Spoken Query Processing for Information Access in Digital Libraries

Fabio Crestani

Department of Computer Science
University of Strathclyde
Glasgow G1 1XH, Scotland
fabioc@cs.strath.ac.uk

Abstract We briefly outline the ongoing research at Strathclyde University on the use of spoken query processing for information access in digital libraries.

1 Spoken Access to Digital Libraries

The widespread access to Internet and the high bandwidth often available to users makes them assume that there is nothing easier than remotely connecting to and searching a large information repository. Nevertheless, a very large part of the world population does not and will not have for a long future access to computers or Internet. There are cases in which the only available (or the most convenient) communication mean is a telephone or mobile phone. In addition, there are also situations and users for whom the availability of computer screen and keyboard is not useful. Blind or partially-sighted users (e.g. users who have problems due to disabilities, protective clothing or working environment) may only have access to information if this is accessible via voice and/or sound. In all these cases, if we want users to take advantage of the large amount of information stored in digital repositories, it is necessary to enable access to them solely via voice. This implies the design and implementation of systems capable not only of understanding the user's spoken request, finding the required information and presenting it as speech, but also capable of interacting with the user in order to better understand his/her information need, whenever this is not clear enough to proceed effectively with the searching.

At Strathclyde University we are investigating how we can improve current information access technology to enable user to access digital libraries using voice. At the current stage, the main objective of the project is to enable a user to interact with a probabilistic Information Retrieval (IR) system over a telephone line. This implies enabling the user to submit queries, commands and relevance assessments via voice and be presented with the content of retrieved documents using an interactive vocal interface. A first prototype system enabling journalist to access newspapers archives using a telephone or mobile phone, is currently being implemented and tested [2]. In this short paper we outline some of the ongoing research on improving the effectiveness of spoken query processing for information access in digital libraries.

2 Spoken Query Processing

While the effectiveness of spoken document retrieval (SDR) has been well studied and well documented in recent IR literature, very little research work has been devoted to studying the effectiveness of spoken query processing (SQP). It has to be recognised that SQP poses a number of additional challenges compared with SDR. The most important ones are:

- query processing needs to be performed on-line and "almost" real time, while spoken document recognition and indexing can be performed off-line;
- queries are usually much shorter than documents and word recognition errors (WRE) may have more serious effects on them;
- we may have very little training data on the voice of each user and we may have a large number of different users in different acoustic conditions.

Given these challenges, we decided that our first priority was to test the limitations of current information access technology in the SQP task, given realistic levels of WRE rates.

3 Effects of Word Recognition Errors in SQP

In [4] we reported the results of an experimental study of the effects of WRE in spoken queries on the effectiveness of an IR system. The document collection we used was generated at Dragon Systems Inc. [1], and this is currently the only test collection available with spoken queries. A large number of experiments using a standard IR system were carried out. In these experiments some of the parameters of the IR process (i.e. stemming, stoplist, weighting schema, etc.) were changed to study their effects on the effectiveness on the IR task in relation to the different levels of WRE rates. For reason of space we cannot report here the results obtained, but we can say that the results show that classical IR techniques are quite robust to considerably high levels of WRE rates (up to about 40%), in particular for long queries.

However, our experimentation falls short in a number of ways. First, the queries used were too long and not really representative of typical user queries (although some initial unpublished user studies on spoken queries indicate that spoken queries are usually longer than written queries). Second, the WRE rates of the queries used in this experimentation were typical of "dictated" spoken queries, since this was the way they were generated. Dictated speech is considerably different from spontaneous speech and easier to recognise. Spontaneous spoken queries have higher levels of WRE rates and different kinds of errors. Finally, the queries used were generated artificially from queries spoken in a laboratory environment. Telephone speech is more difficult to recognise than laboratory speech and produces higher WRE rates.

Initial experiments with a more realistic test collection show that standard IR techniques are not so robust to the kind of WRE found in spontaneous speech. More complex techniques for dealing with SQP of spontaneous speech are therefore needed.

4 Techniques for Improving SQP

We are currently experimenting with the a number of techniques for obtaining acceptable levels of effectiveness of SQP of spontaneous speech for information access. Here, for reasons of space, we can only outline the first results obtained and the directions of the work in progress.

