Combining Content Analysis of Television Programs with Audience Measurement

David Gibbon, Zhu Liu, Eric Zavesky, DeDe Paul,
Deborah F. Swayne, Rittwik Jana, Behzad Shahraray
AT&T Labs - Research
{dcg,zliu,ezavesky,dedepaul,dfs,rjana,behzad}@research.att.com

***Abstract*— Combining content analysis of television programs with quantitative audience measurement can provide insights into customer reactions to advertisements and program content. This work introduces a system architecture that incorporates anonymous audience metrics from an operational IPTV environment with metadata from a content-based analysis of recorded programs. Evaluated on a collection of news programs, the system verifies that events derived from the audience metrics data stream correspond to media segmentation boundaries such as commercial breaks and topic changes. An automated system for executing multimodal media segmentation algorithms for commercial break and topic change detection is also discussed. Better understanding of audience reaction can help IPTV service providers plan infrastructure investments and help in managing multimedia content delivery networks.**

***Index Terms*—audience metrics, media processing, video segmentation, commercial detection**

I. Introduction

Media and entertainment companies and advertisers have a vested interest in gathering accurate information about the viewing preferences of their audiences. A great deal can be inferred from their viewing history, supplemented by demographic and market data. Additional details can be added by analyzing the content of programs audiences watch, like discovering the boundaries between program segments, the introduction of a new topic within a news program, or the arrival of a new guest on a talk show. Given the amount of content being produced and delivered on a daily basis, automated systems for collecting and processing media and audience data are increasingly valuable.

Traditionally, program descriptions from electronic program guide (EPG) sources constituted the only available metadata describing program content. This information, while extremely valuable, can be inaccurate and often lacks sufficient detail. For example, information about program segments, while supported by standards such as TV Anytime [1], is seldom available in practice because it is costly to produce and difficult to maintain throughout the content lifecycle. Broadcast monitoring systems [2-4] allow analysts to browse the nature of the content at a given point in time in much greater detail through the use of recorded Closed Captions, thumbnail images and proxy resolution video retrieval once points of interest in the broadcast have been identified. To determine the audience reaction to content being delivered, audience metrics services can provide estimates of the audience size for each second of the broadcast by anonymously aggregating channel changes of set top boxes.

This work proposes the application of media processing techniques such as multimodal topic segmentation to augment the EPG information with detailed time-indexed metadata describing the content. This information can begin to shed light on unexplained changes in audience size during the course of a program. To begin to explore these ideas, this work focuses on live broadcasts of news programs and show that periods of increased tuneaway correspond to news story boundaries as well as commercial breaks.

This work introduces a system architecture that incorporates audience metrics with content-analysis techniques for precisely timed metadata in section II, provides results and illustrations of this analysis in section III, addresses scalability issues in section IV, and draws conclusions from this study and indicates future work in section V.

II. System and Methods

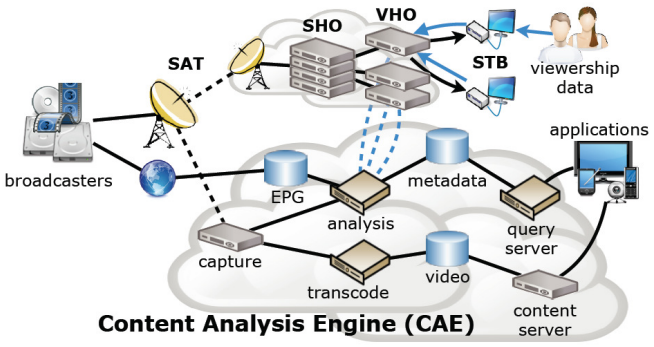

Figure 1: High-level system architecture of the Content Analysis Engine (CAE) and IPTV infrastructure

A. High Level Architecture

Figure 1 shows a high level system architecture of the Content Analysis Engine (CAE) where broadcast content and metadata along with corresponding audience metrics data is acquired, processed with a set of algorithms, and the resulting metadata for each program is stored in an index. The

