Samachaar Vaani: A Framework for providing Automated Spoken News Service

Rohit Kumar
Department of
Computer Sc. & Engineering
Punjab Engineering College
Chandigarh, INDIA
+91-0172-714809
rohit@pec.ac.in

Anuj Singla
Department of
Computer Sc. & Engineering
Punjab Engineering College
Chandigarh, INDIA
+91-0172-714809
a.singla@sify.com

Sanjeev Sofat
Department of
Computer Sc. & Engineering
Punjab Engineering College
Chandigarh, INDIA
+91-0172-746894
chestasofat@yahoo.com

ABSTRACT
Several web based news services provide news in spoken format also. The current processes involved in recording the spoken news are time consuming and costly. This has restricted several service providers from providing news in spoken form. A new framework that uses automatic speech synthesis to enable existing web based news services to provide spoken news services to the users at no recurrent costs has been presented in this paper.

Keywords
Samachaar Vaani, News Server, News Reader, TTS Systems, Spoken News Service

1. INTRODUCTION
Recently there have been several advances [1] in the field of automatic speech synthesis. These advances have enabled automated synthesis of natural sounding speech. The availability of such automated speech synthesis systems has enabled creation of applications which can replace a lot of manual work which is involved currently in recording of speech.

In this paper, one of the applications of speech synthesis in the web services domain has been presented. The Internet is vastly used as a medium for accessing news services. The web based news services provide primarily displayed news in form of text articles and pictures. Some services also provide news in spoken format. The process involved in generating spoken news is time consuming, manual (hence error - prone), un-economical as well as rigid. Section 2 discusses the details of the manual process.

Automatic speech synthesis can be used for replacing the existing process of generating spoken news by a process involving much less costs and almost no human effort. Through this paper, a framework for automation of spoken news services is presented. The framework and the various issues involved are discussed in Section 3. A system for automated spoken news generation for a Hindi news website is described in Section 4.

It must be mentioned here that there have been previous efforts [2] in the direction of generating news through automated speech synthesis. The genuinity of our effort lies in that it is aimed at enabling existing news service providers to provide more service with the infrastructure typically available with them at no recurrent costs.

2. NEED FOR AUTOMATED GENERATION OF SPOKEN NEWS

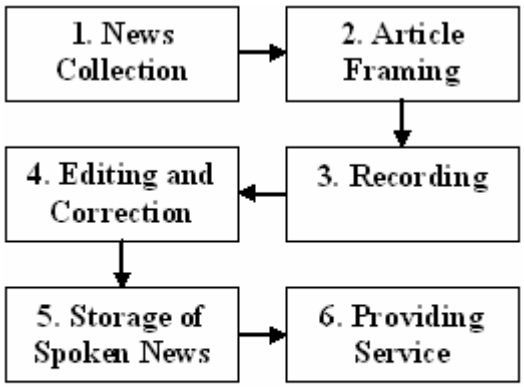

Figure I. Manual Process of Spoken News Generation

Several web based news services provide news in spoken format besides the usual typeset version. The current processes involved in generating the news in spoken format are fully manual and add heavily to the costs involved in news production. Also the problems related to error correction and editing are overwhelming. These problems and costs have restricted several news services from providing news in spoken form. The processes currently involved in generating spoken news are shown in figure I.

The various steps and costs involved in the manual processes of generation of spoken news are:

2.1. News Collections
News Collection involves deployment of reporters at various locations and providing them with appropriate infrastructure. News is also collected through other agencies in which case the expenditure involved is payments made to the agencies. Generally news collection is by a combination of these modes.

2.2. Article Framing
Once the news is collected by the news service provider, news reports and articles are framed on the news made available from the various sources. This process involves manual skill and expenditure in terms of payment to journalists and editors. Also this process overlaps considerably with the News Collection process and to some extent expenditure in these two processes merge together.

2.3. Recording

The finalized news reports are then recorded manually in a studio environment. Payments to the voice talents are involved in this process besides the cost of infrastructure and its maintenance. Though the cost of infrastructure is one time cost, the payments to voice talents is recurrent. Also this manual process of recording adds in the possibility of lot of mistakes leading to re-recording which accumulates into the overall costs.

2.4. Editing and Correction

Since the manual process of recording is error prone. Due to this post – recording editing and in some cases re-recordings are required. Editing requires more infrastructure and specialized manual talent. Also re-recording is both time consuming and costly.

