Reconstruction of Apollo Mission Control Center Activity

Douglas W. Oard
College of Information Studies and UMIACS
University of Maryland, College Park, MD USA
oard@umd.edu

Abhijeet Sangwan, John H. L. Hansen
Center for Robust Speech Systems (CRSS)
The University of Texas at Dallas, Richardson, TX, USA.
{abhijeet.sangwan,john.hansen}@utdallas.edu

ABSTRACT

Some cultural heritage institutions have large and growing spoken word collections, the items of which often live in isolation from each other and from other parts of those collections. This paper describes the design process for construction and evaluation of a system to automatically construct links between spoken conversations that address different aspects of the same event. We see this as one step, among many, towards building richer interconnections between part of the same collection, and between collections. Our development environment is a collection of several thousand hours of recordings made in the Mission Control Center during the Apollo space missions of the 1960's and 1970's.

Categories and Subject Descriptors

H.3.7 [**Information Systems**]: Information Storage and Retrieval – digital libraries.

General Terms

Design, Experimentation.

Keywords

Cultural heritage, content linking.

1. INTRODUCTION

As physical access to archival collections becomes increasingly available, both as a result of digitization initiatives and as a result of substantial investments in making those digitized assets available on the Web, providing equally facile and capable "intellectual access" – the ability of users to find and use what they want -- has risen in importance. Intellectual access to archival collections raises a number of issues; in this paper we address one specific format issue: building links between different items in massive spoken word collections. At present, few collections are dominated by spoken word materials, but as prices drop, recording devices proliferate, and digital sustainability investments level the playing field between media, we can expect that to change [6]. We therefore believe the time is right to begin to explore these questions.

Many complex multi-party activities are coordinated using speech. Examples include air traffic control, military command centers, and human spaceflight. Common to all of these settings is that no single person can listen to everything that is happening, and thus no single person can actually come to know precisely how all that that happened was actually interconnected. Indeed, the quantities of recorded speech can become so vast that in some cases no single person could ever even listen to all of it, much less make sense of it all. As the cost of recording and storing audio continues to decline, scholars who seek to make sense of our past will therefore need new tools to help focus their attention on the small parts of this cornucopia that that they most need to hear, and on which parts of that they need to actually hear together. In this paper, we describe the design of a system for this exploration of the speech recorded in the National Aeronautics and Space Administration (NASA) Mission Control Center (MCC).

2. THE MCC AUDIO

A total of 38 people flew in Apollo spacecraft on 15 missions between 1968 and 1975. Of those, 24 flew to the moon on 9 missions; 12 of those people walked on the moon. Together, the flights spanned more days than a typical person works in a year, with 30-track audio recorders running continuously during that time in the MCC. The result is about 100,000 hours of recorded audio, with perhaps about 10% being speech and 90% silence.

The MCC was organized hierarchically, with one flight director, a dozen or so flight controllers, and a corresponding set of "back rooms" (more properly, "staff support rooms") that supported each flight controller. One "loop" (i.e., intercom circuit) connected the flight director with the flight controllers, and each back room had a separate loop to connect them with the flight controller who they supported. There were also several additional loops that two or more flight controllers could select when necessary to facilitate coordination activities that did not need to be heard by the full flight control team. Two special loops were also recorded, one between the spacecraft and the MCC, and a second for the news media that included those communications along with public affairs commentary. These circuits were recorded using a 30-track tape recorder that ran continuously; specific loops could be assigned to specific channels on the tape recorder, and it was common to record at least all of the circuits mentioned above.

During low workload periods, flight controllers would typically monitor three circuits simultaneously: (1) the flight director loop, (2) the loop to their own back room, and the (3) space-to-ground communications. They would typically alternate between talking on the first two of those; only the CAPCOM (a title derived from the older name "capsule communicator") would normally talk to the astronauts. We can, therefore, trace the flow of information from a specific flight controller's back room to that controller, then from that controller on the flight director loop to the CAPCOM, then (with the flight director's concurrence) from the CAPCOM up to the spacecraft. Indeed, during certain mission phases there were voice recorders running on board the spacecraft so we can hear how the astronauts discussed and acted on information that they received from the ground. The entire system included several dozen people, and during high workload periods there were many more simultaneous conversations going on that any one person could listen to. Flight controllers were able to monitor the back room loops of other controllers, and often did when specific activities on those loops might affect their own decisions.

