It's About Time: Projecting Temporal Metadata for Historically Significant Recordings

Douglas W. Oard[1,2], Min Wu[2,3], Kari Kraus[1], Adi Hajj-Ahmad[3], Hui Su[3] and Ravi Garg[3]
[1] iSchool, University of Maryland, College Park
[2] Institute for Advanced Computer Studies, University of Maryland, College Park
[3] Electrical and Computer Engineering, University of Maryland, College Park

Abstract
Twentieth century audio recordings and motion pictures are important sources, both for scholarly analysis and for public history. In some cases, important metadata has not reached the collecting institutions along with the materials, which are now in need of richer description. This paper describes a novel technique for determining the date and time on which a recording was made based on analysis of incidentally captured traces of small variations in the electric power supply at the time the recording was made.
Keywords: temporal metadata, electric network frequency, audio recordings, motion pictures
Citation: Oard, D. W., Wu, M., Kraus, K., Hajj-Ahmad, A., Su, H., & Garg, R. (2014). It's About Time: Projecting Temporal Metadata for Historically Significant Recordings. In *iConference 2014 Proceedings* (p. 635–642). doi:10.9776/14302

Acknowledgements: The authors would like to thank the ADVANCE Program for Inclusive Excellence at the University of Maryland, which has provided generous support through an Interdisciplinary and Engaged Research seed grant.
Contact: oard@umd.edu, minwu@umiacs.umd.edu, kkraus@umd.edu, adiha89@gmail.com, hsu.eceumd@gmail.com, ravig@umd.edu

1 Introduction

The twentieth century produced two remarkable innovations in information technology: (1) the widespread creation of digital content, and (2) the ability to record sound and motion pictures in very substantial quantities. In our time, these technologies have since converged, but they initially grew up separately. As a result, many of the twentieth century recordings that now form an important part of our cultural heritage lack important metadata (Bamberger and Brylawski, 2010). In this paper, we focus on recovering one specific type of metadata: the date and time at which a recording was made. We do this by leveraging a third great innovation of the twentieth century: electrical power.

Audio engineers have long known that without proper isolation and filtering, electrical noise can cause an annoying low frequency hum in a recording that is audible to the human ear. The reason for this is that the electrical network now almost universally uses alternating current with a frequency of 60 Hz (in most of North America) or 50 Hz (in much of the rest of the world). This signal typically enters the recording device by induction, a process by which an electrical variation in one component propagates to another component using electromagnetic coupling. It is difficult to eliminate this coupling completely, so audio engineers typically seek to design equipment in a way that will reduce the resulting hum to an inaudible level.

This undesirable hum actually turns out to be useful, however. In the last decade of the twentieth century, we learned that this hum varies in detectable ways, and it does so in a pattern that rarely repeats for a sufficiently long recording (Grigoras, 2007). The cause of that variation is the somewhat random process by which generating capacity and electrical loads are added to and removed from the electrical network as demand for electricity changes over time. These activities result in small but detectable fluctuations in the Electric Network Frequency (ENF). Because these variations propagate through the electrical network extremely quickly, the variations follow very similar patterns everywhere in the network. We can use this phenomenon in two ways. First, we can tell when a recording was made by comparing the

ENF signal in the recording with intentionally recorded ENF traces from the world's major electrical networks, which have been collected for this purpose for the past decade or so. Second, we can detect whether a recording has been edited because insertions and deletions will result in sudden changes in the ENF that do not match the reference ENF trace to which the contiguous segments align.

For earlier times, where we lack ENF references, we can at least determine whether two recordings were made at the same time by comparing their ENF traces. This will only work if they were recorded in the same electrical network, but these networks typically have a very large geographic extent. For example, most of North America is covered by one of four electrical networks, one west of the Rockies, one for Texas, one for Quebec, and one for the remainder of North America east of the Rockies. That idea leads to a novel approach for recovering reference ENF traces for earlier times. If we could find a set of recordings that had been made at known times in the past from which we can recover an ENF trace, then we can use those recordings to establish a set of reference traces.

Interestingly, ENF traces can also be detected in some motion picture recordings, using techniques that are based on detecting the imperceptible (to the eye) flicker produced by indoor lighting (Garg, Varna, and Wu, 2011; Garg, Varna, Hajj-Ahmad, and Wu, 2013). We therefore need not limit our attention solely to audio recordings. We do, however, need to work with relatively long recordings, since the random variations in the ENF do (randomly) repeat over short times. It has been shown that about 10 minutes worth of recording of average quality suffices to obtain a good match (Huijbregtse and Geradts, 2009), and with an hour of recording, errors in determining whether two recordings were made at the same time become very low.

