

## **Reconstruction of Recorded Sound from an Edison Cylinder using Three-Dimensional Non-Contact Optical Surface Metrology**

Vitaliy Fadeyev and Carl Haber  
Lawrence Berkeley National Laboratory

Christian Maul  
TaiCaan Technology

John W. McBride  
Department of Engineering Sciences, University of Southampton

Mitchell Golden

The authors can be contacted at [CHHaber@LBL.GOV](mailto:CHHaber@LBL.GOV), [VAFadeyev@LBL.GOV](mailto:VAFadeyev@LBL.GOV), [C.Maul@taicaan.com](mailto:C.Maul@taicaan.com), [jwm@mech.soton.ac.uk](mailto:jwm@mech.soton.ac.uk), [mgolden@mitchgolden.com](mailto:mgolden@mitchgolden.com)

### **Abstract**

Audio information stored in the undulations of a groove in a mechanical sound carrier such as a cylinder or disc phonograph record may be reconstructed, without contact, by measuring the groove shape using precision optical metrology methods and digital image-processing. The viability of this approach was recently demonstrated on a 78-r.p.m. shellac disc using two-dimensional image acquisition and analysis methods. The present work reports the first three-dimensional reconstruction of recorded sound from a mechanical carrier. The source material, a celluloid cylinder, was scanned using color-coded confocal microscopy techniques and resulted in a faithful playback of the recorded information.

### **1. Introduction**

The preservation of recorded sound is of considerable current interest [1-3]. A recent report by two of the authors [4] demonstrated that certain methods of two-dimensional (2D) optical metrology and image analysis could be used to recover sound from mechanical carriers without making contact to the media. These methods also offer the prospect of sampling regions of the surface which have less stylus wear and of reconstructing damaged samples. In this paper the first such audio reconstruction using a full three-dimensional (3D) measurement of the record surface is presented. This work extends significantly the earlier results because a 3D measurement is required to reconstruct vertically modulated carriers such as a cylinder and also to sample the full groove cross section in a laterally cut record.

This paper is organized in the following way. The concepts and principles of the method are discussed in Section 2. Section 3 presents a description of the measurement

technique. Section 4 describes the analysis and signal processing. The results and prospects are discussed in Section 5 with final conclusions presented in Section 6.

## **2. Methodology**

The use of image processing to reconstruct sound, which has been recorded on mechanical carriers, has been described in some detail already [4]. The key points will be summarized here followed by a discussion of specific issues which concern 3D reconstruction.

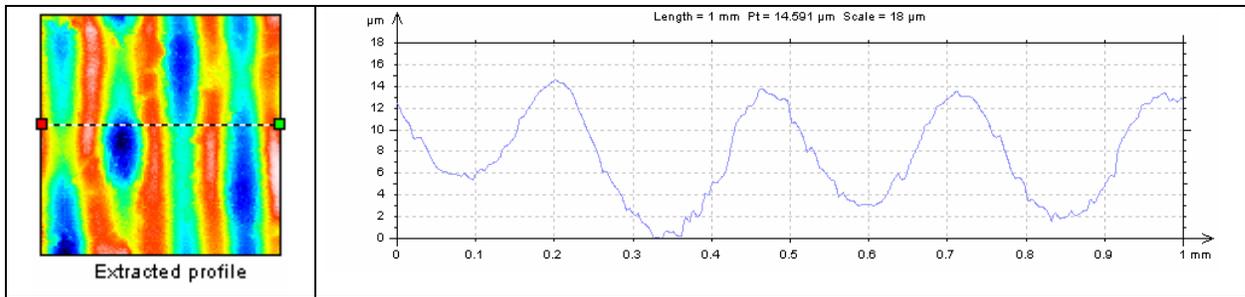
In the method, techniques of digital imaging are applied to construct a detailed map of the surface structure of a mechanical sound carrier. An example of a 3D image and a cross-section are given in Fig. 1. Once these images are acquired analysis techniques can be applied to both emulate the motion a stylus would have made while in contact with the surface and to recover damaged or worn regions. In this way, modern data acquisition and processing capabilities can be brought to bear on the recovery of sound from historical recordings. The aim here is not to (necessarily) provide a real-time playback but rather to extract the maximum information from the sample. The viability of this approach relies upon the use of imaging methods with sufficient resolution and accuracy to sense the minimum undulations of the surface due to recorded sound.

The test of these methods already presented utilized 2D electronic imaging to scan a region of a 78-r.p.m. shellac disc with lateral groove modulation. In that test, the groove bottom was readily imaged with an electronic camera and used for the sound reconstruction. To reconstruct audio from sound carriers with vertical modulation, such as a cylinder or an Edison "Diamond Disc" 3D imaging is required. This is also the case if the full groove profile in a laterally cut record is to be recovered.

Other workers have used optical methods to recover audio from mechanical carriers but these either replace the stylus with a reflected light spot [5-7], reflect light off a low mass mechanical contact [8,9], or are restricted to 2D imaging [10,11]. In the present work, full 3D metrological information is extracted from the surface structure. The general relevance of imaging to the study of cultural artifacts was also discussed by Stanke and Paul [11].

In the methods described here, the physical action of the groove upon a stylus is replaced by a mathematical algorithm executed on digitized surface metrology data. It is therefore important to consider both how the surface structure relates to the audio content, and the surface digitization process itself.

Sound has been recorded both electrically and acoustically on mechanical carriers. In the case of an electrical recorder, sound waves are converted to electrical impulses, amplified, and the groove is cut by an electromagnetic transducer. In the case of an acoustic recorder, sound waves directly drive the cutting stylus through a diaphragm. Some of the characteristics of mechanical sound carriers are summarized in Appendix A1.



Extracted profile from the altitude measurement

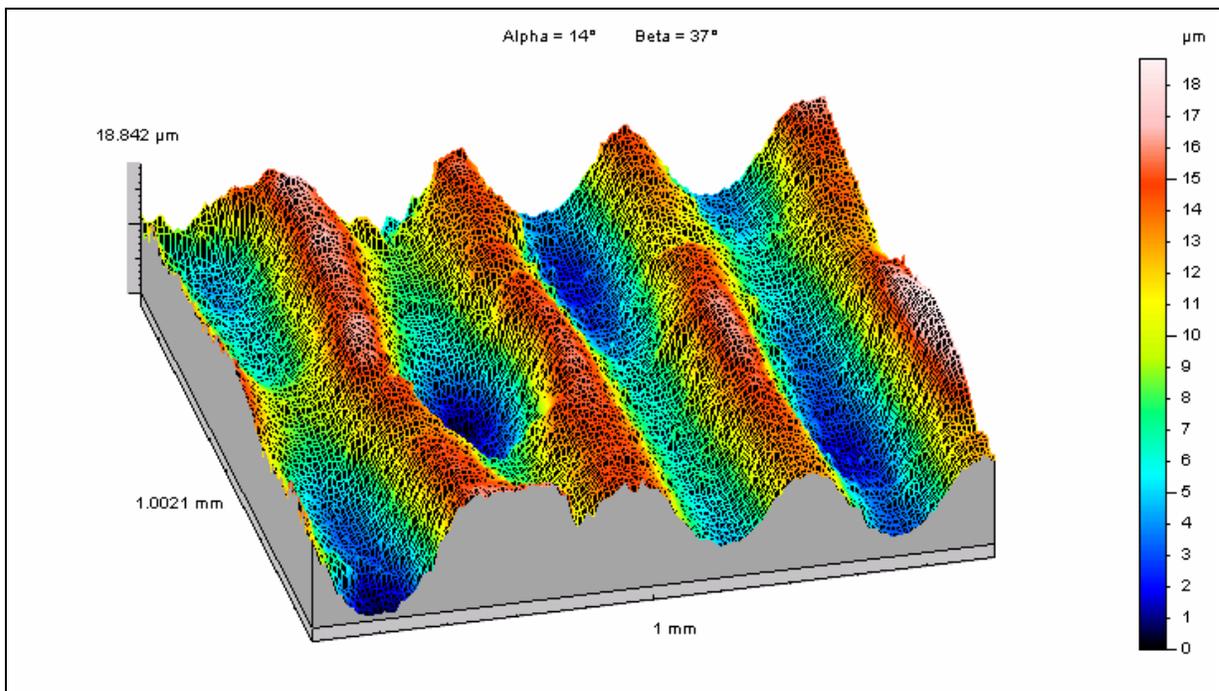


Figure 1: Lower plot shows a 3D surface map of a 1 x 1 mm<sup>2</sup> region of a wax cylinder. The upper plot shown a sectional view across the indicated line. This image was acquired with a CHR450 color-coded confocal microscope, manufactured by STIL SA, Aix en Provence, France. The data points were measured on a 6 μm grid, with a 300 μm depth of field probe, at a rate of 4000 points/s. Figure used by permission of STIL SA.

In an electrical recording, the action of the transducer is to transform sound amplitude into stylus velocity transverse to the groove direction. This proportionality is usually referred to as “constant velocity response”, and means that constant sound amplitude yields constant stylus velocity, independent of frequency.

Signals are compared on the basis of velocity amplitude rather than power.

$$dB_v = 20 \log \left( \frac{v}{v_{REF}} \right) \quad (1)$$

where  $v$  is the stylus velocity and  $v_{REF}$  is some defined reference level (specified in Table A1-1).

For a sinusoidal modulation, the signal amplitude will be maximal at the zero crossings of the groove, where the stylus velocity is largest. The maximum lateral displacement of the groove corresponding to a zero crossing velocity  $v_{MAX}$  is,

$$A_{MAX} = \frac{v_{MAX}}{2\pi f} \quad (2)$$

where  $f$  is the frequency of the recorded tone.

Eq. 2 states that for constant stylus velocity (sound amplitude) the maximum groove displacement depends inversely upon frequency. Because of this, in many electrical disc recordings the lower frequency sound levels are deliberately attenuated to increase the range of signals which will fit in the allocated track spacing. In addition, higher frequency sound levels are often boosted to overcome the drop off of Eq. 2 and to raise the signals above a high frequency noise floor. For recordings before the early 1950's this process of equalization was not standardized [12].

Because of the constant velocity response, the time derivative of the groove modulations, in an electrical recording, will be interpreted as the original sound, with additional modification due to any imposed equalization.

For acoustic recordings, which include all cylinders, the playback signal is approximately proportional to the stylus velocity, but the situation is more complicated than in the electrical case. This proportionality will be assumed in the present work. Deviations from it can be compensated for with different equalizations. The conditions leading to this approximation can be understood from the basic physics of sound propagation, the properties of acoustic horns, and the design of the mechanical sound transducers used in the acoustical recorders and players. The details of this discussion are in Appendix A2.

The term "constant velocity" has historically been associated with the era of electrical recordings, however it is meaningful for acoustic recordings as well. It merely expresses the condition that for fixed sound amplitude, stylus velocity is constant as a function of frequency.

From a digitized image of the mechanical carrier, the lateral or vertical displacement of the groove, with respect to the un-modulated trajectory, is measured on a sequence of points. The measurement of stylus velocity, at each point, is extracted from this displacement waveform by numerical differentiation. When the constant velocity condition applies, this numerical derivative can be interpreted as the originated sound. Additional low frequency boost can be applied digitally, as a matter of aesthetic judgment, lacking a definite model of the components used in recording. The resulting data may also be digitally filtered to remove frequencies typically outside the bandpass shown, for example, in Fig. A2-3.

Having established the physical characteristics of the mechanical carriers, the effects and byproducts of digitization may be considered next. A mechanical carrier holds an analog representation of the source audio waveform. All digital imaging methods acquire object data on a series of discrete points or within discrete pixels. The imaging of the groove pattern is equivalent to a digital time sampling in the spatial direction analogous to time and a digital pulse height sampling in the direction analogous to amplitude. The digital time sampling is determined by the dimensions of the pixel or spot and the spacing between sequential measurements. Let the pixel or spot size be  $W$  and the spacing between measured points be  $D$ . If  $W = D$ , the sampling is dense. If  $W > D$  the surface is over sampled in that adjacent samples overlap in the direction analogous to time. If the linear surface speed of the record is  $S$ , the sampling frequency is given by,

$$f_{SAMPLING} = \frac{S}{D} \quad (3)$$

In order to satisfy the Nyquist criteria, the digital sampling should be at a frequency at least twice the highest signal tone. In addition, aliasing of higher frequency noise must be avoided. Here there are two possibilities. The sampling could be done at a frequency sufficiently high that the medium is noise free, relative to the baseline noise in the audible range. Alternatively, a low-pass filter could be implemented in the primary image acquisition, below the Nyquist frequency,  $f_{SAMPLING}/2$ . With an imaging element of length  $W=2D$  in the time analog direction, groove wavelengths shorter than  $W$  will be attenuated according to the filtering function,

$$X(k) = \sqrt{\frac{2}{\pi}} \frac{\sin\left(\frac{kW}{2}\right)}{k} \quad (4)$$

Eq. 4 represents an imperfect low-pass filter for wave number  $k < 2\pi/W$  or  $\lambda > W$ . This corresponds to the frequency  $f_{SAMPLING}/2$  and therefore satisfies the Nyquist criterion (to the extent Eq. 4 can be considered a low-pass filter). This approach requires adjacent time analog samples (pixels or spots) to overlap by  $D$ . In this case it may be possible to sample at a lower overall rate at the expense of a more complex imaging and analysis procedure.

In the direction analogous to amplitude, the measurement system must have sufficient *resolution* and *accuracy* to capture the full range of the audio signal from the intrinsic noise baseline to the maximum groove displacement. Resolution is the smallest change in input signal that can be detected. Accuracy is the maximum deviation of the measured value from the true value.

To set the scale, Table A1-1, of Appendix A1, shows the known parameters of certain mechanical carriers, as determined from a number of sources [13-16]. Figs. 2 and 3 relate these parameters to the configuration of the groove on cylindrical and disc record respectively. When relevant they are defined at a specific frequency (1000 Hz) where equalization is generally not applied. Typically the required amplitude measurement resolution will be on the submicron scale.

