Segmentation, Content Extraction and Visualization of Broadcast News Video using Multistream Analysis

Mark Maybury, Andrew Merlino, James Rayson

Advanced Information Systems Center
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730, USA
{maybury, andy, jrayson}@mitre.org

Abstract

This paper reports the development of a broadcast news video corpora and novel techniques to automatically segment stories, extract proper names, and visualize associated metadata. We report story segmentation and proper name extraction results using an information retrieval inspired evaluation methodology, measuring the precision and recall performance of our techniques. We briefly describe our implementation of a Broadcast News Analysis (BNA™) system and an associated viewer, Broadcast News Navigator (BNN™). We point to current efforts toward more robust processing using multistream analysis on imagery, audio, and closed-caption streams and future efforts in automatic video summarization and user-tailored presentation generation.

1. Problem and Related Research

Content based video access is a valuable capability for several important applications including video teleconference archiving, video mail access, and individualized video news program generation. The current state of the art for commercial video archives focuses on manual annotation (Davis 1991), which suffers from problems of accuracy, consistency (when performed by more than one operator), timeliness, scalability, and cost. Just as information retrieval techniques are required to manage large text collections, video sources require similar indexing and storage facilities to support real-time profiling as well as retrospective search. Within the imagery stream, techniques performing in the ninety plus percent accuracy range have been developed to index video based on visual transitions (e.g., dissolve, fade, cut) and shot classification (e.g., anchor versus story shots (Zhang et al. 1994). Others have investigated linguistic streams associated with video (e.g., closed captions, transcripts), indexing keywords and associated video keyframes to create static and hypertext depictions of television news (Bender & Chesnais 1988, Shahraray & Gibbon 1995). Unfortunately, inverted indices of keywords (even when supplemented with linguistic processing to address complexities such as synonymy, polysemy, and coreference) will only support more traditional information retrieval tasks as opposed to segmentation (required for higher level browsing), information extraction, and summarization. More complex linguistic processing is reported by (Taniguchi et al. 1995), who use Japanese topic markers such as "ex ni tsuite" and "wa" ("with regard to", "as for"), subject/object markers, as well as frequency measures to extract discourse structures from transcripts, which are then used to provide topic-oriented video browsers.

It has become increasingly evident that more sophisticated single and multistream analysis techniques will be required not only to improve accuracy but to support more fine grained access to content and to provide access to higher level structure (Aigraine 1995). (Brown et al. 1995), for example, provide content based access to video using a large scale, continuous speech recognition system to transcribe associated audio. In the Informedia™ project, (Hauptmann & Smith 1995) perform a series of multistream analyses including color histogram changes, optical flow analysis, and speech transcription (using CMU's Sphynx-II System). Similarly, we have found a need to correlate events such as subject changes and speaker changes to improve the accuracy of indexing and retrieval.

In (Mani et al. 1996), we report our initial efforts to segment news video using both anchor/reporter and topic shifts identified in closed-caption text. In this paper, we report our more recent results, which include the correlation of multiple streams of analysis to improve story segmentation performance, the extraction of facts from the linguistic stream, and the visualization of extracted information to identify trends and patterns in the news. Through the creation of a manually annotated video corpora representing ground truth and an associated set of evaluation metrics and methods, we are able to report statistical performance measures of our algorithms.

2. Video Corpora

Unlike the image processing, text retrieval, and message understanding communities, there exists no standard set of data and evaluation benchmarks for content based video processing. We have therefore created two data sets from major network news programs from two time segments: 17.5 hours during one week in 1991 (7/15/91 to 7/23/91) and approximately 20 hours over three weeks in 1996 (1/22/96-2/14/96). Our database includes the audio, video, and closed captions associated with programs such as CNN Prime News, PBS's MacNeil Lehrer News Hour (now the Jim Lehrer News Hour), ABC World News Tonight, ABC Nightline, CBS Evening News, and NBC Evening News. We have manually annotated representative programs with information about anchor/reporter handoffs, story start/end, story inserts within larger stories, and commercial initiation and termination. We are in the process of semi-automating the markup of this corpora to include not only story segments but also all mentioned named entities (e.g., people, organizations, locations, moneys, dates). By using training and test subsets, we can automatically measure the performance of our automated segmentation and information extraction algorithms using scoring programs.

3. System Overview and Multistream Analysis

Figure 1 illustrates the overview of our Broadcast News Analysis BNA™ system and our associated viewer, Broadcast News Navigator BNN™. BNA consists of a number of multistream analyses. Analysis tools process the imagery, audio, and textual streams and store the gathered information in a relational and a video database management system (collectively referred to as the video/metadata library in the Figure 1).

For the imagery stream associated with video, we use commercial scene transition detection hardware/software (Scene Stealer ™ 1996) together with keyframe selection heuristics we have engineered based on the structure of individual video programs. For example, keyframes for a typical story containing a reporter segment are selected from the center of the reporter segment. For audio analysis, we are experimenting with speaker change detection algorithms created at MIT Lincoln Laboratory. For the closed-caption textual stream, we automatically detect commercials, story segments, and named entities, details and results of which we present below. We are currently collaborating with CMU (Hauptmann & Smith 1995) to measure how spoken language transcription will effect performance results of story segmentation and named entity tagging.

