Toward Access to Multi-Perspective Archival Spoken Word Content

Douglas W. Oard,[1] John H.L. Hansen,[2] Abhijeet Sangawan,[2]
Bryan Toth,[1] Lakshmish Kaushik[2] and Chengzhu Yu[2]

[1]University of Maryland, College Park, MD USA
[2]University of Texas at Dallas, Richardson, TX USA

Abstract. During the mid-twentieth century Apollo missions to the
Moon, dozens of intercommunication and telecommunication voice chan-
nels were recorded for historical purposes in the Mission Control Center.
These recordings are now being digitized. This paper describes initial ex-
periments with integration of multi-channel audio into a mission recon-
struction system, and it describes work in progress on the development
of more advanced user experience designs.

1 Introduction

In the four years between December of 1968 and December of 1972, nine Apollo
missions flew to Earth's moon; six of those missions landed, two orbited, and one
flew by [4]. While it's true that astronauts flew those missions in outer space,
the vast majority of the people who participated in each mission never left the
ground. Three astronauts flew on each Apollo mission, but each mission also
demanded the expertise and engagement of more than one hundred flight con-
trollers and other support personnel. As the movie *Apollo 13* illustrated, those
flight controllers managed the enormous complexity of each mission, and much
of their coordination was conducted using voice intercom circuits [2]. Dozens of
those intercom circuits were recorded, continuously for the many days it took
to conduct each mission, but most of those recordings have never been heard.
We are now working to change that by digitizing nearly the full set of Mission
Control Center recordings from an Apollo mission, and by integrating some of
those recordings into an interactive system for mission reconstruction.

2 Mission Control

In the Apollo era, the Mission Control Center (MCC) in Houston, Texas con-
sisted of a Mission Operations Control Room (MOCR) where the flight con-
trollers worked, and several Staff Support Rooms (SSR) where technical special-
ists worked [3]. Three types of intercom circuits, known colloquially as "loops,"
were used. The most important of these, monitored by nearly everyone, was the
Flight Director loop. The Flight Director had the ultimate authority for all de-
cisions made during a mission, and the flight controllers in the MOCR used the

Flight Director loop to speak with the Flight Director. The most numerous set of loops were those used by each flight controller in the MOCR to speak with the technical experts in the SSR who supported their function. For example, during launch the Booster flight controller in the MOCR used the "Booster loop" to speak with SSR experts on (rocket) Engines and on Propellant (i.e., rocket fuel). The third type of loop was a "meet me" loop that flight controllers and SSR experts who were not normally on the same loop could use to have side conversations about specific issues.

Four 30-track tape recorders were used to record much of this audio, both to support subsequent engineering analysis and for historical purposes [7]. Two recorders ran simultaneously, with the other two being started just before the tape on the first two ended. Each of the two recorders was set up to record different channels, so a total of 56 channels could be recorded simultaneously (56 rather than 60 because one channel on each recorder was set to record a code indicating the time and a second was used for voice annotations of the tape itself). Some of these channels were set to record specific loops, but many channels were configured instead to record the headset audio of specific flight controllers. Flight controllers typically listened to many channels (including at least the Flight Director loop, the radio communication with the astronauts, and their own loop with their SSR experts), with some channels set to be loud (demanding their attention) and others softer (to provide awareness of other things going on at the same time). In addition to loops and headset audio, the radio communication with the astronauts was also typically recorded on one channel, and explanations of mission activity for the television audience that was provided by a Public Affairs Officer (PAO) was typically recorded on another.

3   Digitization

The tapes are stored by the United States National Archives and Records Administration, but they are difficult to replay because the 30-track reel-to-reel recording format is no longer in used. One SoundScriber player for these tapes does exist, but before our project it was able to play only one track at a time. But these old tapes are rather fragile (as is the SoundScriber tape player!) and it is simply not possible to play every tape 28 times to get every channel. We therefore constructed a new 3-track tape head to prove it was possible to capture multiple tracks at the same time without encountering cross-channel interference. Once we verified this was possible, we then set out to build a 30-channel digitization pipeline that now makes it possible to play each tape just once.

To verify the audio quality and to gain experience with the digitization process, in September and December 2014 we first conducted a pilot study of our digitization and processing pipeline using our newly built 3-track tape head. We selected portions of six tapes for six high-interest periods during the Apollo 11 mission (launch from Earth, lunar landing, start of the moonwalk, two other periods while on the Moon, and lunar liftoff) and digitized three tracks at a

time. We digitized a total of 52 twenty-minute segments that together span 29 different headset-audio channels and 6 different intercom lo¡op channels.

