Collaboration and the Knowledge Economy: Issues, Applications, Case Studies
Paul Cunningham and Miriam Cunningham (Eds)
IOS Press, 2008 Amsterdam
ISBN 978–1–58603–924-0

Using Language Technology to Improve Interaction and Provide Skim Reading Abilities to Audio Information Services

Arne JÖNSSON[1], Bjarte BUGGE[2], Mimi AXELSSON[3], Erica BERGENHOLM[3], Bertil CARLSSON[3], Gro DAHLBOM[3], Robert KREVERS[3], Karin NILSSON[3], Jonas RYBING[3], Christian SMITH[3]
[1]*Santa Anna IT Research Institute AB, 581 83 Linköping, Sweden, E-mail: arnjo@ida.liu.se*
[2]*Audio To Me AB, Linköping Sweden, E-mail: bjbu@audiotome.se*
[3]*Linköping University, 581 83 Linköping, Sweden*

Abstract: In this paper we present language technology enhancements to audio-based information services (i.e. services where information is presented using spoken language). The enhancements presented in the paper addresses two issues for audio-based services: 1) interaction with the service is rigid and 2) the ability to listen to summaries is limited. Our developments allow for more natural and efficient control of the service and means that facilitates skim reading. Using speech dialogue instead of traditional buttons provides means for more advanced navigation in the audio material. Vector space techniques are used to collect the most relevant sentences in a text and allows for skim reading of varying depth.

1. Introduction

Today most interactions with technical products are realized using variants of visual user interfaces. The same holds for most techniques for information distribution. Many persons, e.g. persons with visual disabilities or dyslexia, or various age related disabilities, have limited access to such information. The need for audio-based information services is significant for these groups. Considering that in a recent survey 98% answer "No" to the question "Did you manage to read what you wanted to read yesterday?" and that more than 90% of today's information is available in text format only, the user group above is extended enormously. For instance, persons listening to PM:s or reports while driving.

The company AudioToMe has developed a large archive of human voice accessible information based on an open Service Oriented Architecture (SOA) enabling various audio based services support for multiple consumer channels/devices, giving consumers access to audio services using standard consumer technologies.

2. Objectives

The objective of this research is to improve interaction of the AudioToMe information services utilizing methods from language technology. Today, users select information to listen to using the keypad on a cell phone or a special device. This provides a robust and easy to use navigation. However, it is not that natural, flexible or efficient.

2.1 Interaction

If an information system instead were controlled through speech interaction, it would be possible to provide flexible and efficient navigation even without visual elements. Speech interaction allows users, not only to skip articles and navigate on subject level but also to

search for specific articles using natural language. Natural language dialogue systems are today commercially used for a variety of tasks, and generic dialogue systems are available (e.g. [1,2]) that provide a repository of frameworks and tools in the form of software code that can be shared amongst researchers and that is ready to be used and re-used in industry. To illustrate consider the dialogue excerpt below:

User: Any news about Nokia today?
System: There are stock quotes and an article entitled "Nokia releases new GPS-phone". Which one are you interested in?
User: Read the article
System: <Reads article>
User: Stop. What will the weather be like in Linköping tomorrow?

This short excerpt illustrates how users can navigate a newspaper in a much more natural and efficient way than pressing buttons for moving forward or skipping articles.

2.2 Skim Reading

It should also be possible to efficiently browse information, including skim reading at various levels and present summaries of texts.

Summarizing techniques can already be quite useful when applied to written text, but with such a static, visual media you still have the option of skimming through the area in any direction and at any pace you'd like. Audio, however, is a strictly linear, non-static media where you can only go in one direction while gathering information. Since most of us are able to make out words if they are read to us in normal or a little above normal pace, the shortened material produced by a summariser would be immensely useful to people skimming an audio file

3. Methodology

Different methods will be used in different sub-projects. Software development is carried out in parallel with the empirical investigations. Thus, agile software development methodology is utilised as it facilitates rapid prototyping and user involvement [3,4].

3.1 Interaction

There are a number of well-explored dialogue phenomena that can be utilised to direct speech interaction, such as clarification request, contextual interpretation, and topic shifts c.f. [3], but these are merely techniques and as such only useful if correctly used. Understanding which features to use in various situations, and how they are realized in language, e.g. various prompts, is therefore of utmost importance for speech controlled interaction to be useful.

To address this, we need to investigate the behaviour of users interacting with a sound-based medium in various situations and consequently we utilise qualitative methods, such as interviews and open prototype evaluations.

3.2 Skim Reading

Skim reading also involves various techniques depending on situation, user and information content. For instance, a politician driving and skim reading a report before a meeting would probably prefer a short summary of the report, whereas a visually impaired person listening to the daily newspaper might prefer keywords reflecting the content of a story. Investigations on user opinions on various techniques for skim reading in different situations are consequently important, combining qualitative and quantitative methods.