All communication with the spacecraft was transcribed twice, once in near real time for use by the press and the second time (more carefully) for use in post-mission analysis. None of the other loops were routinely transcribed, however. Indeed, with the exception of the flight director loop, most of the other recorded loops have never even been replayed. Today, the tapes are stored in the National Archives and Records Administration (NARA), which has no system capable of playing them. There is such as system at the NASA Johnson Space Center, however, although this machine can currently play only 2 of the 30 tracks at a time. We are currently working with NASA to create or to otherwise gain access to a 30-track replay (and digitization) capability.

3. THE SPEECH LINKING TASK

Because the Apollo missions were carefully choreographed using a meticulously planned timeline, Mission Elapsed Time (MET) since launch provides a natural means for organizing access to the resulting flood of information. Time-based reconstruction has is widely used, include in aircraft accident investigations and for mission replay on military training ranges, and it is well matched to the linear nature of audio. We have, therefore, constructed a mission replay system for the Apollo missions that we call the Apollo Archive Explorer in which audio, video, transcripts, maps, planned and actual event timelines, and post-flight commentary are presented as a unified time-synchronized reconstruction of a mission [7]. Our focus in this paper is on the design of an additional capability that will add multi-channel audio scene reconstruction to the Apollo Archive Explorer.

The basic capability is simple; we can mix audio from multiple sources, some in one ear, some in the other, some in both; some at higher volume, some at lower. Such a capability replicates to a degree the (monophonic) capability that flight controllers had at the time to listen to multiple loops at once. Fight controllers were, however, highly trained to manage that complexity, and experienced in recognizing the voices on those loops. For modern users, we will need to provide some tools to help them manage and make sense of the complexity.

We envision three kinds of tools. First, we can use speech activity detection and speaker identification to identify who is speaking when and then to indicate that graphically in some way. Our initial design for this uses a sketch of the MCC console layout to simply indicate which flight controller is speaking (by lighting up the depiction of that console); in future work, we expect to build a similar visualization showing the back rooms. We do not yet have the back room loops digitized, so we are focusing initially on integrating the flight director loop, the space-to-ground communications, and the audio recorded aboard the spacecraft.

The speaker identification problem is simplified somewhat in this setting because flight controllers are typically the only people who speak on both the flight director's loop and (usually one) back room loop. There are some times when more than one flight controller is present at the same console, but there are also long periods in which only a single flight controller is present. We can therefore cluster speakers across one back room loop and the flight director loop, thus easily identifying which speaker is the flight controller who owns that specific back room loop. Once we know that, we can listen for first-name references to specific members of the back room staff, thus labeling most of the remaining clusters. Because we don't yet have the back room loops digitized, we are initially training a speaker identification system for the Apollo 11 flight director loop by hand-annotating a portion of the recording in a more conventional way. One early result from this work is that the shortness of some utterances (e.g., polling flight controllers for their agreement to proceed to the next mission phase) is challenging for conventional speaker identification techniques. We therefore plan to build in some interaction models that leverage specific forms of stylized interactions that often result in short utterances in this setting, and we plan also to leverage limited-vocabulary isolated-word speech recognition because it is common for an interaction to begin with the statement of a name or a position title that indicates who is being addressed.

Although the initial use we will make of speaker identification will be to indicate to the user of the Apollo Archive Explorer who is speaking when, our most important use of speaker identification will be as a basis for speaker-dependent Large-Vocabulary Continuous Speech Recognition (LVCSR). Our initial experiments with speaker-independent LVCSR (trained on other sources) yielded results that are not sufficiently accurate for content linking, so improving LVCSR accuracy is on our critical path. In addition to creating speaker-dependent models for each flight director and each flight controller (about 60 people total, because flight control teams worked in shifts), we are also now building domain-specific language models. The Apollo program is one of the most extensively documented undertakings in all of human history, so there is no shortage of text that can be used for language modeling. Much of this text was originally in printed form and is now available from the NASA Technical Reports Server (NTRS) as scanned PDF, for which Optical Character Recognition is easily performed. Of course, OCR introduces errors, so there will surely be a quality-quantity tradeoff to explore.

