OPTICAL AUDIO RECONSTRUCTION FOR STEREO PHONOGRAPH RECORDS USING WHITE LIGHT INTERFEROMETRY

Beinan Li
Music Technology Area
Schulich School of Music
McGill University
beinan.li@mail.mcgill.ca

Jordan B. L. Smith
Music Technology Area
Schulich School of Music
McGill University
jordan.smith2@mail.mcgill.ca

Ichiro Fujinaga
Music Technology Area
Schulich School of Music
McGill University
ich@music.mcgill.ca

ABSTRACT

Our work focuses on optically reconstructing the stereo audio signal of a 33 rpm long-playing (LP) record using a white-light interferometry-based approach. Previously, a theoretical framework was presented, alongside the primitive reconstruction result from a few cycles of a stereo sinusoidal test signal. To reconstruct an audible duration of a longer stereo signal requires tackling new problems, such as disc warping, image alignment, and eliminating the effects of noise and broken grooves. This paper proposes solutions to these problems, and presents the complete workflow of our Optical Audio Reconstruction (OAR) system.

1. INTRODUCTION

OAR has proven to be an effective contactless approach to digitizing monophonic phonograph records [1] [2] [3] [4]. Furthermore, it is an available solution for restoring broken records. Li et al. previously presented a theoretical framework for optically reconstructing audio with a white-light interferometry (WLI) microscope and image processing [5]. A few cycles of stereo sinusoidal signal, extracted from a small number of images, illustrated that their approach is capable of extracting stereo signals from LPs. To reconstruct a few seconds of audio, however, the scanning region must be scaled up to a much larger disc area, resulting in thousands of images. A sophisticated image acquisition and post-capture processing workflow is thus desired to tackle the challenges that emerge from large-scale scanning: e.g., disc surface warping, image alignment errors, groove damages, and unwrapping the grooves into a one-dimensional audio signal.

In Section 2, we review previous OAR systems. Our system to acquire record groove images is introduced in Section 3, followed in Section 4 by our image processing procedures for extracting audio from the scanned images. The reconstructed result is illustrated and discussed in Section 5.

2. EXISTING OAR APPROACHES

In this section, four previous OAR approaches are described. Although they operate on recordings of different formats, most OAR frameworks follow the same high-level three-step procedure for reconstructing an audio recording: first, the grooves are scanned; second, the groove undulations are isolated and extracted; third, these undulations are converted into audio. Approaches vary significantly in terms of the hardware used, some using a general-purpose commercial product such as a confocal microscope, others using a custom installation. The hardware, in turn, affects how the grooves are scanned and thus how groove undulations must be extracted. By contrast, the audio conversion step (which may include post-processing, such as equalization) depends solely on the record production procedures that were used for the particular item being scanned. This step almost always includes filtering the signal to undo the RIAA equalization used in production and obtain the audio.

The systems developed by Iwai et al. and Nakamura et al. use a ray-tracing method to obtain the groove contour of a phonograph record [6] [7] [8]. The groove is illuminated with a laser beam, and the groove undulations are measured by detecting the angle at which the beam is reflected. In this way the laser functions as a simulated stylus—a replacement for the mechanical stylus—and can output an analog audio signal directly.

Unfortunately, since such systems must trace out the grooves, they are unable to handle broken records. In addition, two types of errors limit this approach: the errors caused by the finite laser beam width, which leads to echoes and high- and low-frequency noise in the extracted audio signals, and the tracking errors that may occur when the beam misses the groove entirely.

Fadeyev and Haber built an OAR system for 78-rpm records based on confocal laser scanning microscopy [1]. With the help of a low-mass probe, they built another one for wax cylinders [2]. Their system is capable of scanning the record in 3D with a vertical accuracy of around 4.0 microns. However, in their work on 78-rpm records only 2D imaging is emphasized, at a resolution of 0.26 x 0.29 microns per pixel. It takes their system 50 minutes to scan about 1 second of recorded audio, corresponding to 0.5–5 GB of image data. The groove bottom is obtained using 2D edge detection on the pixel illumination data, and the groove undulation is defined as the radial deviation of the groove bottom with respect to an unmodulated trajectory about the centre of the record.