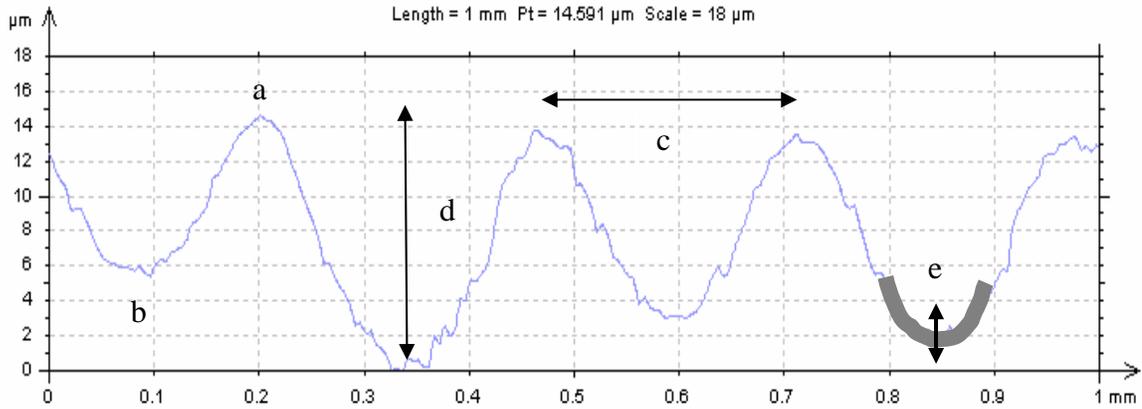


Figure 2: Cross section of Fig. 1 showing vertical grooves on the surface of a cylinder. Features are indicated for schematic purposes only: a) ridge between grooves, b) groove bottom or valley, c) groove spacing, d) maximum groove depth, and e) noise is represented by some variation envelope about the groove profile. Horizontal and vertical axes are not on the same scale.

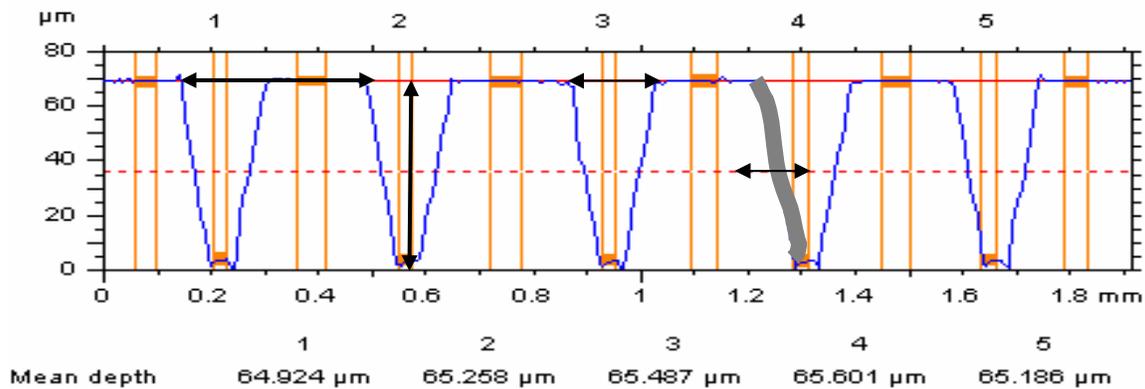


Figure 3: Cross section of lateral grooves on the surface of a disc record. The horizontal arrow between groove 1 and 2 indicates the groove to groove spacing. The vertical arrow in groove 2 indicates the groove depth. The arrow at groove 3 indicated the groove width at the top surface. The grey thick diagonal contour in groove 4 indicates some envelope of variation about the groove profile due to noise, signal, or wear. Note the significant difference between the groove depths here as compared to Fig. 3. Horizontal and vertical axes are not on the same scale. Figure (without arrows) is courtesy of STIL SA, used by permission.

Methods of high resolution 3D imaging have been developed to serve the need for precision inspection in such fields as semiconductor wafer processing, micro-machining, optics, and paper and fabric processing. Some of the key approaches are based upon either scanning confocal microscopy principles [17,18] or white light interferometry

[19,20]. The capabilities of these methods are similar but various trade-offs exist due to speed, cost, and flexibility of measurement.

The basic principle of scanning confocal microscopy [17] is shown in Fig. 4. A light source is defined by a pin-hole and brought into focus on the surface of a sample. The reflected light is then refocused on a second pin-hole in front of a photo-detector. The focusing optics are moved and the position corresponding to the maximal detected signal is recorded. As this apparatus is scanned over the surface of a sample the in-focus position of the optics is used to determine the surface topography. Commercial systems feature resolution of 10-100 nanometers, minimum spot sizes of  $\sim 1 \mu\text{m}$ , and data rates of  $\sim 1000 \text{ Hz}$ . Depending upon the surface quality it may be necessary to average a number of measurements at each point. A variant of this method is shown in Fig. 5. For color-coded scanning confocal microscopy [18] a polychromatic pin-hole source is used and the optics has an exaggerated chromatic aberration. Now each wavelength comes into focus at a different depth and the reflected in-focus signal is analyzed by a spectrometer. This method features similar resolution and spot size as the monochromatic approach but may acquire points at a higher rate with stationary optics. Maximum data rates are 4000 Hz and depth of field can vary from 20  $\mu\text{m}$  to millimeters. Fig. 1 is an example of an image acquired by this method.

These scanning methods rely also on precision mechanical stages to translate the sample from point-to-point. If the sample is moving during the measurement the sensor will average over an elongated region of width  $w$  ( $w = \text{spot size}$ ) and length  $l = v/f$  ( $v = \text{stage velocity}$  and  $f = \text{data rate}$ ). The minimum time required to scan a region of size *Area* using a grid of points of with spacing  $g_X$  and  $g_Y$  is given by,

$$T_{SCAN} = \frac{1}{f} \frac{Area}{g_X g_Y} + \text{reset} \quad (5)$$

If  $x$  is the first scan direction, and the  $x$  stage is moving, its velocity  $v$  must be  $fg_X$ .

As an example, the minimum time required to cover a 2 inch cylinder with  $g_X = 10 \mu\text{m}$ ,  $g_Y = 5 \mu\text{m}$  (a reasonable choice as discussed in below) and  $f = 4000 \text{ Hz}$  is 81000 s or 22.5 h. Factors of 2 or more improvement come with a coarser grid size if allowed. In practice it may be necessary to stop and start the scan periodically in order to control positioning systematics, re-establish a home position, or gather data. In this case an extra term (*reset*) should be added to Eq. 5 to account for this overhead.

The basic principle of white light interferometry is shown in Fig. 6a. A white light source is used in order to maintain a short coherence length and to localize the interference fringes to the narrow envelope shown in Fig. 6b. The system is then configured as, for example, a Michelson Interferometer (or alternate). Light reflected from the sample is combined with that reflected from the reference mirror at the focal plane of a Charge Coupled Device (CCD) or other pixelated imaging array. As the sample, or the interferometer, is scanned along an axis, perpendicular to the sample surface, a series of fringes evolve on the image plane. Coincidentally, a series of image frames is collected and then processed to recover surface structure. The number of

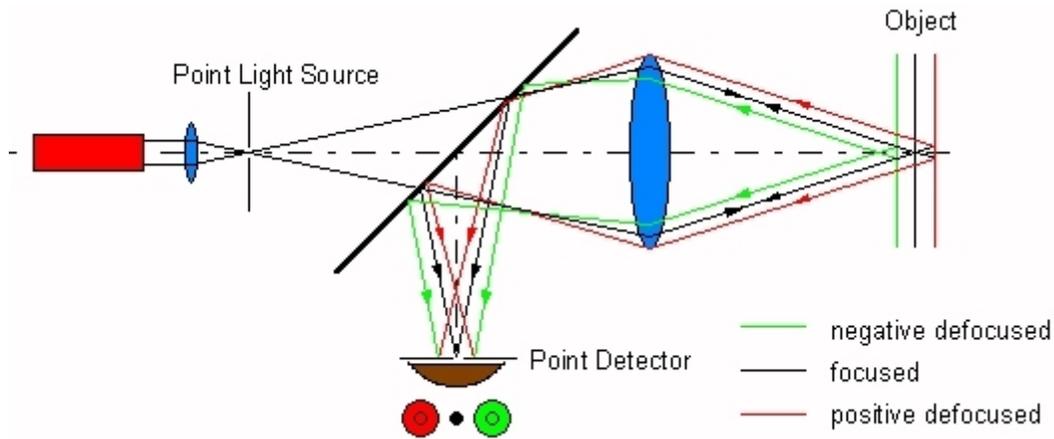


Figure 4: Basic concept of scanning confocal microscopy. The signal in the point detector is correlated with a sweep of the large lens position. Figure is courtesy A.Ruprecht, University of Stuttgart, used by permission.

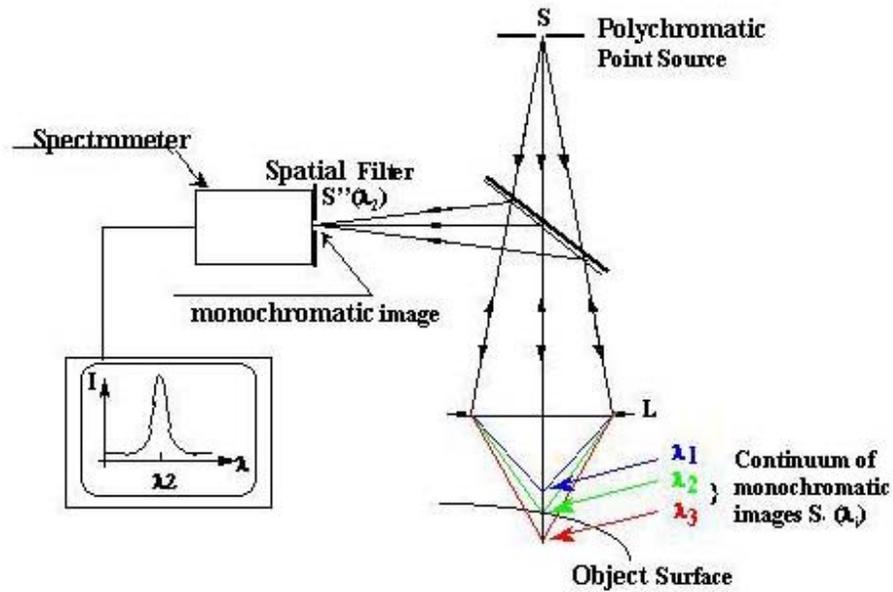


Figure 5: Basic concept of color-coded confocal microscopy. The lens at position L has a large chromatic aberration causing the three wavelengths shown to focus at different depths. Figure is courtesy of STIL SA, used by permission.

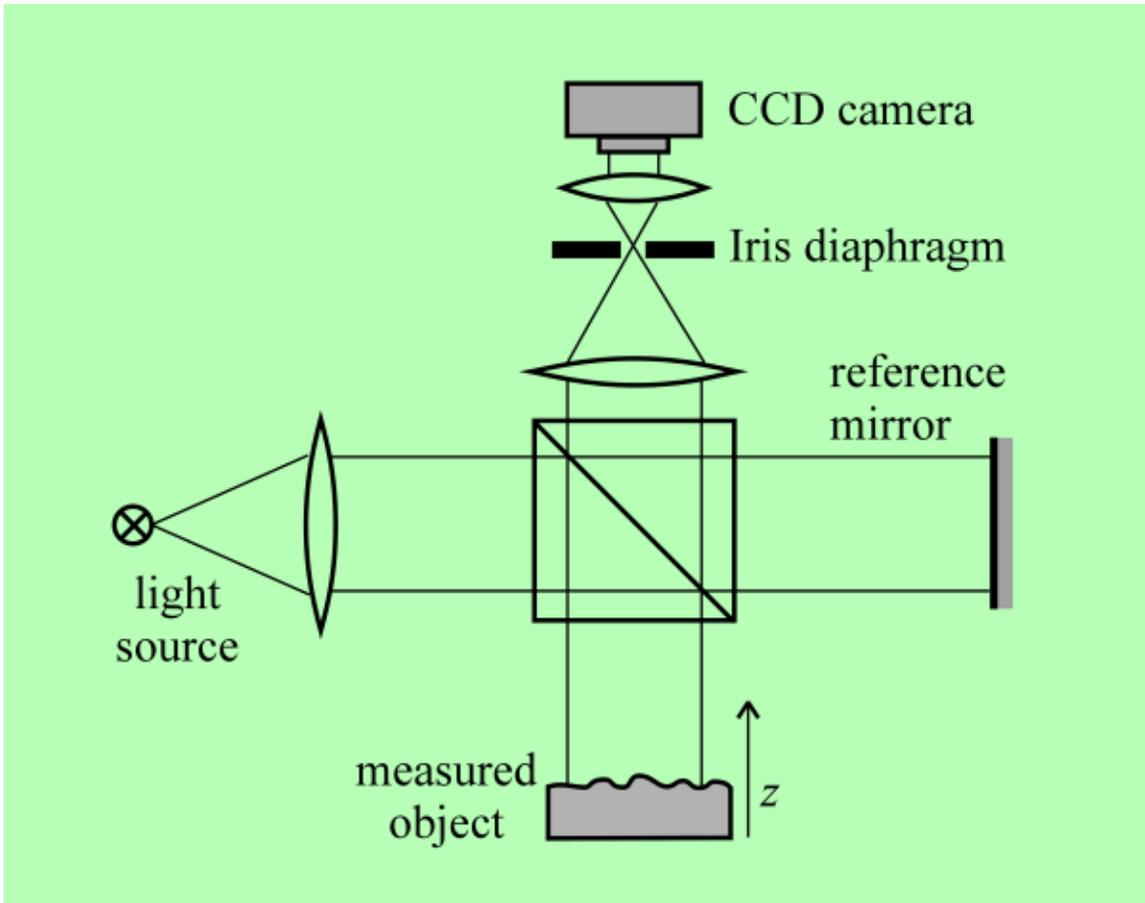


Figure 6a: Basic concept of white light interferometric surface measurement. As measured object moves in indicated  $z$  direction the fringe pattern on the CCD camera evolves. Due to the short coherence length of the white light it is not necessary to count fringes. For each pixel on the CCD the envelope shown in Fig. 7b is determined as a function of  $z$  position. Figure courtesy P. Pavlicik, RCO Czech Republic, used by permission.