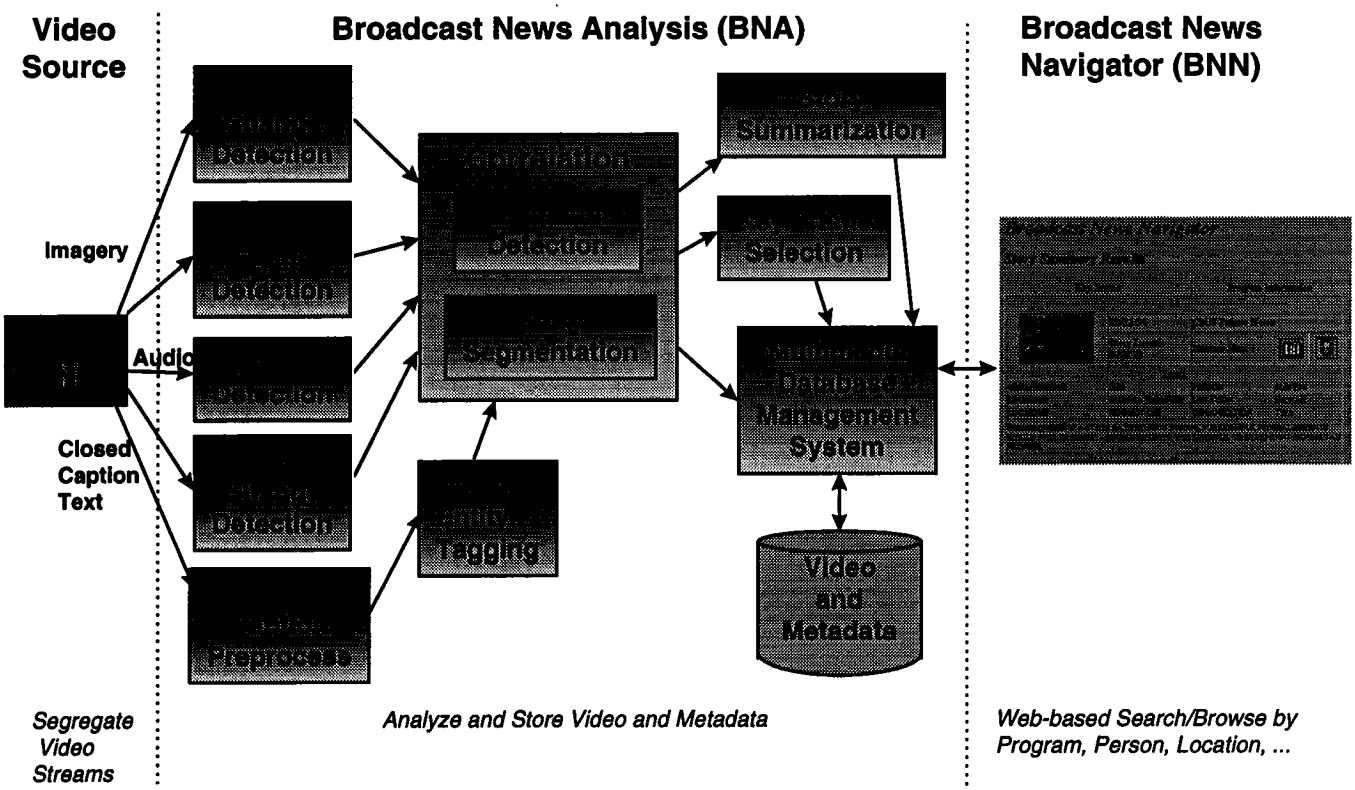


Figure 1. System Architecture

The closed-caption text and the results from the imagery, audio, and text analysis (metadata) are stored in a relational database management system with associated audio and video stored in MPEG-1 format in a video management system to provide real-time, streamed access. To support evaluation, the relational database includes both manual annotations of ground truth as well as automatically generated analysis. To enable cross stream referencing, data in the database management system is both temporally annotated and semantically correlated (e.g., speaker changes in the audio channel are related to speaker and story changes detected in closed-caption text).

Figure 2 provides a temporal visualization of cross-channel processing results applied to the ABC World News Tonight broadcast of 22 January 1996. The X-axis is measured in seconds from the start of the news program. Human annotations (ground truth) appear in the first row of the Y-axis and include the start and end points of anchor, reporter, and commercial segments as well as inserts. Start points appear slightly above end points for anchor, reporter, and commercials. An insert is a 10-second or shorter segment of supporting information (typified by a new speaker or new scene) inserted within an anchor or reporter segment. Commercials (circled in ovals) are simply indicated by a start and end point per contiguous sequence of commercials, with no annotation to indicate the start and end of each commercial nor speaker changes within commercials. The "all cues" line is the union of the above.

Row two of Figure 2 indicates the results of speaker change detection software applied to the audio stream. The audio stream is subdivided into series of non-overlapping windows that are 2, 5, 10, and 20 seconds in length. Feature vectors are derived for each window and an entropy distance metric is used to measure the change from window to subsequent window. Changes exceeding a threshold are illustrated.

The third row of the figure indicates scene changes of image content derived from video frames extracted at one second intervals from the video stream for a threshold value of 10,000. The scene change line illustrates that, without detailed models of anchor and reporter shots as in (Zhang et al. 1995), our image stream segmentor over generates. The scene change data appears to lag the other data by several seconds. This is due to some limitations in how clock initialization is performed in the scene change detection system and will be corrected by our current transition to SMPTE time code as the common time source. The fourth row of the figure shows anchor, reporter, and commercial segments as well as operator cues and blank lines extracted from the closed-caption stream, algorithmic details of which we describe next.

