Play it again: a study of the factors underlying speech browsing behavior

Steve Whittaker, Julia Hirschberg, Christine H. Nakatani
ATT Labs-Research
180 Park Avenue
Florham Park, NJ 07932, USA
+1 (973) 360 8339
stevew/julia/chn@research.att.com

ABSTRACT
Several recent UIs support access to recorded speech archives, but these have not yet been systematically evaluated. We describe a laboratory study of speech archive browsing using a GUI. We evaluate the effects of four factors: *task type*, *familiarity*, *structure*, and *play operation duration*. We found that while users learnt the overall layout of topics in the archive, they experienced major problems in learning the internal structure of archival topics. Contrary to our expectations, we also discovered that structural information and fixed duration play operations were less useful for browsing than anticipated. We discuss the impact of our results for speech archive UI design, and describe a new UI which supports navigation within topic.

Keywords
Speech archives, browsing, search, retrieval.

MOTIVATION AND METHOD
The amount of personal and public data stored in speech archives has greatly increased over the last few years, and several recent prototype UIs have been built to enable browsing of these archives [1,3]. However little is known about: (a) how people access and search speech data; (b) design principles for UIs to speech archives. In this study, we therefore examine how access is affected by two factors: (a) *task type;* and (b) *familiarity of material*. While previous research has suggested that these factors affect browsing, no detailed evaluation has been done. Second we investigate the impact of two browser features: (c) *topic structure;* (d) *play duration.* Although these features have been implemented in previous browsers, their impact on browsing and their interaction with *task* and *familiarity* has not been systematically tested. Our hypotheses were:

- Tasks: Search efficiency (i.e. number of search operations and search time) depends on the amount of speech information users must access: summary tasks requiring access to an entire topic will less efficient than search for two specific facts, which in turn will be less efficient than search for one fact.
- Familiarity: familiar material will elicit more efficient search.
- Topic: providing information about where topics begin will increase the efficiency of search.
- Play duration: Short duration fixed play intervals will be used for identifying relevant topics, whereas longer fixed play durations will be used for search within a topic.

Fourteen people were given a speech archive, consisting of 8 voicemail messages ('topics') appended together in one audio file lasting 236.3s. Message retrieval is an example of a real situation where users access speech archives, which we felt would be familiar to our users. Users accessed the archive to answer 16 questions about the 8 topics. The questions were based on retrieval tasks identified in a naturalistic study of use of a real speech archive [2].

There were 3 types of *task*: 1fact, 2fact and summary. Four questions required users to access one specific fact, e.g. a date or phone number from a topic (1fact), a further four required access of two such facts (2fact), and eight questions required users to reproduce the gist of a topic (summary).

The first 8 questions required users to access each of the 8 topics once, and questions 9-16 required each topic to be accessed again. To investigate the effects of *familiarity* we compared users' performance on the first 8 versus the second 8 of the 16 questions.

Users were given one of two GUI browsers: "basic" and "topic". The topic browser allows the user to select a given topic by serial position (e.g. topic 1). Play will then begin at the start of that topic. Both browsers support random access: the whole archive is represented as a horizontal bar and users can select any point in the archive (e.g. inserting the cursor halfway across the bar begins play halfway through the archive). For both browsers, users then select one of three play *durations*: "play short" (3s), "play long" (10s) and "play to end" (unrestricted play until the user stops it). We used a simple GUI rather than testing complex search features. This was motivated by three factors. First, data on accessing a real speech archive indicate that even highly experienced users make little use of sophisticated features such as scanning, speed up/slow down, jump forward/back [2]. Second, informal evaluations of complex speech UIs reveal that advanced browsing features are often not well understood by users, and do not necessarily benefit search [1,3]. Given the unclear benefits of complex features, we wanted to establish baseline data for speech retrieval using a simple prototype. Finally, the features we tested will be part of any browsing interface.

Users were given 5-10 minutes on practice tasks before the experiment. After it, we gave users a memory test, asking them to recall the content, name of caller and serial position of each topic. We then administered a questionnaire eliciting reactions to browser features and comments about the tasks.