For our initial LVCSR experiments have focused on 11 hours of spacecraft communications with the MCC [8]. We trained a language model using text from a number of sources including transcripts, books, and technical reports. Although there was existing OCR for all of these materials, we obtained better results from rerunning OCR with a more modern system We used the resulting text to train a word trigram language model with a 38,000 word vocabulary. We ran forced alignment on the 11 hours of audio, finding that only 3 hours aligned well enough to be used; we rejected the remaining 8 hours due to (i) inaccurate transcripts, (ii) inaccurate timestamps in the transcripts, and (iii) poor quality audio. The remaining 3 hours of audio was split equally into adaptation and evaluation sets. Using a conversational telephone acoustic model (trained on a mixture of the switchboard and Fisher datasets) as baseline, the adaptation set was used to perform MLLR followed by MAP adaptation (resulting in adapted acoustic models). Upon decoding the evaluation dataset, we obtained a word error rate (WER) of 77% for the adapted system, which was a substantial improvement over the 92% word error rate of the baseline system without adaptation (i.e., trained only on Fisher and switchboard, but with our Apollo language model). We are now working on using the in-band transmitter keying tones ("quindar tones") to improve the time alignment and thus gain access to additional training data, and of course we can ultimately train on one entire mission and then test on subsequent missions. Moreover, much more material exists from which richer language models could be built. And, of course, the MCC loops that are our principal focus are far more tractable acoustically than the spacecraft communication circuits. We therefore foresee little difficulty in eventually sufficiently accurate transcripts to support content alignment.

Other issues that may affect the accuracy of speaker identification and speech recognition include unmodeled variations such as (i) background noise (such as side conversations in MCC or pumps running aboard the spacecraft), (ii) band limited recording equipment (particularly for recordings that were later transmitted from the spacecraft to the ground and recorded there), and (iii) time-varying channel characteristics (which can result both from the characteristics of the analog take recorders used at the time and from the need to replay these tapes on much older equipment today) [1,9]. Additionally, despite the image of “right stuff” astronauts and flight controllers when they are speaking on the radio, the “off the record” recordings of the mission control loops and the interaction among the astronauts exhibit clear variations in speech production due to the whole range of human emotions such as stress, anxiety, and joy. Physical, emotional and cognitive state are well known to influence speech production, and the resulting variations can adversely affect both speech recognition and speaker identification [2]. Moreover, the Apollo astronauts spoke in an exceptionally diverse range of physical environments, including under extreme g-forces during launch and reentry, in low pressure pure oxygen during moonwalks, and (later in life) in oral history interviews. This exceptional range of diversity in working environments in itself offers some remarkable research opportunities for speech processing systems. Indeed, those unique opportunities were one of our principal initial motivations for undertaking this project.

There is a large body of research that has focused on problem of “speech under stress.” The usual approaches are to attempt to remove variability in speech (introduced due to environment, channel and speech production) in either the feature domain or the model domain [3], and we should be able to apply some of those techniques as well. Unlike the speech corpora on which much of this earlier research has been performed, we have very large amounts of speech from a relative small number of people. That offers us an unprecedented opportunity to investigate long-term adaptation techniques that could ultimately have broad applications beyond this specific task (for example, personal speech systems such as Siri face similar challenges).

Once we have adequately accurate LVCSR (for which prior work in a query-based ranked retrieval setting suggests requires will require word error rates below about 50%), we can begin to build content linking systems. We already have some experience with content linking in this setting from an experiment we reported in the main conference in which we linked mission events from the transcript of the communication between the spacecraft and the MCC to question-answer pairs from the oral history interviews with the same astronauts that were recorded many years later [5]. In that work, simple sliding window bag-of-words techniques yielded a mean reciprocal rank at 3 of about 0.5 for mission events for which a substantial mention existed in the oral history. Importantly, however, we have not yet tackled the important problem of automatically determining when no link should be made. Such a capability will be essential for content linking between the MCC loops. For this we will need to switch from a ranked retrieval design to one based on supervised machine learning for text classification, and for that we will need training data. Thus system design naturally leads us to the question of test collection deign.

4. TEST COLLECTION DESIGN

Our goal is to discover when two loops should be linked, so our test collection must contain some ground truth for those kinds of links. The synchronized nature of our task greatly simplifies the search space – we seek only to link two loops at the same time, not to build links that span different mission phases or that span different missions. This constraint results in a simple form for a link, it is specified by a start time, an end time, and a pair of loops to be linked during that interval.

In our initial thinking, we can see two employment scenarios that lead to two link types. In one scenario, the listener hears something being discussed on the flight director loop and wishes to hear the conversations between the flight controller involved and his (flight controllers were all men in Apollo) back room. In the other, more challenging, scenario, the listener hears something on one back room loop and wishes to know if there are related conversations ongoing on other back room loops. We plan to identify some ground truth links of each type.

