A Dynamic Content Summarization System for Opportunistic Driver Infotainment

Barbara Rosario
Intel Labs
2200 Mission College Blvd.
Santa Clara, CA 95054
+1.510.495.3010
barbara.rosario@intel.com

Kent Lyons
Intel Labs
2200 Mission College Blvd.
Santa Clara, CA 95054
+1.408.765.2940
kent.lyons@intel.com

Jennifer Healey
Intel Labs
2200 Mission College Blvd.
Santa Clara, CA 95054
+1.408.653.8904
jennifer.healey@intel.com

ABSTRACT

The in-vehicle experience offers a unique challenge for delivering the right amount of information to the driver at the right time. The level of attention required to successfully manage the driving task is often in variable. An ideal in vehicle information delivery system would deliver content to the driver only during low task demand times, such as waiting at a stop light, when the driver's safety would be minimally compromised. The system would also have to respond to sudden changes in the situation such as driver interruption or distraction and terminate gracefully, allowing the driver to refocus on the driving task. In this paper, we present an embedded natural language processing (NLP) system that delivers speech synthesized summarized text content into tailored time slices. The system is also designed to respond dynamically to interruptions. We anticipate that this system could safely deliver speech synthesized content to drivers and allow them to make the most of their time on the road. We have implemented this system on an Atom Z530 processor with 1GB of RAM, a processor comparable to those found in factory installed In-Vehicle Infotainment (IVI) systems and have evaluated it in a laboratory test using a standard NLP corpus to demonstrate this potential.

Categories and Subject Descriptors

H.5.2 **[User Interfaces]:** Natural language, Voice I/O, Prototyping; I.2.7 **[Natural Language Processing]**: Text Analysis; Language parsing and understanding;

General Terms

Algorithms, Design, Human Factors

Keywords

Automotive Speech Interfaces, Natural Language Processing; Text Summarization; In-Vehicle Information Systems; Text to Speech Applications, Interruption Management

AutomotiveUI'11, November 29-December 2, 2011, Salzburg, Austria

1. INTRODUCTION

We present an automatic text summarization algorithm designed to deliver tailored content into specified time slices. The system dynamically responds to interruption and can re-summarize text to fill any remaining time. We envision that this system could be integrated with in-vehicle sensing and communication technology to take advantage of opportunistically discovered low task demand times to deliver audio content safely to drivers. The main contributions of this paper are the design, implementation and testing of an embedded NLP system to deliver summarized content into specified time windows.

Currently, most intelligent automotive systems assume a purely reactive interaction model in which the system constantly monitors and reacts to the driver's attention [1]. This interaction model assumes that the driving experience is constantly in unpredictable flux and fails to leverage opportunistic periods of low task demand time that can be reasonably predicted during a drive, such as when the driver is stopped at a traffic light. To take advantage of these opportunities, our system adopts an interaction model based on both planning for predicted low demand times (initial text summarization) and dynamically reacting to emergent events that interrupt these times (dynamic text re-summarization). Our model for time availability has been described as "plastic time" reflecting the idea that periods time that are available to interact with information systems expand and shrink to fit opportunistic gaps [2].

Recent estimates show that people commute an average of 86 minutes per day in the U.S. and 43 minutes per day in Europe [3]. To make the most of their time, drivers might want to access engaging content; however, if this content is delivered indiscriminately, the cognitive load presents a safety risk. For example, interactions with speech-based e-mail delivery systems can induce up to a 30% (310 msec) delay in driver response time [4]. Such a delay could have serious consequences in a high demand situation, but be relatively low risk in low velocity, low traffic situations.