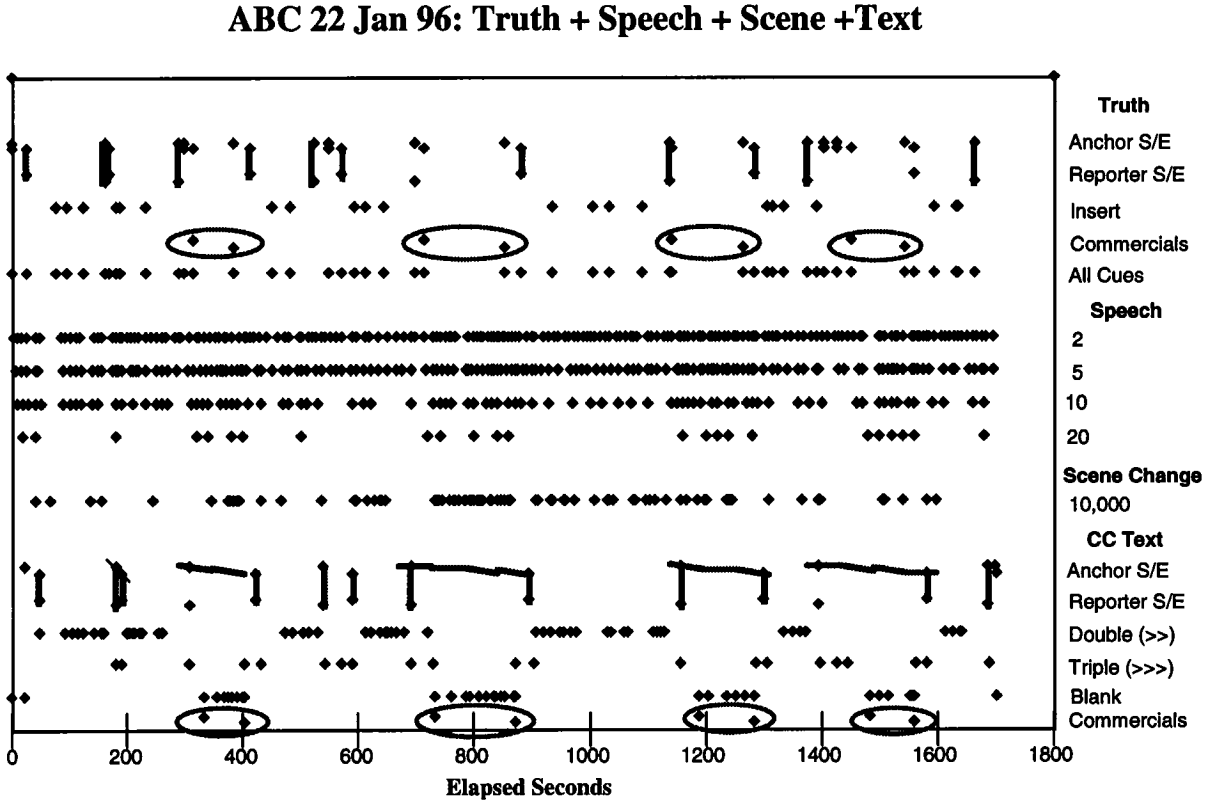


Figure 2. Cross Channel Processing Results

The upside down U-shaped lines show the anchor segments paired with their associated reporter segments. In the ground truth they are interspersed by blank space corresponding to commercials which, of course, contain no anchor or reporter segments. In the closed-caption data these "U" shapes are connected by dashed lines which erroneously show long anchor segments that incorporate the commercials that have not yet been removed.

Note the correlation of ground truth segments to automatically derived segments from the closed captions. There is also correlation of both of these to the speech data, particularly when using clusters of events identified in the 20-second windows to identify commercials. There is also some correlation with the scene change detection and the "all cues" line in the human annotation. In future work we intend to exploit additional information in the audio channel, such as emphasis detection (via intonation) and music detection. We are experimenting with various methods of improving this processing, including computing confidence measures in segments based on the number and strength of individual channel analyses.

4. Discourse Analysis for Story Segmentation

Human communication is characterized by distinct discourse structure (Grosz & Sidner 1986) which is used for a variety of purposes including mitigating limited attention, signaling topic shifts, and as a cue for shift in interlocutor control. Motivated by this, we have investigated the use of discourse structure to analyze broadcast news. In (Mani et al. 1996) we report our initial experiments with topic segmentation and story segmentation of closed-caption text, the former using thesaurus based subject assessments (Liddy & Myaeng 1992), the latter on explicit turn taking signals (e.g., anchor/reporter handoffs). We have investigated additional news programs, refining our discourse cue detectors to more accurately detect story segments. Closed-caption text present a number of challenges, as exemplified in Figure 3, including errors created by upper case text and during transcription (e.g., errors of omission, commission, and substitution).

We have identified three parallel sources of segmentation information within the closed-caption stream: structural analysis, closed-caption operator cues, and discourse cues. Together they provide redundant segmentation cues that can be correlated and exploited to achieve higher precision and recall. For ABC World News Tonight the structural analysis shows that:

- broadcasts always start and end with the anchor,
- each reporter segment is preceded by an introductory anchor segment and together they form a single story, and
- commercials serve as story boundaries.

In MacNeil-Lehrer, the structure provides not only segmentation information but also content information for each segment. The series of segments is consistently:

- preview of major stories of the day or in the broadcast program
- sponsor messages
- summary of the day's news (only partial overlap with major stories)
- four to six major stories
- recap summary of the day's news
- sponsor messages

There is often further discourse structure with the major story segments as well as specialized types of stories (e.g., newsmaker interviews). The above program structure is illustrated in Figure 3, which starts with a news/program preview followed by a sponsor message, followed by the news summary, and so on. This means that after segmenting the broadcast we can identify summaries of the day's news, a preview which serves as a summary of the major stories in the broadcast, and the major stories themselves. A user wishing to browse this one-hour newscast needs only to view a four-minute opening summary and a 30-second preview of the major stories. Similarly, sponsor messages can be eliminated. These are the building blocks of a hierarchical video table of contents.