Lengths of the groove are skipped when dust and debris occlude the image and no edges are detectable, but no solution to data restoration is provided; such skipping may thus cause a loss of data. Fadeyev and Haber compared their optically reconstructed audio sample, a turntable-digitized audio sample, and a remastered CD sample of the same recording. The quality of the reconstructed sample was judged to be better than the turntable-reconstructed version, but poorer than the CD version: while the OAR system produced fewer clips and pops than the turntable, and had a lower continuous noise level, it also contained background hissing and low-frequency modulation absent from the CD version.

Stotzer et al. created a system that performs a multi-step optical reconstruction for 78-rpm records [3]. First, a 2D analog camera is used to rapidly photograph the records and the images are preserved on film. This film is then scanned and digitized into 2D digital images, which are further analyzed to extract the audio. During the scanning, the film is placed on a rotating tray with an overhead stationary camera that carefully captures the groove images passing through its field of view (FOV). The rotation of the tray simulates the rotation of the record during playback, and effectively unwraps the groove segments to become uniformly oriented in the resulting images. Similar to Fadeyev and Haver's strategy, edge detection is then used on these images to extract the groove undulations, which are simply described by the edge that separates the groove valley from the space between grooves.

In this system, the imaging resolution is compromised by shading blur, motion blur, and sampling blur introduced by the illumination and the rotating scanner. The blur is estimated to be roughly 24.6 microns along the direction of rotation. The system also suffers from various acquisition artifacts, such as the very low-frequency noise caused by the off-axis placement of the film.

In an effort to restore damaged grooves, smoothing and corrupted-pixel-map-based enhancement are performed. The robustness of the damage detection is nevertheless questionable due to the simplifying assumption that damages such as scratches are solely perpendicular to the grooves. The blurring described above also makes it difficult to reliably detect scratches in the grooves.

The reconstructed sound quality achieved by their system was evaluated according to several standard audio engineering parameters; for example, the signal-to-noise ratio of the system was found to be roughly 16dB.

Tian used 3D scanning based on dual-focal microscopy for reconstructing audio from 78-rpm records. The ability to handle LPs is claimed, but not yet implemented [4]. Contrary to the aforementioned approaches, the groove undulation is defined by the groove sidewall orientation at each tangential increment relative to the disc center, instead of by the edges of either the top or bottom of the groove. Ray-tracing is used to create a 3D image of the entire record groove surfaces, including the sidewalls. The stylus movement across the grooves is represented by the optical flow derived from groove image intensity derivatives. The sidewall orientations are obtained by using dense depth maps and projecting complex surface onto the groove cross-section plane. The microscope in use has a lateral resolution of 1μm per pixel. Tian's optical flow approach requires 1390 x 36 images of the FOV 640 x 480 pixels to represent a two-second audio signal, although on which groove revolution the reconstruction is performed is not reported. It takes three days for four workstations to generate the image representation of a three-second audio. The image acquisition time is unclear in the literature. The equivalent audio sampling rate in their experiment is about 2404.71 Hz.

Although laser turntables can serve as a solution to optically retrieve audio from phonograph records, they require emulating the exact groove-following behavior of a turntable. We would like to find a general image-acquisition-based preservation solution without mimicking turntable behavior to derive digital audio directly from images. We also wish to obtain 3D information: although Fadeyev and Haber claimed their system can be adapted to a 3D groove profile, they did not implement it, while the system of Stotzer el al. does not retrieve 3D information. Tian's system is 3D-based, but his experiments are not performed on stereo LPs, the target considered here. Thus, in contrast to these works, our research focuses on optically reconstructing a digital stereo audio signal from LPs by extracting the lateral and vertical groove undulations from 3D groove information.