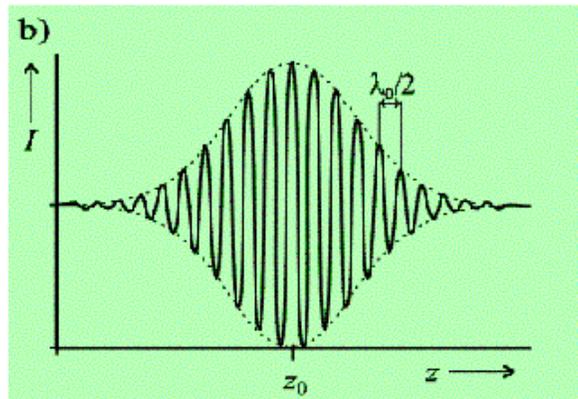


Figure 6b: Amplitude of interference pattern seen in each pixel of Figure 7a. Figure is courtesy of Veeco Instruments Inc.

planes required, and the spacing between planes, is determined by the slopes and height differences on the sample. This method can image an entire area at once of size determined by the usual constraints of 2D imaging; field of view (FOV), magnification, pixel geometry, and required transverse resolution in the plane. The vertical resolution of this method is typically 10 nanometers. The image acquisition time is determined by camera frame rates, the number of frames per view required, and scanning overhead such as step-and-repeat shift and settling times. For a surface with continuous variation, such as a mechanical sound carrier, typically 1 – 15 s is required, per view, depending upon depths and slopes.

An example of WLI measurement of the surface of a black-wax Edison cylinder is shown in Fig. 7. That image was acquired with a Veeco Instruments NT1100 WLI system using a magnification of 2.5x for a pixel size on the sample of 3.96  $\mu\text{m}$ . The time required to measure the field was  $t_{MEAS}=7$  s. Higher performance versions of these systems exist. The Veeco Instruments NT8000 can scan the same field in  $t_{MEAS}= 1-2$  s.

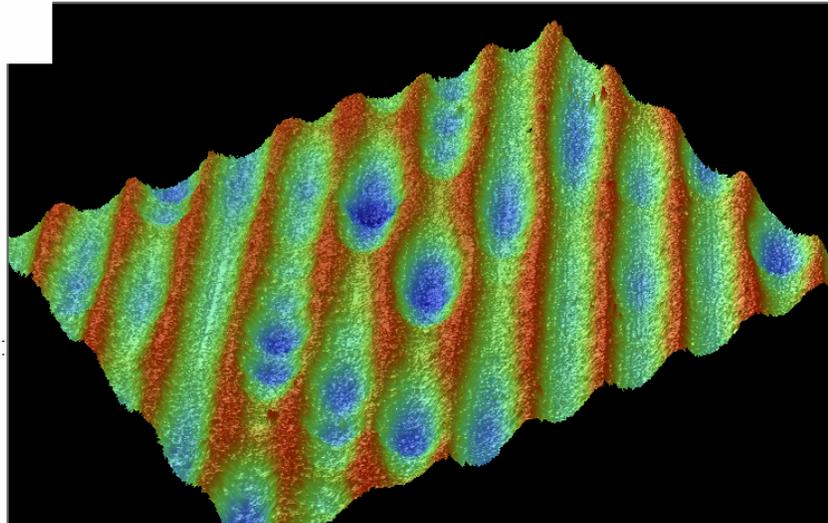


Figure 7: Example of an image field acquired with the Veeco NT1100 white light interferometry system. Image is from surface of a black wax cylinder. Figure is courtesy of Veeco Instruments.

To cover a large surface of area =  $Area$ , the measurement would be repeated on a regular pattern and adjacent processed images stitched together. The minimum time required to cover the region is determined by the FOV,  $t_{MEAS}$ , and the time required to *shift* to the next position.

$$T_{SCAN} = \frac{Area}{FOV} t_{MEAS} + shift \quad (6)$$

The FOV, combined with the pixel size will determine the transverse resolution of the measurement. As an example, to cover the surface area of a cylinder, as given in Table A1-1, with a pixel size projected onto the surface of the same 3.96  $\mu\text{m}$  requires a FOV=1.8 x 2.5 mm, With  $t_{MEAS}= 2$  s, and  $shift=0$ , the minimum  $T_{SCAN}= 2$  h. To cover the

surface area of 10 in shellac disc record is correspondingly longer due to surface area and groove depth (Table A1-1), typically 4-7 h, minimum. Factors of two or more improvement can be made with a larger FOV, if the resolution requirements can be relaxed. Further study would be required to determine this.

Frame rates in the WLI approach, and sensor rates in the confocal scanning approach are fundamentally limited by the reflected light from the sample. Improvements can come with brighter light sources, more sensitive photo-detectors, faster frame-rate cameras, and more sophisticated motion control systems. The commercial motivation driving further developments is large-area, high-resolution, inspection for flat panel displays, semiconductor wafers, micro-electrical and mechanical devices (MEMs) and other new technologies. Due to the value of these markets, improvements are to be expected.

The confocal and WLI imaging methods also present different approaches to the low-pass filtering strategies discussed in relation to Eq. 4, when the sampling is done at the highest noise frequency. In the confocal case the spot is conveniently elongated if the faster scan direction is parallel to the time axis on the record. To achieve the low-pass filter effect each position needs to be scanned twice with a half-step shift between. In the WLI case each field would be imaged twice with a half pixel offset between. In some custom implementation the WLI imager could be designed with rectangular pixels projections in order to optimize the scanning strategy.

Table 1 compares the confocal and white light imaging approaches.

### **3. Measurement Process**

In the study described here the color-coded confocal imaging method was used because it was available and already in practice by the authors. As discussed in Section 2, the parameters which can vary in such an approach include the size of the light spot, the measurement range, the resolution, the data rate, and the spacing between points in a scan. In the case of a mechanical sound carrier, the point spacing along the groove direction determines the ultimate sampling frequency. The spacing across the groove contributes to the accuracy with which the groove profile position can be determined at any one time slice.

The Model CHR150 probe, manufactured by STIL SA, was used here. This probe was coupled to custom configured stage movement and read out through a computer. The stages were controlled by DC servo motors and read out by linear encoders. The linear stage resolution was 100 nanometers and the accuracy was 2  $\mu\text{m}$ . The operating characteristics of the scanner and stage movement used are listed in Table 2. The setup is shown in Fig. 8.

The sample studied was an Edison Blue Amberol cylinder and is described in Table 3. The artist is shown in Fig. 9.

The surface scanning strategy used was based upon the parameters of the probe and the record. As a matter of convention, “lateral” will refer to the direction across the tracks and parallel to the axis of the cylinder. As a matter of convention, “temporal” or “azimuthal will refer to the direction along the tracks.

Table 1: Comparison of confocal and white light imaging technologies

<b>Parameter</b>	<b>Confocal</b>	<b>White Light Interferometry</b>
Acquisition	Circular or elongated point	Pixel field n x m elements
Transverse resolution	1.5 - 10 $\mu\text{m}$	Projected pixel size 1-10 $\mu\text{m}$
Vertical resolution	10 nanometers	10 nanometers
Points/measurement	1	n x m $\sim$ 480 x 540 = 259200
Max Time/measurement	250 $\mu\text{s}$	1-10 s
Effective time/point	250 $\mu\text{s}$	4-40 $\mu\text{s}$
Low-pass filtering?	Two passes with offset	Image field twice with offset
Depth of field	20 $\mu\text{m}$ – millimeters	Depth is scanned
Cost of probe only	$\sim$ \$30K	$>$ \$100K

Table 2: Parameters of the confocal probe used in this study.

<b>Parameter</b>	<b>Value</b>
Probe Model	STIL CHR150
Measurement Range	350 $\mu\text{m}$
Spot size	7.5 $\mu\text{m}$
Sampling Frequency	300 Hz
Vertical Resolution	10 nanometers
Vertical Accuracy	100 nanometers
Step size across grooves	10 $\mu\text{m}$
Step size along grooves (circumferential)	$0.01^\circ$ (= 5 $\mu\text{m}$ on circumference)
Linear scan speed (parallel to cylinder axis)	3 mm/s

Table 3: Description of the cylinder sample studied here.

<b>Parameter</b>	<b>Value</b>
Cylinder issue	Edison Blue Amberol
Diameter	2 in (2.1875 in)
Artist	Will Oakland and Chorus
Title	“Just Before the Battle, Mother”
Composer	George F. Root
Serial number	1516(..76; 4M-297-2)originally as Amberol #297 1909
Date of original recording	1909
Date of manufacture	$\sim$ 1920’s
Tracks per inch (t.p.i.)	200
Track spacing	127 $\mu\text{m}$

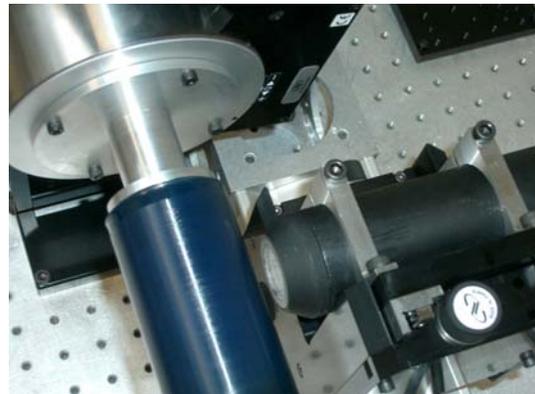
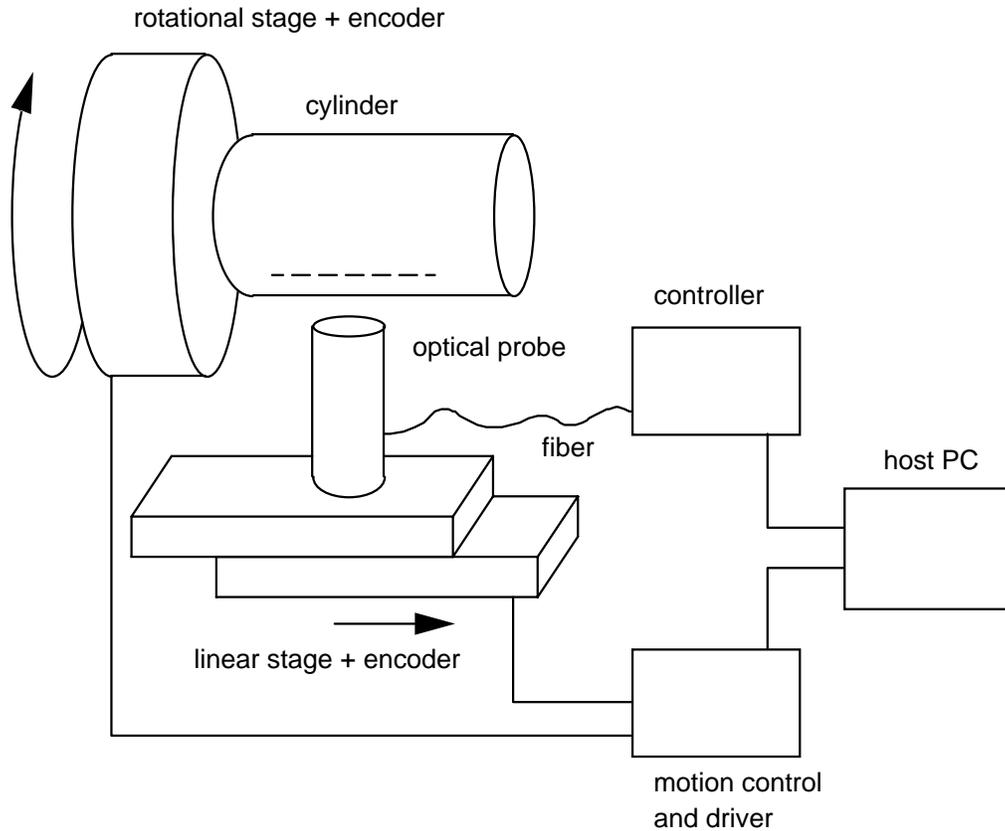


Figure 8: Top image is a schematic view of the scanning apparatus.. Lower left image shows overall setup including linear and rotational stages. In lower right image, confocal probe enters diagonally from the center right. Cylinder enters diagonally from the bottom center.

The result of a cylinder scan will be a data set consisting of heights at various lateral and azimuthal positions. From the lateral point set acquired at each azimuthal (temporal) position, a best estimate of the local groove depths can be derived. The lateral sampling interval is determined such that sufficient accuracy is found on the groove depth. The azimuthal or temporal sampling is determined to provide for sufficient audio time sampling.



Figure 9: Will Oakland.

For the 200 t.p.i. cylinder technology of the present sample, the groove width is  $127\ \mu\text{m}$  ridge to ridge. The confocal probe spot size used is  $7.5\ \mu\text{m}$  and the probe signal averages over a region of area  $7.5\ \mu\text{m}$  by  $f_{g_x}$  during a single measurement. Sufficient points must be acquired across the profile to assess its shape. For the  $127\ \mu\text{m}$  pitch,  $10\ \mu\text{m}$  between points was judged sufficient, assuming complex damage structures did not need to be resolved.

From the specified surface speeds the wavelength of a given tone can be computed for a cylinder. The results are given in Table 4 for a 2 inch cylinder in microns around the circumference. The corresponding angular interval can be scaled from  $360^\circ$  by the circumferential fraction for each tone. Sampling in  $0.01^\circ$  increments is convenient for the stages and results in a 96 kHz rate for 160 r.p.m.

The strategy used was to scan along the axis and then increment the azimuth. At 300 Hz sampling the lateral velocity was therefore the 3 mm/s indicated in Table 2. The data from a segment of one lateral scan is shown in Fig. 10. The advantage here is that the data is well organized for the depth determination of each groove at fixed time. The (slight) disadvantage is that the data needs to be reorganized later into a monotonic time series. A time penalty is also incurred (Eq. 5) since the linear stage should translate back to the same starting position for each sweep.

