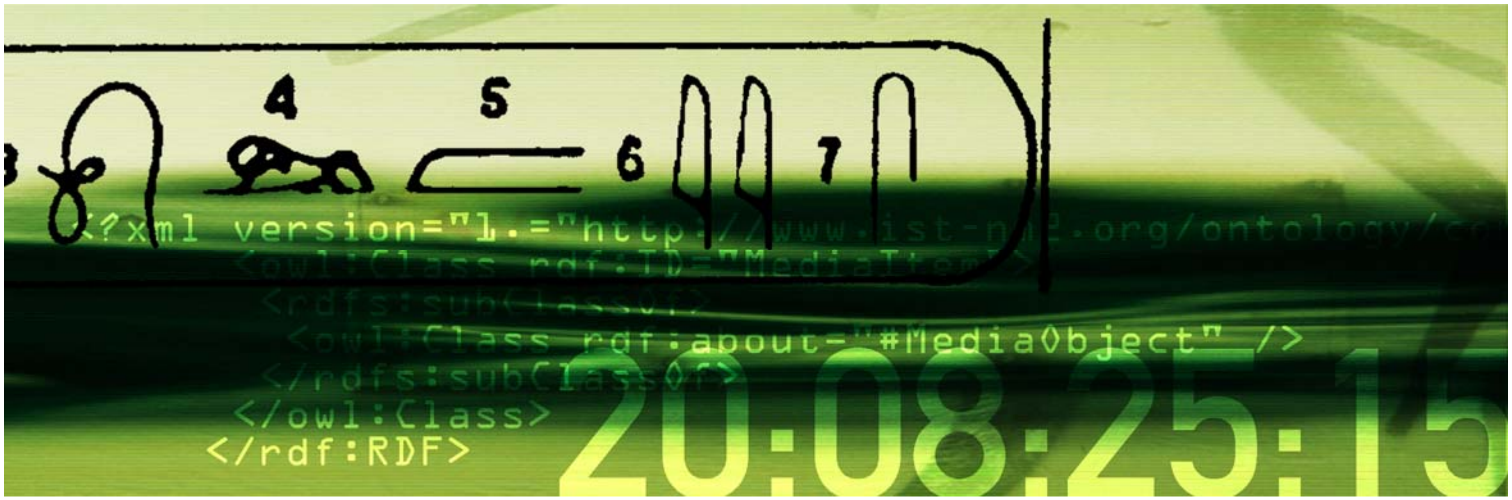
JOANNEUM RESEARCH

Institut für Informationssysteme & Informationsmanagement



Comparison of Content Selection Methods for Skimming Rushes Video

Werner Bailer, Georg Thallinger

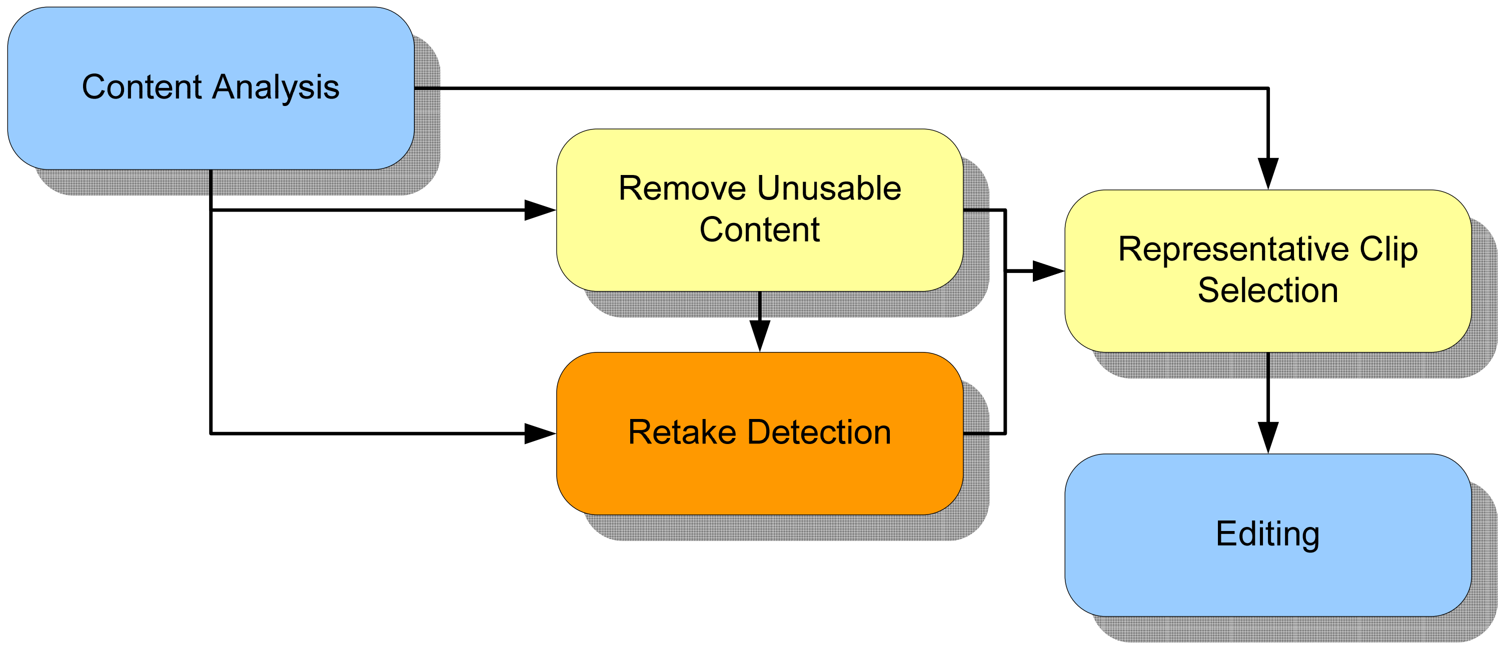
TRECVID Video Summarization Workshop @ ACM MM, 2008-10-31

www.joanneum.at

Outline

- Summary creation process
- Content selection
  - a closer look to the problem
  - rule-based approach
  - HMM based approach
- Results and comparison
- Conclusion

