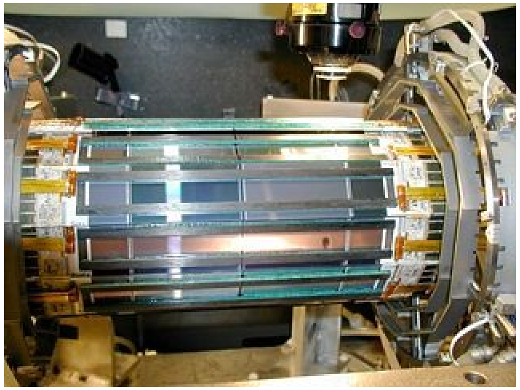


The Reconstruction of Mechanically Recorded Sound by Image Processing

Vitaliy Fadeyev and Carl Haber
Physics Division
Lawrence Berkeley National Lab
Berkeley, California USA

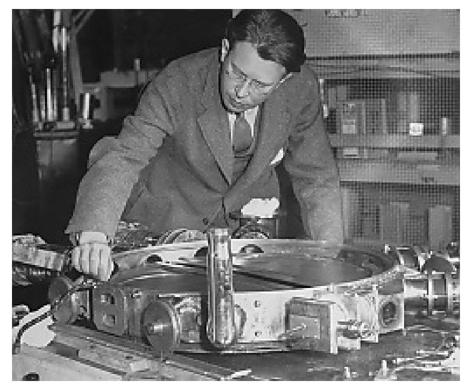
Outline

- Introduction
- Background
  - History
  - Issues
- The Imaging Method
  - Basic Idea
  - Advantages
  - Detailed Discussion of Elements
- Context
- Proof of Concept Test
- Conclusions

Lawrence Berkeley National Lab

www.lbl.gov

- Founded in 1931 by E.O.Lawrence
- Oldest of US National Labs
- Operated by the University of California for the US DoE
- 4000 Staff, 800 Students, 2000 Guests
- 14 Research Divisions including
  - Physics, Nuclear Science
  - Materials, Chemical Science
  - Life Sciences, Physical Bioscience
  - Energy and Environment, Earth
  - Computing
- Major user facilities-
  - Advanced Light Source
  - Nat. Center for Electron Microscopy
  - Nat. Energy Research Super Computer Center

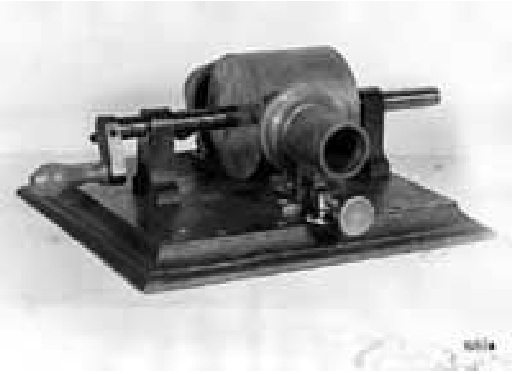


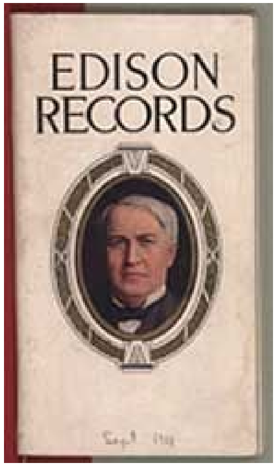
Introduction

- We have investigated the problem of optically recovering mechanical sound recordings **without contact** to the medium
- This work may address some concerns of the preservation, archival, and research communities:
  - The reconstruction of damaged media
  - The playback of delicate media
  - Mass digitization and storage
- Message to take away from today's presentation:
  - Optical techniques can produce acceptable reproductions and some improvements
  - Measurement, data storage, and computing technologies may be approaching performance levels required for this application
  - Cross disciplinary interactions can be of real value here

History

- Recorded sound was introduced by Edison in 1877 who embossed audio data onto metal foil
- A variety of media and methods used since then
  - Wax cylinders with vertical modulation
  - Shellac disks with vertical or lateral modulation
  - Vinyl disks with lateral or 45/45 (stereo) modulation
  - Acetate instantaneous recordings, lateral modulation
  - Metal reversed stampers (disks) and galvanos (cylinders)
  - Magnetic tape and wire
  - Compact digital disks (CD)
- Essentially all pre-1948 recordings were mechanical