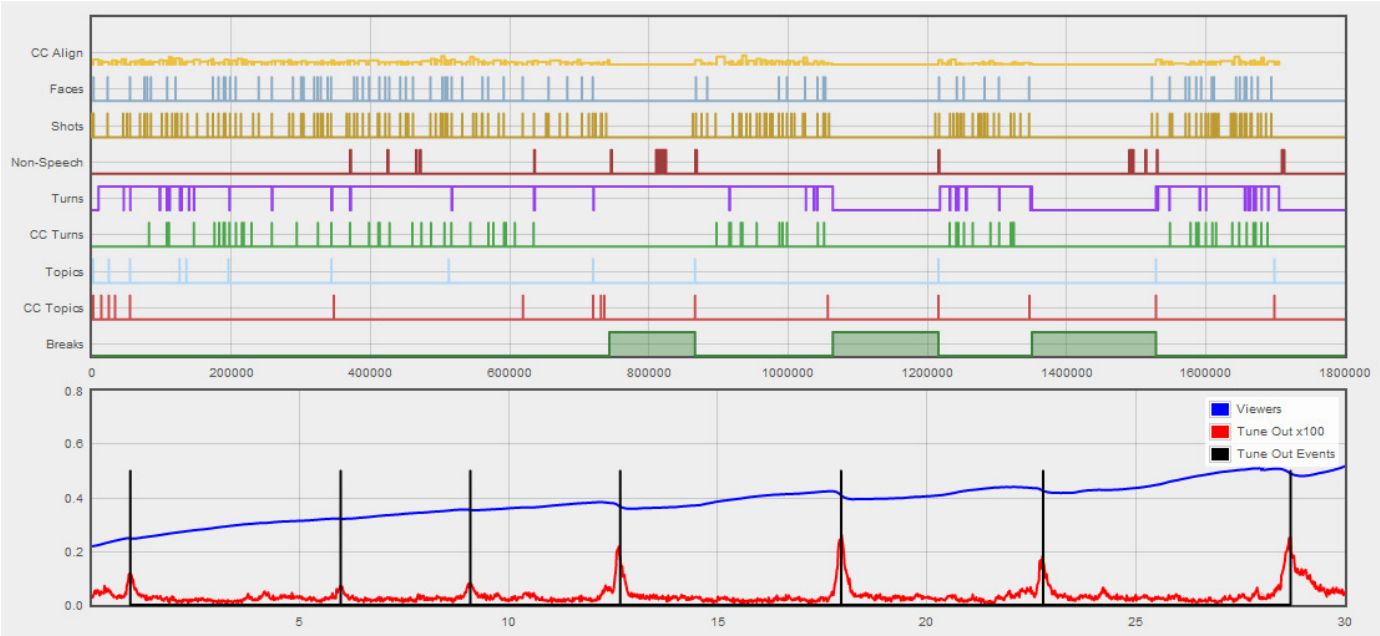

Figure 2: Analysis results for a 30 minute program showing media processing and audience measurement on a common axis.

architecture supports a range of applications such as broadcast monitoring or audience measurement research; each of these applications has its own user interface, allowing users to select programs of interest, render results and provide quantitative measurements. Content processing includes video and audio segmentation, topic segmentation, face detection, speech to text conversion, mid-level semantic classification of key frames and other operations [5]. The system utilizes a multimodal approach for locating topic boundaries in program segments (between commercials) that includes linguistic processing of the closed caption text, audio processing for speaker segmentation and anchorperson detection [8].

In most IPTV systems, IP multicast is used to deliver the live streams, with each set top box joining only one or two streams at any given time, and interframe video coding is used for bandwidth efficiency. To support fast response to channel change requests, a short burst of media may be delivered using unicast. IPTV service providers must manage the network to handle multiple simultaneous channel changes while maximizing video quality and bandwidth utilization. Characterizing the audience's collective channel change behavior can be useful in designing and managing these services.

While the metadata generated by such a content analysis can be as granular as the words spoken or the number of shots (i.e. same scene, same speaker segments), this work focuses on high-level content boundaries that indicate either commercial breaks or topic boundaries. This content analysis is used to explain observed audience tuneaway behavior, but it is also possible to utilize this data as a source of information for automated media segmentation algorithms.

