* <https://www.scitepress.org/papers/2018/65977/pdf/index.html>

***Why MIDI?***

The proposed model uses MIDI (Musical Instrument

Digital Interface) ﬁles to generate music. MIDI for-

mat is one of the best options present today to gen-

erate new music as the ﬁle contains only synthesizer

instructions hence the ﬁle size is hundreds of times

smaller than the WAV format (Waveform Audio For-

mat) which contains digitized sound and hence has a

very large size (hundreds of megabytes).

One advantage of WAV format over MIDI is that

the WAV format has a better quality of sound because

sound depends on the sampling rate and hence will be

same for different computers. This is not the case for

MIDI ﬁles, and hence the quality of sound is different

for different computers. Hence we have a trade-off

* <http://vis.berkeley.edu/papers/emotionscores/emotionscores.pdf>

The tradition of oral storytelling is thousands of years old and still endures today through digital recordings like audiobooks and podcasts. At audiobook sites like Audible [2] and LibriVox [13], these recordings number in the hundreds of thousands. In 2013, users downloaded over 500 million hours of audiobooks from Audible alone [2], nearly 15% of Americans listened to an audiobook [37], and 14% of Americans listened to podcasts every month [32]. Highly-produced audiobooks, podcasts, and children’s stories combine a speech recording, which vocally conveys the story, with a musical score that serves to emphasize and add nuance to the emotions of the speech. Yet, crafting such a musical score involves smoothly re-sequencing, looping and timing the music to match the emotions in the story as they change over the course of the narrative. This process requires significant audio production expertise and is challenging even for expert producers. As a result, most of the recorded stories available today focus on providing the speech track and do not contain a musical score.

* <http://vis.berkeley.edu/papers/audiostories/>

Audio stories are an engaging form of communication that combine speech and music into compelling narratives. Existing audio editing tools force story producers to manipulate speech and music tracks via tedious, low-level waveform editing. In contrast, we present a set of tools that analyze the audio content of the speech and music and thereby allow producers to work at much higher level.

* <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-237.pdf>

Audio stories—ranging from audio books to radio stories and podcasts—combine speech and music into compelling, engaging narratives. New technology is continually simplifying the process of capturing and distributing multimedia content.

* Automatically scoring audio and video stories is a longstanding problem in multimedia research. Foote et al. [9] automatically create music videos by editing video clips to match a piece of music. Their method finds suitable locations to cut video and then splices video to match points in music where audio features change. Our work instead edits the music to create a soundtrack tailored to the speech by matching emotions rather than low-level audio/video features