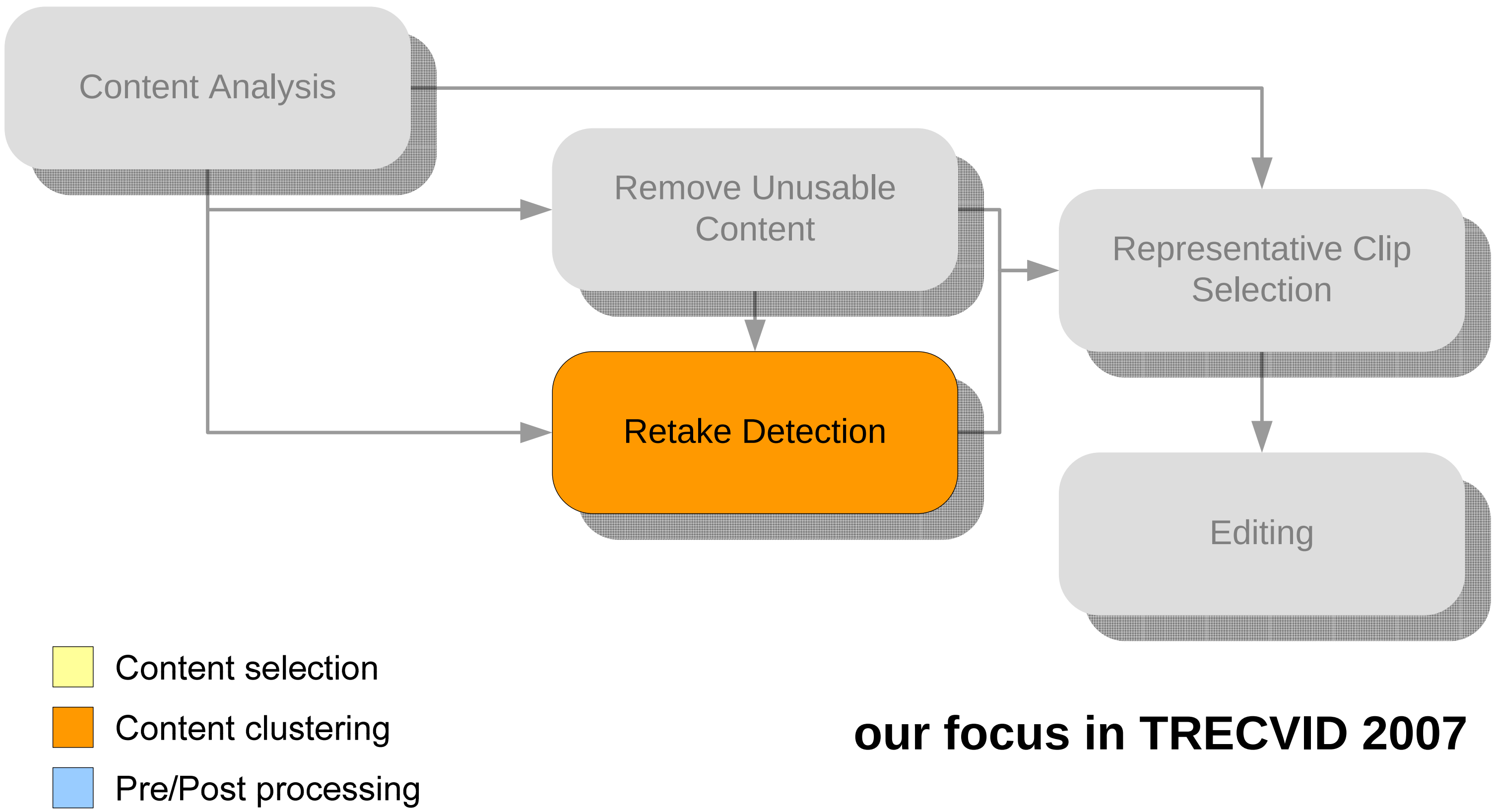
Summary Creation Process



- Content selection
- Content clustering
- Pre/Post processing

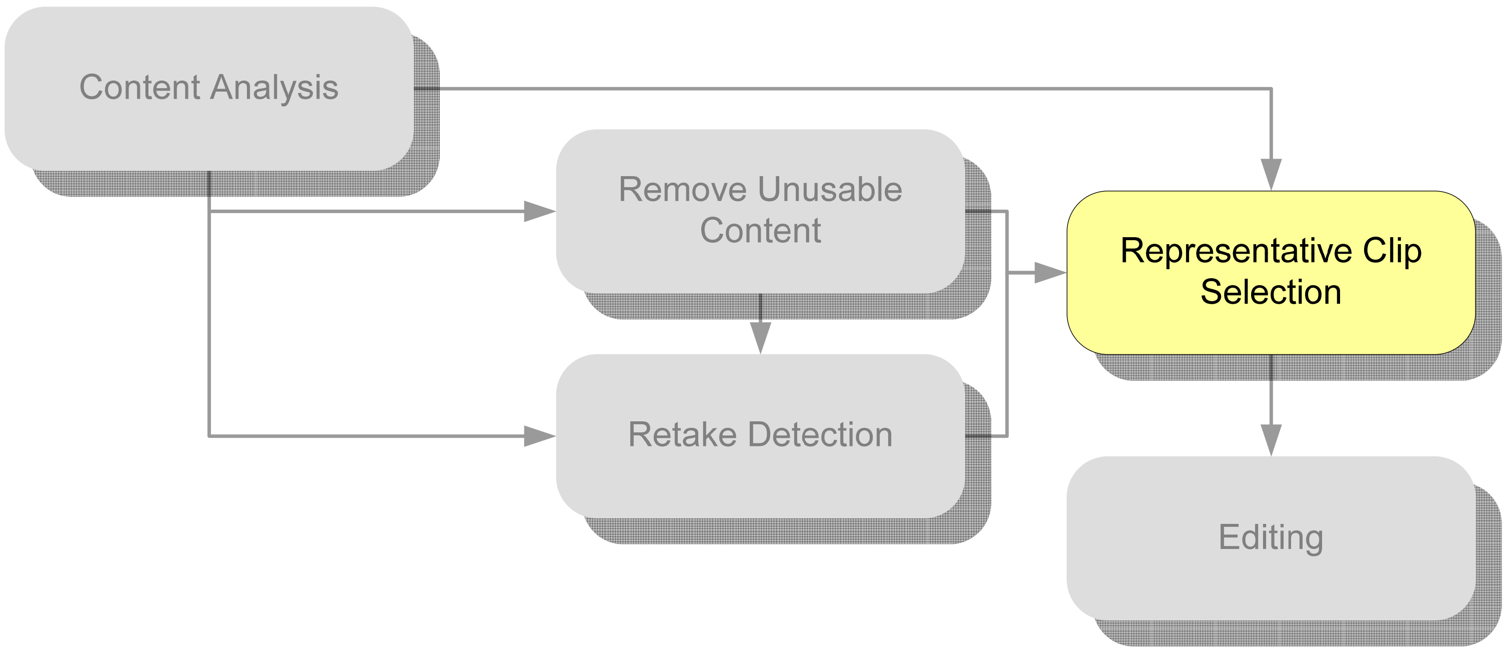
www.joanneum.at

Summary Creation Process



our focus in TRECVID 2007

Summary Creation Process



Content selection
Content clustering
Pre/Post processing

our focus in TRECVID 2008

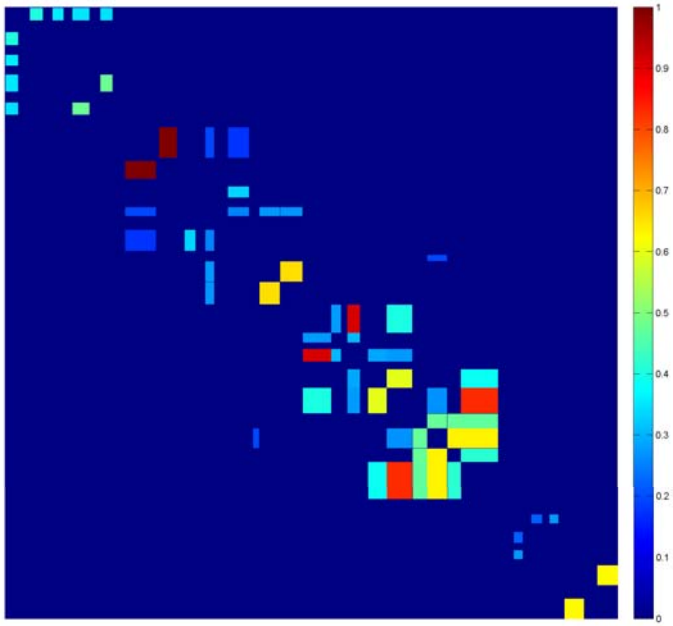
JOANNEUM RESEARCH

Content Analysis & Junk Content Removal

- Shot boundary detection (hard cuts)
  - frame differences, SVM classifier trained on TRECVID 2006 data
- MPEG-7 Color Layout and EdgeHistogram
  - descriptors extracted from every 10th frame
- Visual activity
  - averaged over 10 frames
- Face detection
  - Viola/Jones, OpenCV implementation
- Junk content removal
  - skip short shots: duration < 10 frames
  - remove color bars and monochrome frames: standard deviation in columns < 15 intensity levels in each channel

Repeated Take Detection

- Take of same scene, from same camera
- Split shots into parts (subshots)
- Pair-wise matching of parts
  - match extracted colour, texture and visual activity descriptor sequences of the parts (temporally sub-sampled by 10)
  - modified Longest Common Subsequence (LCSS) algorithm
  - remove contained and overlapping matches
  - result is a similarity matrix of the take candidates
- Cluster take candidates
- Determine relevance
  - based on overlap with takes in the same cluster



www.joanneum.at

Representative Clip Selection

- Content selection problem for BBC rushes 2007 test data
  - values based on ground truth provided by NHK
- Relevant content
  - mean 38.02% (min. 11.13%, max. 87.75%)
  - all ”meaningful“ content
- Non-redundant content
  - use longest take of all takes of a scene
  - mean 15.20%
- Summarization goal of 2% requires
  - discarding ~87% of non-redundant content
  - or using 7.6x acceleration