RESULTS

We logged the number and type of each play operation, duration and location of played speech within the archive, and time to answer each question. The results for each hypothesis follow and all differences discussed are statistically significant at $p < 0.05$, using ANOVA.

Tasks: As expected 1fact were answered more efficiently than both other tasks (see Table 1). However, contrary to expectations, summaries were more efficient than 2fact, despite requiring access to more information. The results indicate that performance depends both on the *type* and the *amount* of information users must access. User comments revealed why 2fact were so difficult: with summaries it was possible to remember several pieces of approximate information. 2fact questions required complex navigation within topic and the additional precision required to retain verbatim information often meant that users forgot one fact while searching for the second. They then found it hard to relocate the fact they had just forgotten. The user logs reveal problems of forgetting and relocating prior facts. In the course of answering each 2fact question users actually played the two target facts a combined total of 7.9 times. In contrast target facts for 1fact tasks were only accessed 1.5 times and topics 2.9 times for summary tasks.

	# operations		solution time	
1fact	2.4		23.0	
2fact	4.1		37.6	
summary	2.9	(F = 7.43)	32.3	(F = 11.7)
familiar	2.1		22.5	
unfamiliar	4.1	(F = 35.5)	40.1	(F = 36.6)
topic	3.7		30.0	
no topic	2.5	(F = 5.09)	32.5	(F = 6.60)

Table 1: Effects of task, familiarity and topic structure on retrieval efficiency, with relevant F ANOVA values.

Familiarity: Overall familiar material elicits more efficient search. We then separated overall search operations into: (a) *identification of the relevant topic*; (b) *information extraction* i.e. finding the answer within the target topic. Familiarity only helped with topic identification, it had no effect on information extraction.

Topic: Users made frequent use of topic boundary information. Although random access was available with the topic browser, users only used it for 33% of their access operations. Furthermore, users' comments about the topic boundary feature were highly positive. Despite this, however, we found that topic based access may be less efficient: users took more operations although less time to answer questions when topics were provided. Why was this the case? Post-hoc tests showed that topic browser users had worse memory for the 8 topics than simple browser users. Users of the simple browser reported making strenuous efforts to learn a mental model of the archive. In contrast, reliance on topic structure may lead topic browser users never to do so.

Sampling: Play duration was independent of whether search was within or outside topic. Furthermore, there was little use of either fixed play operation: all users preferred unrestricted play. In the survey, users reported that fixed duration options reduced their comprehension by truncating topic playback in unpredictable places. Users preferred the greater control of unrestricted play even though this meant the overhead of stopping play, when they had heard enough.

DESIGN IMPLICATIONS

What are the design implications of these results? First, for this relatively small archive, users readily learnt the overall archive layout and gist of the various topics. They were much poorer at operations within topic however: they were unable to relocate previously accessed information within topic for 2fact tasks, and showed no familiarity effects for search within topic. In future work we will test a new *transcript-based* UI which uses speech recognition to display a transcript of the current topic to support local navigation and memory. Second, our sampling results argue against designs of UIs which present fixed duration 'skims' of salient speech information, where questions have been raised about the optimal fixed interval for skimming [1]. Our data indicate there should be no fixed intervals. Instead skims should access salient speech, but allow users control over the precise playback duration. Finally providing topic boundaries may be of limited value: although users like this feature, heavy use of it may prevent them from learning the contents of the archive.

REFERENCES

1. Arons, B. Interactively skimming speech. Unpublished PhD thesis, MIT Media Lab, 1994.
2. Hirschberg, J. and Whittaker, S. Browsing and searching a real voice archive. In Proceedings of AAAI Spring Symposium on Multimedia Indexing, AAAI Press, CA, 1997.
3. Hauptmann, A. and Witbrock, M. Informedia: News-on-Demand Multimedia Information Acquisition and Retrieval. In Maybury, M., ed. Intelligent Multimedia Information Retrieval, AAAI Press, 1997.