3. WLI-BASED IMAGE ACQUISITION

WLI is a powerful scanning technique based on physical optics, as opposed to the geometrical-optics-based approaches that include confocal microscopy and ray-tracing. In WLI, a broadband light source simulates an ensemble of narrow-band interferometers to make high-precision measurements. We used a Wyko NT8000 WLI microscope equipped with both Michelson and Mirau interferometers. Adjusting the vertical focus of these allows one to perform vertical scanning interferometry with a vertical resolution of better than 1 nm [9]. (The amplitudes of the groove depth are typically on the order of 25–100 μm [3].) In the current experiment, only the Michelson interferometer (10X magnification) is used, with a 0.644 x 0.483 mm (or 640 x 480 pixel) FOV. Using the 3X vertical scan speed with 20% overlap between fields of view, it takes roughly 27 minutes to scan one second of audio content. This configuration provides a reasonable tradeoff between acquisition quality and time cost.

Due to warping of the disc surface, the phonograph record is not perfectly flat. Expanding the scanning range to span all possible groove depths is unfeasible because of the greatly increased time cost. On the other hand, re-estimating for each FOV the range of groove depths to scan takes too long as well. To minimize the time cost without risking unfocused images, a hierarchical scanning scheme was chosen: the entire grid of FOVs is divided into sub-regions, and the scanning depth is adjusted only for each sub-region. Compared to the approach with a global range, this adaptive scheme reduces the scanning time by half.

4. AUDIO EXTRACTION BY IMAGE PROCESSING

Once the images have been acquired, the next step is to extract from them the groove undulations, which can then be converted into audio. A stereo phonograph audio signal is encoded in lateral and vertical groove undulations, so both must be extracted. To do so, first the entire grid of FOVs must be realigned using a dynamic programming algorithm. Next, the structures that define groove undulations are identified (and restored, where damaged). Finally, the groove is traced over the entire field and unwrapped, permitting the straightforward extraction of the groove undulations. The extracted undulations may then be decoded into stereo audio. This workflow is illustrated in Figure 1.

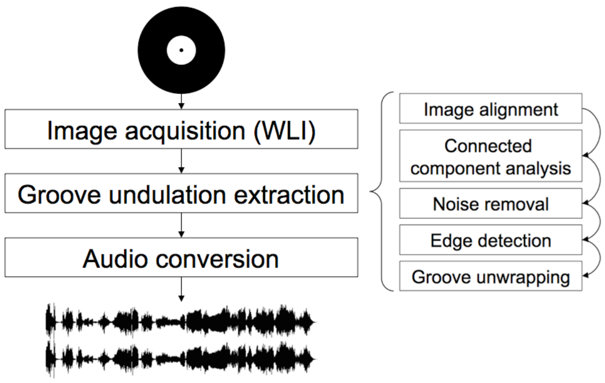

Figure 1. Diagram of the implemented system.

4.1. Image Alignment

As in most OAR systems, the FOVs are scanned with a degree of overlap. Unfortunately, mechanical translation during the image acquisition usually results in the images being misaligned. Before the information in the images can be extracted and combined, the FOVs must be realigned with each other.

This realignment can be achieved with an iterative frame-by-frame image registration approach, using either of two algorithms: a greedy one and a dynamic programming one [10]. In the greedy algorithm, the alignment of local FOV pairs is optimized, ignoring the precision of the overall grid alignment. It therefore can suffer from cumulative registration errors, resulting in unrecoverable gaps between the last few rows of the grid. By comparison, dynamic programming can be used to achieve a globally optimal alignment by forcing the elimination of any inter-row gaps as a constraint. This latter approach was selected.

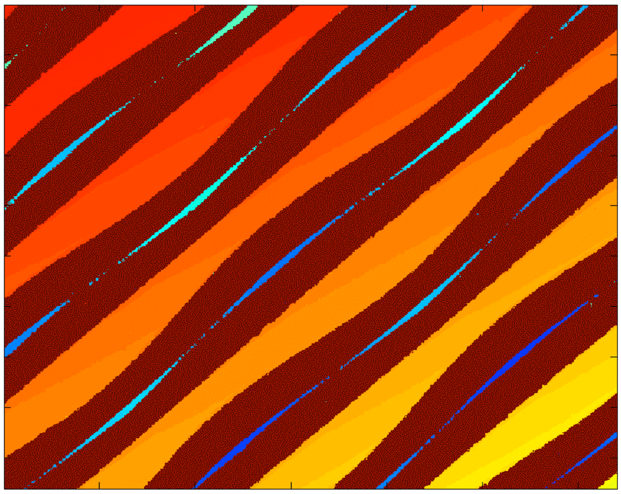

Figure 2. The top view of a typical groove image. The false colors represent vertical coordinates, except for the valleys, which are left monochromatic for clarity. The light color bands are the spaces between the grooves, and the light, thin lines are the valley bottoms.