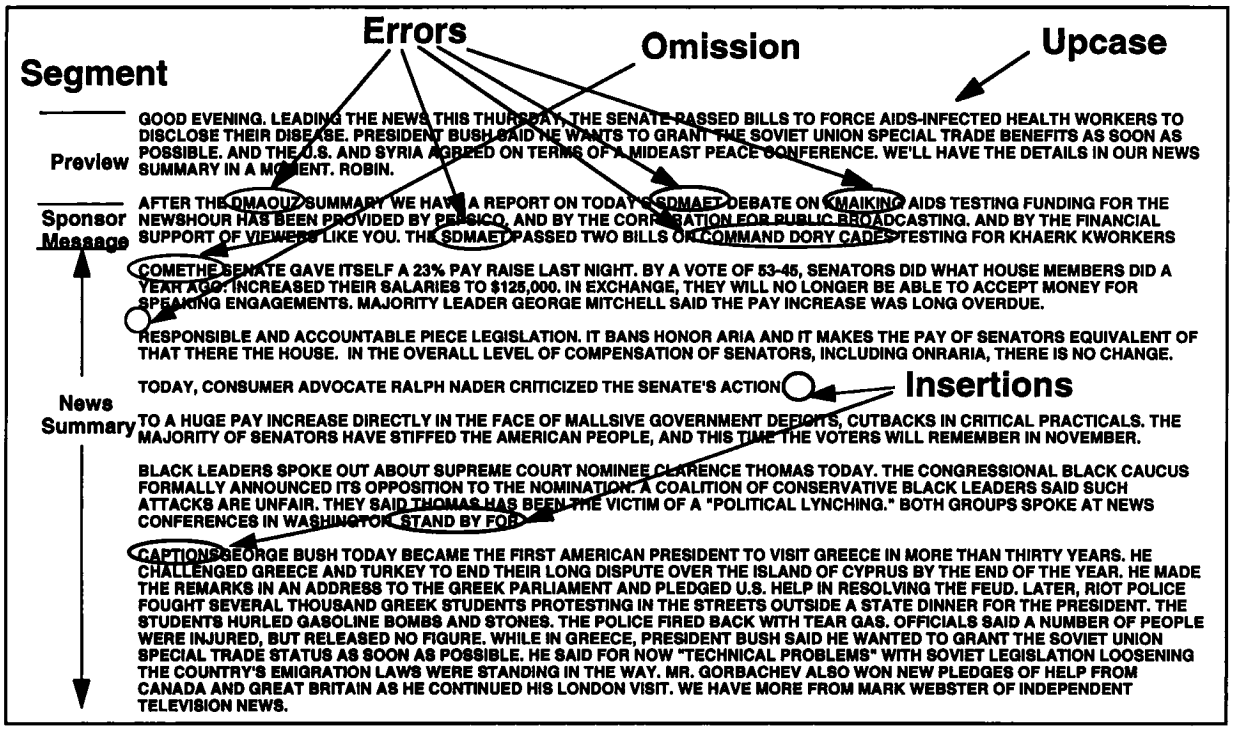


Figure 3. Closed-Caption Challenges
(MacNeil-Lehrer, July 19, 1991)

The closed-caption operator cues are supplementary information entered by the closed-caption operators as they transcribe the speech to closed-caption text as an aid to viewers. Most notable is the ">>" cue used to indicate a change in speaker and ">>>" cue used to indicate a change in story. When available, the closed-caption operator cues are a valuable source of information; however, their use by closed-caption operators are not standard. While most broadcast news sources provide the speaker change indicator, the story change indicators are present for some sources (ABC, CNN) and absent for others (CBS, MacNeil-Lehrer).

It is important to note that the three types of segmentation information identified for closed-caption text also apply to broadcast sources with no supporting closed-caption data. After speech-to-text processing, the resulting text stream can be analyzed for both structural and discourse cues, of course subject to transcription errors. In addition, speaker change identification software can be used to replicate the bulk of the closed-caption operator cues.

4.1 ABC News Segmentation

The ABC news closed-caption segmentation illustrated at the bottom of Figure 2 is performed using Perl scripts that exploit a combination of all three types of segmentation information. These build on techniques applied by (Mani et al. 1996) in which the anchor/reporter and reporter/anchor handoffs are identified through pattern matching of strings such as:

- (word) (word) ", ABC NEWS"
- "ABC'S CORRESPONDENT" (word) (word)

The pairs of words in parentheses correspond to the reporter's first and last names. Broadcasts such as ABC News have very consistent dialog patterns which ensure that a small set of these text strings are pervasive within and across news programs. Combining the handoffs with structural cues, such as knowing that the first and last speaker in the program will be the anchor, allows us to differentiate anchor segments from reporter segments. We intend to generalize the text strings we search for by preprocessing the closed-caption text with MITRE's part of speech tagger. It will allow us to identify proper names within the text in which case the strings we search from will become:

- (proper name) ", ABC NEWS"
- "ABC'S CORRESPONDENT" (proper name)

A combination of structural and closed-caption operator cues allow us to identify commercials. The end of a sequence of commercials in the closed-caption text is indicated by a blank line followed by a line containing a change of speaker (>>) or new story (>>>) indicator. We then search backwards to find the beginning of the sequence which is indicated by a ">>" or ">>>", followed

by any number of non-blank lines, followed by a blank line. The blank lines correspond to the short (0.5 to 3 seconds) silences which accompany the transitions from news to commercials and back to news. This approach to commercial detection works well as shown in Figure 2. Having the ability to identify and extract commercials also benefits our proper name extraction and summarization routines.

4.2 MacNeil-Lehrer News Segmentation

Hand analysis of multiple programs reveals that regular cues are used to signal these shifts in discourse, although this structure varies dramatically from source to source. For MacNeil-Lehrer, discourse cues can be classified into the following categories (numbers in parenthesis indicate frequency and/or location of examples in one week worth of data, items in square brackets indicate variations):

• Start of Preview (of news and program)
- GOOD EVENING. LEADING THE NEWS THIS [MONDAY, TUESDAY, WEDNESDAY ...] (almost every day)
- GOOD EVENING. THE OPENING OF THE BUSH- GORBACHEV SUMMIT LEADS THE NEWS THIS TUESDAY

• End of Preview (also start of News Summary)
- BY THE CORPORATION FOR PUBLIC BROADCASTING AND BY THE FINANCIAL SUPPORT OF VIEWERS LIKE YOU.,
(every day)

• End of News Summary, Start of Main Stories
- THAT ENDS OUR SUMMARY OF THE DAY'S TOP STORIES. AHEAD ON THE NEWSHOUR, LIFE INSURANCE PROBLEMS,
- THAT'S IT FOR THE NEWS SUMMARY TONIGHT. [NOW IT'S ON TO THE MOSCOW SUMMIT]
- ON THE NEWSHOUR TONIGHT, THE SUMMIT IS OUR MAIN FOCUS.

• Newsmaker Interview
- WE GO FIRST TONIGHT TO A NEWSMAKER INTERVIEW WITH THE SECRETARY OF DEFENSE, DICK CHENEY...
- WE TURN NOW TO A NEWS MAKER INTERVIEW WITH NICHOLAS BRADY WHO JOINS US FROM LONDON.
- NOW IT'S ON TO NEWSMAKER INTERVIEWS WITH
- In interview segments, "DR. FISCHER IN BOSTON?, MR. SCHULMANN., DR. SHELTON?, R. FISCHER?