2.5. Storage of Spoken News

The recorded news needs to be stored and requires a lot of storage infrastructure which is another recurrent cost. Moreover maintenance costs are involved. Spoken news cannot be discarded as it may be needed in future and re-recording is often undesirable.

2.6. Providing Service

The news service provider needs high performance streaming servers to provide spoken news service. Maintenance and personnel costs are involved in it. Moreover higher bandwidth is required for such services adding to the costs.

It is noted that existing manual processes of providing spoken news services are undesirable due to the high costs and efforts involved and the slow response time. This has led several news service providers to restrain from providing spoken news service as the costs and benefits do not break even for small news service providers.

We observe that an automated process for generating spoken news service that would enable existing news service providers to provide spoken news service with their existing infrastructure is desirable.

Further in this paper a framework for developing such an automated process by the use of Automatic Speech Synthesis is proposed. Also an actual system developed by following this framework is described.

3. SAMACHAAR VAANI: A FRAMEWORK

Samachaar Vaani is the result of our efforts to come up with a framework for enabling existing news service providers to provide spoken news service with the infrastructure typically available with them. Figure II shows the Samachaar Vaani framework.

Samachaar Vaani framework comprises of two components namely the News Server and the News Reader. The model of deployment as depicted in Figure II would be that the News Server will be installed at the news service provider's end and the News Reader will be distributed to the users through which they can access the news service. The News Reader can be distributed by the News Service provider or by a freeware organization on the Internet.

Now the two components of the Samachaar Vaani Framework and their blocks are discussed. The various issues related to each are mentioned.

3.1. News Server

The News Server accesses News collected and framed into articles by the News service provider through various sources as mentioned in section 2.1 and 2.2. There is no extra cost involved in this as the news service provider is already bearing the expense of collecting and framing articles. These news articles are passed through a set of filters and converters as shown in figure 2. The role of filters is to process the news to extract only the content in a suitable form depending upon the way it will be used in succeeding blocks.

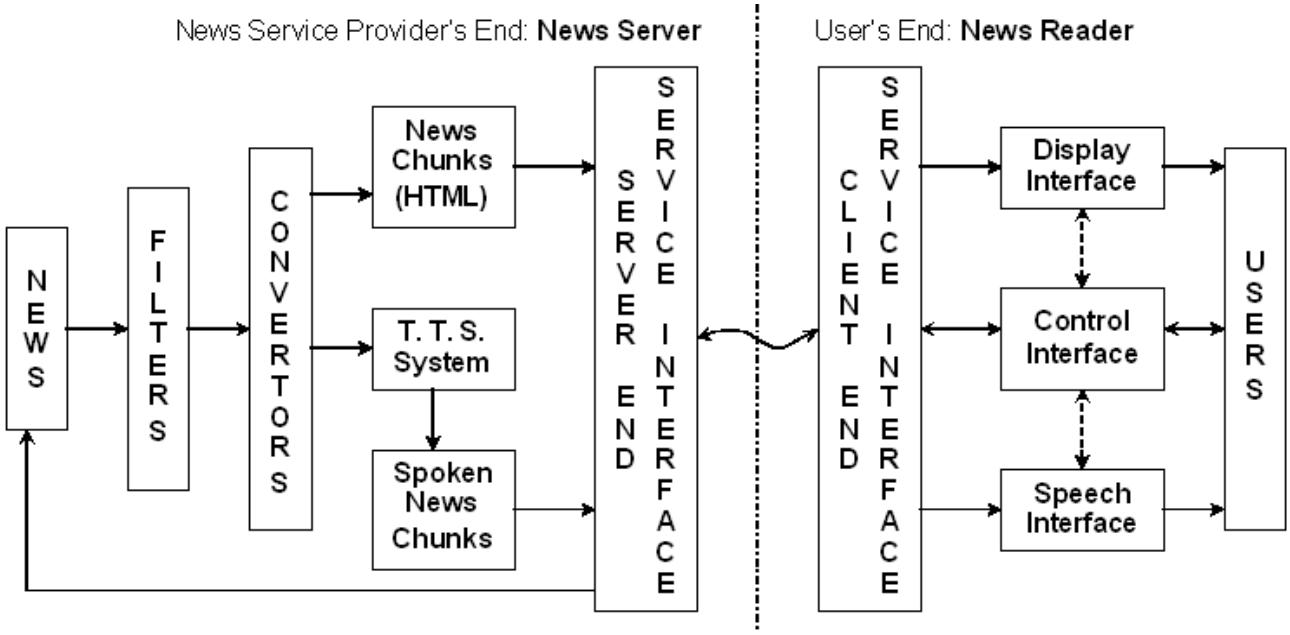

Figure II. The Samachaar Vaani Framework

3.1.1. Filters

Primarily the filter for selecting the news to be broadcasted in spoken format breaks up the news into chunks, collects the headlines into the start of the news, arranges the news in a sequence and assign each chunk an identity number, adds additional messages like *"Good Morning, Welcome to <XYZ> news service and this is the News at 10:00 AM on 12 March 2003. The headlines first."* and *"That is the end of this bulletin. Please reconnect at 5:00 PM for the next update. Thank You."*