2. NLP SYSTEM CONSIDERATIONS

Creating speech content from text to fill opportunistic gaps requires many summarization considerations. First, the available time is likely to be short [3]. Furthermore, human listening speed, 180 words per minute (wpm), is already much slower than the human reading rate (400 wpm) [5]. Also, audio is linear and non-persistent which makes it difficult to scan [6] and listening to synthesized speech while on-the-go can impede comprehension [7]. To meet these constraints and provide the most salient information to the user in a given time, text summarization has a clear advantage over compressed audio alone. Due to the dynamic and life critical nature of the driving task, such a system must also be able to respond to driver interruption. If a driver is interrupted during content delivery, the system should seamlessly terminate content delivery, direct the driver's attention to the road if necessary, be able to bookmark content at the interruption and be able to re-summarize the remaining text on demand and deliver it into new opportunistic gaps. In this paper, we demonstrate the capabilities of an NLP system designed to meet these specifications. We implemented the on a platform using the Atom Z530 processor with 1GB of RAM and testing it using on a standard NLP research corpus [8]. The NLP corpus results were evaluated using two different text summarization systems and the results are presented in Section 4.5. We additionally enabled the system to access content from the Internet to enable just in time information delivery in response to on road situations, such as finding out more information about a local landmark from a Wikipedia page.

3. PREDICTING LOW DEMAND TIME

We designed our system to be integrated with an intelligent in-vehicle system for predicting low driver task demand time. Although no commercial systems currently exist to meet this requirement, several research efforts have shown that estimating driver downtime is possible and could potentially be a commercial reality in the next decade. For the purpose of content delivery, these predictive systems do not need to be as accurate as they would for a safety critical application, they could be used to take advantage of as many low task demands they could find, even if some opportunities are missed.

3.1 Predictive Systems

We are aware of two systems that have been developed to predict the time a driver will have available to consume content. In 1998, Schmandt and Davis developed a smart navigation system that predicted the amount of time a driver would have to receive short audio instructions using an only an analysis of driving dynamics from sensors that already existed within the vehicle[9]. In 2010, Alt et al. developed a system for estimating the amount of time a driver was likely to wait at a traffic light based on a modeling of recorded GPS data and the driver's current location and time [3]. It is this second type of system that inspired the idea of using time tailored audio to fill low demand times as a form of micro entertainment. Beyond these systems, we may soon see the widespread implementation of both smart infrastructure such as traffic lights that communicate exact wait times to vehicles via Discrete Short Range Communication (DSRC) and infrastructure-free Vehicle to Cloud (V2C) services that provide an aggregate of wait time information from user driven traffic services such as "waze" (www.waze.com).

3.2 Reactive Systems

In addition to being able to predict low task demand times, we designed our system to react to the driver's attention by gracefully terminating and resuming content delivery at the point of interruption. We anticipate that a driver distraction system will provide the input to terminate the system. There are currently commercially available system detect distraction with over 90% accuracy (www.seeingmachines.com) which researchers have improved using the Seeing Machines faceAPI in conjunction with their own learning algorithms to 92% [10]. Other vision

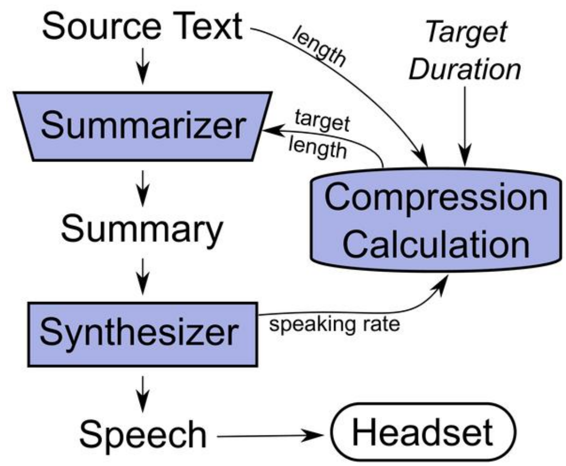

Figure 1. Summarization architecture showing synthesis pipeline and dynamic compression rate for a target time.

