Summary of image to audio conversion from mechanical media presentation

Simon de Leon

McGill University, MUMT611 3/30/06

1.0 Introduction

This presentation is an overview of previous work in image to audio conversion from various mechanical media such as wax cylinders and 78 r.p.m. records. In its current state, image to audio conversion is much slower than realtime recording, computer intensive, and expensive. Motivation and potential for research in this field outweighs these cons and will also be outlined. At the time of this writing, there are three research groups working directly in this area (Lawrence Berkeley National Laboratory, McGill Music Technology, and the Fribourg University of Applied Sciences).

2.0 Saving the music

Millions of recordings of cultural and historic significance are gradually decaying in today's libraries. Not only do media such as wax cylinders and shellac records deteriorate upon playback, but they are also susceptible to dust, fungus, and lacquer fracturing. User mishandling and improperly calibrated tonearms can also cause permanent damage.

Digitizing the audio from the media via traditional playback methods is too time consuming and risky. Furthermore, there is no standardized equipment for creating archive quality digital recordings from such media. Recognizing the need to develop standards for preservation, the United States government established the National Recording Preservation Board (NRPB). Their objective is to formulate the selection criteria for recordings deemed worthy of archiving, as well as to develop standards of preservation and access. The NRPB is a major financial driving force in this field of research in the United States.

3.0 Vinyl physics and standards

In order to fully understand the pros and cons of image to audio conversion, the basic characteristics of mechanical recordings are presented.

3.1Cutting standards

The recording needle creating the grooves in the record can undulate with constant amplitude or constant velocity (Galo 1996). A constant amplitude cut will maintain equal maximum stylus displacement over all frequencies, and consequentially linear stylus velocity will increase with frequency. In contrast, a constant velocity cut will maintain equal linear stylus velocity over all frequencies and vary displacement amplitude inversely with frequency.

Constant velocity cut is the natural choice since the magnetic transducers in phonograph cartridges are velocity sensitive. However, the stylus displacement amplitude at low frequencies can be quite large and violate groove spacing restrictions. Furthermore, high frequencies result in small stylus displacements that can be lost in the noise floor. To combat this effect, several recording and playback equalization curves were developed. These equalization curves work by attenuating the bass and amplifying the treble upon recording, and then performing the opposite operations upon playback to achieve a flat overall frequency response. The

Recording Industry Association of America (RIAA) standardized a recording equalization curve of 6 dB/octave attenuation below 500 Hz and 6 dB/octave amplification above 2122 Hz (Galo 1996). It is important to note that optical systems will need to explicitly take these effects into account to achieve faithful audio reproduction.

3.2 Cutting orientation

Mono mechanical media typically have lateral stylus grooves (parallel to the surface of the record), although there are early mono records with vertical stylus grooves. Stereo recordings use both lateral and vertical techniques by storing the two channels in the undulations of the groove sidewalls. To avoid complications with mono recordings cut in stereo format, the left and right channels are recorded into the media out of phase.

3.3 Other vinyl anomalies

It is useful to consider other flaws in mechanical media playback and how they translate to optical playback. Transient impulse noise such as dust flakes and scratches are virtually unavoidable in mechanical playback. Optical playback offers the possibility of removing these transients at the source rather than post-filtering than audio. Sonic imperfections due to wow and flutter are virtually eliminated with optical methods since there are no realtime dependencies on motors. Finally, the wear zone areas of grooves can potentially be avoided with optical methods with high enough resolution.

4.0 Previous work in optical audio reconstruction

4.1 VisualAudio

VisualAudio works by taking a high resolution analog photograph of both sides of a 78 r.p.m. record and digitizing it with a rotating scanner (Stotzer et al 2004). Curve trajectories are extrapolated from the observed edges and the undulations from them used to generate audio.

Although this procedure is relatively simple and cost effective compared to white-light interferometry, VisualAudio suffers from several drawbacks when archive quality results are required. For example, the random granularity of analog film is known to add white noise to the recovered audio. Resolving jitter from the rotating scanner causes vibration and distortion in the digitally captured images, resulting in further quality decrease. From preliminary calculations, Stotzer et al. estimate a required edge noise standard deviation of less than 1.28 μm to achieve 40 dB SNR from RIAA recordings.

4.2 Lawrence Berkeley National Laboratory (LBNL) Optical Metrology

The LBNL has successfully demonstrated two-dimensional audio reconstruction from 78 r.p.m. records (Fadeyev and Haber 2003) and three-dimensional audio reconstruction from wax cylinders (Fadeyev et al. 2004). The two-dimensional data capture was done with a high resolution digital camera, while the three-dimensional data capture was performed with a confocal laser scanning probe. Audio examples demonstrate superior sound quality to that of traditional mechanical playback.

In comparison with VisualAudio, the LBNL systems are extremely computer intensive and resource hungry. The obvious tradeoffs with field of view, resolution, and magnification exist, with data requirements for a typical 78 r.p.m record approaching terabytes.

5.0 Issues

Optical methods share a number of common issues and are actively being explored today. Firstly, speed and automation need to be vastly improved before these methods can be used to efficiently archive the worlds' mechanical recordings. Secondly, image processing techniques need to be explored in the context of damage removal. Finally, there is currently no way of detecting the unworn section of the groove on a record.

6.0 Conclusion

In order to preserve our rapidly deteriorating musical and cultural heritage stored on mechanical recordings, research into optical capture has received much attention as of late. However, image to audio conversion from mechanical media is in its infancy and is still far from achieving an automated, faster-than-realtime system. It seems as though this is a very solvable issue in the foreseeable future since there does not appear to be any real theoretical problems. Many of the obstacles described in the previous section seem to be only technological, and will undoubtedly improve over time as more powerful computers and imaging systems become available.

References

Fadeyev, V., and C. Haber. 2003. Reconstruction of mechanically recorded sound by image processing. *LBNL Report* 51983.

Fadeyev, V., C. Haber, C. Maul, J. McBride, and M. Golden. 2004. Reconstruction of recorded sound from an edison cylinder using three-dimensional non-contact optical surface metrology. *LBNL Report* 54927.

Stotzer, S., O. Johnsen, F. Bapst, C. Sudan, and R. Ingold. 2004. Phonographic sound extraction using image and signal processing. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* 4: 289-92.

Galo, G. 1996. Disc recording equalization demystified. *Journal of the Association for Recorded Sound Collections* 27:2.