Ideally the filter for displayed news organizes the displayed news into a sequential format in accordance with the spoken news content and inserts markers to achieve synchrony between the spoken and the displayed interfaces to provide the user with a multimodal interface.

The filters are not required to be news extraction or language processing modules. Since the news articles are being framed by the news service provider itself, the news service provider can annotate the news by putting in some special markers to indicate the type of content like headlines, comments, captions, etc. Though has been proposals to standardize this kind of annotation of news content [3] by international bodies [4], none has yet been adopted worldwide. With the annotated news being available, the filter is a parser of these annotations. Standardizing will help in generic filters which can be used for several websites. Even without standardization, a website specific filter can be used.

3.1.2. Converters

Each type of news content from the filters is passed through content specific converters for use with the blocks succeeding the converters. The chunks of spoken news content are passed through a converter that converts the notation of the news content to a notation acceptable by the text to speech system described ahead. This may not be required for English news. But while working with languages like Hindi and other languages for which no widely adopted standard for information interchange exists, the converters become indispensable. The kind of converter we have used is mentioned in section 4. The converter for spoken news content acts as a preprocessor to the language processing module of the text to speech system.

As in the case of converters for spoken news content, for chunks of displayed news content, which is generally in HTML, the content is converted to be compatible with the fonts being supported by the client end i.e. the news reader.

3.1.3. Text to Speech System

A Text to Speech (TTS) System based on Concatenative Synthesis by optimal unit selection from a large annotated corpus can be used to generate natural sounding spoken news content. TTS Systems based on such approaches provide the flexibility of plugging in new voices without requiring a change in the synthesizer, besides the high quality of output. But such TTS Systems require a large corpus leading to distribution problem.

We have decided to keep the TTS System on the news service providers' end in the Samachaar Vaani framework. The place of deployment of the TTS System is an arguable issue. Keeping the TTS System at the user's end saves a lot of bandwidth involved in transfer of speech content over the Internet. But this is at the loss of plugability of new voices due to the large sizes of the corpus. The size of corpuses for speech engines based on concatenative synthesis varies between 100 – 250 MB (without compression). Moreover the user will need to download speech databases for different languages if he/she is interested in listening to News content in more than one language. Finally the News service providers itself may not be interested in allowing the speech engine to be available with the user as this would lead to his loss of control over News content being read out using the system, ultimately leading to loss of revenue, directly or indirectly. Owing to all these issues the Samachaar Vaani Framework places the Speech Engine on the service provider's end.

The speech synthesis module of the TTS System is succeeded by a speech coder to compress the speech. A GSM coder can produce sufficient compression of speech.

3.1.4. Server End Service Interface

The server end service interface implements the application layer protocol for communication between server and reader over the network (generally Internet). The protocol allows authentication or identification of connecting user, request for current news or news of a previous date, listing of size of each chunk in the news content requested and downloads of the chunks in any order as per news reader's requirement. Also the server end service interface has streaming server for streaming the spoken news content to the reader in synchrony with the requests received through the protocol.

3.2. News Reader

The News Reader acts as the client to the News Server and is deployed on the user's end. Owing the Internet based distribution, the News Reader is built using platform independent tools and is easily portable to a large number of software as well as hardware platforms. Further the News Reader is a small, easy to use, easy to download and easy to install software. To achieve these characteristics it is further desirable to keep the Text to Speech System on the service provider's end.

3.2.1. Client End Service Interface

The client end service interface implements the client end application layer for communication with the News Server using a protocol common with the service interface on the server end. As in the case of server end service interface, the client end service interface implements authentication/identification of the user and requesting/receiving of news content.

3.2.2. Control Interface

The control interface is the primary interface through which the user interacts with the News Reader. This interface allows the user to send his/her authentication/identification information, to request news from the server through the reader and to view/hear to downloaded news content through the display and the speech interfaces respectively. The control interface brings about the synchrony between both the output interfaces providing the user with a multimodal interface. The control interface also stores the downloaded news content for offline access to the latest downloaded news.