The remainder of this paper is organized as follows. In section 2, we identify some sources of historically significant audio, some of which were recorded at known times and others of which were not. Section 3 then presents some initial experiment results with a few of these sources. Section 4 concludes the paper with some observations about the implications of this work for archival practice.

2 Twentieth Century Recordings

In attempting to reverse engineer reference ENF traces using surviving audiovisual materials, we must contend with a patchwork landscape of recordings that cannot match the degree of coverage provided by more recent forensic databases. This problem is compounded the further back in time we go; the rich repositories of sound available to us from the 1970s, for example, contrast with the comparative scarcity of sound from just a decade earlier, in the 1960s. In order to compensate for these disadvantages, we have started to identify promising sources of audiovisual materials that are of sufficient size and scope to begin to lay down a twentieth-century ENF reference resource. To be useful as a reference, these materials must be accompanied by trusted metadata. Good candidates include live radio and television broadcasts--such as sports, news, and entertainment (think Johnny Carson or Saturday Night Live)—that have either been preserved within the organization that originally produced them or acquired by a collecting institution, such as the Library of Congress. In the case of historic radio broadcasts, it is often private collectors rather than large institutions that have endeavored to save them. NASA space missions and college radio broadcasts are also possible sources for reference ENF traces.

To illustrate some of the challenges and opportunities inherent in this work, consider the Vanderbilt Television News Archive, founded in 1968, which contains over 40,000 hours of news footage, including the nightly news broadcasts of ABC, CBS, and NBC from 5 August 1968 to the present (1968-2013). Spanning nearly five decades, the collection is thus a tremendously rich source of audio and video, with trusted date and time stamps that exist on a sufficiently large scale to be of value as a reference. These strengths notwithstanding, this archive also has some potential drawbacks. One relates to provenance: although the weekday shows were usually recorded from local network affiliates at the time they aired in Nashville, Tennessee, there are some cases in which these were supplemented by recordings captured in other locations, not all of which are on the same electrical network (Lynch 2013). Since this information is not necessarily available in the archive's online catalog, it adds an element of uncertainty that must be factored into the ENF analysis. Other collections offer complementary coverage, including a large collection of analog tape of National Public Radio (NPR) programs dating back to the 1970s (Ottalini, 2013). Because the NPR audio has not yet been fully digitized, it may be possible to instrument the digitization process. As explained below, that ability can be useful in some cases.

3 Initial Experiments

Analog recordings must be digitized before we can extract ENF traces. As a result, ENF traces may be present both from the original analog recording process and from the later digitization process. The spectrogram of a digitized audio recording from President Kennedy's White House conversations (Miller Center) is shown in Figure 1. The original recording was made in 1962 using an analog tape recorder, and digitized later. We observe two different ENF signals near 240 Hz,[1] one of which disappears well before the end of the digitized audio file. Listening to the audio, we note that the original recording was turned off at this time, leaving no recording on the remainder of the tape. We can therefore reasonably conjecture that the 239 Hz signal is the ENF trace due to the original recording, and the 240 Hz signal is the ENF trace due to the digitization process. The ENF trace from the original recording would be expected to deviate from its nominal value (240 Hz) if the tape were playing somewhat more slowly than it should have during digitization. Although undesirable for the fidelity of the audio content, this small error works to our benefit for ENF analysis. Indeed, since the error can be corrected during digital replay, we might actually prefer to use a slightly off-speed analog replay device when seeing to capture ENF traces.

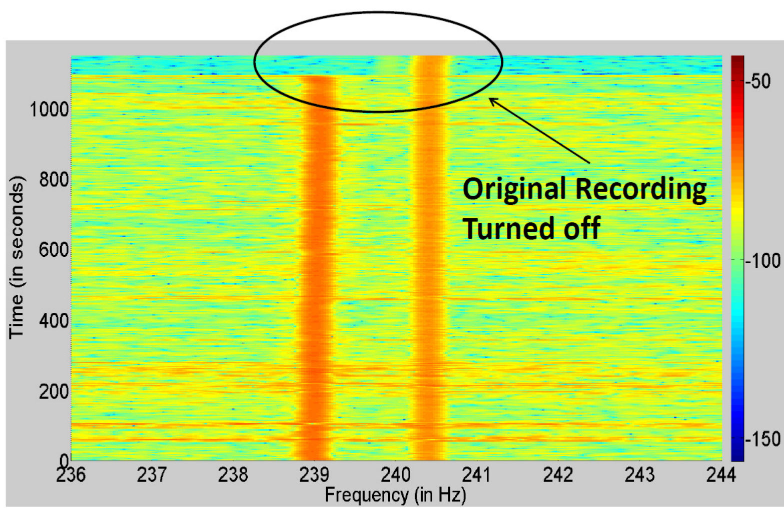


Figure 1: Spectrogram of a digitized Kennedy recording.