Table 4: Wavelength for various tones on cylinders for various r.p.m.

diameter (inches)	2.1875				
RPM	80	90	120	144	160
surface velocity (mm/s)	232.740	261.832	349.109	418.931	465.479
frequency	wavelength ( $\mu\text{m}$ )				
10	23274.0	26183.2	34910.9	41893.1	46547.9
100	2327.4	2618.3	3491.1	4189.3	4654.8
500	465.5	523.7	698.2	837.9	931.0
1000	232.7	261.8	349.1	418.9	465.5
5000	46.5	52.4	69.8	83.8	93.1
10000	23.3	26.2	34.9	41.9	46.5
15000	15.5	17.5	23.3	27.9	31.0
20000	11.6	13.1	17.5	20.9	23.3
44100	5.3	5.9	7.9	9.5	10.6
88200	2.6	3.0	4.0	4.7	5.3
100000	2.3	2.6	3.5	4.2	4.7

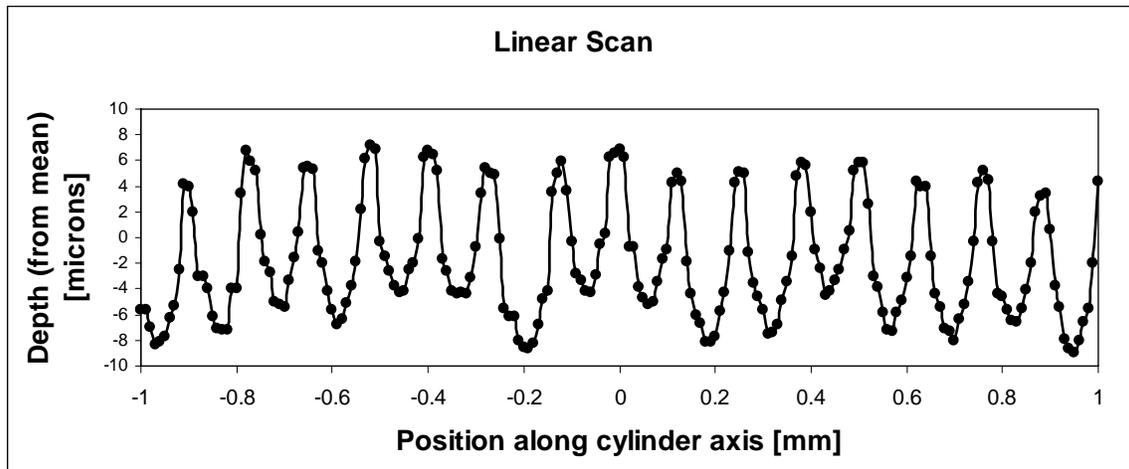


Figure 10: Segment of linear scan along the cylinder axis for fixed angular position. Horizontal scale is millimeters and vertical scale is microns.

The scan covered a lateral region of 10 mm length and the full  $2\pi$  rotation. Approximately 78 tracks were covered in this scan. About 20% of the data set is shown as a surface plot in Fig. 11. The vertical scale is exaggerated but a clear deviation from a cylindrical form is seen. The overall deviation is  $\sim 250 \mu\text{m}$ . This can be due to surface imperfections and off-center rotation. The structure is at low frequency since the cylinder rotates at 2.67 Hz during a stylus-based playback, and can be reduced by global fitting, low-pass filtering, or other corrections to the data. This is discussed further in Section 4.

#### 4. Analysis of Measured Data

As indicated in Section 3, the data scans were taken in slices along the cylinder axis with incremental rotation in azimuthal angle. The grid spacing was  $10\ \mu\text{m}$  along the cylinder axis ( $x$ ). The segment of length  $10\ \text{mm}$  scanned yielded  $m=1001$  points along the axis. With the azimuthal ( $\phi$ ) spacing of  $0.01^\circ$  ( $5\ \mu\text{m}$ ), the full circumference was covered by  $n=36000$  slices. The raw data points consisted of triplets  $x_{ij}$ ,  $\phi_{ij}$ , and  $h_{ij}$  (surface height) as follows,

$$\text{data point } 1 : x_{11}, \phi_{11}, h_{11}$$

$$\text{data point } m : x_{m1}, \phi_{m1}, h_{m1}$$

$$\text{data point } m+1 : x_{12}, \phi_{12}, h_{12}$$

$$\text{data point } m \times n : x_{mn}, \phi_{mn}, h_{mn}$$

where for fixed index  $i$ ,  $x_{ij}$  is independent of  $j$  and for fixed index  $j$ ,  $\phi_{ij}$  is independent of  $i$ . The data were processed as follows.

1. The analysis initially looped over index  $i$  to find local groove minima (valleys) and maxima (peaks) for each angular slice (fixed index  $j$ ). A segment of one such slice is shown in Fig. 10. Slices were then looped over for each angular position  $j$ .

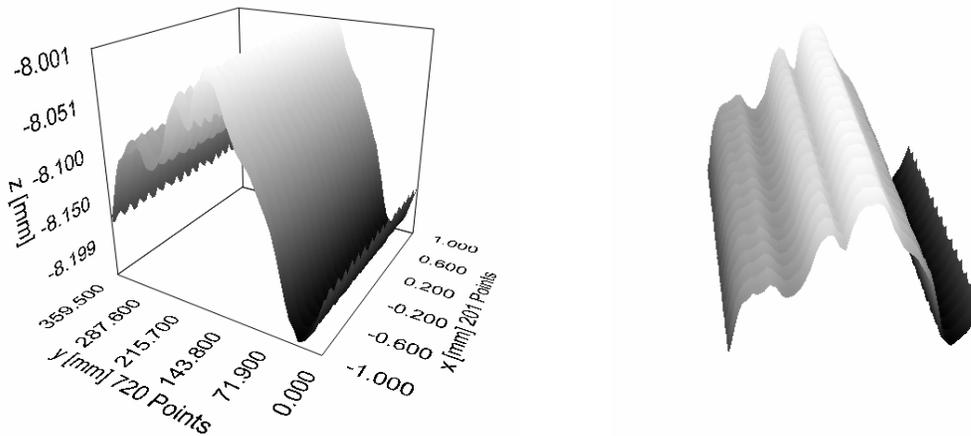


Figure 11: Plots shows 20% of the data scan “unrolled”, two views. Left-right axis ( $y$ ) is angle from  $0$  to  $360$  degrees, axis into the page ( $x$ ) is position along the cylinder axis in  $\text{mm}$ , vertical axis ( $z$ ) is surface height (from a reference point) in  $\text{mm}$ . A perfect cylinder would be flat at a fixed  $z$  position. Instead the surface differs from flat by about  $250\ \mu\text{m}$ . This low frequency structure can be due to a combination of shape imperfection, and off-axis rotation.

- 1.1. For each angular slice, minima (maxima) candidates were identified as data points which were not higher (lower) than the four nearest points. Some of these candidates can be false or undesirable, due to material damage and extraneous particles (dust or debris).
- 1.2. Bad candidate extrema were removed (filtered) using the inherent data periodicity of the known track structure. The correct groove bottoms (and tops) form a grid with fixed distance between neighbors. For the scanned sample this distance was found to be  $T = 0.1271$  mm, close to expectations for this type of cylinder (200 tracks/inch, see also Table 3). A relative phase, or offset, of the grid along the cylinder length axis was found. For this, a “chi-squared” like quantity was formed,

$$Rsum = \sum_{k=1}^N IR\left(\frac{x[k]}{T} - phase\right)^2 \quad (7)$$

where  $IR(y)$  is the difference between  $y$  and the nearest integer value,  $N$  is the number of extrema in the slice,  $k$  an index over extrema, and  $x[k]$  the position of the  $k^{th}$  extrema along the cylinder axis. This quantity was minimized, with respect to  $phase$ , to find  $Rsum_{min}$  and  $phase_{min}$ . Then the set of candidates was scanned and the points with

$$\left| IR\left(\frac{x[k]}{T} - phase_{min}\right) \right| > 3.0 \times \sqrt{\frac{Rsum_{min}}{N}} \quad (8)$$

were identified as outliers. They were removed from the list of extrema.

- 1.3. The groove ridge heights were taken from the list of maxima identified in the previous step. To find a best estimate for the minimum height the groove shape in the valley is used. For the  $k^{th}$  valley, a list of nearby points was formed from the found minimal point and 10 nearest neighbors in the slice. Points were removed from this list if they belonged to the list of maxima or maxima adjacent points. This set of points was fit to a parabolic function,  $H_k(x)$ , with a fixed quadratic term using a  $\chi^2$  minimization method,

$$H_k(x) = A_k x^2 + B_k x + C_k \quad (9)$$

$$A_k = \text{constant}$$

$$\chi_k^2 = \frac{1}{\sigma^2} \sum_{i=l-5}^{l+5} \left[ h_{ij} - H_k(x_{ij}) \right]^2 \quad (10)$$

where  $l$  is the raw data index of the  $k^{th}$  valley and  $j$  is fixed for the present angular slice. The cutting tools used to imprint the vertical undulations on a particular

cylinder are assumed here to have negligible wear over the course of the transcription. Therefore, only the groove position and depth will change. The parabolic shape is an approximation to a short circular segment, and is found to fit the local groove shape well. For this sample, the quadratic term was fixed at  $A = 3.853 \text{ mm}^{-1}$ , corresponding to a  $130 \text{ }\mu\text{m}$  radius for the cutting tool. (This best estimate of the quadratic term was determined by first fitting a large sample of grooves, with no fixed term.) The measured curvature also agrees well with historical data on cylindrical record cutting tools [21]. The individual point error was assumed to be  $\sigma=0.22 \text{ }\mu\text{m}$  based upon the width of the distribution of fit residuals. Points were compared to the results of the fit. If an outlying point deviated by more than  $3\sigma$ , it was removed and the fit was re-iterated. The fit iterations were stopped if no points deviated by more than  $3\sigma$  or only 5 points remained. As a result of the fit the value of groove position  $x_k = -B_k/(2A)$  and depth  $d_k = C_k - B_k^2/(4A)$  were obtained. Constraining the data to lie on this parabola is analogous to the simple edge-to-edge distance selection for 2D scanning discussed in the prior work [4].

- 1.4. The list of fit minima was filtered on the basis of periodicity similar to step 1.2 above. This was done to remove spurious fit results, some of which came from the region near the end of the scan, where the full groove cross-section was not available.
- 1.5. As described in Section 3, the scan was composed of a set of linear sweeps parallel to the cylinder axis. After each sweep, the probe returned to a starting position. Due to imperfections in the mechanical movement there could be some jitter in the distance between the probe and the cylinder from sweep to sweep. Such “common-mode” offsets were removed by averaging the  $\sim 78$  ridge maxima across the sweep and subtracting a single number from all fit minima for that sweep. In practice, an average and an r.m.s. value were found, and iterated, to remove outlying maxima beyond 3 time the r.m.s.
2. With a set of groove extrema at fixed azimuth, the data was then re-organized into a time series which followed the spiral trajectory of the groove bottom. This was done by establishing the “seed” groove bottom positions off the first slice. For each position, the groove bottom (ridge) positions in the next slice were searched for in the minimal distance along the cylinder axis. If the minimal distance was less than  $30 \text{ }\mu\text{m}$ , then the position was attributed to the same groove, and the seed value was updated to that of the next slice. The groove position on the cylinder was traced in this way. As the result of the iterative fitting and periodicity filtering described in steps 1.2 and 1.3, the data would naturally be missing when the groove shape is corrupted. The distance matching criterion is sufficiently loose to continue tracing the groove in case of missing data
3. The data, fitted and re-ordered, should contain 36,000 points per cylinder rotation, corresponding to the 96 kHz sampling for the 160 r.p.m. sample. Any missing points, due to the cuts discussed above, are linearly interpolated from the neighboring points. At least three data streams can be derived from this set.

- 3.1. The groove bottoms can be considered as a stand-alone data stream.
  - 3.2. The groove top (ridges) can be considered as a stand-alone data stream.
  - 3.3. The groove bottom position can be considered with respect to the top position. In this way the local surface imperfections seen in Fig. 11 can be subtracted out. The groove top can be used to track the generic cylinder surface shape. The groove top determination for a given slice is less precise than the bottom, because less measurement points went into its height value, and because of potentially more wear. Therefore, a point-by-point bottom-top subtraction would add noise. Instead a low sampling frequency template was made by averaging every 100 top points corresponding to one degree of azimuthal angle. Each bottom height value was referenced to a parabolic interpolation of the local top value taken from three nearest points of the top template. Due to the low sampling frequency of the template, the correction for the surface irregularity was restricted to low frequencies.
4. Excess frequency content was filtered using a Discrete Fourier Transform (DFT) method, implemented as a Fast Fourier Transform (FFT). This was done both to satisfy the Nyquist criteria before re-sampling to a lower frequency digital audio rate and because the acoustic recording technology has only a limited frequency content as discussed in Appendix A2. In addition, the sound amplitude is proportional to the stylus velocity, rather than displacement. The measured groove depth distribution is therefore differentiated to determine the stylus velocity. The filtering and differentiation can be accomplished in single step using the DFT as follows. Define  $A(nT)$  as the unfiltered waveform in the time domain and  $C(k)$  as the  $k^{th}$  element of its DFT, ( $k$  is not to be confused with the continuous wave-number used earlier),  $n=n^{th}$  time sample, and  $T$ =sampling period. Define  $A_M(nT)$  to be the filtered form of  $A(nT)$ . The total number of samples is  $N$ , and  $2\pi f=k\Omega$ , where  $f$  is the frequency in cycles per second and  $\Omega$ =radian-frequency sampling interval. Let  $M(k)$  be an imposed filtering function. As Eq. 11 shows, filtering and differentiation are caused by the  $(-ik\Omega)M(k)$  multiplier applied to the DFT coefficients.

$$\begin{aligned} \frac{d}{d(nT)} A_M(nT) &= \frac{d}{d(nT)} F_D^{-1}[C(k)] = \frac{1}{N} \sum_{k=0}^{N-1} \frac{d}{d(nT)} M(k)C(k)e^{-ik\Omega nT} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} (-ik\Omega)M(k)C(k)e^{-ik\Omega nT} \quad (11) \end{aligned}$$

The filtering factor  $M$  is defined as follows (the frequency  $f$  is related to the index  $k$ , as above ( $2\pi f=k\Omega$ )).

$$M = \left. \begin{array}{l} 0 \text{ for } f < 20\text{Hz} \\ 1 \text{ for } f \in [20\text{Hz}, 4.8\text{KHz}] \\ \left(1.0 - \frac{(f - 4.8)}{0.4}\right) \text{ for } f \in [4.8\text{KHz}, 5.2\text{KHz}] \\ 0 \text{ for } f > 5.2 \text{ KHz} \end{array} \right\} \quad (12)$$

- 4.1. The cut below 20 Hz removes the low frequency structure in the bottom-only data due to the cylinder shape irregularity.
- 4.2. The 400 Hz wide transition to zero at 5.0 kHz was used to avoid the interference-like pattern triggered by jumps in the data.
- 4.3. The cut above 5.2 kHz satisfies the Nyquist criteria before re-sampling to a lower digital audio standard.

All the data from the scanned sample could not be easily processed simultaneously, due its size. Instead, 1 s long samples were treated sequentially. The filtering was done in 100,000 points space, of which 96000 points corresponded to the scanned sampling frequency. Two “zero-padding” regions of 1000 points lengths were placed before and after each such chunk of data, with the same length linear transitional regions in between. This was done to avoid mismatching jumps between sequential samples.

5. The list of filtered height measurements was rescaled to 16 bit dynamic range, and converted to WAV format with 22.05 kHz sampling.