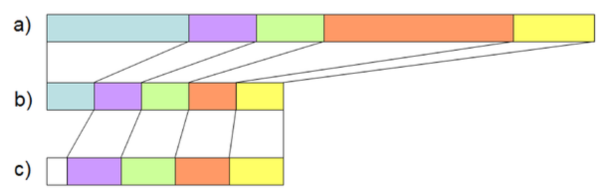

Figure 2. Document lengths through summarization and re-summarization: (a) source document lengths (b) initial compression for time target (c) re-summarization after interruption of the first document.

algorithms using yaw and pitch angles to track gaze directions report up to 86% accuracy [1]. In addition to vision systems, researchers have explored using body worn sensor such as head and leg mounted accelerometers [11] and electrocardiograms (heart rate) [12] to detect driver distraction at rates with greater than 90% accuracy. Distraction has also been detected using only information from the car's own embedded sensors such as steering and braking information [13, 14].

Our design anticipates using such a predictive system to define the initial time slot for summarization and a reactive distraction detector to terminate content delivery.

4. NLP CONTENT SUMMARIZATION

The main contribution of this paper is an embedded NLP system that generates speech synthesized audio content from summarized text to fit specified time periods. There is a wide range of content of interest to drivers available in text form including: email, news, blogs, wikis and messages from social networking sites. While on the road, it is dangerous and in many cases illegal to access this information via visual display. Our system utilizes NLP algorithms together with a speech synthesizer to dynamically produce audio content from text. The goal of the summarization is to provide the driver with as many important facts as possible in the anticipated time. The algorithm tailors content to an initial predicted time and also responds to interruptions and user requests to expand or re-summarize content to fill remaining time.

Several previous methods have been used to adjust the amount of time needed to consume audio content, most of which focus on time-compressed speech [15]. This can provide moderate increases in speaking rate (e.g. 1.3 - 2 times faster than normal). In contrast, by combining text summarization with speech synthesis, we can achieve order of magnitude compression rates.

There has been a significant amount of work on automatic text summarization. The goal of summarization is to extract content from an information source and distill it to “the most important information” that can be consumed within the length specified [17]. To determine what information is important, the summarizer needs to understand both the semantics of the original text and what is most relevant to the user and his or her context. Most automatic summarization work is text input to text summary. To the best of our knowledge, there is no previous work on text to time targeted text summarized speech synthesis.

4.1 Text Summarization Algorithm

The summarization algorithm begins with a traditional pipeline that obtains source text material, summarizes it, and then renders it to audio using a speech synthesizer (See Figure 1). Our system retrieves the HTML documents, extracts the text and sends the text to the NLP summarization component. We targeted two different general purpose summarization systems for comparison: Open Text Summarizer and Copernic. The summarized text is then converted to audio using the AT&T speech synthesizer. The resulting audio is then played either to the driver’s headset or the car’s speaker system. We anticipate that e-mail and web pages such as CNN and Wikipedia will be of greatest interest to drivers; however, to run a well structured test of the system and comparison of the summarizers, we utilized a standard, labeled NLP corpus to estimate our system’s performance [8].

4.2 Dynamic Summarization

To deliver tailored content into the specific time periods identified as low task demand times, we added an additional Compression Calculation to adjust the speaking rate of the synthesizer to fine tune the target duration audio. Our system can fill the target time with either a single article or a collection of articles. For multiple articles, we chose to compress each document to fill an equal portion of the available time, as shown in Figure 2. In the event of a sudden interruption, for example an incoming phone all or a distraction, content delivery can be stopped and then re-summarized to fill the remaining low task demand time (Figure 2 (b) (c)). In addition, if a user finds a particular summary very interesting, he or she can a request more detailed version and a longer summary can be delivered on demand.

4.3 Implementation

We implemented our system on with a 1.6GHz Intel Atom Z530 processor with 1GB of RAM using a Fujitsu U820 netbook. This class of processor is the same as those found in higher end embedded IVI systems. We initially tested the system using documents from either CNN or Wikipedia to evaluate to system’s ability to dynamically generate speech. We generated synthesized speech from the original documents and from four summaries based on different compression ratios. Each synthesis was conducted 10 times sequentially. From these trials, the mean speaking rate of the synthesizer was calculated to be 137.9 wpm. On average, the system required 5.47ms to synthesize each word (SD=0.34ms).