B. Commercial Break Alignment

Figure 2 shows the extracted metadata for a particular 30 minute news program along with the audience metrics for each second. Plots like these are generated dynamically in an interactive tool that supports a cursor indicating the current video replay position in the plots, so that analysts may select a particular point in the plot to play the corresponding video. In Figure 2, a correlation is visible between viewer tuneaway events (shown in red) and the onset of commercial breaks as detected by content processing. Further, smaller scale tuneaway events correspond to changes in the news topics during the first program segment. This finding demonstrates that the proposed system is consistent with other works that have found a strong correlation between tuneaway events and commercial onset times.

None of the plots shown in Figure 2 represent ground truth annotation, however the traces labled 'CC Turns' and 'CC Topics' are derived from manually entered, realtime captioner data primarily for the benefit of hearing impaired viewers. While they are not intended to be a comprehensive labeling, they do have correlation with the traces labeled 'Turns' and 'Topics' which are the result of speaker segmentation and multimodal topic segmentation respectively. The other traces at the top of the figure are lower level extracted features that are used in the multimodal processing stages: 'CC Align' represents the corrected delay between the appearance of the text and the spoken word, 'Faces' are video segments with one or more detected faces, 'Shots' are the video shot boundaries, and 'Non-Speech' are segments containing silence or music.

C. Commercial Break Detection

The best way to determine when commercial breaks occur is to capture signaling information from the broadcaster, however in the majority of broadcast content analyzed, SCTE-35 [7] or

other signaling information for ad insertion was not available. Instead, multimodal processing combines the results of audio and video processing cues including silence, black frames, station logos and aspect ratio change detection to estimate the location of commercial breaks. For this work, the publically available tool, Comskip [6] was found to be highly accurate. For genres such as news programs, it is accurate enough to alleviate the need for collecting ground truth when estimating the performance of commercial break detection derived from viewing totals at large scale.

D. Tuneaway Event Detection

For a particular airing of a television news program, audience reaction is captured through a measurement of the number of viewers tuning away as a function of time. Anonymous aggregated tune-in, tuneaway, and total viewer information from an IPTV service provider is collected from a number of geographically distributed markets and processed to form a data stream for analysis. While the data is derived from channel change events, it is treated as a continuous signal sampled at one second intervals. In this work detecting periods of increased tuneaway on the order of 10 seconds allows for variations in the time for viewers to make a decision and change the channel. Small scale variations are filtered out by applying a non-linear filter (3 tap median filter) to first smooth that signal prior to detecting local maxima using a 10 second window. A post processing step rejects minor peaks and those that occur with a separation of less than twenty seconds which are deemed to belong to a single prolonged tuneaway event. The value of the parameter used in the rejection step is critical for accurate detection of small scale tuneaway events, but is immaterial for the large scale events that occur at the beginnings of commercial breaks and is addressed further in Section III C. The automatic event detections for a particular program are indicated in Figure 2 as black impulses labeled "Tune Out Events".

III. Audience and Topic Alignment Study

A collection of 15 nationally broadcast half-hour news programs from December 2010 was analyzed to evaluate the effects of program segment boundaries determined by automated media processing as well as audience reaction. Ground truth annotation identified the positions of 275 discontinuities in the program narrative at topic boundaries and commercial breaks. The content index for these programs was produced by an offline analysis of each program after capture from over the air broadcasts in a single market. While each regional market may air different content in the commercial segments of this program, this work focused on the actual program content (i.e. the news stories), so possible sub-region influences on audience information can be ignored.

A. Annotation of News Program Segments

Discontinuities in the narrative flow of broadcast television news programs may be characterized as transitions between:

1. program content and ad pods (sets of commercials)
2. main news stories, 90 seconds or more in duration
3. related stories, or between field material and studio commentary within a given story
4. brief news stories, less than 90 seconds in duration
5. headlines
6. program content and promotions for upcoming stories

This work postulates that some of these events will generate measureable audience reaction; in particular, occurrences of tuneaway are expected to be more likely at these transition points. Manual annotation of some of the transitions listed above was performed on the corpus of news programs. Transitions to headline or promotional segments were annotated, but individual headlines were not marked; and while transitions between related main stories were indicated, transitions between field and studio within a story were not. Additionally, points at which the closed captioner indicated a change of topic with the ">>>" convention were also logged. Finally, this data was compared with the results of an event detection algorithm operating on second by second audience tuneaway data.