The audio data is shown in Fig. 12a, b, and c for the three streams described in Step 3 above and can be accessed at an Internet URL<sup>1</sup>. Note the scale differences in the figures. Interestingly, the groove top stand-alone stream contains audible, albeit noisy sound, with some interference from different times in the recording. This may be due to the residual localized pressing of the cutting tool on the nearby surface. The pressing on the two tracks on both sides of a ridge can create the mentioned interference.

Because of the sound content in the top stream, and the possibility of some “self-subtraction”, the top/bottom subtracted method is biased. Due to the logarithmic perception of the sound amplitudes, the amplitudes in samples would have to be very close for the audible degradation to occur and as practical matter there is, however, little effect. The implicit correction for global cylinder shape is valuable and therefore this approach is shown as well. It effectively removes structures below about 100 Hz which are in-fact only marginally audible.

To make a comparison with stylus playback, the same cylinder was digitized on an (modern) Archeophone player. This player spun the cylinder with a motor and captured the sound with a Shure V14 cartridge and a KAB Souvenir Pre-amplifier set for mono, a flat frequency correction, and vertical movement. After digitization, frequency cuts were applied to the stylus data to match Eq. 12. The comparison is shown in Fig. 13a and b for the optical and raw stylus versions respectively. Figs. 14a and b show an expanded view of these same clips. This sample is available at the aforementioned URL<sup>1</sup> as well.

As shown in Figs. 12 and 13, the bottom-only and bottom/top subtracted samples are indeed very similar and also similar to the stylus versions. However, the bottom-only sample has some periodic clicking sound content, similar to a stylus striking an obstacle. This feature is similar to what can be heard more clearly on the “raw” sound from the stylus playback and may therefore be inherent to this record. The comparison is also illustrated in Fig. 15a-d which show FFT spectrum analysis for the bottom-only, top, bottom/top, and stylus versions respectively. From these it is clear that the stylus version contains significantly more low frequency artifacts.

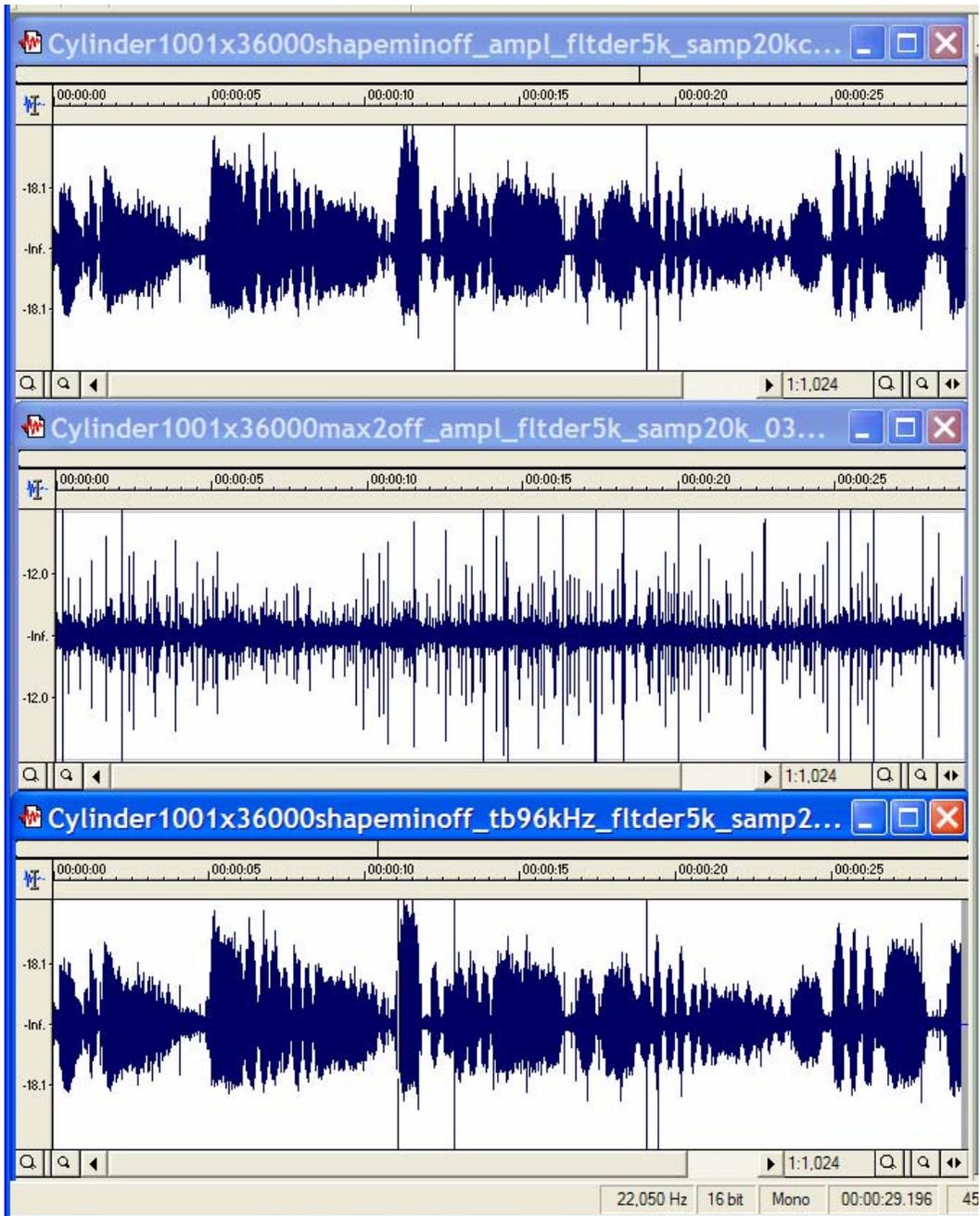


Figure 12: a) top shows audio stream from bottom only extraction, b) middle shows audio stream from ridge top only extraction, and c) lower shows top-bottom extraction (line at 10.2 s is a cursor). Approximately 29 s of audio is shown.

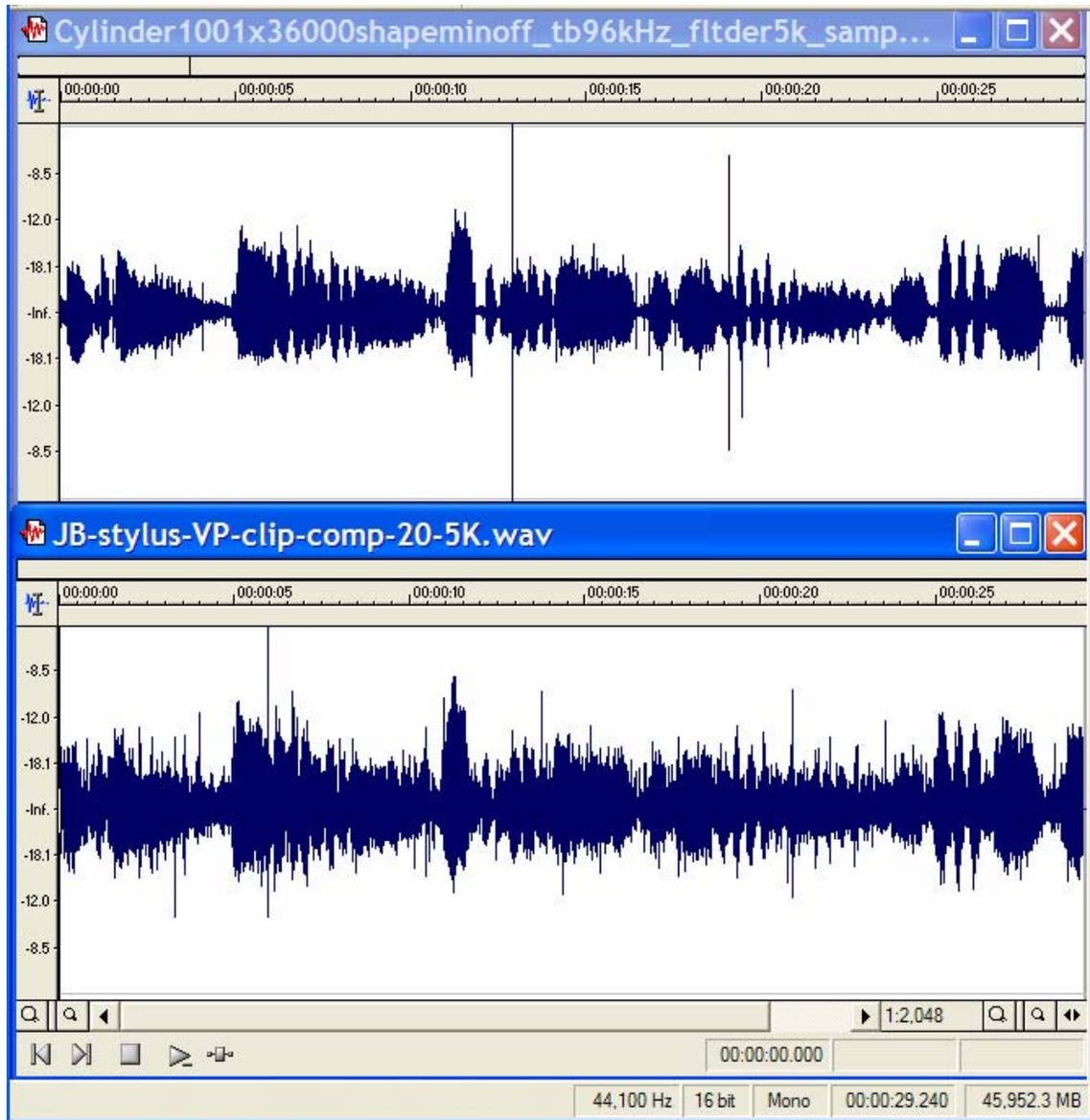


Figure 13: a) top shows top-bottom audio extraction, identical to Fig. 12c and b) lower shows stylus playback of the same cylinder with only Eq. 12 cuts applied. Approximately 29 s of audio is shown.

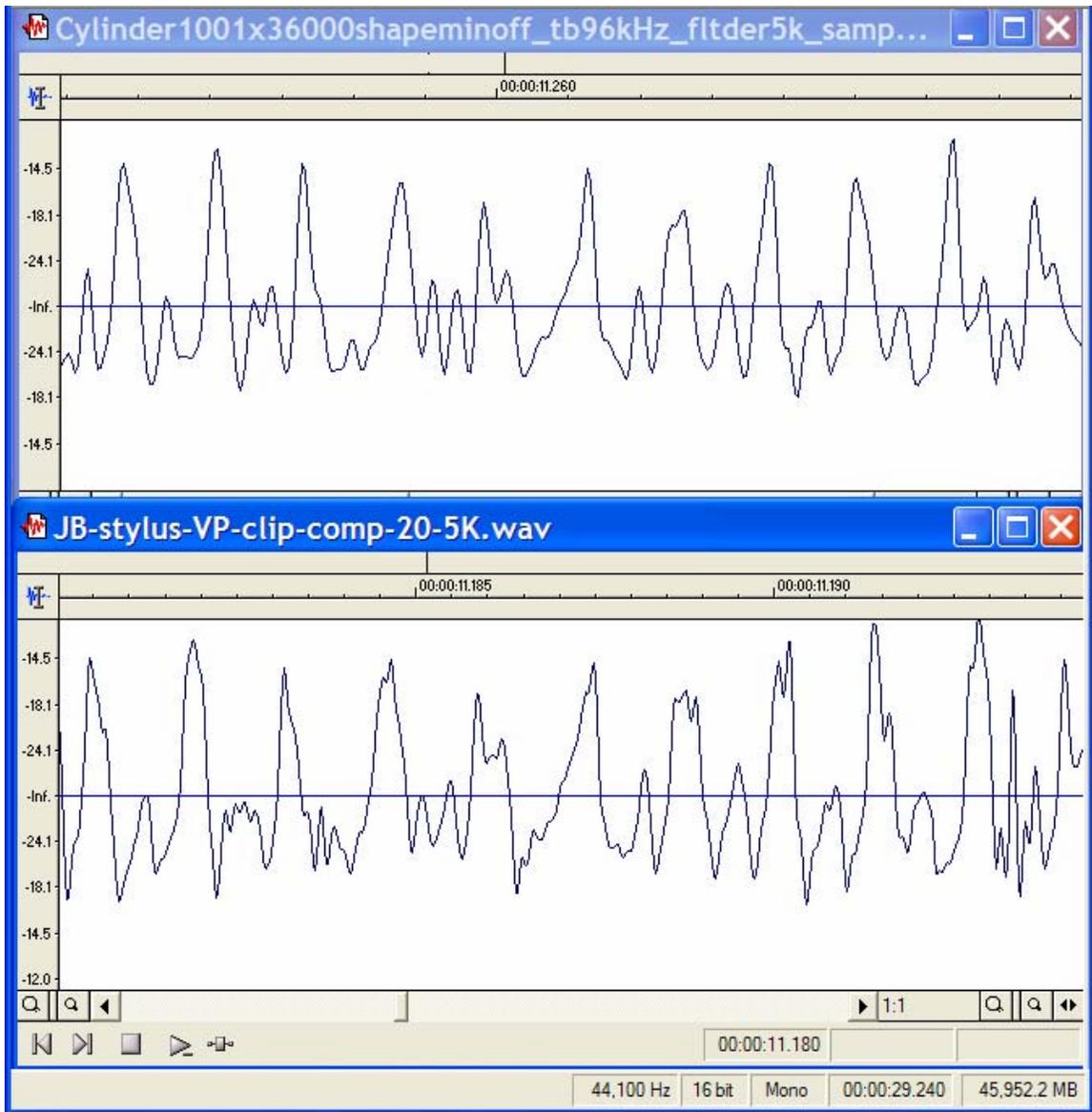
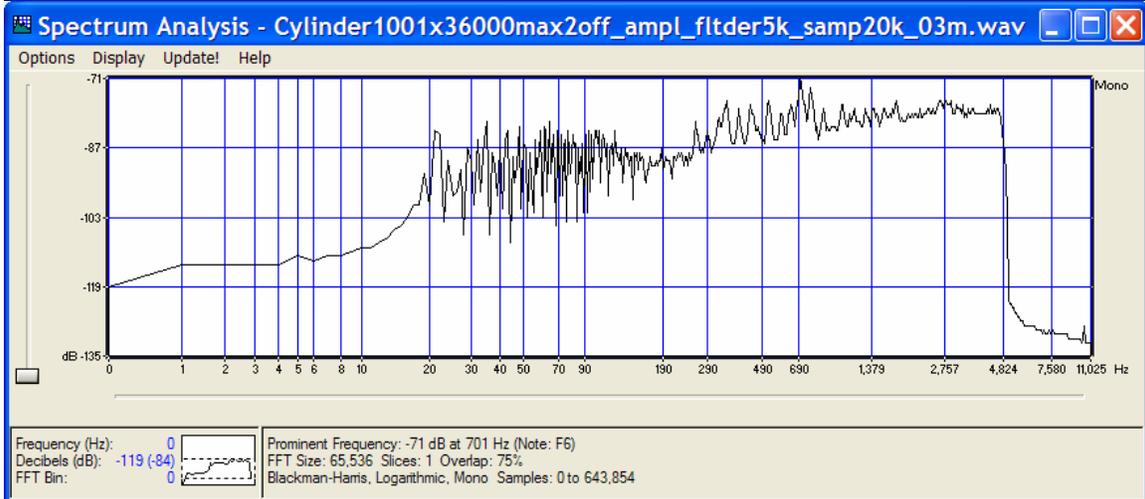
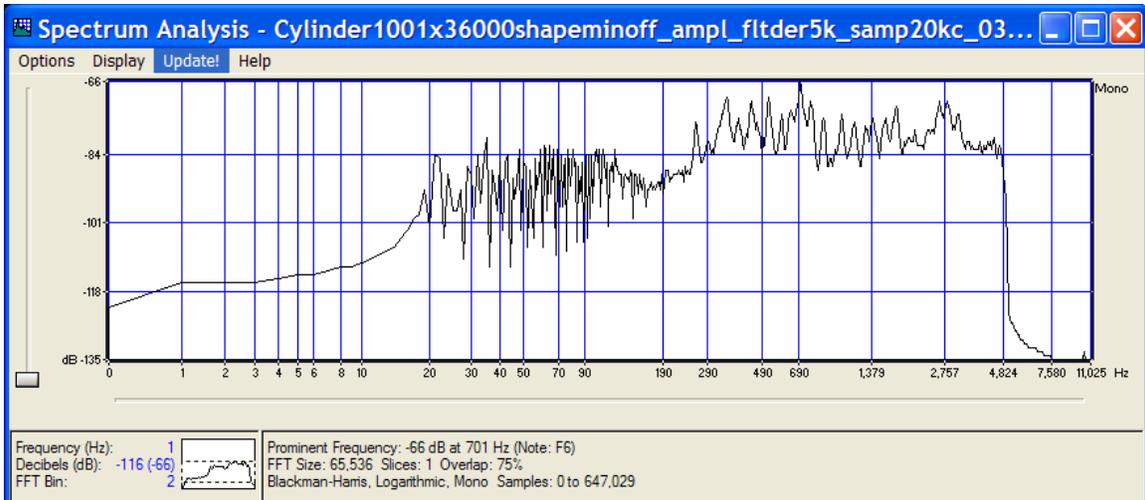