Table 1. Information retentions results: Target Lengths are the goal word count for each summary; Mean Result Lengths are the average word count of the obtained summaries; Mean Times are mean seconds of the synthesized speech generated from the summaries; Facts Found are the percentage of the total number of facts retained in the summaries.

Copernic Summaries

Target Lengths	Mean Result Lengths	Mean Times (s)	Facts (%)
100	118.9	51.7	13.8
250	243.5	105.9	23.3
1000	671.7	292.3	37.1

OTS Summaries

Target Lengths	Mean Result Lengths	Mean Times (s)	Facts (%)
100	132.2	57.5	16.7
250	281.8	122.6	23.9
1000	950.2	413.4	37.1

4.4 Content Retention Analysis

We wanted to quantify the performance of our system in terms of the trade-off between overall duration and the number of *important* facts retained. For this test we used a standardized, labeled corpus of data freely provided by the Text Analysis Conference (TAC). In particular, we chose data from the TAC 2008 summarization track which contains original HTML documents of blogs on 25 topics [8]. We considered 206 blog files from nine topics with a total of 573 facts. For each topic, a set of questions are provided by TAC. The TAC08 task was to produce summaries that contain answers to these questions. The TAC data set also contains a set of predetermined ground-truth answer snippets (facts) to these questions. The average number of facts per blog was 2.7 with the number of facts ranging from one fact per document to a maximum number of 79 facts for a longer blog (SD = 5.9). Here, we make the assumption that these provided answers are the facts that are most important for each document which an ideal summary should retain.

4.5 Processing and Summarization

We began processing the blogs by removing non-content text such as HTML tags, lists of archives and recent posts. After these steps, the average blog length was 1623.3 words (SD=4718.7). The important facts were relatively evenly distributed throughout the documents [8], making automatic summarization particularly challenging as most summarizers tend to focus on the early parts of a document. The text was then processed using both OTS and Copernic to generate 100, 250 and 1000 word summaries that were then synthesized and correspond to approximately 1 minute, 2 minute and 7 minute summaries respectively.

4.6 Text Summarization Results

For the two summarizers used and for each compression rate, we calculated how many of the facts identified by the TAC 2008 challenge were still present in the summaries. Table 1 shows the results for the corpus in terms of the mean summary word length achieved vs. the target word length, the duration of the target audio and the number of important facts found. Our shortest summaries (about 1 minute) take an order of magnitude less time than the original documents and retain 14-17% of all facts. This test is particularly challenging given the specificity of the information requested. For example, "facts" were answers to questions such as *"What features do people like about Vista?"* and "*Why do people like Starbucks better than Dunkin Donuts?"* Moreover, often there are multiple identical or similar facts for the same question but our evaluation measures the ability of capturing *all* facts. Furthermore, the facts we are looking for in the summaries are not necessarily the most salient aspects of the blogs. Rather, they are the answers to particular questions that were asked for the TAC08 task. Therefore, the results reported here provide a lower bound on the performance of the summarizers for this task and we think we may be confident that for "easier" source texts and less specific information needs (i.e. less specific than the questions and facts of the TAC08 dataset) the information retentions rates would be higher. The ultimate test will be the user's satisfaction with the delivered summaries. Our system provides extensive opportunities for customization that we hope will satisfy most drivers, for example, the ability to skip past uninteresting content and the ability to request longer summaries of content that is more interesting.

5. CONCLUSIONS

We have developed an embedded NLP system that delivers customized audio content into pre-specified time windows. The system is dynamic and can re-summarize content to different time lengths when interruptions occur and new opportunities emerge. We presented an overview of two types of in-vehicle systems currently being developed: systems for estimating low task demand time for drivers and systems for detecting driver distraction. Our system is designed as an advanced IVI system that would integrate with these time estimation and interruption detection systems. To test the ability of our system to deliver salient facts from text based content into various time durations, we performed a summarization analysis using a standardized NLP corpus. Our analysis quantified the tradeoffs between degree of content compression and the percentage of facts retained. The results show that the system is accurate in achieving the desired time slice of audio content.