• Anchor/Reporter Transition, Speaker Change
- WE'LL HAVE DETAILS IN OUR NEWS SUMMARY IN A MOMENT. [JIM, ROGER MUDD IS IN WASHINGTON TONIGHT. ROGER]
- WE'LL HAVE THE DETAILS IN OUR NEWS SUMMARY IN A MOMENT. [ROBIN, JUDY WOODRUFF IS IN NEW YORK TONIGHT.]

• Start of a New Main Story
- WE TURN NOW TO [ROGER MUDD, THE MAN, A LOOK AT], WE TURN FIRST TONIGHT TO
- OUR LEAD [STORY, FOCUS] TONIGHT
- FIRST TONIGHT, WE FOCUS ON ...
- STILL AHEAD, STILL TO COME ON THE NEWSHOUR TONIGHT,
- NEXT (2), NEXT TONIGHT (2)
- NOW, AN OVERVIEW
- ROGER MUDD HAS THE DETAILS
- MARK, I'M SORRY, WE HAVE TO GO
- FINALLY TONIGHT (3), WE CLOSE TONIGHT WITH A

• Recap
- AGAIN THE MAIN STORIES OF THIS [MONDAY, TUESDAY, ...]
- AGAIN, THE MAJOR STORIES OF THIS [TUESDAY, WEDNESDAY, THURSDAY, ...]

• Close
- I'M JIM LEHRER. THANK YOU, AND GOOD NIGHT (2)
- GOOD NIGHT [ROBIN (3) ROGER]
- GOOD NIGHT, JIM. THAT'S THE NEWSHOUR TONIGHT. WE'LL SEE YOU TOMORROW NIGHT. I'M ROBERT MACNEIL. GOOD NIGHT. FUNDING FOR THE NEWSHOUR
- WE'LL SEE YOU TOMORROW NIGHT. (3)
- THAT'S OUR NEWSHOUR FOR TONIGHT
- FUNDING FOR THE NEWSHOUR HAS BEEN PROVIDED BY [2] ... AND BY THE FINANCIAL SUPPORT OF VIEWERS LIKE YOU

The regularity of these discourse cues from broadcast to broadcast provides an effective foundation for discourse-based segmentation routines. However, as mentioned, this is not the only effective mechanism for segmentation. Figure 4 shows the results of applying a different type of closed-caption segmentation information -- that derived from the structural analysis -- to segment the broadcast. We were able to not only segment the broadcast, but also identify each segment as a non-news preamble, preview, summary, story, or non-news epilogue. By viewing several broadcasts and simultaneously scanning a hard copy of the

corresponding closed captions, we discovered a trivial but quite effective segmentation cue. In the broadcast, a standard musical jingle and static graphic are used to separate stories, summaries, and previews. The music is ignored by the closed-caption operator and appears in the closed-caption text as one more blank line. This provides segmentation information which we then combine with segmentation indicator is not entirely reliable because, for four MacNeil-Lehrer broadcasts, one of them had almost no blank lines. We have not yet determined if this was due to variations in the broadcast or variations in our technique for capturing the closed caption text. A more robust mechanism for exploiting this indicator would incorporate music detection to identify the jingle in the audio track and/or video pattern matching to identify the static graphic. We are in the process of pursuing segmentation through the use of audio information.

In Figure 4 we see that the automatic segmentor has accurately recognized all of the MacNeil-Lehrer segments except for one. What it identified as a single story on legislative activity was actually two stories addressing two issues before the legislature: a farm bill and Whitewater hearings. The producers of the newscast decided to treat them as a single story and did not separate them with the jingle, causing the segmentor to miss the boundary. This serves as an example of where additional audio or video analysis, discourse cues, and/or alternative linguistic analyses (e.g., Hearst 1994) might improve performance while also enhancing closed caption based segmentation through the use of discourse cues.

By comparing manual annotations of the video corpora (which indicate when viewers believe stories start and stop) with automatically generated tags from our algorithms, we can measure system performance on the example in Figure 4 and other stories using the following standard information retrieval metrics:

- Precision = # of correct program tags / # of total program tags.
- Recall = of correct program tags / # of hand tags.

Figure 5 illustrates the results of applying these measures to three broadcasts, including that in Figure 4, giving an indication of how many correct segments we identified and how many we missed.