Figure 14: a) top is the same data as Fig 13a but expanded to show a segment of length 14 ms beginning at 11.254 s from the start of the clip, b) bottom is the corresponding section from the stylus version.



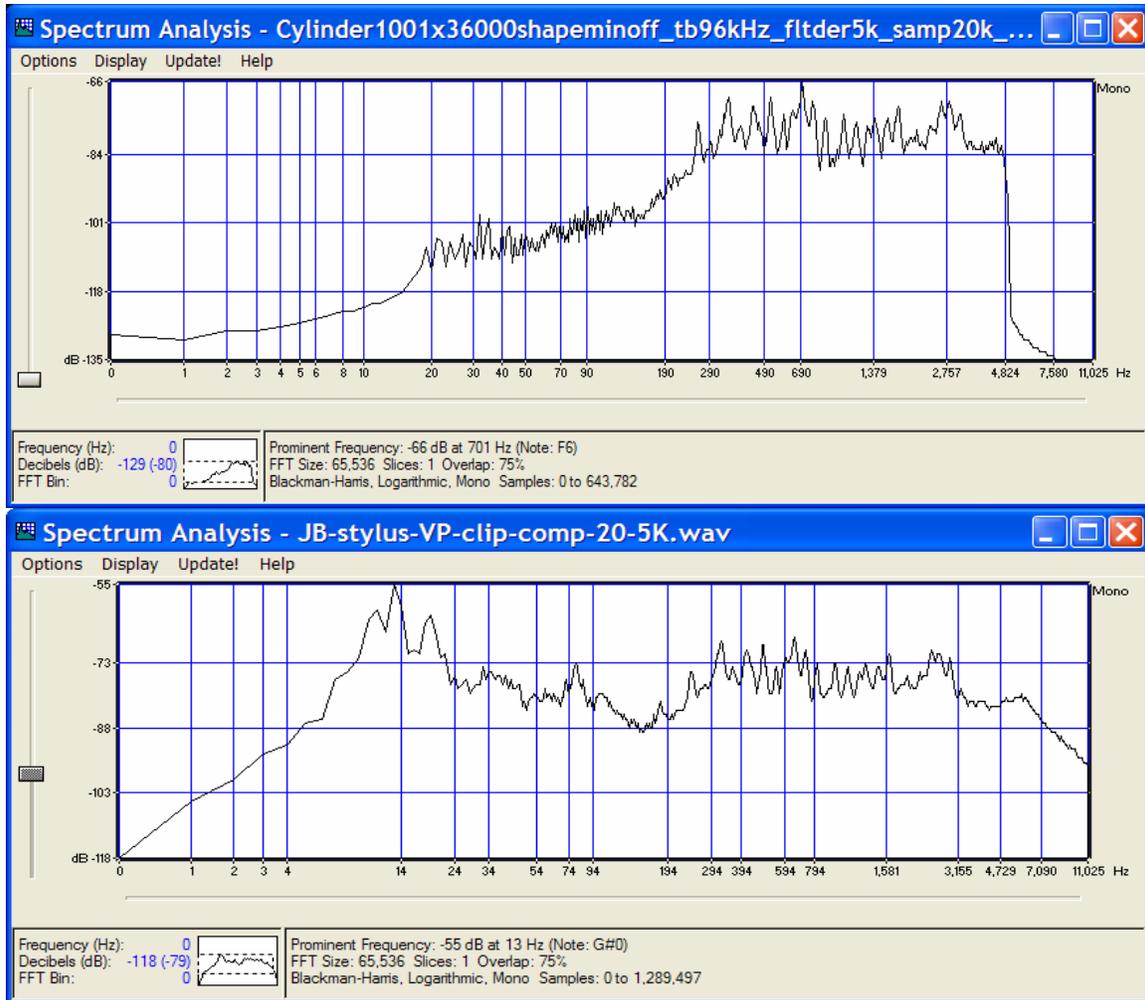


Figure 15 a) top Fast Fourier Transform spectrum analysis of data of Fig. 12a groove bottom only, b) FFT of top ridges only as in Fig. 12b, c) FFT of top-bottom subtracted data as in Fig. 12c, and d) FFT spectrum analysis of stylus playback data of Fig. 13b.

## 5. Discussion

It is clear from the figures and clips that the 3D optical method provides an accurate audio transcription as compared to a stylus playback. There are a number of issues for further consideration.

No post-processing has been applied to the samples presented here. In principle the sound quality could be enhanced by additional bass boost, and digital de-clicking or noise reduction. (Further examples are presented on the aforementioned website.)

An example of a damage site is shown in Fig. 16. This figure illustrates that damage structures can be of significant size compared to the characteristic amplitudes of the cylinder technology. The fixed grid, time-slice-based approach, while simple to implement, and effective in the noise-reduction when used in shape-based filtering, may not be optimal for certain complex damage sites. For those sites, the full 3D data should

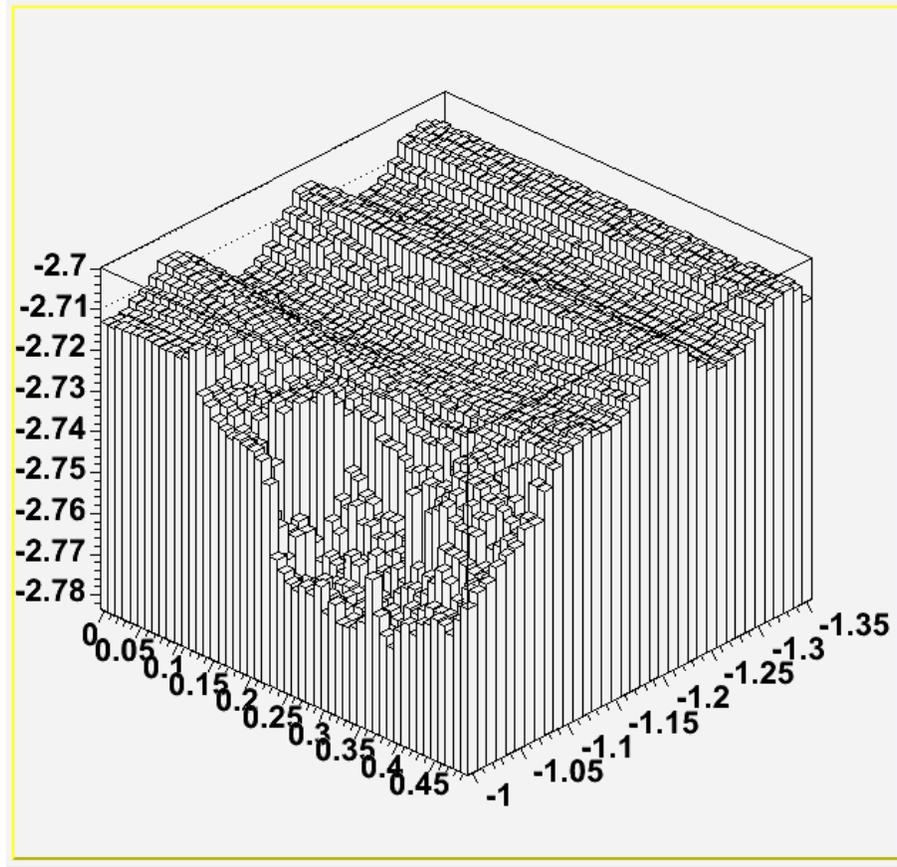


Figure 16: Example of a damage site on the cylinder. Scale is in millimeters. Each 2D bin represents a measurement point of size  $5\ \mu\text{m}$  along the groove and  $10\ \mu\text{m}$  across the groove.

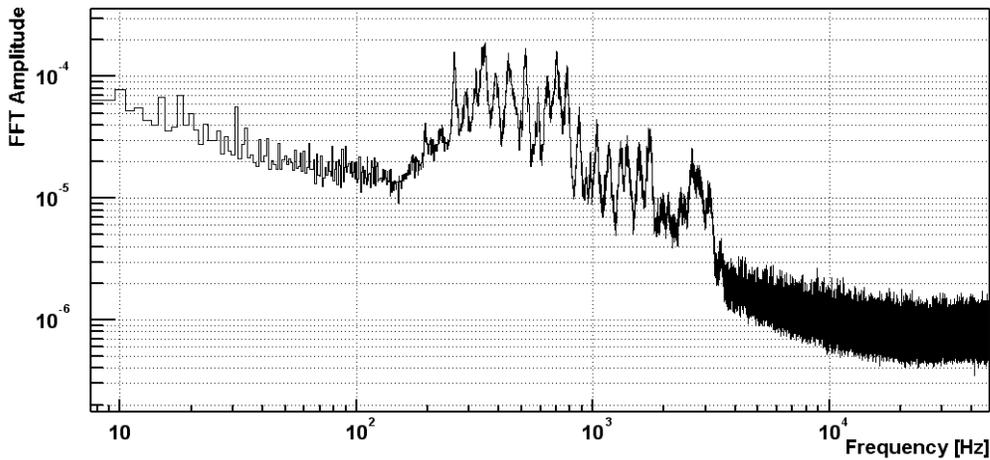


Figure 17: The FFT spectra of the reconstructed sample groove amplitudes before differentiation. For the initial sampling at  $96\ \text{kHz}$ , the Nyquist frequency is  $48\ \text{kHz}$ . Any noise which is aliased to lower frequency is below the baseline noise at audible frequency.

be used to “delete” or correct the damage site. The adjacent undamaged regions would be used as a basis for such an interpolation which may be statistically determined. A denser sampling of the surface may be required to fully correct the “clicks” and “pops” which manage to pass the initial filtering procedure. This could be accomplished by returning to damage sites after an initial scan and re-measuring with a finer grid. The usefulness of this approach could be the topic of a subsequent study.

Another related area concerns samples with surface contamination. An important example is brown or white mold which is known to grow on the surface of wax cylinders. Due to the inherent multiple sampling across the groove profile in this method, it may be possible to overcome these effects at some level. Again, the grid density is a variable which may require further optimization. If the mold is known to always be additive or subtractive of the surface a shape fitting constraint could be developed. This will be the topic of a subsequent study.

In Section 2 the aliasing issues were discussed. From the results presented, it seems unlikely that a significant aliased signal is folded into the audible region. As a further demonstration of this, Fig. 17 shows an FFT spectrum of the data at the raw 96 kHz sampling, before any differentiation, filtering, or re-sampling has been applied. Any noise content above 48 kHz should be a continuation of the level present just below 48 kHz, or if it is already aliased, then less. The level seen is already a sizable factor below the signal baseline in the audible region. These issues could be explored further with more complex optical sampling strategies as discussed in Section 2.

A clear area for improvement is in scan time. Confocal probes can run at faster sampling rates than the 300 Hz applied here, at some potential loss of reflected signal. A subsequent study will consider the effect of a coarser baseline grid and faster confocal rates (up to 4000 Hz) on signal quality. These modifications could yield significant improvements in scan time. A scan at 4000 Hz, with temporal sampling of 48 kHz (rather than 96 kHz), and lateral sampling of 20  $\mu\text{m}$  (rather than 10  $\mu\text{m}$ ) would require 8 h to scan a 4 min cylinder.

The discussion of acoustic recorders, diaphragms, and horns in Sec. 2 and App. 2, suggests that improved modeling of their characteristics could aid in a higher fidelity restoration with respect to the original audio source waveform. Such an effort goes beyond the issues of surface metrology but does speak to the signal processing methods used to analyze the measured data.

## **6. Conclusions**

The technical basis for 3D surface metrology applied to audio recovery from mechanical sound carriers has been discussed. A proof-of-principle was demonstrated on a celluloid cylinder and results were compared to stylus playback. Further development of these methods could enable the systematic recovery of delicate, damaged, broken, or moldy samples. Optimized hardware, software, and analysis tools could further enhance the impact of these methods on audio preservation.