Relevance Feedback. Given the level of IR effectiveness with spoken queries obtained in [4], we can conclude that it will be quite likely that in the first n retrieved documents (with n dependent on the user's preference and usually less than 10) there will be at least one relevant document. In a previous study we have successfully tested the ability of the user to perceive if a document is relevant to an information need when the document is presented in the form of a short spoken summary [8]. Such scenario is consistent with our current target prototype application. The above two results enabled us to conclude that *relevance feedback* (RF) could be a good strategy to improve effectiveness in a SQP task.

Indeed, in [6] we showed that both standard RF and pseudo RF enable to improve the IR effectiveness with spoken queries, in particular for short queries. In fact, the use of RF (and in particular of standard RF) enables to reduce considerably the difference in effectiveness between long and short queries previously found in [4].

Use of Prosodic Stress for Topic Detection. In [7] we investigated the relationship between acoustic stress in spoken sentences and information content. On one side, the *average acoustic stress* was measured for each word throughout each utterance. On the other side an IR index (the $tf{-}idf$ weight) was calculated. The scatter plots of the two measures showed a correlation between higher values of average acoustic stress with increasing values of the IR index of the word in the majority of the analysed utterances. A statistically more valid proof of such a relationship was derived from the histogram of the words with high average acoustic stress vs. the IR index. This confirmed that a word with high average acoustic stress has also a high value of the IR index and, if we trust IR indexes, also high informative content.

This study confirmed our hypothesis of a direct relationship between acoustic stress and information content as identified by IR weighting of spontaneous spoken sentences. The next stage of this work will be the integration of prosodic stress and IR weighting evidence into an new IR weighting algorithm for spontaneous sentences. This weighting schema will take into account both word acoustic and statistical clues to characterise the document/query informative content. It will be extremely useful in a number of tasks and in particular for SQP where the short length of queries requires every possible clue to fully capture the user's information need.

Combination of Semantic and Phonetic Term Similarity. A fundamental problem of IR is "term mismatch". A query is usually a short and incomplete description of the user's information need. Users and authors of documents often use different terms to refer to the same concepts. This fact produces an incorrect relevance ranking of documents with regards to the information need expressed in the query [5]. A similar problem can be found in SDR and SQP, where terms misrecognised by the speech recognition process are found not matching in query and document representations. Naturally, this hinder the effectiveness of the IR system. We called this problem "term misrecognition", by analogy to the term mismatch problem.

In [3] we presented a model for dealing with the term mismatch and the term misrecognition problems in SDR and SQP. Term similarity is used at retrieval time to estimate the relevance of a document in response to a query by looking not only at matching terms, but also at non-matching terms whose semantic and/or *phonetic similarity* are above a predefined threshold. Semantic similarity can help solve the term mismatch problem. It can be estimated using Expected Mutual Information Measure or some similar measure. Phonetic similarity, on the other hand, can help tackle the term misrecognition problem. It can be estimated using Error Recognition Confusion Matrices, for example. An experimental investigation is currently being carried out. The experimental results will provide useful feedback on the effectiveness of the models proposed and on how to effectively combine semantic and phonetic similarity.

References

1. J. Barnett, S. Anderson, J. Broglio, M. Singh, R. Hudson, and S.W. Kuo. Experiments in spoken queries for document retrieval. In *Eurospeech 97*, volume 3, pages 1323–1326, Rodhes, Greece, September 1997.
2. F. Crestani. Vocal access to a newspaper archive: design issues and preliminary investigation. In *Proceedings of ACM Digital Libraries*, pages 59–68, Berkeley, CA, USA, August 1999.
3. F. Crestani. Combination of semantic and phonetic term similarity for spoken document retrieval and spoken query processing. In *Proceedings of the 8th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, Madrid, Spain, July 2000. In press.
4. F. Crestani. Effects of word recogntion errors in spoken query processing. In *Proceedings of the IEEE ADL 2000 Conference*, Washington DC, USA, May 2000. In press.
5. F. Crestani. Exploiting the similarity of non-matching terms at retrieval time. *Journal of Information Retrieval*, 2(1):23–43, 2000.
6. F. Crestani. Word recognition errors and relevance feedback in spoken query processing. In *Proceedings of Fourth Internation Conference on Flexible Query Answering Systems*, Warsaw, Poland, October 2000. In press.
7. R. Silipo and F. Crestani. Acoustic stress and topic detection in American English spoken sentences. Technical Report TR-00-005, International Computer Science Institute, Berkeley, CA, USA, February 2000.
8. A. Tombros and F. Crestani. Users's perception of relevance of spoken documents. *Journal of the American Society for Information Science*, 2000. In press.