4. Technology Description

AudioToMe services include audio mailbox, pod casts, RSS, banking services, landlord information, health care information, and newspapers. The sound files are easy to access and users can create their own profile reflecting the order in which they want to listen to the information, e.g. a newspaper.

4.1 Interaction

Newspaper information is stored in m3u or DAISY (Digital Accessible Information SYstem) files generated every day and contain subject area tags and within each subject area the relevant articles are ordered based on an individual user's normal reading order. The parsed m3u-files provide a basic structure for each user and are used for browsing data.

We will use the Nuance speech recognizer, which is a speaker independent recognizer that allows for full sentence recognition. Interaction can be accomplished using more or less sophisticated commands. Some commands are domain independent, such as READ <x>, where <x> can be the name of a paper, an article, a report etc., NEXT, PAUSE, STOP, BACK etc. Others are domain dependent, e.g. `DELETE` makes sense in an e-mail system but not when reading the newspaper.

Control in the AudioToMe service system is done using various spoken commands, based on results from the empirical investigations. The commands must be natural to use, but at the same time technically easy to identify. Unfortunately, words that are easy to identify for the speech recognizer need not be the ones that users prefer to use.

4.2 Skim Reading

Skim reading techniques depend on the situation, user, information needs etc. Sometimes whole sentences are preferred, for instance as a summary of a report, but key words, reflecting the content of a text, can also be used as a basis for deciding if an article is interesting to hear.

For skim reading we utilise the DAISY format, which defines how textual information links to corresponding sections in sound files. DAISY also defines a number of standard tags for marking up sections, depending on the source text. For instance, in fiction novels there are tags for chapters, whereas scientific texts have tags down to the level of single sentences. Newspapers are somewhere in between depending on the publisher.

It is, thus, possible to use standard text search techniques to navigate in the information and once selected play the corresponding sound file.

Vector space techniques[5], such as Latent Semantic Analysis[6], and Random Indexing[7], on whole sentences create the set of sentences in a text that best resembles the original content[8]. Using whole sentences means that no speech synthesis is needed, instead sentences are assembled from the original corpus and the sound files are delivered.

Vector space techniques used in a similar way on words provide content information. In such cases pre-processing is used to remove stop words, build term lists (with synonyms), and perform stemming. Finally, speech synthesis is needed to deliver the information.

5. Developments

For skim reading we are using two techniques, PageRank and Random Indexing[9]. Random Indexing is performed using use the Java-toolkit developed by Martin Hassel[10]. Based on Hassel´s toolkit we have developed a test bed allowing us to pick out words or sentences from the text and also vary the number of items to present. The program also uses the Snowball-stemmer[11].

Random Indexing is language independent and it is thus possible to produce summaries for many languages. However, stemming and stop word lists are language specific and new such are needed for high quality summaries.

6. Results

This section presents results from experiments on interaction and skim reading.

6.1 Interaction

We have done initial investigations on how people would like to use speech to navigate audio-based newspapers. The pilot tests have been performed on four students from Linköping University in Sweden, three men and one woman. During these tests the experimenter had a computer with a program for audio based newspapers. The newspapers are recorded by the newspaper company and distributed over the Internet or with the radio to the subscribers. The investigations were exploratory. The participants were to pretend that they had access to audio-based newspapers in their daily lives and were instructed to use language freely to control the interaction. We also assigned tasks to the participants like "Read the domestic news" and "Read the second letter in 'Ordet fritt'" (Ordet fritt is the section with letters to the newspaper).

Three of the four participants started immediately by saying the name of the section they wanted to read. The fourth participant said the name of the newspaper and then continued with the name of the section he was interested in. To start reading an article subjects either wanted to say the article name or number of the article in numerical order, the persons that wanted some kind of table of contents read when they entered a section suggested this. Three of the four participants wanted to hear some kind of table of contents.

Two of the four participants talked spontaneously about searching with key words. For instance, searching for special words in the text or headlines to find an interesting article.

The difference in expected response from the system varied. One subject wanted to be certain the system had understood his question and preferred paraphrasing like "Do you want to go to domestic news?" One subject preferred if the system started to read articles, or if the command was intended to direct to a section, to read the name of the section.

The study on speech-controlled interaction has provided a corpus that will be used for initial development of the speech-controlled interface.

Further investigations will be carried out to develop the dialogue system and to overcome the problem that it can be hard to make people pretend that they are talking to a computer or in this case something that don't even exist. To handle this, we will use a method similar to[12], where participants will talk to a human that in turn is accessing the current button-based system. These dialogues will be distilled[13] and analysed to provide further knowledge on how persons want to talk to a computer, what kind of commands they want to use and what kind of functions they need.