B. Media Segmentation and Audience Reaction Alignment

In Figure 2, the top plot time axis in is milliseconds and the bottom plot is in minutes. The trace marked "Topics" represents the results of multimodal topic segmentation described above. In this figure, three false positives are found in the second and third minutes due to reporters interviewing different people. The trace marked "CC Topics" indicates the points where the closed captioner used the ">>>" convention to indicate a change of topic. Note that they did not mark the topic change in the eighth minute, while this was correctly detected by the multimodal processing and by the tuneaway event detection. The captioner marks topic changes for very short (e.g. a single sentence) segments such as headlines at the start of the program or just prior to commercial breaks as teasers for upcoming stories. By design, these are not detected by the multimodal processing, and coincidentally they are not detected in the tuneaway data. The other traces in the figure represent lower-level features that are used for computing the topic segmentation and for other applications. These features include changes of speaker ('CC Turns' and 'Turns'), presence of speech, duration of shots and occurrences of shots containing faces. The delay of the closed caption with respect to the speech is also shown ('CC Align').

C. Results

In all news programs, four ad pods were found, including the break at the end of the program. An average of 18.3 discontinuities in the narrative flow, (e.g., topics, ad pods, promotions) per program was observed. The closed caption indicated 91% of these with 11.3% spurious indications occurring within a story. For captions, a 15 second latency was allowed to account for real-time delay. Captions indicated topic changes at headline boundaries but these were not annotated, so they were not counted. Captions did not indicate the beginning of ad pods, so these were also not counted.

The performance of the audience-metrics-based event detection is as follows: 48.5% of the main breaks in the narrative were found which consisted of 100% of the ad pod

breaks being correctly identified, and 19.1% of the main topic transitions detected with a false alarm rate of 5.6%. In this case, a 20 second latency was allowed to account for audience reaction delay plus the time elapsed from the onset of increased tuneaway until it reaches a local maximum.

Table 1 summarizes the program segment detection results as indicated by the closed caption and processing of audience metrics. Captions have fewer possible main segments listed since the onset of commercial breaks is not included. Excluding commercial pod breaks, the audience metrics based detection method found 20 out of 107 possible topic changes or 18.7%. Detections that corresponded to minor topics (types 4,5,6 listed in section 3.A) were not counted as false positives.

Table 1: Program segment detection results

	Main Program Segments	Detections	Minor Topic Detections	False Positives
Captions	107	97 (91%)	167	12
Audience	167	80 (48%)	15	9

Note that the values listed in Table 1 correspond to the point of maximum F-measure for the tuneaway peak detector. As shown in Figure 3, the detector may be tuned for applications where improved recall (or detection rate) is desirable, but with an increased rate of false alarms (lower precision.)

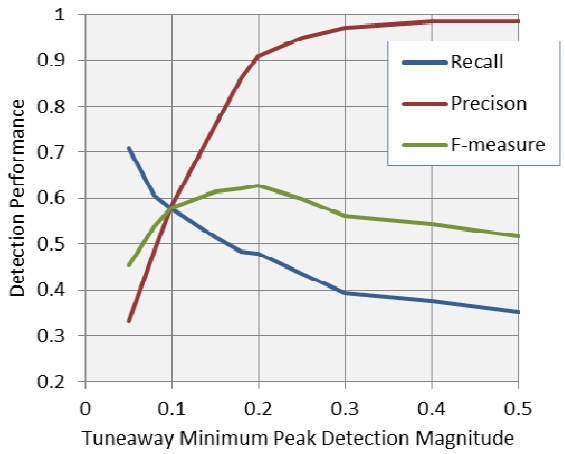

Figure 3: Effects of variying tuneaway peak detection sensitifity on detection performance.