4.2. The Image Model

Figure 2, a sample FOV, is a top view of a typical scanned region on a phonograph. The three parts of a groove can be seen readily: the tops (*T*), the valleys (*V*), and the bottoms (*B*). The goal is to extract the undulations of these grooves, but first the three parts have to be identified and separated. This is done using connected-component analysis (CCA). Similar to standard CCA in 2D image processing, a binarization process (thresholding) is performed as pre-processing to separate the global height levels of the *T*'s and *B*'s in the FOV; then with CCA, individual *T*'s, *V*'s and *B*'s are recognized.

The original 2D Cartesian coordinates are converted into polar coordinates (r, θ), with the origin being the disc center. We then distinguish three types of connected components (CCs): the top edges $T(r, \theta)$, the valleys $V(r, \theta)$, and the groove bottoms $B(r, \theta)$. Two useful properties are used in the identification of these regions: First, each region should be tangentially continuous, meaning that only one CC is supposed to be found on a single revolution of a *T*, *V*, or *B*. Second, the geometrical relationship between them is known: the *B*'s are completely contained by the *V*'s, while the T's and V's do not contain one another because they both stretch across the entire FOV.

4.3. Noise Removal and Groove Restoration

The raw CCs usually include noise from dust and dirt on the record and must be cleaned. Although heuristics based on the size of the CCs may seem intuitive, they turn out not to be robust in practice. Instead, we simply resort to the theoretical geometrical properties of the CCs, described in Section 4.2: since the tops and the valleys do not contain one another, any top found

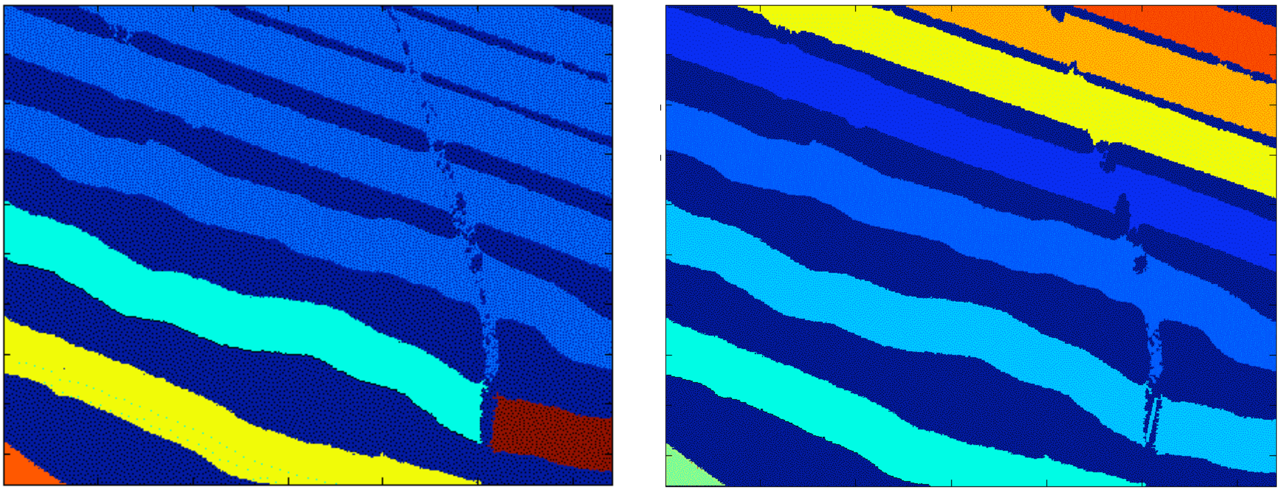

Figure 3. Groove restoration: CCA fails to identify the correct grooves (left) because the scratch connects certain grooves and disconnects others. These faults are corrected in the restored image (right).

contained by a valley (or vice versa) should be noise. This appeared to be a robust noise removal method.