Of course, we are not always so lucky. To explore the worst possible case, completely overlapping ENF traces, we first made a recording in an acoustic chamber, and then we later played it back a speaker and re-record it in the same chamber. From the spectrogram of the recaptured audio shown in the bottom half of Figure 2, we observe that the two resulting ENF traces overlap completely, creating a single rather

[1] ENF traces in audio recordings made in the United States generally have nominal value of 60 Hz, and higher *harmonics* with the same patterns are found at small integer multiple of that value. Which harmonic is most useful as an ENF trace depends on both the characteristics of the recording equipment and the sound being recorded. In this case, we are seeing the fourth harmonic, near 240 Hz.

confused pattern.[2] At the top of the figure (the time after the original audio was switched off) the signal becomes clean again, as only a single ENF trace is now being captured. We have elsewhere shown that if a reference ENF trace is available from the time of the recording, we can essentially subtract that reference from the combined ENF traces, recovering a usable ENF trace from the original recording (Su, Garg, Hajj-Ahmad, and Wu, 2013).

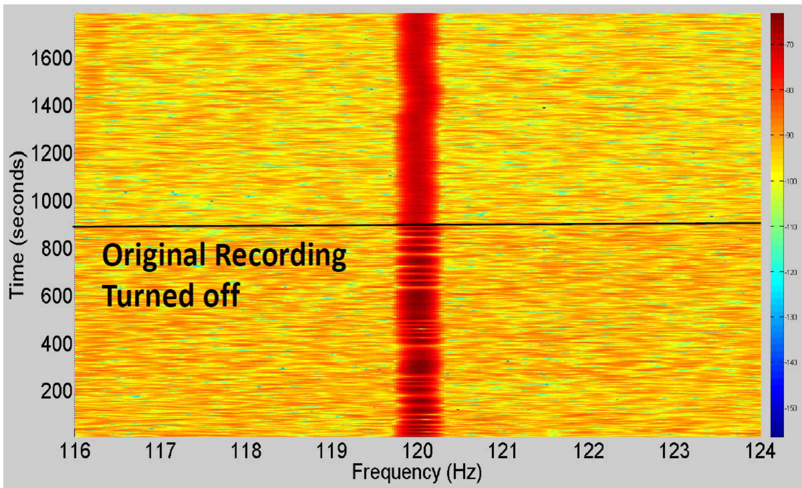


Figure 2: Spectrogram of a re-recorded audio signal with overlapping ENF traces.

Figure 3 shows an expanded and cleaned view of the ENF traces from two historical recordings that we know were recorded at the same time and in the same electrical network (Houston Audio Control Room). One ENF trace is from the landing of the Apollo 11 spacecraft on the Moon, as released to the broadcast media by the Public Affairs Officer (PAO) in Houston. The other ENF trace is from the intercom loop used by the Flight Director in the Mission Operations Control Room to coordinate the activities of the flight controllers during that landing. Because the voices of the astronauts appear in both recordings, we can be sure of the accuracy of our time alignment, which we performed manually. ENF traces from different sources might be stronger or weaker; for plotting purposes we have normalized the amplitude scale from each source to facilitate visual comparison. Despite some minor deviations (which result from recorded sounds that at some times happen to be at a frequency near that of the ENF trace), the ENF signals extracted from the two recordings present very similar variation patterns, sufficient to yield a high correlation coefficient (around 0.7) and thus a high confidence automatic match. This example also illustrates that ENF traces might also prove useful when seeking to align recordings made in different places at the same time, as might be the case, for example, in time-synchronized reconstruction of complex events.

[2] The effect is most easily seen when this figure is viewed in color.