D. Observations

The tuneaway events at breaks due to ad pods are so pronounced that it is possible to detect these with 100% accuracy and very few false positives. The false alarm rate indicated above is higher since the detection process also attempted to detect local maxima at topic boundaries which are close to the noise floor of the tuneaway data. It's clear why commercials drive viewers away, but it is less obvious that topic changes would also generate measurable tuneaway. While it is possible that in some cases the viewers may know that the subsequent program segment is not of interest, perhaps by dint of previous promotional descriptions, it is also possible that viewers are simply waiting for the conclusion of the current story before tuning away. Most frequently, viewers are tuning to other programs to avoid commercials and they tend to dwell on program segments of interest. Given these factors, a detection rate of 19% for intra-content topic change tuneaway events shows promise. In fact, although this work is framed in terms of a detection problem, the important point is whether or not a given topic break prompts a high level of tuneaway. In many cases, news programs successfully avoid tuneaway by artfully ordering related stories and creatively transitioning between them. For other genres such as comedies, increased tuneaway at the onset of opening credits after a cold opening can be observed [9].

The high rate of closed captions indicating main topic transitions (91% of the time) suggests that this data stream may be useful in a larger data collection scenario without the need for complete manual annotation. Further processing would be required to remove headlines and promotional topic marks, and given that ad pod breaks are not marked, this data would have to be added. Fortunately, commercial detection algorithms have good success and in some cases, metadata such as SCTE-35 may be available.

IV. Series Information from Automatic Segmentation

With the CAE system illustrated in Figure 1 and the potential to predict in-program tuneaway events, alternative strategies are proposed for positioning advertising pods within programs of different genres. Topic segmentation and commercial detection algorithms can be evaluated on any television program and aggregated en-masse to produce insights about its structure as the viewer sees it. As an example, Figure 4 illustrates results of a commercial detection algorithm aggregated over six months for two programs, "Late Night with Jimmy Fallon" and "The Tonight Show with Jay Leno" on the left and right respectively. This visualization demonstrates that there is high amount of variance in the start and duration of commercial segments during the "Jimmy Fallon" show compared to the relatively stable segment position and duration during "Jay Leno". Although some plot locations may be attributed to detector error, the likely explanation for these differences is the requirement for the content provider to provide relatively specific time slots for both local- and nation-wide advertising campaigns.

While it is possible for advertisers to collect ad insertion information for each market from ad verification services, this bottom-up content-based generation of metadata can produce reliable timing information with a fully automated system. Combining lessons from this analysis with the topic segmentation analysis above, content analysis systems like the CAE can help both advertisers and broadcasters to rapidly gain insight into pod location distributions over time for particular programs to better inform ad placement decisions.

V. CONCLUSIONS

This work has shown that audience reaction, as indicated by increased tuneaway events, can be attributed to discontinuities in the narrative content of television programs as determined by automated media processing. While most pronounced at commercial breaks, this phenomenon was observed within program content such as at news story boundaries. This insight suggests that product placement advertising or sponsorship of a program segment with an in-frame graphic can be intelligently placed within a program segment. One example of this intelligent placement would be a news story where the news anchor hands off the story to the local correspondent or when graphics from different areas are incorporated. Findings in this work demonstrate that if viewers are interested in a particular story, they do not tuneaway until another story (or a commercial segment) begins.