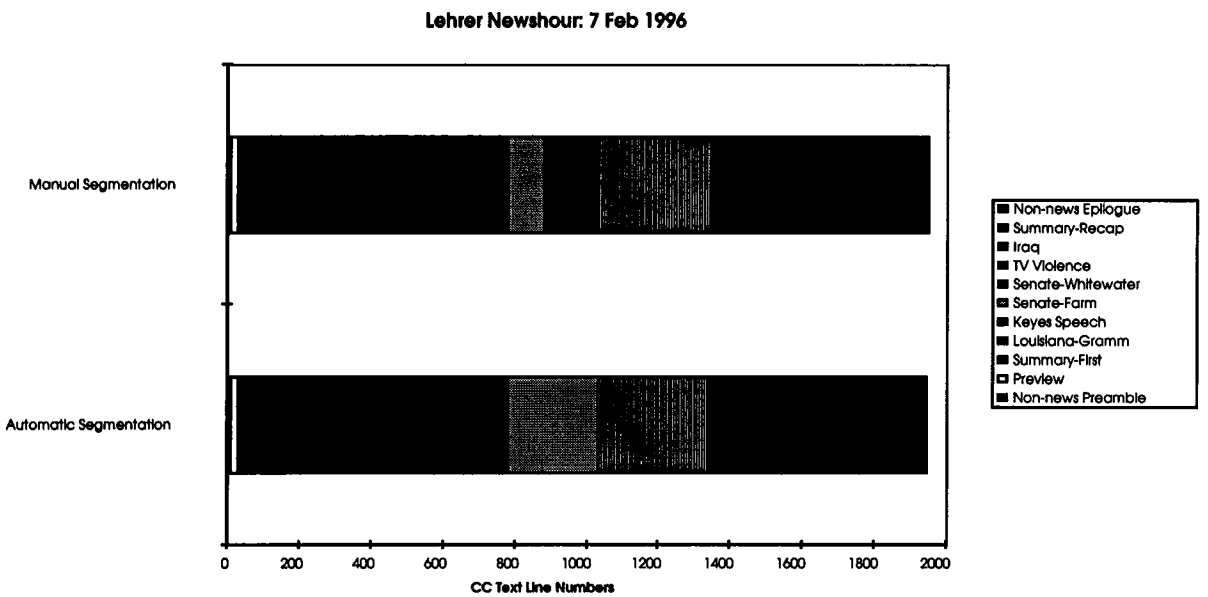


Figure 4. MacNeil Lehrer NewsHour Segmentation

Date	Precision	Recall
7 February 1996	87%	87%
8 February 1996	88%	100%
12 February 1996	100%	100%
Average	92%	96%

Figure 5. MacNeil-Lehrer Precision and Recall

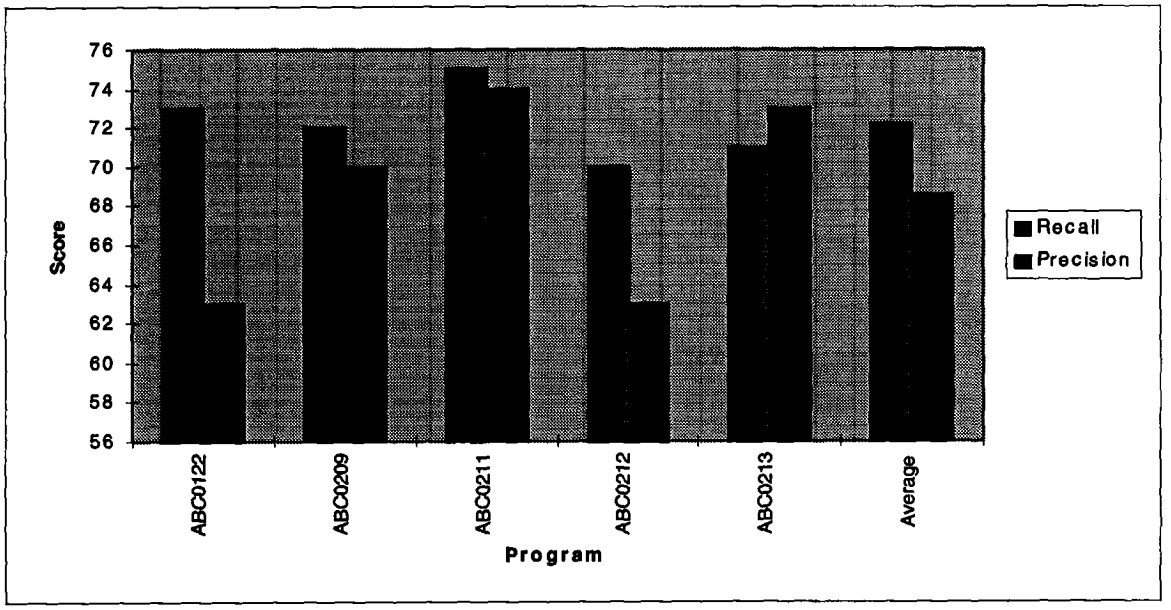
The only errors in boundary detection were due to erroneously splitting a single story segment into two story segments, or merging two contiguous story segments into a single story segment. That means that in all cases for all broadcasts the segment type labels, which identify the content of the segment (e.g., preview, news summary, story) are correct. The precision and recall values are already quite good for this small sample size, especially since we have only applied the segmentation analysis cues. Applying the discourse and closed caption operator cues should improve these results further (a current implementation). Our perfect results for segment types classification enables us to derive a table of contents for the broadcast, along with a high quality summary of the entire broadcast, without the use of complex automated summarization algorithms.

5. Named Entity Extraction and Story Summarization

In addition to detecting segments and summarizing stories (for search or browsing), it is also important to extract information from video, including who did what to who, when, where and how. For this problem, we have applied a proper name tagger from the Alembic System (Aberdeen et al. 1995). This trainable system applies the same general error-reduction learning approach used previously for generating part-of-speech rules designed by (Brill 1994) to the problem of learning phrase identification rules.

Figure 6 illustrates initial precision and recall results of applying the phrase finder to detect people, organization, and locations in the closed-caption transcripts for five ABC World News Tonight programs. These results are promising given that we are applying a part of speech and named entity tagger trained previously on scientific texts to a new genre of language (broadcast news) with no modification to either the lexicon or grammatical rules, despite the significant differences in language, grammar, noisy data, and upper case only text.

Whereas some broadcasts contain embedded summaries (e.g., Jim Lehrer), others (e.g., ABC) contain no such summary. In collaboration with (Mani 1995), we have been experimenting with several heuristics for text summarization. The first is simply to pick the first N words in a document or segment, assuming that the text was created in journalistic style. Since we are extracting proper names out of the text, however, another strategy is to use the most frequently occurring named entity within each segment to serve as its label. We still have the remaining problem of choosing not only a sample video frame and perhaps also a representative set of key frames (e.g., a video clip). Perhaps most challenging, we need to choose the most visually informative and/or pleasing keyclip.



	P&R	2P&R	P&2R
F-MEASURES	70.53	69.53	71.55

Figure 6. Named Entity Extraction Performance

6. Application to Network News Access

Using the above segmentation and information extraction techniques, we have developed BNN to enable a user to search and/or browse the original video by program, date, person, organization, location or topic of interest. Within BNN, a web user can first specify particular broadcast news agencies and time periods they are interested in searching. BNN will generate a list of named entities with their frequencies sorted in descending order. Figure 7 illustrates such a response to a user query indicating, among other things, that there were many references to the "FAA", "GINGRICH" and "ZAIRE".