The National Aeronautics and Space Administration (NASA) must review all materials for public release to comply with U.S. law, so we have worked with NASA to develop a scalable review process. Listening to dozens of channels for many days of audio would be infeasible, so we used two technologies to accelerate the review process. First, we developed a Speech Activity Detection (SAD) system that is able to accommodate long periods of silences (which are common on many channels) and that can handle headset audio channels that include radio communication with the astronauts (as many do) [8]. We then manually transcribed the entire recorded radio communication between MCC and the Astronauts for the eight-day Apollo 11 mission to the precise timing standards required for training a Large-Vocabulary Continuous Speech Recognition (LVCSR) system, and we used that transcribed data to train such a system [6].

Another challenge that we encountered was the pervasive use of acronyms to facilitate efficient communication during the Apollo program. This required that we supplement the term list used by our LVCSR system. To accomplish this, we searched all available NASA and Apollo related sites for (usually scanned) documents from which acronyms could be extracted. Pronunciations were then developed for each acronym, another challenging task because some acronyms were by convention spoken as a word (e.g., "fido" for FDO) while others were by convention spelled out (e.g., "c s m" for CSM). Next, we adapted acoustic models developed in our earlier work to match the statistical characteristics of the Apollo radio transmissions on which we trained the LVCSR system.

We then ran the resulting LVCSR system on each of the 52 twenty-minute digitized audio files from headset audio and intercom files and we provided the resulting automatically generated transcripts to NASA along with the audio files. NASA used the transcripts to identify portions of the audio that might require detailed review, and then they conducted a detailed review of those portions using the digitized audio. The net effect was a more efficient review process than would have been possible with the audio alone. NASA completed the review of the 52 twenty-minute segments in August, 2015, and those initial digitization results are now available for our use in multi-channel audio experimentation.

In the meantime, we designed and and installed a new 30-track tape head on the one existing SoundScriber player. We initially tested that installation, and our new 30-channel digitization pipeline, using analog calibration test tapes that had been created to test the original 30-channel SoundScriber recorders (which no longer exist) back in the 1960's. We have to date digitized the entirety of the first lunar landing mission (Apollo 11), and portions of the Apollo 13 mission (which is of historical interest because of an explosion in space that prevented a lunar landing on that mission), obtaining more than 19,000 hours of digitized audio over a three-month period in late 2015. We have run our SAD and LVCSR systems on that audio, and it is now being reviewed for release by NASA.

4 Mission Reconstruction

We originally developed the Apollo Archive Explorer (AEX) to serve as a platform for experimenting with time-synchronized replay of the multimedia records of an Apollo mission [5].[1] AEX performs time-synchronized replay of four types of media: audio, transcripts, video, and photographs. Additional synchronized content includes an animated map (showing where on the Moon the astronauts are during moonwalks), flight plans (showing what the astronauts had been planning to do at that time), and post-flight interviews (which are topic-linked rather than time synchronized). Three transcripts are available, one for the radio communication and one each for the (intermittently operated) tape recorders aboard the two Apollo spacecraft—the Command Module (CM) and the Lunar Module (LM). AEX is a Java application.

The original design goal of AEX was to provide a multi-perspective immersive experience that would give users a richer experience than any single source could provide in isolation. In the initial AEX design, one audio channel was available at a time. Initially this was radio communication with the astronauts (with PAO commentary), although we also have experimented with instead presenting audio recorded aboard the CM (which is available for parts of several missions). An ability to integrate audio from additional sources offers the potential for constructing different immersive experiences (e.g., from the perspective of an individual flight controller), but it also offers the potential to construct perspectives that no participant at the time could actually have experienced. For example, we might hear the astronauts talking among themselves in one ear, while we hear discussions on Earth about the same topic in the other.

While there's no practical way that users could hope to listen to dozens of channels at once, it is possible to play more than one channel. As a first step, we can use stereo replay to play different channels in each ear. As a second step, we can (as the flight controllers did) allow the user to make some channels loud and other channels softer. This capability is now implemented in the current AEX release. Presenting a mission reconstruction in which users could potentially access dozens of audio channels poses several new challenges, however.

One challenge revealed by our initial multi-channel audio implementation is that as the number of channels grows it becomes more difficult to clearly indicate to the user what they should expect to find on each channel. Another challenge is that when different channels (in different ears) contain some of the same audio, small timing differences in the replay can create a unpleasant "echo" effect. For demonstrations we can preselect channels that have no shared content, but for unrestricted mission replay we will need some way for users to see what's on each channel before making selections and to manage their selections in ways that minimize content overlap. Our initial design for this (using the spatial layout of the MCC as a visualization for the available headset audio channels) is useful as a starting point, but we will also need good ways of showing what's being

---

[1] AEX can be obtained from http://www.umiacs.umd.edu/∼oard/aex

listened to on each of the available headset channel and we will need some other way of indicating the availability of separately recorded loops.