JOANNEUM RESEARCH

Input to Content Selection

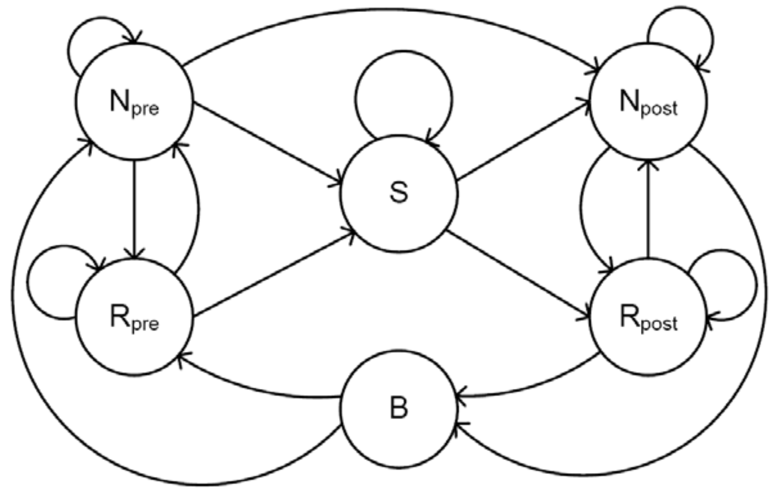
- List of arbitrary segments
  - relevance value
  - redundancy information
    - absolute: probability that this segment is useless
    - relative: list of segments w.r.t. which the current segment is redundant, and a similarity value for each of these segments
- In our experiments
  - retakes: relative redundancy information + similarity values
  - junk content: absolute redundancy information
  - motion activity: selected segments with relevance
  - presence of faces: selected segments with relevance

Two Approaches to Content Selection

- Rule-based approach
  - merge relevant and redundant segment lists into one relevance function over time
  - adaptive thresholding yields list of segments (takes length constraint into account)
  - optimize by removing/adding parts of segments
- HMM based
  - vector of relevance/redundancy values for each time instant
  - selected/not-selected etc. are hidden states
  - training
    - extract relevance/redundancy vector sequences from test set
    - create state sequence from ground truth
  - content selection
    - find ML path for given sequence of relevance/redundancy vectors