Grooves may also be interrupted by scratches, or appear to be so due to occlusion by dust. Dust may also cause neighbouring grooves to appear attached to each other. These conditions create difficulties for extracting the groove undulations. To restore such grooves, two heuristics based on the tangential continuity property are used. First, by locating discontinuous $V(r, \theta)$ on the same revolution, broken grooves are detected; these may be restored by simply interpolating and reconnecting them. Similarly, attached grooves are detected and restored by locating and tangentially reconnecting discontinuous $T(r, \theta)$ that exist on the same revolution. Both situations are illustrated in Figure 3. Note that cleaning the records before scanning does not guarantee a better noise condition, because dust accumulates throughout the long time span of the scan. (Hosting the microscope in a dust-free environment may be one way to reduce this source of noise.)

4.4. Extracting Groove Undulations

The lateral undulations of the inner and outer groove top edges (with respect to the centre of the disc), and the vertical undulations of the groove bottom are the results of the recording stylus cutting into and across the disc surface. Following CCA and groove restoration, each FOV has been segregated into groove top, valley, and bottom regions. The next step is to trace the undulations, defined by the oscillations of both edges of the groove, as well as by the depth of the groove. Using a search-based edge detection algorithm, the inner and outer (with respect to the centre of the disc) groove top edges can be located; these are denoted as $T_i(r, \theta)$ and $T_o(r, \theta)$, respectively. The edges $T_i(r, \theta)$, $T_o(r, \theta)$, and the groove bottom $B(r, \theta)$ are iteratively overlapped and matched across adjacent, properly-aligned FOVs. In doing so, and in unwrapping the spiral-shaped grooves, the three 1D undulation sequences are obtained. The lateral undulations T_i and T_o are then averaged to obtain a single sequence T corresponding to the center of the groove. Note that phase unwrapping needs to be performed to obtain a continuous "time line" of the undulations, because the polar coordinates have the range of $[0, 2\pi)$. This is due to the fact that the derivation of polar coordinates is done to individual FOVs before their temporal topological order in the audio signal is clear.

4.5. Converting Groove Undulation to Audio

The raw, unwrapped groove undulations need to be resampled at an audio sampling rate to be converted to digital audio. Since the digital image format used here uses rectangular pixel tessellation, the pixel density along the undulation varies. It is therefore necessary to interpolate when sampling the image. We chose a reasonably high industry standard sampling rate (96kHz).

According to the constant velocity cutting scheme of LP production, the tangential velocities contain the audio undulation. This is achieved by performing numerical differentiation on the unwrapped undulations. The stereo audio is derived according to the following equations:

$$Channel_{left}(t) = \Delta T(t) - \Delta D(t)$$

$$Channel_{right}(t) = \Delta T(t) + \Delta D(t)$$

where $\Delta T(t)$ and $\Delta D(t)$ are the resampled and differentiated sequences of the center points and the depths of the groove, respectively. Finally, a counter-RIAA equalization filter is applied to the audio.

5. RESULTS AND DISCUSSIONS

In our experiment, the OAR system presented above was used to extract a roughly 1.8-second stereo audio signal. The result was compared to a turntable-digitized version of the same signal. Waveforms and frequency responses for both signals are displayed in Figures 4 and 5.