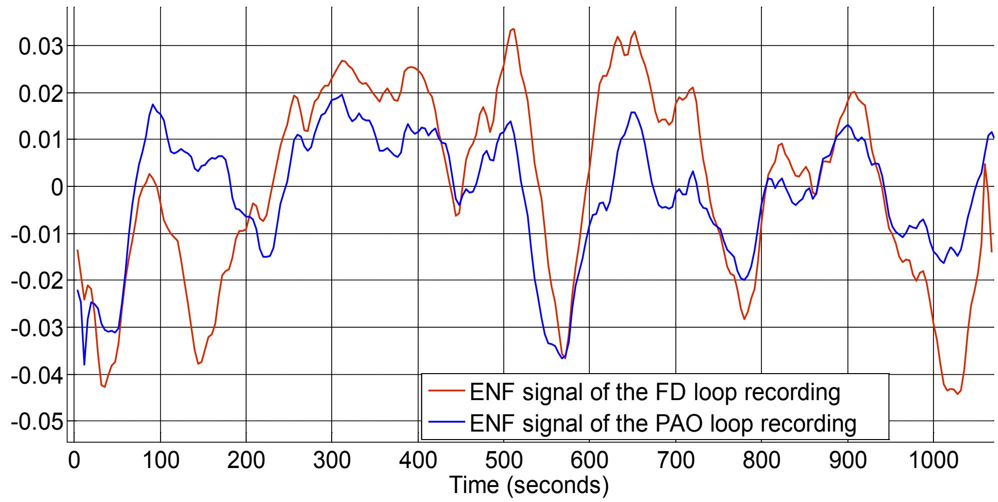


Figure 3: ENF traces extracted from two concurrent Apollo 11 audio recordings.

The cumulative coverage of the early space missions in the 1960s and 1970s amounts to about a year of audio in total, with continuous coverage of between a few days and a few months for any one mission, mostly recorded in the Texas electrical network. For more nearly continuous coverage, broadcasting provides one possible source. We therefore have also experimented with evening news programs from two different networks that had been broadcast at the same time. For this purpose, we chose the 5:30 PM news programs from CBS and NBC that were broadcast on August 9, 1974, the day Richard Nixon resigned and Gerald Ford became President of the United States.

As Figure 4 shows, each recording has significant energy around 60 Hz and its harmonics. No match between the ENF traces could be found, however. Figure 5 clearly illustrates the reason: the live broadcast from news announcers accounts for only a small part of each program, and those parts did not often occur at the same times. Other parts of the programs include field reports (which might have been pre-recorded) and commercials (which were usually pre-recorded). With some effort we might be able to reconstruct some longer alignments from all of these short segments, but we now believe that our time would better be spent with alternative sources for long-period live broadcasting, including for example sports events and CSPAN.