Our initial work with multi-channel audio has focused on mission events such as launch from Earth and the lunar landing during which there is a lot of activity, and selecting almost any channel during such times will result in some audio content. But as we integrate multi-channel audio from less busy mission periods (e.g., the crew sleep periods, which lasted several hours), we will surely find that many channels the user might choose will have long periods of silence. We therefore also need some form of visualization to indicate to the user which channels will have activity in the near future. This is easily done using speech activity detection, although we do not yet have that capability implemented.

Indicating who is talking, and when they are talking, may suffice for less active mission phases, but during particularly intense periods (e.g., the lunar landing) users may need assistance in navigating among the available cacophony. One way in which me might seek to facilitate that navigation would be to visually indicate to users which of the available channels have activity that is related to what they are listening to now. Apollo flight controllers were skilled in selecting which channels to listen to, but casual users of the AEX will likely require more support from the system to perform that task well.

A challenge that we had not anticipated is the need to help users think differently about channels that record loops (which the users can then combine as they wish) and those that contain headset audio (for which the selections might change as the flight controller being recorded selected and deselected specific loops for their headset). We may be able to use automated speaker identification to recognize which channels are available in a headset, and in the longer term we may be able to perform channel separation to isolate the content of specific loops that had been mixed on a single recorded channel. We might also use content-based alignment techniques to perform precise time alignment to at least suppress the annoying "echo" phenomenon that now arises when the same content is played from different channels. As we work through these alternatives, we will need to give thought to how best to indicate to the user the capabilities and limitations of each selection that they might make.

5   Conclusion

One major result of our project will be the creation of a newly digitized collection containing tens of thousands of hours of recorded audio that will potentially be of interest to people as diverse as historians, speech processing researchers, scholars who study decision making under stress, and educators. Another anticipated contribution will be the use of portions of that collection in the Apollo Archive Explorer. Although reconstruction of the Apollo missions is an interesting challenge in its own right, it is not hard to imagine other applications of similar techniques. The Apollo program was not unique in creating a centralized coordination activity that supported time-critical decision making—similar things happen every day in control centers for the electric grid, cell phone net-

works, stock markets, the Internet, newsrooms, police forces, and many other kinds of physical and social infrastructure. If we are able to go beyond creating immersive experiences and support productive analysis of multi-channel audio, then recording that audio might become more common.

Another potential application of similar techniques is to the product of what is colloquially referred to as "lifelogging," where people seek to capture information about events in their life, sometimes from multiple perspectives [1]. We presently think of lifelogging as an egocentric activity, but of course the lifelog of a family, or of a work group, would raise many of the same issues that we see in the archival Apollo materials. Indeed, we can think of the remarkable records available from the Apollo program as a sort of prehistoric lifelogging. Learning to reconstruct events from multiple perspectives may thus help to shape how we think about lifelogging in the future.

Many people know a little about what happened during Apollo, and some people know a lot. But no human alive at the time or since has ever heard every word that was recorded in the Mission Control Center. It therefore seems reasonable to expect that capabilities of the type we are developing in the AEX will ultimately make it possible for historians to gain new perspectives, for engineers designing systems for a return to the Moon to analyze the Apollo experience in new ways, and for schoolchildren around the world to imagine themselves in that room at that moment.

Acknowledgments

References

1. Gurrin, C., Smeaton, A.F., Doherty, A.R.: LifeLogging: Personal big data. Foundations and Trends in Information Retrieval 8(1), 1–107 (2014)
2. Kranz, G.: Failure is not an Option: Mission Control from Mercury to Apollo and Beyond. Simon and Schuster (2009)
3. NASA: MCC Operational Configuration: Mission J1 (Apollo 15). NASA (1971), http://klabs.org/history/history_docs/jsc_t/mcc_operational_configuration_as15.pdf
4. NASA: Apollo Program Summary Report. NASA Johnson Space Center (1975), http://history.nasa.gov/alsj/APSR-JSC-09423.pdf
5. Oard, D.W., Malionek, J.: The Apollo archive explorer. In: Joint Conference on Digital Libraries. pp. 453–454 (2013)
6. Oard, D.W., Sangwan, A., Hansen, J.H.: Reconstruction of Apollo Mission Control Center activity. In: SIGIR Workshop on Exploration, Navigation and Retrieval of Information in Cultural Heritage. pp. 1–4 (2013)
7. Swanson, G.: We have liftoff!: The story behind the Mercury, Gemini and Apollo air to ground transmissions. Spaceflight 43(2), 74–80 (2001)
8. Ziaei, A., Kaushik, L., Sangwan, A., Hansen, J.H., Oard, D.W.: Speech activity detection for NASA Apollo space missions. In: Interspeech. pp. 1544–1548 (2014)