10 of these cases there is a logo change indicating topic change.

Story segmentation: CNN versus ABC



CNN:2030 ABC: 1075