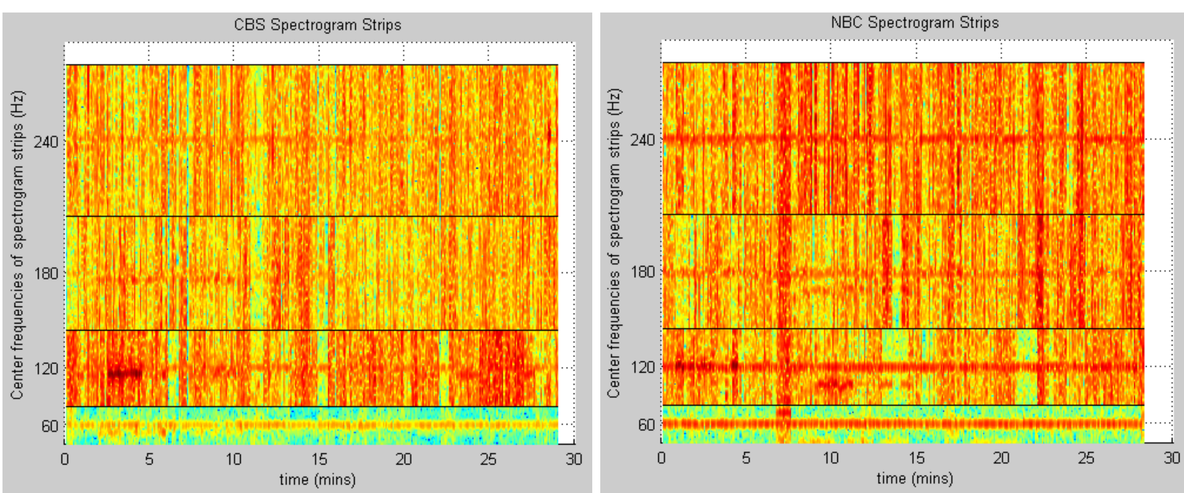


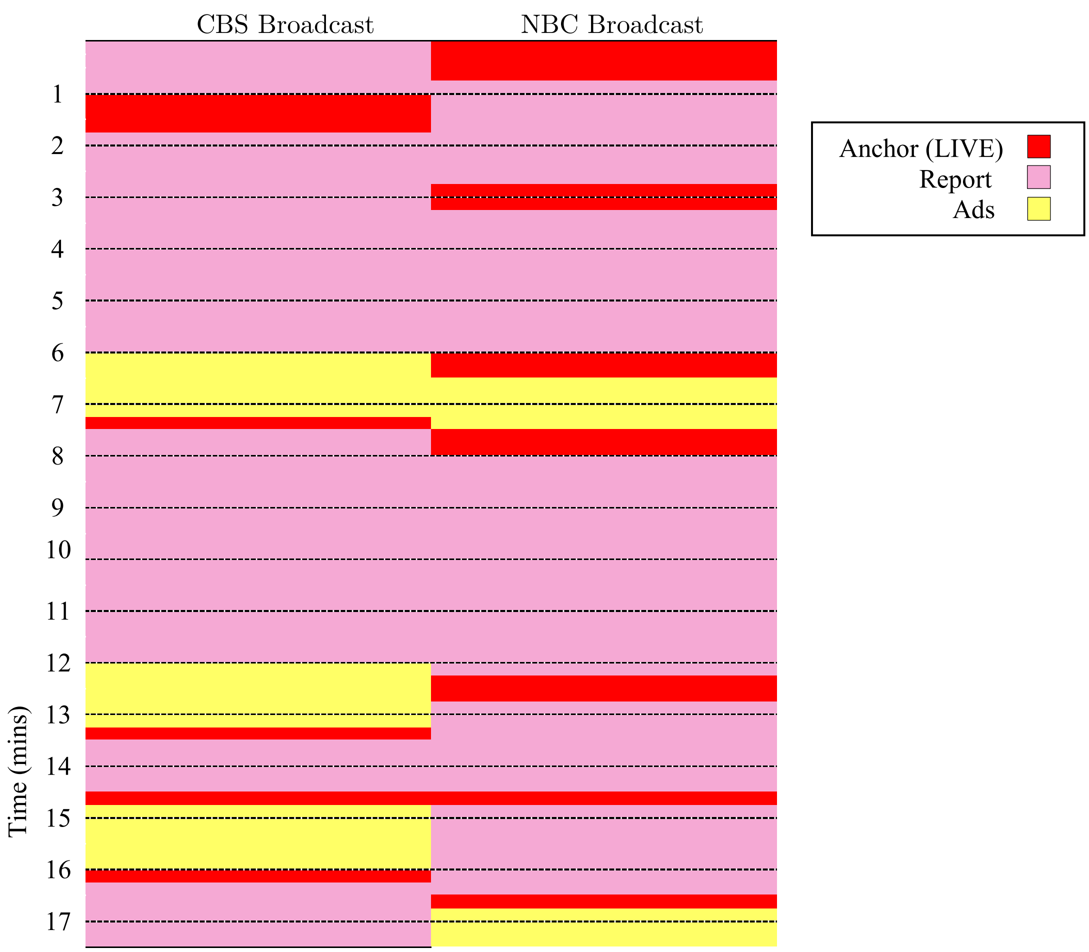
Figure 4. Spectrograms for the August 9, 1974 CBS and NBC evening news recordings.

4 Conclusions and Future Work

We have shown that recovering ENF traces from mid-twentieth century audio recordings is indeed both possible and practical. Moreover, our survey of twentieth century recordings indicates that substantial amounts of content that was recorded at known times are available for the second half of the century, with far spottier coverage before that.

Of course, much remains to be done. For one thing, research on recovery of ENF traces from motion pictures has not yet been tried with motion picture film or videotape, so the applicability of video-based techniques to the era before digital video remains to be determined. For another, a more complete set of reference ENF traces might be constructed by chaining together overlapping recordings, only the first (or last) of which has a known time. However, we do not yet know how common sufficiently long overlapping segments will be.