6.2 Skim Reading

We have also conducted experiments on two types of summaries for skim reading, presenting only words or presenting whole sentences. In the study, 20 students, between 20-30 years old, not visually impaired, listened to sound files of either complete sentences or words. Humans, not synthetic speech, were used to produce the sound files.

The subjects were presented a varying number of sentences, or words, representing the "best" 10, 25, 50 or 70% of the total number of sentences, where "best" is based on the Random Indexing ranking. Stop words were removed when words were presented to the subjects. The order in which sentences, or words, occurred in the original text were preserved when presented to the subjects.

Our results show that subjects prefer whole sentences to words on all four levels (10, 25, 50 or 70% of the sentences or words). 17 out of 20 preferred sentences to words. Note that in the instructions we informed our subjects that 10% should be seen as a way of deciding if an article is worth reading whereas 70% should be seen as a summary of the text. One could assume that on the 10% or 20% level subjects would prefer words to sentences as an indication to whether the article is interesting to listen to or not, but that was not the case.

Thus, our initial experiments conclude that for un-experienced users whole sentences are preferred for skim reading. Further research will include using experienced users of audio-based services.

7. Business Benefits

The core business of Audio To Me is to give consumers access to written information through the use of audio. This includes all aspects from service design and content production to multi channel access and user experience. The research described above is a direct result of field tests and requests from our customers. Adding this to our portfolio will directly enhance the customer value proposition of our services.

Both enhancements, natural dialogue and advanced skim reading, will improve our market position significantly and gives us a strong position to address local, regional and a global market. Our services address a wide range of consumers, from people with defined reading disabilities to the executive enable to find time for reading in a busy schedule.

An example of customer value is how advanced skim reading will help a political active person with dyslexia get an grip of the extensive text material required in an the democratic process.

A segment of special interest is consumers with low technical adoption where audio services can be a new way to give access to media based on RSS, PodCast and blogs etc. We have found the demographic and structurally prerequisites for these type of new audio services positively homogenous within major parts of the European market.

8. Conclusions

We have presented language technology enhancements to support navigating large collections of information material available as audio services.

The research addresses two of the major issues that have been identified in investigations of various uses of audio services. The physical user interface with buttons of various sizes is often a challenge for non-technically skilled users or users with motor dysfunctions. Providing means to use spoken interaction to direct the information services greatly enhances the usability of audio-based services. Furthermore, active persons with reading disabilities face the challenge to orient in and select information from large text masses. Giving such users the possibility to skim read greatly enhances their productivity and hopefully enjoyment using audio-based services.

Currently our results are mainly from investigations on how users would like to use language technology based software in audio-based services. We have not yet integrated the software components for speech-controlled interface or the skim reading techniques to AudioToMe's products.

References

[1] Michael F McTear, Spoken dialogue technology: toward the conversational user interface. Springer Verlag: London, 2004.

[2] J., Hochberg, N., Kambhatla, S., Roukos, A flexible framework for developing mixed-initiative dialog systems. In: *3rd SIGdial Workshop on Discourse and Dialogue*, Philadelphia, Pennsylvania, 2002.

[3] Pontus Johansson, Lars Degerstedt and Arne Jönsson, Iterative Development of an Information-Providing Dialogue System, *Proceedings of the 7th ERCIM Workshop "User Interfaces for All"* Paris, France, 2002
[4] Lars Degerstedt and Arne Jönsson, A Method for Iterative Implementation of Dialogue Management, *IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Seattle, 2001.
[5] Lars Eldén, *Matrix Methods in Data Mining and Pattern Recognition*. Society for Industrial & Applied Mathematics (SIAM), 2007.
[6] Thomas K. Landauer and Susan T. Dumais, A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240, 1997.
[7] Magnus Sahlgren, An Introduction to Random Indexing. *Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering*, 2005
[8] Martin Hassel, *Resource Lean and Portable Automatic Text Summarization,* PhD Thesis, ISRN-KTH/CSC/A--07/09—SE, KTH, Sweden, 2007.
[9] Niladri Chatterjee, Shiwali Mohan, Extraction-Based Single-Document Summarization Using Random Indexing, *19th IEEE International Conference on Tools with Artificial Intelligence*, 2007
[10] http://www.csc.kth.se/~xmartin/java/JavaSDM/
[11] http://snowball.tartarus.org/
[12] Pontus Wärnestål Modelling a Dialogue Strategy for Personalized Movie Recommendations. *Proceedings of Beyond Personalization 2005 Workshop (Intelligent User Interfaces 2005)*. San Diego (CA), U.S.A., January 9, 2005. pp. 77-82.
[13] Arne Jönsson and Nils Dahlbäck, Distilling dialogues - A method using natural dialogue corpora for dialogue systems development *Proceedings of 6th Applied Natural Language Processing Conference,* Seattle, 2000, pp. 44-51.