There are at least three ways of identifying such events in a mission. First, NASA prepared a post-flight mission report in which every engineering anomaly that occurred during the mission was identified. For example, there were water leaks in one of the two Apollo spacecraft during both the Apollo 11 and the Apollo 15 missions. The resulting database (present in multiple scanned documents, not actually yet as a database) offers one possible source for events that would have prompted discussion on one or more loops. Second, over the past several decades, authors and documentary film makers have mined the records of the Apollo program for compelling human interest events. For example, the commander of the first lunar mission (Apollo 8) became ill between the Earth and the Moon. These events, less well codified but now nonetheless well known, offer an alternative source of events that could have prompted discussion on multiple loops. A third obvious source for events is the sequence of planned mission events from the pre-flight flight plan. For example, a planned television broadcast from lunar orbit would require coordination among flight controllers responsible for spacecraft attitude and communication systems. From these three sources, it seems reasonable to conclude that identifying a broad range of events for which links might be built should be straightforward.

The more interesting question will be whether people can agree on the proper span for a link. Here we are helped a bit by the fact that the onset time of a link might reasonably be of greater importance to the listener than its termination time, for the simple reason that once the listener chooses to listen to something they can make the decision of when to stop listening on their own. We will, therefore, ask annotators to mark both onset and termination times, but we will initially evaluate (and assess inter-annotator agreement) based solely on the onset time errors. As we have done previously for retrieval of unsegmented speech, we plan to initially use a one-sided linear penalty function as our evaluation measure [4].

Ultimately, of course, we will need to conduct user studies to learn which kinds of links users are most interested in seeing, and what kinds of errors actually prove to be most troublesome for them. But this is a chicken-and-egg sort of problem, in that we cannot study how users would use a system until such a system exists. So we necessarily anticipate a spiral development model in which we first build a plausible system, and then we iteratively refine that system as we learn more about how it will be used.

5. CONCLUSION

We often think of cultural heritage as involving things that are centuries old, and often at best incompletely documented. As time progresses, however, we will surely encounter more collections like that created by the Apollo missions, where our problems will be not how best to make the most of that which is scarce but rather how best to make the best use of that which is abundant. The Apollo missions, flown as they we now nearly half a century ago, offer an outstanding laboratory with which to begin that quest.

6. ACKNOWLEDGEMENTS

The author is grateful to the rest of "Houston We Have a Solution" project team for the rich discussions and initial implementation and experimental effort that have brought us to this point, including Lakshmish Kaushik, Chengzhu Yu, Joseph Malionek and Bryan Toth. This material is based upon work supported by NSF under Grant 1218159. Opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

7. REFERENCES

[1] Murat Akbacak and John H.L. Hansen. Environmental sniffing: noise knowledge estimation for robust speech systems. *IEEE Transactions on Audio, Speech, and Language Processing,* 15(2) 465-477, 2007.

[2] John H. L. Hansen, Abhijeet Sangwan and Wooil Kim. Speech under stress and Lombard effect: Impact and solutions for forensic speaker recognition. *Forensic Speaker Recognition*, Springer, New York, 2012, pp.103-123.

[3] Wooil Kim and John H.L. Hansen. Time–Frequency Correlation-Based Missing-Feature Reconstruction for Robust Speech Recognition in Band-Restricted Conditions. IEEE Transactions on *Audio, Speech, and Language Processing,* 17(7)1292-1304, 2009.

[4] Baolong Liu and Douglas W. Oard, One-Sided Measured for Evaluating Ranked Retrieval Effectiveness with Conversational Speech, in SIGIR 2006, pp.673-674.

[5] Joseph Malionek, Douglas W. Oard, Abhijeet Sangwan and John H.L. Hansen, Linking Transcribed Conversational Speech, in SIGIR 2013, 4 pages.

[6] Douglas W. Oard, Unlocking the Potential of the Spoken Word, *Science*, 321(5897)1787-1788, 2008.

[7] Douglas W. Oard and Joseph Malionek, the Apollo Archive Explorer, in JCDL 2013, 2 pages.

[8] Abhijeet Sangwan, Lakshmish Kaushik, Chengzhu Yu, John H.L. Hansen and Douglas W. Oard, Houston we Have a Solution: Using NASA Apollo Program to Advance Speech and Language Processing Technology, in Interspeech 2013, 5 pages.

[9] Umit Yapanel and John H.L. Hansen. A New Perceptually Motivated MVDR-Based Acoustic Front-End (PMVDR) for Robust Automatic Speech Recognition. *Speech Communication,* 50(2)142-152, 2008.