## Appendix A1

The known parameters of certain mechanical sound carriers, as determined from a number of sources [13-16] are tabulated here. Figs. 2 and 3 relate these parameters to the configuration of the groove on cylindrical and disc records respectively. When relevant they are defined at a specific frequency (1000 Hz) where equalization is generally not applied.

- 1) Groove width: distance across the top of the groove.
- 2) Tracks per inch ( $G_d$ ): the number of tracks cut in the surface per radial inch.
- 3) Track spacing: center-to-center distance between two adjacent tracks.
- 4) Fixed Groove Depth: for a disc record with a lateral groove, the depth of the groove below the surface. Not applicable to vertically cut records.
- 5) Reference signal level: the peak transverse velocity used to set a baseline for the recorded signal. This quantity is in principle arbitrary but is key to defining the noise and dynamic range discussed in the literature.
- 6) Maximum groove amplitude: the maximum displacement of the groove from an un-modulated path.
- 7) Noise level below reference level (signal to noise ratio): noise levels or limits are usually expressed as dB below the reference signal. This is taken to mean the standard deviation of any random noise source, such as the underlying surface noise source discussed above, or the maximum allowed deviations due to the low frequency systematic effects.
- 8) Dynamic range: a measure of the range of audible signals up to the maximum peak recorded signal level, defined here with respect to the noise level at 1000 Hz
- 9) Groove amplitude at noise level: maximum amplitude deviation from a signal free path corresponding to the noise level in item 6) above and defined in Eq. 2.
- 10) Maximum and minimum radii: the respective radii at which audio data is specified to begin ( $R_{MAX}$ ) and end ( $R_{MIN}$ ).
- 11) Area: the area covered by audio data. For a cylinder  $L$  is the length along the axis.

$$Area = \pi(R_{MAX}^2 - R_{MIN}^2) \quad disc \quad (A1-1)$$

$$Area = 2\pi RL \quad cylinder \quad (A1-2)$$

- 12) Total length: the path length along a complete groove between the two radial extremes.

$$L = G_d \times Area \quad (A1-3)$$

Table A1-1: Parameters of various mechanical sound carriers.

<b>Parameter</b>	<b>Coarse</b>	<b>Micro-Groove</b>	<b>Cylinder</b>
Diameter inches	10-12	12	2-5
Revolutions per minute	78.26	33.3333	80-160
Groove width at top	150-200 $\mu\text{m}$	25-75 $\mu\text{m}$	variable
Tracks/inch $G_d$ (mm)	96-136 (3.78-5.35)	200-300 (7.87-11.81)	100-200 (3.94-7.87)
Track spacing	175-250 $\mu\text{m}$	84-125 $\mu\text{m}$	125-250 $\mu\text{m}$
Fixed Groove depth	40-80 $\mu\text{m}$	25-32 $\mu\text{m}$	NA
Ref level peak velocity@1kHz	7 cm/s	7 cm/s (11 $\mu\text{m}$ )	unknown
Maximum groove amplitude	100-125 $\mu\text{m}$	38-50 $\mu\text{m}$	~10 $\mu\text{m}$
Noise level below ref, S/N	17-37 dB	50 dB	unknown
Dynamic range	30-50 dB	56 dB	unknown
Groove max ampl@noise level	1.6 - 0.16 $\mu\text{m}$	0.035 $\mu\text{m}$	< 1 $\mu\text{m}$
Max/Min radii mm	120.65/47.63	146.05/60.33	fixed
Area containing audio data	38600 $\text{mm}^2$	55650 $\text{mm}^2$	16200 $\text{mm}^2$ (2'')
Total length of groove meters	152	437	64-128

## Appendix A2

As discussed in Sec. 2, and applied in the analysis of Sec. 4, the sound amplitude was taken to be proportional to the stylus velocity rather than displacement (“constant velocity” versus frequency characteristic). For electrical recordings this follows directly from the action of the electromagnetic cutting transducer. The purpose of this appendix is to discuss the assumptions and conditions leading to the constant stylus velocity condition for acoustic recordings.

The basic arrangement, used in both acoustic recording and playback, is shown in Fig. A2-1. A horn, or tube, is coupled to a diaphragm which either drives (recording) or responds to (playback) a stylus contacting the surface of the mechanical sound carrier. The sound, either generated by a source in recording, or received by the listener in playback is a pressure wave. In order for the constant-stylus-velocity, flat-frequency condition to apply, the stylus velocity should be proportional to, and in phase with, the sound pressure at the source/listener at any frequency. In this case, as in the case of an electrical recorder, the time derivative of the mechanical carrier surface modulation can be directly interpreted as the sound at the source. Two conditions should be satisfied.

- C1. The recording horn supports the propagation of plane waves from the source to the diaphragm.
- C2. The diaphragm vibrates with a velocity in-phase, and a displacement, 90 degrees out-of-phase, with the driving pressure wave.

For plane waves [22], the equation of motion and continuity condition, for the gas, are given, to first order by,

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial u}{\partial t} \quad (A2-1)$$

$$\kappa \frac{\partial p}{\partial t} = -\frac{\partial u}{\partial x} \quad (A2-2)$$

where  $p = p(x,t)$  is the pressure,  $u = u(x,t)$  the velocity,  $\rho$  the density, and  $\kappa$  the compressibility of the gas. Combining these equations to eliminate either  $p$  or  $u$  yields the equation for either a velocity or pressure wave traveling at the speed of sound  $c = (\rho\kappa)^{-1/2}$  with solutions of the form,

$$p(x,t) = p_M \exp[i(\omega t - kx)] \quad (A2-3)$$

$$u(x,t) = u_M \exp[i(\omega t - kx)] \quad (A2-4)$$

where  $p_M$  and  $u_M$  are constants,  $\omega$  is an angular frequency and  $k$  a wave number. These solutions must obey the condition

$$\frac{k}{\omega} = \sqrt{\rho\kappa} = \frac{1}{c} \quad (A2-5)$$

which is just the dispersion relation  $c = \lambda f$ , where  $f = \omega/2\pi$  is the frequency and  $\lambda$  is the wavelength. The coefficients in Eqs. A2-3 and A2-4 are related by  $u_M = \kappa c p_M$  and

therefore,

$$u(x, t) = \kappa c p_M \exp[i(\omega t - kx)] \quad (A2-6)$$

which is the required proportionality and phase relationship for the velocity and the pressure.

For practical reasons, the diaphragm is typically small and therefore not a good transducer for wavelengths  $\lambda > 2\pi R_D$ , where  $R_D$  is the diaphragm radius. For this reason the diaphragm is coupled to the throat-end of a horn which can receive or radiate sound from a much larger effective radius,  $R_M$ , at the mouth end. The issue is then whether the general arrangement of Fig. A2-1 supports plane waves between the source or listener and the diaphragm [22-27].

A detailed description of the behavior of horns is complicated and a number of simplifications and compilations have been presented. The properties of a horn are equivalent in transmission and radiation. In general, the behavior of a horn is determined by its profile (conical, exponential, catenoidal, etc.) and its length,  $L$ . The profile can be described by a cross-sectional area function  $S(x)$ , where  $x$  is position along the axis. In the limit of increasing length and mouth diameter, and/or frequency, the throat impedance of the horn is mostly real and also approaches  $\rho c$ , meaning that plane waves are transmitted, and resonances are suppressed. In this limit the horn is a good conduit for plane waves between the open environment and the throat. As stated, the use of a horn extends the region of approximate plane wave propagation to lower frequencies than would be possible with just a small diaphragm coupled directly to an open volume. None-the-less the low frequency response will still degrade at some point and differently so depending upon the horn geometry. At that point the plane wave approximation also breaks down.

To illustrate the deviation from plane wave response, the finite conical and exponential horns, shown in Fig. A2-1, may be considered as an example. Both will be limited to some overall length  $L$ . The conical horn has a opening angle  $\vartheta$  and an apex distance  $x_0$ , such that,

$$S(x) = S_0 \left[ 1 + \left( \frac{x}{x_0} \right) \right]^2 \quad (A2-7)$$

$$S_0 = \pi \left[ 2x_0 \sin\left(\frac{\vartheta}{2}\right) \right]^2 \quad (A2-8)$$

The exponential horn has a cross-sectional area function,

$$S(x) = S_0 \exp\left(\frac{2x}{h}\right) \quad (A2-9)$$

where  $h$  is a parameter. The horn transmission coefficient is defined as the ratio of power, introduced by a diaphragm and radiated by a horn of throat area  $S_0$  to that radiated by the same diaphragm, moving at the same velocity, into an infinite cylindrical

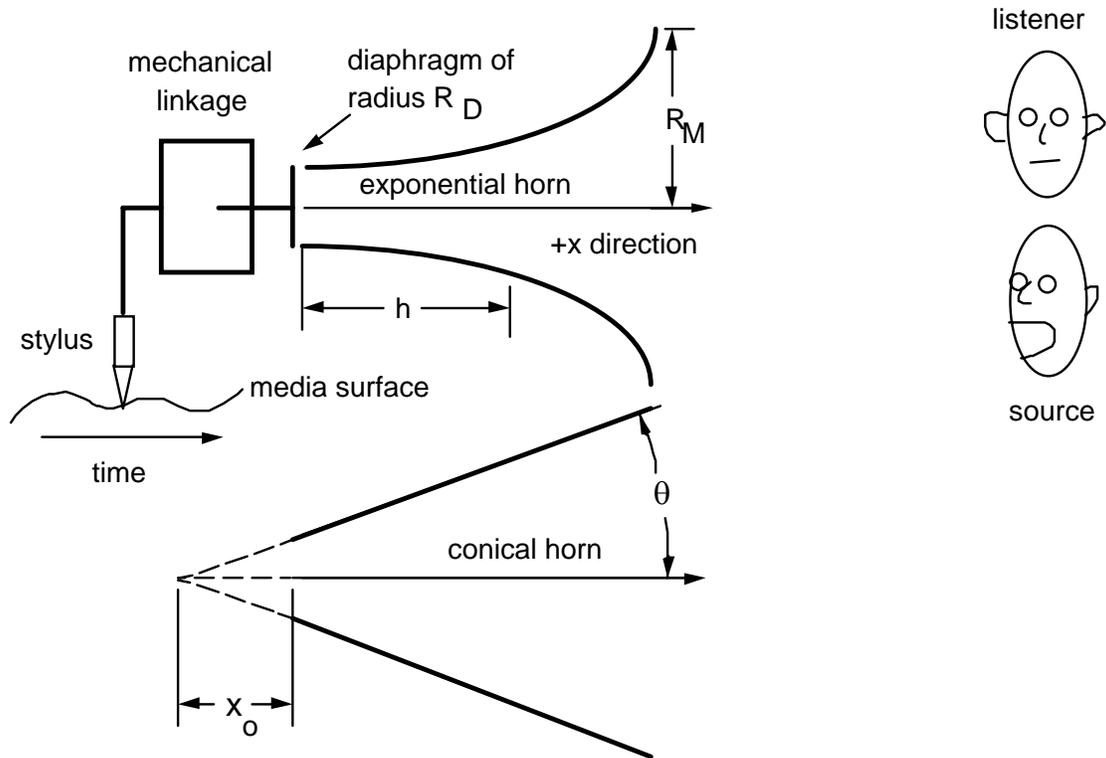


Figure A2-1: Basic arrangement used in acoustic recording and playback. Shown are two variations of horn, the exponential and conical types, with the same overall length.

tube of the same cross-sectional area. For the conical horn, the transmission coefficient takes the form,

$$\tau_{CONICAL} = \frac{1}{1 + (\lambda/2\pi x_0)^2} \quad (A2-10)$$

and for the exponential horn,

$$\tau_{EXPONENTIAL} = \sqrt{1 - \left(\frac{\lambda}{2\pi h}\right)^2} \quad (A2-11).$$

A comparison of these is given in Fig. A2-2 for model horns of the same overall length. The exponential horn has a sharp cut-off at a certain low frequency and a relatively flatter frequency response above cut-off than the conical horn. Where the transmission deviates from unity, the acoustic properties of the horn distort the sound and the plane wave condition fails to apply.

When the plane pressure waves reach the throat of the horn they will excite the diaphragm. To maintain the constant-velocity flat-frequency response, condition C2, stated above, must apply. For the early Edison recorders the diaphragm was held in place by a gasket and coupled to the cutter through a simple pivot. Such an assembly is rather complicated, but it can roughly be understood as a damped, driven, simple-harmonic oscillator. The designer of the diaphragm attempts to make its responsiveness as flat as possible over the frequency range admitted by the horn, while simultaneously making the

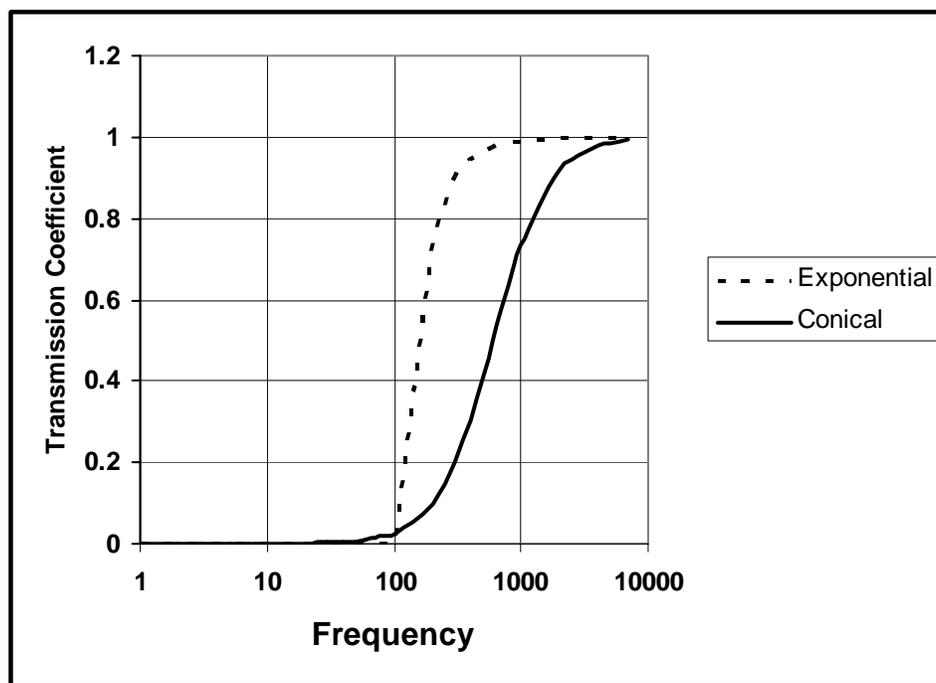


Figure A2-2: A comparison of the transmission coefficient of an exponential (dashed curve) and conical (solid curve) horn of the same overall length. The particular shapes depend upon the horn geometry but the qualitative trend is characteristic.

diaphragm as sensitive as it can be. If an SHO is underdamped, it will have a strongly varying sensitivity in the region of its resonance. On the other hand, if it is overdamped, the friction term dominates the mass or spring terms in the equation of motion, and the velocity is proportional to the driving force. This yields the behavior described in condition C2. However, the greater the damping the smaller the amplitude of the motion, so there is a tradeoff between sensitivity and flatness of response.