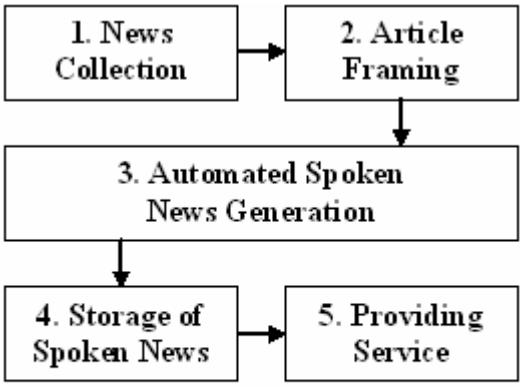

Figure III. Process of Automated Spoken News Generation

Figure III shows the revised processes and costs involved in automated spoken news generation using a system based upon the Samachaar Vaani framework. Here we can see that the manual processes of recording and editing are replaced by an automated process which besides bringing down the costs reduces the response time to generate spoken news to almost real time. Also it should be mentioned that the storage requirements in step 4 of Figure III are lesser than those in step 5 of Figure I because now only the recently synthesized spoken News content needs to be stored and no archives need to be maintained as old news can be re-synthesized whenever required. This further removes the recurrent costs associated with tapes and disks. Moreover the use of above framework for generating spoken news does not require any special infrastructure beyond that typically available with news service providers.

4. A SYSTEM FOR INDIAN LANGUAGE SPOKEN NEWS SERVICE

Based on the framework described in section 3, a system for enabling automated spoken news service for Hindi has been developed. The system is currently setup to work for the Hindi website of BBC World Service. The present system used a generic Indian language TTS System which can be used with large number of Indian languages besides Hindi. The TTS System is able to produce natural sounding speech for Indian languages by using concatenative synthesis approach and implements a Unit Selection algorithm using Syllable as basic unit [5].

On the server end the filters are preceded by a HTML parser which parses the HTML pages available locally on the BBC web server and provides the filters with news. In this way the News Server is essentially just another piece of software running on the service providers' end and it interacts with the web server through a HTML parser. The news content on the BBC web server is annotated with <SPAN> tags to specify the type of news content. The filters use the annotated news to generate chunks of displayed and spoken news content. We have not used any converters for displayed news content because the BBC Hindi web service uses UTF-8 encoded Devnagari script [6] which is a standard encoding and is support on several platforms. But in case of other web services, it will be required to convert the displayed content to a standard notation like UTF-8, Unicode [7] or ISCII [8]. The spoken news content is converted to Z-Notation [5: Appendix A] as per the requirements of the text to speech system.

The service interfaces on the server and the reader ends support identification of user, requests for latest news content and transmission and reception of news content in spoken i.e. audio as well as HTML format.

CONCLUSION

We have proposed a framework for implementing an automated spoken news service using automatic speech synthesis systems. It has been discussed that use of automated speech synthesis can lead to reduced cost and time required for providing spoken news service. The system enables the existing News service providers to provide more service to the users with the infrastructure typically available with them.

ACKNOWLEDGEMENTS

We are thankful to Dr. Rajeev Sangal, Director, IIIT Hyderabad for allowing us to use to the Indian language speech synthesis system developed at LTRC, IIIT Hyderabad.

REFERENCES

[1] A. J. Hunt and A. W. Black, "Unit Selection in a concatenative speech synthesis system for a large speech database," in Proceedings of Intl. Conf. on Acoustics, Speech and Signal Processing, pp 373-376, 1996.

[2] Ananova, the virtual newscaster, http://www.ananova.com/

[3] NewsML (News Markup Language) Specifications, http://www.iptc.org/site/NewsML/index.html

[4] IPTC (International Press Telecommunications Council), http://www.iptc.org/

[5] S. P. Kishore, Rohit Kumar, Rajeev Sangal, "A Data Driven Synthesis Approach for Indian Languages using Syllable as Basic Unit", in Proceeding of Intl. Conf. on Natural Language Processing, pp. 311-316, 2002.

[6] UTF-8 Devnagari Specifications, http://www.unicode.org/charts/PDF/U0900.pdf

[7] Unicode Specifications, http://www.unicode.org/

[8] IS 13194-1991, "Indian Script for Information Interchange", Bureau of Indian Standards, Manak Bhavan, New Delhi