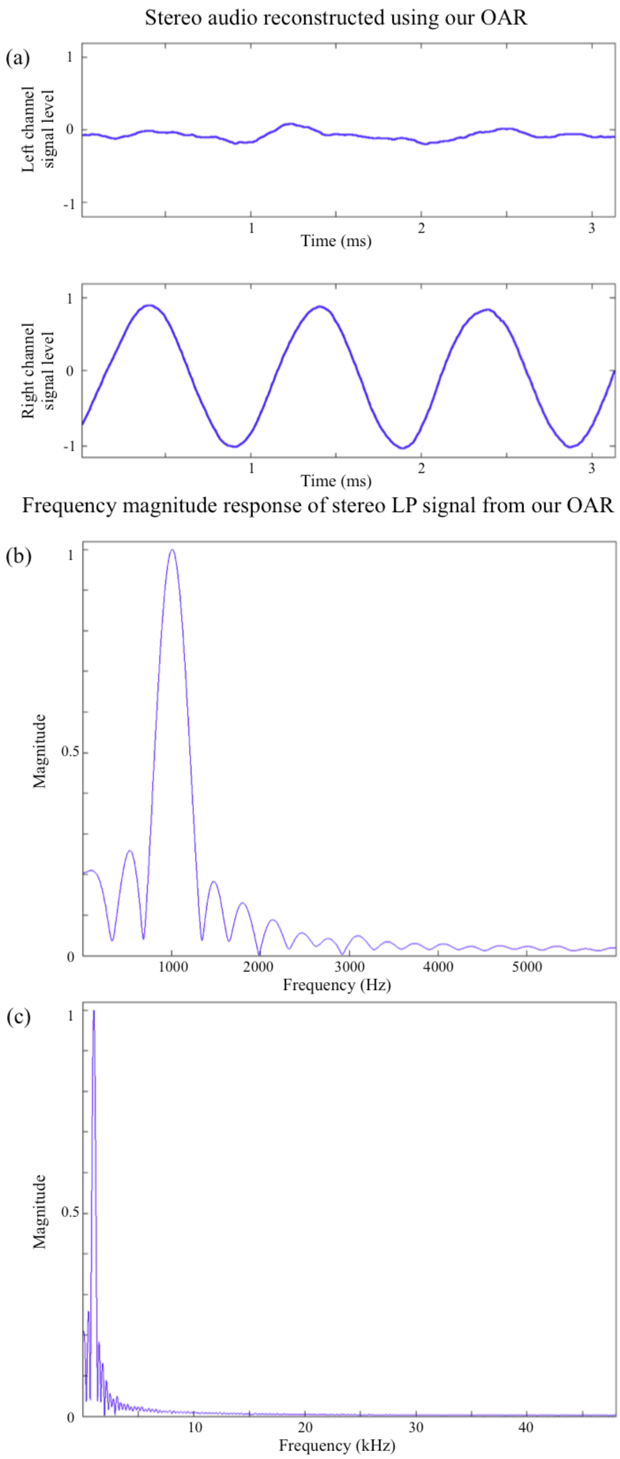

Figure 4. The reconstructed stereo signal: a 1kHz sine wave in the right channel and a silent left channel. A three millisecond segment of the waveform is shown in (a); in (b), the magnitude response of the extracted right channel signal up to 6kHz; in (c), the magnitude response of the same signal up to the Nyquist frequency (48kHz).