An example is given in Fig. A2-3 which shows the amplitude and phase characteristics of an over-damped oscillator with a characteristic frequency of 5000 Hz. A flat amplitude and 90-degree phase shift are seen, to within about 15%, over a significant portion of the audio range.

Near the end of the acoustic era, Maxfield and Harrison [28] described an advanced mechanical transducer which acted as a bandpass filter over a frequency range of at least 100 to 5000 Hz. Figure A2-4 shows the frequency response of a system which coupled that transducer to an exponential horn. (See reference [29] for a comment on this plot). The range covered should be considered a best case, benefiting from some of the understanding of horn and transducer properties which emerged late in the acoustic era [30]

For the broad resonance of the simple Edison recorder, the diaphragm would vibrate approximately 90 degrees out of phase with the in-coming sound pressure wave over a significant portion of the usable frequency range. For the bandpass configuration of the

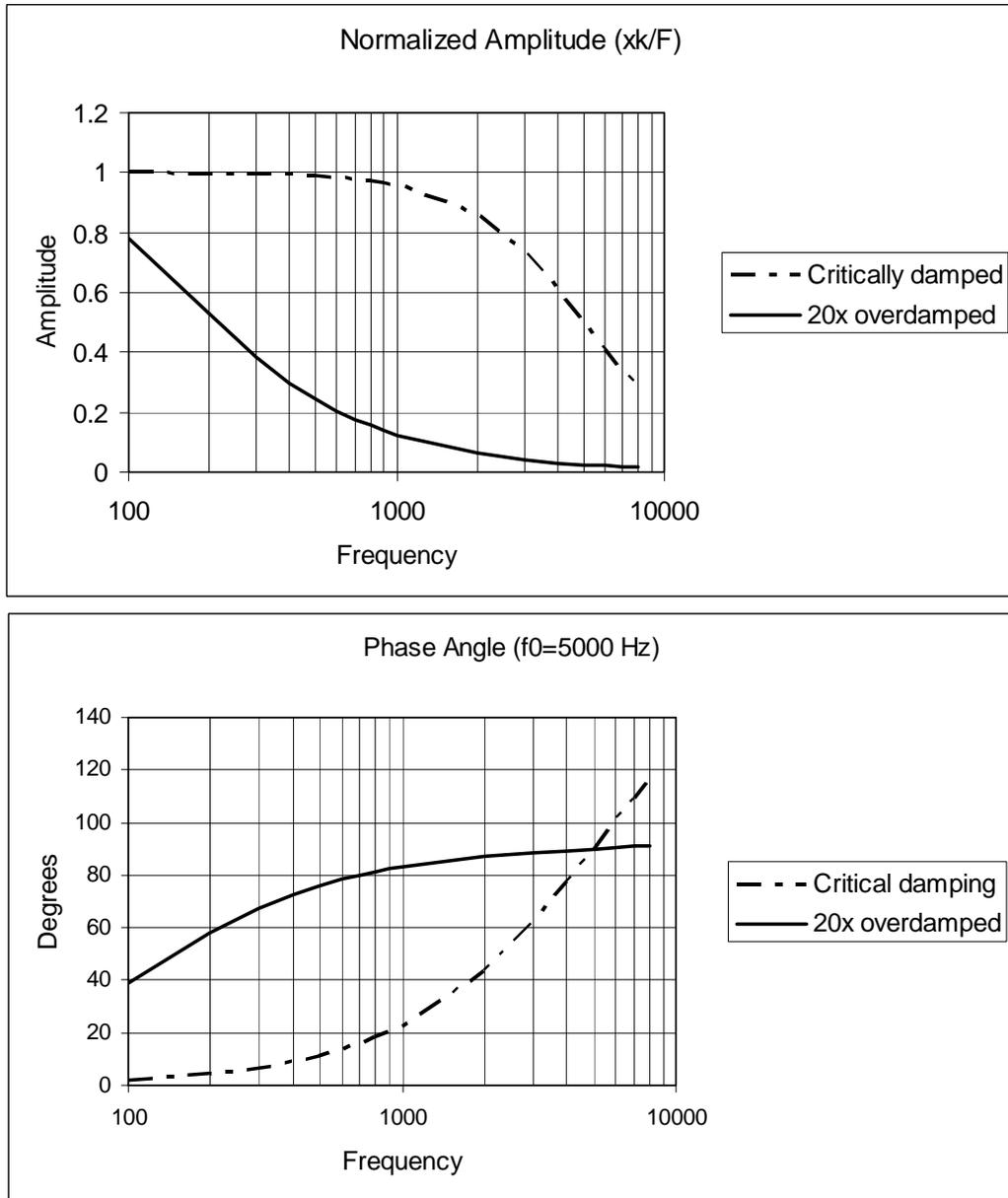


Figure A2-3a (top) Normalized amplitude ( $Xk/F$ ) and A2-3b (bottom) phase shift of a simple harmonic oscillator with stiffness constant  $k$ , driven by a force of magnitude  $F$ , resulting in a displacement amplitude  $X$ . The characteristic frequency is 5000 Hz. Two cases are shown, critical damping, and over-damped by a factor of 20.

advanced transducer, a constant 90-degree phase shift would be expected across the entire pass band.

Above about 5000 Hz, the mass of the diaphragm limits the response of the system and the transmission of sound becomes strongly attenuated. Below some hundreds-of-Hz, due to horn geometry, sound propagation is again attenuated. If the horn is not simple, but has bends, additional high frequency limitations may apply as well. A short horn, or one with discontinuities, may also have resonances.

The common “rule-of-thumb” [23] was that a horn-and-diaphragm system could cover only up to one decade in response. Archival photographs of studio recording equipment from the acoustic period, show conical horns in use. Fig. A2-5 is an example. If this is typical then these recordings would benefit, in playback, by a low frequency boost which reverses the conical horn response [31]. The one decade “rule-of-thumb” would tend to apply in these case due to the low frequency roll-off of the conical horn. The extended low frequency response of the system discussed by Maxfield and Harrison most likely benefited from the exponential horn shape to exceed this limitation.

A more accurate modeling of the properties of *recording* horns, and transducers, used in the acoustic recording era, and applied to the problem of sound reproduction, could be the topic of a future study. Such models could be used to augment the results presented here. An actual improvement in the recovered fidelity might emerge from such an effort.

To the extent that the conditions C1 and C2, stated in this appendix, are not met, sound amplitude may be attenuated, or, in the case of resonances, boosted at particular frequencies.

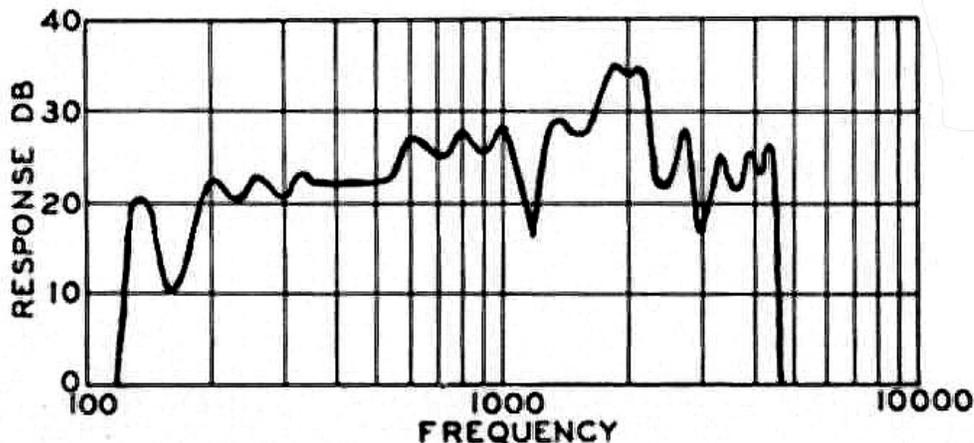


Figure A2-4: The frequency response of a particular mechanical (playback) phonograph discussed and analyzed by Maxfield and Harrison [19]. (The data shown here was re-plotted by Olsen [17]. See reference [20] for a comment on this.). In this case a folded exponential horn was used, leading to a sharp cut off just above 100 Hz. As discussed, no response is seen above 5000 Hz. The input response of a mechanical recorder of similar construction would be similar as well.



Figure A2-5: Recording at the Edison Studio, 79 Fifth Avenue, New York City, Jacques Urlès singing, Sodero conducting, March 30, 1916. Note the narrow conical horn of approximate length 1 m in use. Photo courtesy of the Edison National Historic Site, used by permission.

### **Acknowledgements**

This work was supported by the Laboratory Technology Research Program (SC-32), within the Office of Science, U.S. Department of Energy under Contract No.DE-AC03-76SF00098.

The authors wish to thank and acknowledge a number of individuals for assistance in this work. Bill Klinger provided advice and cylinder samples. Adrian Cosentini of Vidipax and Richard Martin and Meagan Hennessey of Archeophone Records provided stylus transcriptions of source material. Mark Roosa, Peter Alyea, Lawrence Appelbaum, and Samuel Brylawski, of the Library of Congress provided advice, encouragement, and samples. Peter Copeland offered interesting remarks on the aliasing issues. George Horn of Fantasy Records in Berkeley, California critiqued the sound quality and offered valuable advice on equalization. Gerald Fabris of the Edison National Historic Site provided information on horn geometries used in the acoustic era.

Sample scans were provided by STIL SA (Aix en Provence, France), Veeco Instruments Incorporated (Tucson, Arizona), and Microphotonics Incorporated (Irvine, California). The authors wish to acknowledge the assistance of the staff at each of these companies.

### **Footnotes**

1. The audio clips can be accessed at <http://www-cdf.lbl.gov/~av/>

## References

- [1] Hart, M., **J.Audio Eng. Soc.**, Vol. 49, No. 7/8, 2001 July/Aug
- [2] **Folk Heritage Collections in Crisis**, published by The Council on Library and Information Resources, Washington, D.C., 1996, ISBN 1-887334-82-3
- [3] Fox, Barry, "Not Fade Away", **New Scientist**, 1<sup>st</sup> March 2003, pp 40-43
- [4] Fadeyev V. and Haber, C., **J. Audio Eng. Soc.**, vol. 51, no. 12, pp. 1172-1185 (2003 Dec.).
- [5] Fremer, M., "Analog Corner", **Stereophile**, December 2003
- [6] T. Iwai, T. et al., **Appl. Opt.**, vol. 2, no.5, (1986) pp. 597-604
- [7] Penn, W. and Hanson, M.J., "The Syracuse University Library *Radius Project*: Development of a non-destructive playback system for cylinder recordings" **First Monday**, volume 8, number 5 (May 2003)
- [8] Poliak, J., Robert, P., and Goy, J., Optical Fibre Turntable for Archive Records, **Proceedings of the 92<sup>nd</sup> Convention AES**, Vienna, Austria.
- [9] Petrov, V.V., et al., "Optomechanical method of Edison cylinders sound reproduction" **Audio Eng. Soc. Preprint** (Preprint 4491), Munich 1997, March 1997
- [10] Cavaglieri, S., Johnsen, O., and Bapst, F., **Proc of AES 20<sup>th</sup> International Conference**, Budapest, Hungary 2001, Oct 5-7.
- [11] Stanke, G. and Paul, L., **Inform. Serv. & Use**, 15 (1995) pp. 289-301
- [12] Powell, J. R. and Stehel, R. G., **Playback Equalizer Settings for 78 RPM Recordings**, Gramophone Adventures, Portage, MI., 1993
- [13] EIA Specification RS-211-Rev D and Rev C.
- [14] Read, Oliver, **The Recording and Reproduction of Sound**, H.Sams, Indianapolis, 1953
- [15] Langford-Smith, F., editor, **Radiotron Designer's Handbook**, Fourth Edition. The Wireless Press, Australia, 1953.
- [16] Certain cylinder parameters were provided by W. Klinger of the Association of Recorded Sound Archives (ARSC), private communication
- [17] Corle, T.R., and Kino, G.S., **Confocal Scanning Optical Microscopy and Related Imaging Systems**, Academic Press, San Diego, 1996

- [18] Cohen-Saban, J. et al, **Proc. SPIE**, 4449, 178-183, (2001)
- [19] Davidson, M. et al, **Proc. SPIE**, 775, 233-247, 1987
- [20] Caber, P.J., **Applied Optics** 32, 3438-3441 (1993)
- [21] Klinger, W. “Stylus Shapes and Sizes, Preliminary Comments on Historical Edison Cylinder Styli”, private communication.
- [22] Morse, P.M., **Vibration and Sound**, 1<sup>st</sup> Edition, McGraw-Hill, New York, 1936
- [23] Morse, P.M., **Vibration and Sound**, 2<sup>nd</sup> Edition, McGraw-Hill, New York, 1948
- [24] Molloy, C.T., **J. Acoust. Soc. Am.**, 22:551-557 (1950)
- [25] Blackstock, D.T., **Fundamentals of Physical Acoustics**, Wiley-Interscience, Hoboken, N.J., 2000
- [26] Olson, H.F., **Acoustical Engineering**, Van Nostrand, Princeton, N.J., 1957
- [27] Pierce, A.D., **Acoustics**, McGraw-Hill, New York, 1981
- [28] Maxfield, J.P. and Harrison, H.C., **The Bell System Technical Journal**, Volume V, No.3, July 1926, pp. 493-524
- [29] The data plotted in Fig. A2-3 was first published by Maxfield and Harrison [28] and subsequently re-plotted by Olson [26]. Olson’s plot is shown here. However Olson claims the response shown is for a recording with a low frequency attenuation of 6 dB/octave below 800 Hz. Comparing this plot to the original in reference [28], the present authors cannot see how Olson’s claim is supported. In either case the cut-off frequencies are not affected, only the shape from cut-off to 800 Hz.
- [30] Webster, A.G., **Proc. Nat. Acad. Sci. of the U.S.A.**, No. 7, July 15, 1919, pp. 275-282
- [31] George Horn, private communication, emphasized the apparent attenuation of low frequency content in these recordings.