With the frequency screen displayed, the user can view the stories for one of the values by selecting a value, for example "ZAIRE". Upon selection of the value, BNN searches through the 2893 stories discovered in the 163 broadcasts to display the related stories seen in Figure 8. The returned stories are sorted in descending order of key word occurrence. Each returned story contains the a key frame, the date, the source, the six most frequent tags, a summary and the ability to view the closed caption, video and all of the tags found for the story. The summary is currently the first line of the segment. In the future, the system will extract the sentence, which is most statistically relevant to the query.

While viewing the video, the user has the ability to directly access the video related to the story segment. Thus, if the story segment starts at six minutes and twelve seconds into the news broadcast, the streaming of the video to the user will start at that point. While viewing the streaming video, the user can scroll through out the video with VCR like controls.

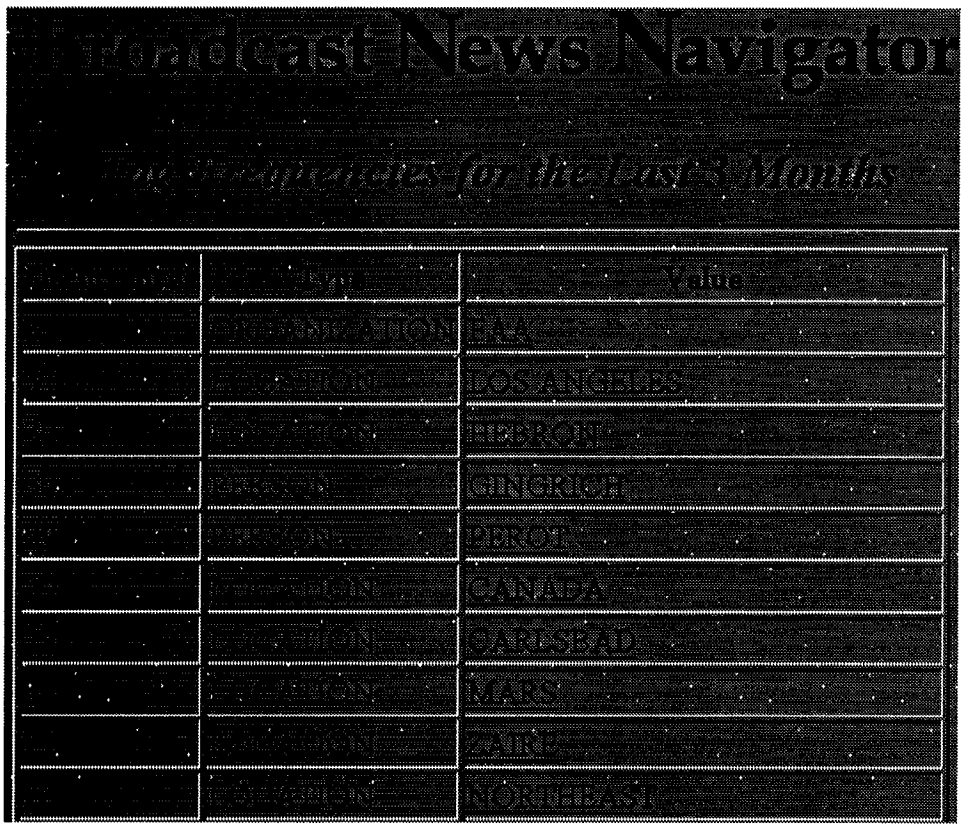


Figure 7. Named Entity (Metadata) Visualization

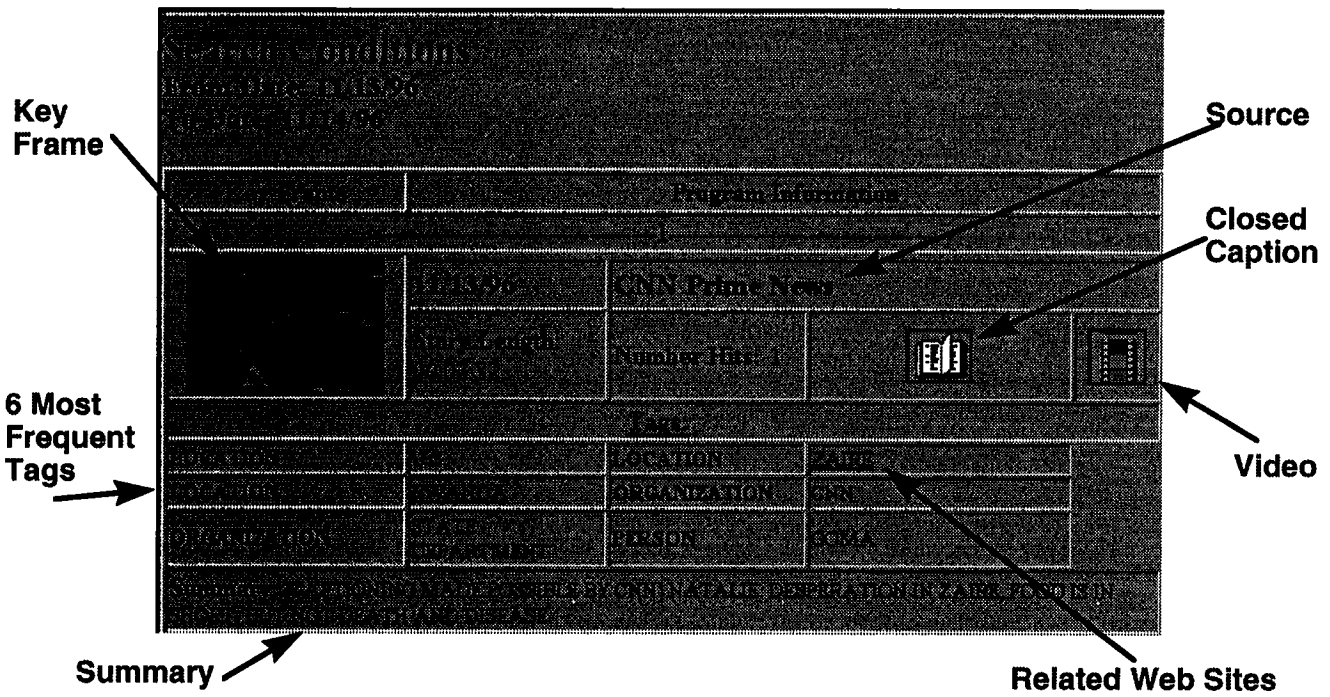


Figure 8. Viewing Story Segments that Meet the Query Criteria

8. Limitations and Future Directions

An important area for further research is to create segmentation and extraction algorithms that are more robust to signal noise and degradation, perhaps using confidence measures across multiple streams as a means for combining multiple sources of evidence. Ultimately one would like to automatically acquire both more general and program specific segmentation algorithms. In our current experiments, we aim to significantly improve the above performance by training both our part of speech and proper name tagger on closed-caption and spoken language transcripts from our video corpus. We are also experimenting with the accuracy of segmentation and extraction from spoken language transcription (Hauptmann & Smith 1995, Brown et al. 1995) and will contrast their performance with that of closed captioned sources. We further seek to evaluate not only segmentation and extraction accuracy, but also efficiency measured in time to perform an information seeking task. We are collaborating with (Mani 1995) to extend our system to incorporate automated story summarization tailored to particular users interests as in (Maybury 1995).

9. Conclusions

This paper reports on extensions to previous efforts, including automatic discourse structure analysis and information extraction techniques to support news story segmentation and proper name extraction/visualization. As in (Taniguchi et al. 1995) work, we report specific performance results in terms of information retrieval metrics of precision and recall. These discourse structure and content analyses are an important step toward enabling more sophisticated customization of automatically generated multimedia presentations that take advantage of advances not only in user and discourse modeling (Kobsa & Wahlster 1989) but also in presentation planning, content selection, media allocation, media realization, and layout (Maybury 1993).

10. Acknowledgments

At MITRE, we are indebted to Inderjeet Mani for providing text summarization routines, to Marc Vilain and John Aberdeen for providing part of speech and proper name taggers, and to David Day for customizing these to our domain. We thank Douglas Reynolds of the Speech Systems Technology Group at MIT Lincoln Laboratory for his speaker change detection software.