6. REFERENCES

[1] Kutila, M., Jokela, M., Markkula, G. and Rue, M. R. 2007. Driver distraction detection with a camera vision system. In *Proceedings of the IEEE International Conference on Image Processing* (San Antonio, Texas, USA. September 16-19, 2007) ICIP 2007. IEEE, VI 201 – VI 204.

[2] Rattenbury, T., Nafus, D. and Anderson. K. Plastic: A metaphor for integrated technologies." *In Proceedings of the 10th international conference on Ubiquitous computing (*Seoul, South Korea September 21 - 24, 2008) UbiComp 2008. ACM, New York, NY, USA. 232-241

[3] Alt, F., Kern, D., Schulte, F., Pfleging, B., Shirazi, A. S. and Schmidt, A. Enabling micro-entertainment in vehicles based on context information. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Vehicular Applications,* (Pittsburgh, PA, USA, November 11-12, 2010) AutoUI 2010 Springer-Verlag, 117-124.

[4] Lee, J. D., Caven B., Haake, S. and Brown T. L. 2001. Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway, *Human Factors*, 43(Winter 2001), 631-640.

[5] Schmandt, C. 1994. *Voice Communication with Computers: Conversational Systems.* Van Nostrand Reinhold.

[6] Whittaker, S., Hirschberg, J., and Nakatani, C. H., Play it again: a study of the factors underlying speech browsing behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA, April 18 - 23, 1998). CHI '98. ACM, New York, NY, 247-248.

[7] Vadas, K., Patel, N., Lyons, K., Starner, T. and Jacko, J. Reading on-the-go: a comparison of audio and hand-held displays. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (Espoo, Finland, September 12-15, 2006) MobileHCI, 2006.

[8] Dang, H. T. and Owczarzak, K. Overview of the TAC 2008 update summarization task. In *Proceedings of the NIST Text Analysis Conference* (Gaithersburg, Maryland, USA, November 17-19 2008) TAC 2008. U.S. DOC 1-16.

[9] Schmandt C. and Davis, J. Synthetic speech for real time direction-giving. In *Proceedings IEEE Transactions on Consumer Electronics*, 35 (August 1998), 649-653.

[10] Azman, A., Meng, Q. and Edirisinghe, E. Correlation between eye movements and mouth movements to detect driver cognitive distraction. In *Proceedings of Brain Inspired Cognitive Systems* (Madrid, Spain, July 14-16, 2010) BICS 2010.

[11] Sathyaranayana, A., P. Angkititrakul, J. H. L. Hansen. Detecting and Classifying Driver Distraction. In *Proceedings of the Biennial on DSP for In-Vehicle and Mobile Systems*, (Istanbul, Turkey, June 17-19, 2007). Springer Publishing 2007. DSP '07.

[12] Masahiro, M., Kawanaka, H. and Oguri, K.. Effect of pattern recognition features on detection for driver's cognitive distraction. In *Proceedings of the13th International IEEE Conference on Intelligent Transportation Systems*, 2010 (Madeira Island, Portugal, September 19-22, 2010). IEEE 2010. ITSC 2010.

[13] Pentland, A. and Liu, A. Towards Augmented Control Systems. In *Proceedings of Intelligent Vehicles Symposium* (Detroit, Michigan, USA, September 25-26). IEEE 1995

[14] Sathyanarayana, A., Nageswaren, S., Ghasemzadeh, H., and Jafari, R. Body sensor networks for driver distraction identification. In *Proceedings of the 2008 IEEE International Conference on Vehicular Electronics and Safety* (Columbus, Ohio, USA September 22-24, 2008). ICVES 2008. IEEE 120-125.

[15] Arons, B. Tools for building asynchronous servers to support speech and audio applications. In *Proceedings of the 5th ACM Symposium on User Interface Software and Technology*,(1992) UIST '92. ACM, New York, NY, USA

[17] Radev, D. R., Hovy, E. and McKeown, K. Introduction to the special issue on summarization. *Computational Linguistics* 28 (December, 2002), 399-408