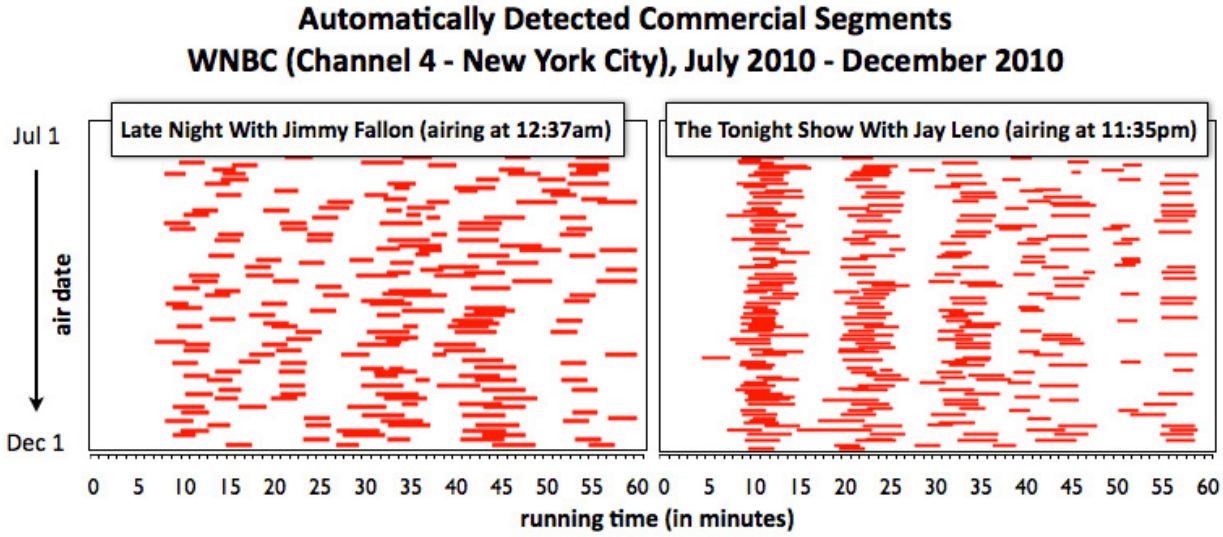

Figure 4: Detected commercial segments aligned to program running time and aggregated over several air dates spanning six months in 2010 visualized to detect commercial location patterns.

Metadata that describes these events of increased tuneaway activity can be derived from cues that closed captioners insert for the benefit of hearing impaired viewers. However, multimodal processing of the recorded program content can find these points and others that are not indicated in the caption. Media processing results could also improve closed caption streams by inserting additional marks at topic or speaker changes.

Further study is warranted to determine the extent to which other metadata streams such as the number of faces or other automatically detected cues can be shown to influence audience metrics. Detailed analysis of the effectiveness of the media segmentation algorithms on a broad range on content genres beyond news such as talk shows and entertainment content should be undertaken, and the effects of other factors such as first run vs. repeated broadcasts and time of day of the broadcast should be quantified.

Using the CAE to process larger collections of broadcast programs and viewer data, other applications of interest to content creators and advertisers can also be explored. The text indexing capability of the CAE also provides an effective framework for studying the effects of linguistic cues (e.g. 'coming up', 'after the break') on early tuneaway at commercial breaks or finding common phrases during topic transition that appear to prompt higher levels of tuneaway. Future automated analysis may also answer localized sociological problems. Is audience retention in a series improved by mixing up its ad pod timing? Do viewers need a cognitive break after X minutes, such that they exhibit tuneaway for a minor topic change that wouldn't have occurred sooner after a commercial break? In non-dramatic (i.e. news) programs, does breaking to a live interview help avoid viewer tuneaway within a story, e.g.?

This work has shown that second-by-second anonymous audience measurement data from set top boxes can improve the performance of content segmentation algorithms. In the same way that other techniques use crowd-sourcing methods, this aggregated feedback provides more accurate content segments that can be used in subsequent applications.

REFERENCES

[1] ETSI TS 102 822-3-1 V1.4.1 (2007-11), Broadcast and On-line Services: Search, select and rightful use of content on personal storage systems ("TV-Anytime"); Part 3: Metadata; Sub-part 1: Phase 1 – Metadata schemas, 2007.

[2] Critical Mention, Inc. http://www.criticalmention.com

[3] BBN Broadcast Monitoring System, 5/14/10 http://www.bbn.com/resources/pdf/bms.pdf

[4] Rentrak, TV Essentials, http://www.rentrak.com/downloads/info/Rentrak_TV_Factsheet.pdf

[5] David Gibbon, Andrea Basso, Lee Begeja, Zhu Liu, Bernard Renger, Behzad Shahraray, Eric Zavesky, " Large-Scale Analysis for Interactive Media Consumption," *TV Content Analysis*, CRC Press, (to appear).

[6] Erik Kaashoek, Comskip, 1/1/10. http://www.kaashoek.com/comskip/

[7] SCTE American National Standard, "Digital Program Insertion Cueing Message for Cable", ANSI/SCTE 35 2007.

[8] Zhu Liu, David Gibbon, Harris Drucker, Andrea Basso, "Content Personalization and Adaptation for Three Screen Services," in ACM International Conference on Image and Video Retrieval, March 5, 2008.

[9] Pamela Douglas, *Writing the TV Drama Series,* Michael Wiese Productions, 2007.