aSeamless Scripting and Recording of Audio Narratives

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

ABSTRACT

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces - Graphical user interfaces

Author Keywords
Audio recording, scripting, transcript-based editing

INTRODUCTION

Audio narratives are a common form of communication used in voice-overs, podcasts, audio books and e-learning. Closest to everyday conversation, audio is a medium with relatively low barriers to entry. So, it is used by many laymen who are not professional producers or writers to publicize their stories.

A common workflow for creating audio narratives involve three main steps: writing a script, recording audio and editing audio. To create a compelling audio recording, producers usually go through several iterations of these steps.

Consider the case of recording a voice-over for a video. The narrator does an initial recording based on a prepared script. Afterwards, while placing it on top of the video, the producer may want to make some changes to parts of the narration e.g. to control the timing of a specific word, or to match changes made in the shots after the audio was recorded. While some of these edits can be done from the existing recording using an audio editing software, others require the narrator to edit the script and re-record the affected parts. Similarly, consider an audio recording of an online lecture. After the initial publication, the lecturer may want to re-record or add parts e.g. to keep examples up to date, or to address common questions that came up afterwards. These are but a few of many scenarios where producers go through several iterations back and forth between script writing, audio recording and audio editing.

Most existing audio editing systems provide functionalities necessary to support the latter two steps: recording and editing audio. However, the first step, writing the script, is usually overlooked or treated as a completely separate step. This is the case even when scripts play a key role in recording and editing speech.

In this paper, we present an interface that supports and links all three steps of the aforementioned workflow. Our system addresses challenges that span the process of creating audio narratives, including (1) associating script to audio recordings, (2) iterating back and forth between script and audio, and (3) combining multiple audio recordings. Our interface is inspired from familiar document editors and text merge tools, which are easy to learn.

RELATED WORK

Adobe Story [1], FinalDraft and Celtx are examples of software applications dedicated to script writing. They support collaboration, automatic formatting, navigation and planning for future production, but they treat the script as a text document that is essentially separate from the recordings. In fact, in our preliminary interview of amatuer and professional producers, we found that many of them use general-purpose document editors like Google Docs or Microsoft Word to prepare their scripts.

At the recording and editing stage, Adobe Audition, Avid ProTools, GarageBand and Audacity are among the most popular digital audio workstations (DAWs). These tools allow users to edit audio by manipulating waveforms in a multi-track timeline interface. They also provide a wide variety of low-level signal processing functions. However, since they are designed to serve as general-purpose audio production systems, they include many features that are not directly relevant for creating audio narratives whose main content is speech. Hindenburg Systems develops tools that are specifically targeted for audio narratives, doing away with much of the complexities. Still, all of these systems are primarily concerned only with the latter two steps of production–recording and editing–and they do not deal with the script directly.

Recently, several researchers have explored text-based navigation and editing of audio. Whittaker and Amento [5] demonstrate that users prefer editing voicemail through its transcript instead of its waveform. Inspired by similar intuition, Casares et al. [3] and Berthouzoz et al. [2] enable video navigation and editing through time-aligned transcripts. Rubin et al. [4] extend this approach to audio narratives, but they focus only on the third step (i.e. editing pre-recorded speech tracks). Our system links and supports all three steps of the production, from script writing to recording and editing, enabling a seamless workflow.

KEY IDEAS

We based our design and implementation on several key insights which we gained from a series of informal interviews with people who regularly created audio narratives.

Scripts play a key role in the recording process. All of the interviewees prepared written materials about their narrative before they started recording. The format and level-of-details of these scripts varied: it could be a list of bullet points outlining the main points to cover, or it could be a word-for-word transcription of what was going to be said. In all cases, during the recording phase, the interviewees kept their scripts within view and depended on them to construct their narrative.

The final audio narrative is created by mixing and matching multiple recordings. Most interviewees recorded multiple takes, and then edited them to produce the final recording. Many of them said that aligning the multiple takes and finding the exact place to seamlessly connect the different cuts were the most time consuming and tedious tasks.

Producers iterate between the planning, recording and editing steps. Even when a thorough word-for-word script has been prepared beforehand, the final audio narrative may not follow exactly. The speaker may add more details while recording or find a more natural way of saying the same sentence. Sometimes a major change has to be made to the script after some recording has happened. In collaborative scenarios, different persons may be working on the script and the audio iterating back and forth. One interviewee who recorded online lectures said he periodically changed and re-recorded parts of the lecture, for instance to include up-to-date examples.

USER INTERFACE

We first describe our interface through an example workflow, and then present the main algorithms used in our system in the 5 section.

Our system displays two types of documents: the master-script and transcripts. The *master-script*, representing the current audio narrative, is the document that users work on. Users start by writing on the master-script what they plan to record. At this stage, the master-script is like an ordinary word document or script. Once the users start recording, the text corresponding to each take appears in a separate *transcript* document tab.

If the *aligned-view* is turned on, the master-script and the currently selected transcript is segmented.

The user can click on a button to use that segment of the audio. Now that part appears in text color to indicate. If the user has multiple takes, they can go back and forth. If they go to the All Tab, they can see multiple takes of each sentence. Also edit the master script, just as they would edit a text document. If the user edits a recorded portion of the script, the relevant section is makred as dirty to indicate that it should be re-recorded.

ALGORITHMIC METHODS

RESULTS

CONCLUSIONS

ACKNOWLEDGMENTS

REFERENCES

1. Adobe Story. (????). `https://story.adobe.com/en-us/` Retrieved April 2, 2016.

2. Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. 2012. Tools for placing cuts and transitions in interview video. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 67.

3. Juan Casares, A Chris Long, Brad A Myers, Rishi Bhatnagar, Scott M Stevens, Laura Dabbish, Dan Yocum, and Albert Corbett. 2002. Simplifying video editing using metadata. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 157–166.

4. Steve Rubin, Floraine Berthouzoz, Gautham J Mysore, Wilmot Li, and Maneesh Agrawala. 2013. Content-based tools for editing audio stories. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 113–122.

5. Steve Whittaker and Brian Amento. 2004. Semantic speech editing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 527–534.