One important implication of our work for archival practice is that what cannot be heard or seen may nonetheless be important. Of course, this broad principle is well appreciated by archivists, and it is the driving force behind archival audio standards such as Broadcast Wave Format. It is surely useful to have an example such as this one at hand, however. A second implication for archival practice is that the temporal resolution of existing metadata is a limiting factor in our ability to project that timing data to new content. This may argue for reconsidering the temporal resolution of metadata for certain types of sources (e.g., live broadcasting) that can be particularly useful as a basis for establishing reference ENF traces. Of course, a third implication of ENF traces for archival practice is that they can provide an additional basis for assessing authenticity of recordings whose authenticity might otherwise be open to question.



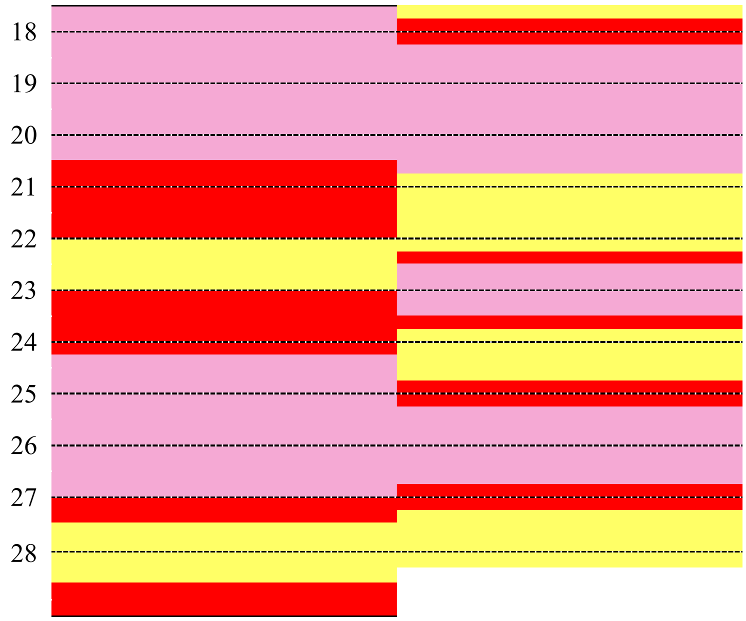


Figure 4: Approximate timeline showing news show breakdown by clip type.

5 References

Bamberger, R. and Brylawski, S. (2010). The state of recorded sound preservation in the United States. Commissioned by the Council on Library and Information Resources and the National Recording Preservation Board of the Library of Congress. Retrieved from http://www.loc.gov/rr/record/nrpb/pub148.pdf

Garg, R., Varna, A., and Wu, M. (2011 November). "Seeing" ENF: natural time stamp for digital video via optical sensing and signal processing. *ACM Conference of Multimedia*, Scottsdale, AZ. doi:10.1145/2072298.2072303

Garg, R., Varna, A., Hajj-Ahmad, A., and Wu, M. (2013 September). "Seeing" ENF: power signature based timestamp for digital multimedia via optical sensing and signal processing. *IEEE Transactions on Information Forensics and Security*, 8 (9) 1417-1432. doi: 10.1109/TIFS.2013.2272217

Grigoras, C. (2007). Applications of ENF criterion in forensics: audio, video, computer and telecommunication analysis. *Forensic Sciences International*, 167 (2-3) 136-145. doi: 10.1016/j.forsciint.2006.06.033

Houston Audio Control Room, Apollo 11 audio. Retrieved from http://archive.org/details/Apollo11Audio

Huijbregtse, M. and Geradts, Z. (2009). Using the ENF criterion for determining the time of recording of short digital audio recordings. *Proceedings of the Third International Workshop on Computational Forensics*, 116-124. doi: 10.1007/978-3-642-03521-0_11

Lynch, J. (2013 August 12). TV News Collection, email to Kari Kraus.

Miller Center, Kennedy presidential recordings, http://millercenter.org/scripps/archive/presidentialrecordings/kennedy.

Ottalini, D. (2013, May 1). NPR donates vast collection to the University of Maryland's National Public Broadcasting Archive, 1 May 2013. Retrieved from http://www.merrill.umd.edu/deadline/?p=1594

Su, H., Garg, R., Hajj-Ahmad, A., and Wu, M. (2013 May). ENF analysis on recaptured audio recordings. *Proceedings of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, Vancouver, Canada, 3018-3022. doi: 10.1109/ICASSP.2013.6638212

Vanderbilt Television News Archive, University of Vanderbilt, Nashville, Tennessee 1968-2013, http://tvnews.vanderbilt.edu

6 Table of Figures