11. References

Aigraine, P., Joly, P., and Longueville, V. 1995. Medium Knowledge-based Macro-Segmentation of Video into Sequences. In Maybury, M. (editor) Working notes of IJCAI-95 Workshop on

Intelligent Multimedia Information, 5-16. Retrieval, Montreal.

Aberdeen, J., Burger, J., Day, D., Hirschman, L., Robinson, P., and Vilain, M. 1995. Description of the Alembic System Used for MUC-6. In Proceedings of the Sixth Message Understanding Conference. Advanced Research Projects Agency Information Technology Office, 6-8. Columbia, MD.

Bender, W., and Chesnais, P. 1988. Network Plus. *SPIE Vol 900 Imaging Applications and the Work World* 81-86. Los Angeles, CA.

Bender, W., Håkon, L., Owrant, J., Teodosio, L., Abramson, N. Newspace: Mass Media and Personal Computing 329-348. USENIX, Summer '91, Nashville, TN.

Brill, E. Some Advances in Rule-Based Part of Speech Tagging. In Proceedings Proceedings of the Twelfth National Conference on Artificial Intelligence, 722-727. Seattle, WA.

Brown, M. G., Foote, J. T., Jones, G.J.F., Sparck-Jones, K., and Young, S. J. Automatic Content-Based Retrieval of Broadcast News. In Proceedings of ACM Multimedia 1995 35–44. San Francisco, CA.

Davis, M. Director's Workshop: Semantic logging with Intelligent Icons. In Maybury (ed.), In Proceedings of the AAAI Workshop on Intelligent Multimedia Interfaces 122-132.

Mani, I., House, D., Maybury, M. and Green, M. Towards Content-Based Browsing of Broadcast News Video. to appear in Maybury (editor) Intelligent Multimedia Information Retrieval. Forthcoming.

Grosz, B. J. and Sidner, C. July-September, 1986. "Attention, Intentions, and the Structure of Discourse." *Computational Linguistics* 12(3):175-204.

Hauptmann, A. and Smith, M. 1995. Text, Speech, and Vision for Video Segmentation: The Informedia Project. In Maybury, M. (editor) Working notes of IJCAI-95 Workshop on Intelligent Multimedia Information 17-22. Retrieval, Montreal.

Hearst, M. A. Multi-Paragraph Segmentation of Expository Text, ACL-94, Las Cruces, New Mexico, 1994.

Kobsa, A., and Wahlster, W. (eds.) 1989. *User Models in Dialog Systems.* Berlin: Springer-Verlag.

Liddy, E. and Myaeng, S."DR-LINK's Linguistic-Conceptual Approach to Document Detection", Proceedings of the First Text Retrieval Conference, 1992, pub. National. Institute of Standards and Technology.

Mani, I. .1995 "Very Large Scale Text Summarization", Technical Note, The MITRE Corporation.

Maybury, M. (ed) 1993. *Intelligent Multimedia Interfaces*, Cambridge, MA: AAAI/MIT Press.

Maybury, M. T. 1995. Generating Summaries from Event Data. *International Journal of Information Processing and Management : Special Issue on Text Summarization.* 31(5): 735-751.

Proceedings of the Sixth Message Understanding Conference. Advanced Research Projects Agency Information Technology Office, Columbia, MD, 6-8 November, 1995.

Dubner, B. Automatic Scene Detector and Videotape logging system, User Guide, Dubner International, Inc., Copyright 1995, 14.

Shahraray, B. and Gibbon, D. 1995. Automated Authoring of Hypermedia Documents of Video Programs. In Proceedings of ACM Multimedia '95 401-409. San Francisco, CA.

Taniguchi, Y., Akutsu, A., Tonomura, Y. and Hamada, H. "An intuitive and efficient access interface to real-time incoming video based on automatic indexing", Proceedings of ACM Multimedia '95 25–34. San Francisco, CA.

Zhang, H. J., Low, C. Y., Smoliar, S. W., and Zhong, D., "Video Parsing, Retrieval, and Browsing: An Integrated and Content-Based Solution", Proceedings of ACM Multimedia '95 15-24. San Francisco, CA.