Issues and Concerns

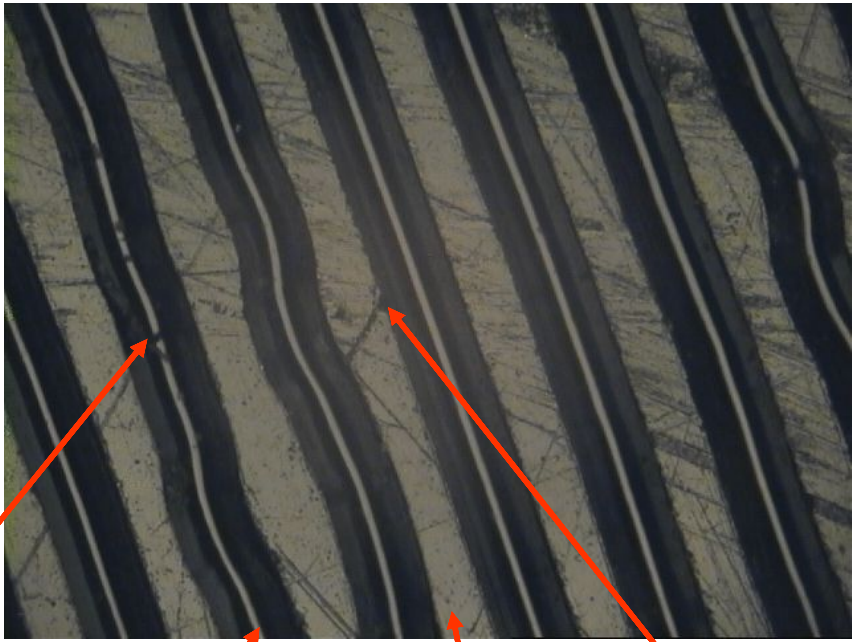
- Can recordings be mass digitized in an efficient way to enable preservation and access for future users?
  - Diverse formats
  - Damaged samples which require intervention or are impossible to play at all
  - Further damage to delicate samples
- Can samples which are of particular value to someone be recovered or improved in a useful way?

A Non-contact Method

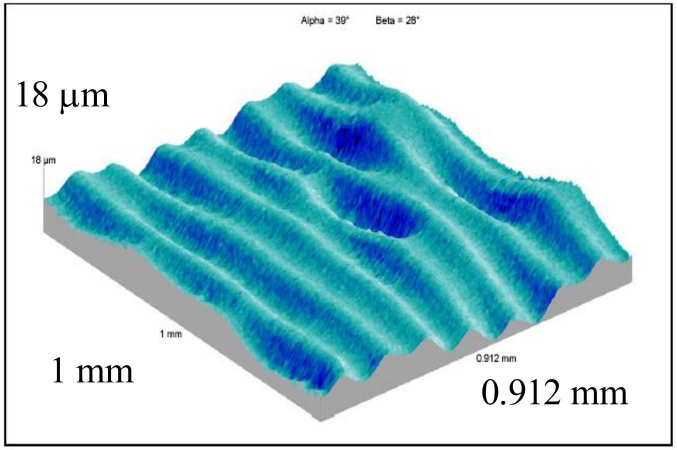
- Using optical techniques, the pattern of grooves or undulations in a recording surface can be imaged.
- To cover a surface (thousands of) sequential views can be acquired.
- Views can be stitched together.
- The images can be processed to remove defects and analyzed to model the stylus motion.
- The stylus motion model can be sampled at a standard frequency and converted to digital sound format.
- Real time playback is not required de-facto, method is aimed at reconstruction and digitization.

Example of Groove Image

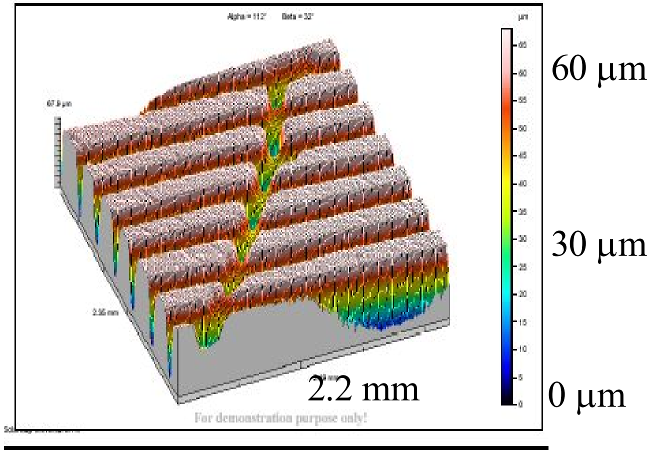
- Two dimensional view (2D)
- Field is 1.39 x 1.07 mm
- Groove width is 160 μm
- Lighting is perpendicular to surface.
- Bright line is groove bottom.
- Acquired with a digital camera and magnification.