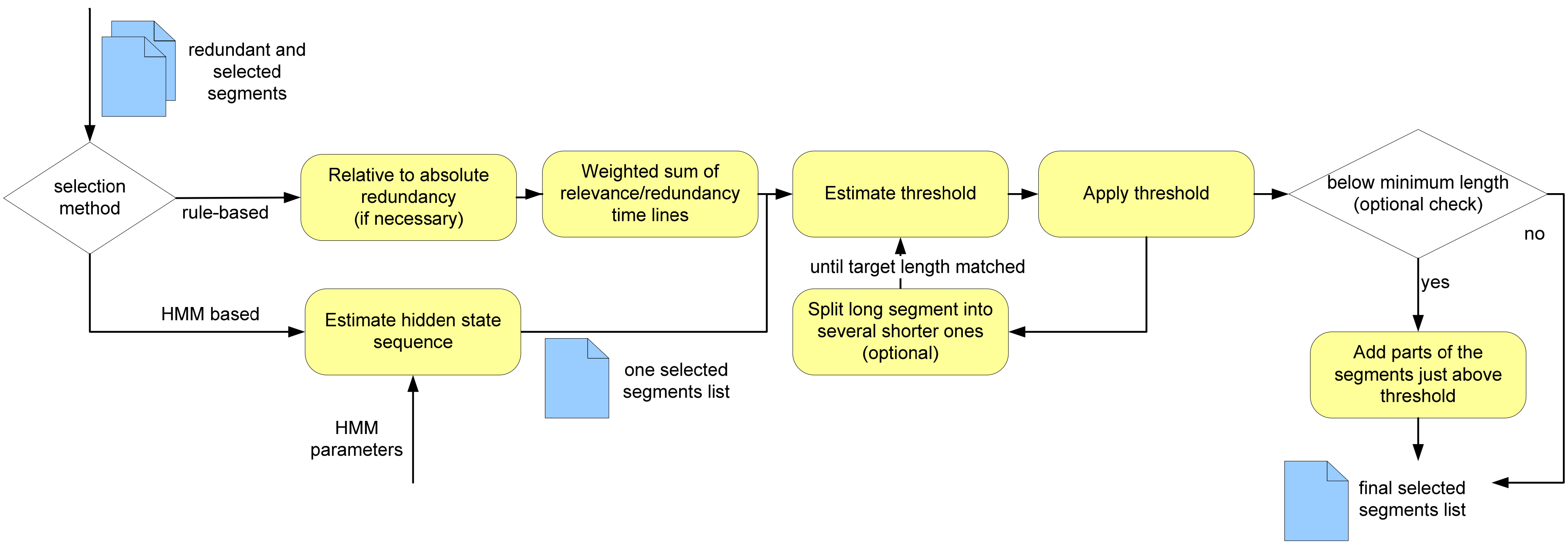
HMM Based Approach

- 6 states
  - non-relevant (Npre)
  - relevant (Rpre)
  - selected (S)
  - scene boundary (B)
  - non-relevant (Npost)
  - relevant (Rpost)



- Parameter λ in state transition matrix
  - control number and length of selected segments
- Limitations
  - not possible to enforce length constraint
  - junk content not deterministically excluded

Approaches to Content Selection - Overview



Results

	JRS1		JRS2	
Parameters	**rule**		**HMM**	
min. segment length	0.5		2.0	
max. segment length	3.0		3.0	
min. segment distance	5.8		5.8	
$w_{rel}=w_{red}$	0.5		n/a	
max. total length	0.02		0.02	
min. total length	0.30		0.30	
rel. to abs. redundancy	longest		n/a	
split long segments	true		true	
Results				
DU (median)	18.50	(0.15)	14.00	(0.03)
XD (median)	13.38	(0.24)	14.20	(0.22)
TT (median)	25.33	(0.09)	26.67	(0.13)
VT (median)	20.00	(0.05)	18.33	(0.00)
IN (median)	0.22	(0.19)	0.28	(0.27)
IN (min)	0.00		0.08	
IN (max)	0.53		0.67	
JU (median)	3.67	(1.00)	3.00	(0.50)
RE (median)	4.00	(1.00)	4.00	(1.00)
TE (median)	3.33	(1.00)	2.33	(0.50)

Results - MS221050

rule

HMM

Results - MS221050

	JRS1 (rule)	JRS2 (HMM)
DU	14.00	4.20
XD	6.19	15.99
TT	17.33	21.00
VT	16.33	7.67
IN	0.28	0.61
JU	4.33	3.33
RE	4.33	4.00
TE	2.67	1.67

below/exactly/above median of this run

Results - MS221050

	JRS1 (rule)	JRS2 (HMM)
DU	14.00	4.20
XD	6.19	15.99
TT	17.33	21.00
VT	16.33	7.67
IN	0.28	0.61
JU	4.33	3.33
RE	4.33	4.00
TE	2.67	1.67

below/exactly/above median of all runs on this video

Results – Comparison

- Both runs yield short summaries, well below the 2% limit
  - the rule based: 58.00% of max. length, 1.20% of original content
  - HMM run: 49.65% of max. length, 0.99% of original content
- HMM based selected method
  - 6% higher inclusion (increase of 27%)
  - duration is 24% shorter
  - lower score for pleasant timing
  - lower score for junk (not causally related to shorter duration or higher inclusion)
  - 47% higher editing time (more and shorter segments)
  - estimation of ML state sequence takes on average 4.75 sec/video
  - evaluation against NHK ground truth supports the results (precision and recall in the range 0.3-0.35)

Conclusion

- Comparison of two methods for content selection
- Both parametrized to yield quite short summaries
  - high scores for pleasant tempo, repeated content and junk
  - low inclusion score
- Comparison
  - HMM slightly higher inclusion at shorter duration
  - HMM difficult to control (junk, length constraint)