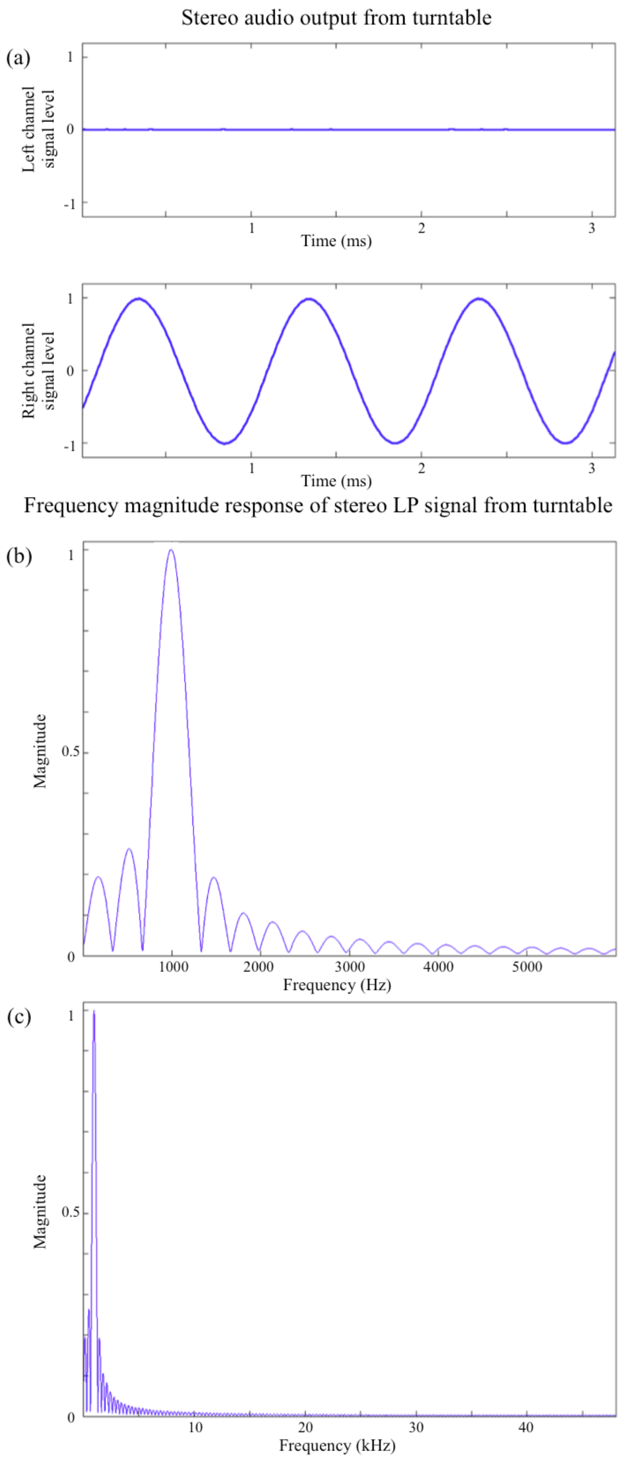

Figure 5. The turntable-digitized version of the same stereo signal.

It can be observed that the reconstructed version very much resembles the turntable-digitized version of the same stereo signal. As expected, the most salient component in the output stereo signals is 1kHz. However, the reconstructed left channel signal is not complete silence. In addition, close inspection of the peaks and troughs of the reconstructed waveform reveals a non-zero DC offset; moreover, the peaks appear to fluctuate slightly over time. This low-frequency wow noise, as in Fadeyev and Haber's system, is partly due to the error in estimating the disc center, which, as the origin of the polar coordinate system, forms the basis of the estimated lateral groove undulations.

6. CONCLUSIONS AND FUTURE WORK

OAR methods have been proven to be effective alternatives in digitizing mono phonograph records. Our WLI-based OAR system has successfully reconstructed digital stereo audio signals from LPs. Future work will be directed to improving the audio quality while decreasing the scanning time. Better center correction strategies will be studied, along with other configurations capable of pushing the audio quality higher. To push down the tremendous time costs, we will also investigate the minimum scanning resolution required to produce an acceptable audio result. On the other hand, the time costs may be reduced as the scanning hardware improves to provide, for instance, a larger field of view and faster vertical scanning speed.

7. ACKNOWLEDGEMENTS

We would like to thank the Canada Foundation for Innovation and the Daniel Langlois Foundation for their financial support.

8. REFERENCES

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

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

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

[4] Tian, B. *Reproduction of sound signal from gramophone records using 3D scene reconstruction*. Ph.D. thesis. Department of Computer Science, The University of Western Ontario, 2006.

[5] Li, B., S. de Leon, and I. Fujinaga. Alternative digitization approach for stereo phonograph records using optical audio reconstruction. *Proceedings of International Symposium of Music Information Retrieval* 165–6, Vienna, Austria, 2007.

[6] Iwai, T., T. Asakura, T. Ifubuke, and T. Kawashima. Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method. *Applied Optics* 25 (5): 597−604, 1986.

[7] Nakamura, T., T. Ushizaka, J. Uozumi, and A. Toshimitsu. Optical reproduction of sounds from old phonographic wax cylinders. *Proceedings of SPIE* 3190: 304−13, 1997.

[8] ELP Corporation. ELP Laser Turntable: plays vinyl records without a needle. http://www.elpj.com/main.html (accessed 10 August 2009).

[9] Veeco Metrology Group. *WYKO NT8000 setup and operation guide*. Tucson: Veeco Instruments, Inc., 2003.

[10] Chow, S., H. Hakozaki, D. Price, N. MacLean, T. Deerinck, J. Bouwer, M. Martone, S. Peltier, and M. Ellisman. Automated microscopy system for mosaic acquisition and processing. *Journal of Microscopy* 222 (2): 76–84, 2006.