Example of 3D Images: Surface Profiles



Edison "Blue Amberol" cylinder



78 rpm disk with large scratch

- Fields are few mm^2
- Acquired with confocal laser scanning probe
- Suitable for detailed groove reconstruction and vertical modulated recordings

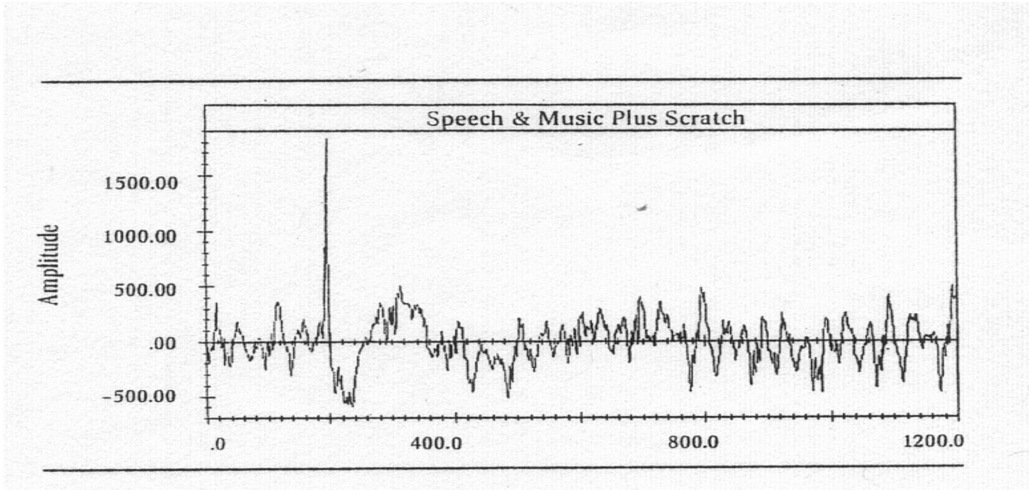
Advantages of Imaging Method

- Delicate samples can be played without further damage.
- Broken samples can be re-assembled virtually.
- Independent of record material and format – wax, metal, shellac, acetates…
- Effects of damage and debris (noise sources) can be reduced through image processing. Scratched regions can be interpolated.
- Discrete noise sources are resolved in the “spatial domain” where they originate rather than as an effect in the audio playback.
- Dynamic effects of damage (skips, ringing) are absent.
- Classic distortions and systematics (wow, flutter, tracing and tracking errors, pinch effects etc) are absent or removed as geometrical corrections
- No mechanical method needed to follow the groove.
- Suggests a method for mass digitization, full 3D maps.

Material and Format Independence

- Image of metal "stamper" used to mold plastic record.
- Molding technology is obsolete.
- Can be played with special "cowboy" stylus which rides ridges
- But easily imaged in 2D

Dynamic Effects



- Continuum of imperfections up to a full skip.
- Mechanical stylus responds dynamically to these imperfections
- Result is a "ringing" which may persist as an artifact if only clicks are removed in a standard digital remastering
- Plot is from Rayner, Vaseghi, and Stickells, FIAF Joint Technical Symp, "Archiving the Audio-Visual Heritage", May 20-22, 1987

Relationship to Other Work

- Laser turntables (ELP): these devices work off a reflected laser spot only and are susceptible to damage and debris and sensitive to surface reflectivity.
- Stanke and Paul, "3D Measurement and modelling in cultural applications", Inform. Serv. & Use 15 (1995) 289-301: use of image capture to read cylinder "galvanos", depth was sensed from greyscale in 2D image – lacks resolution required for good reconstruction.

Relationship to Other Work

- S.Cavaglieri, Johnson, and Bapst, Proc of AES 20th International Conference, Budapest, Oct 5-7, 2001: use of photographic contact prints and scanner to archive groove pattern in 2D – insufficient resolution, no 3D analog.
- O.Springer (http://www.cs.huji.ac.il/~springer/): use of desk top scanner on vinyl record – lacks resolution for useful reconstruction, no notion of magnification nor image processing to improve data.
- Penn et al at Belfer Lab: no information available, some sort of real time interferometric playback system (??), no notion of image processing

Elements

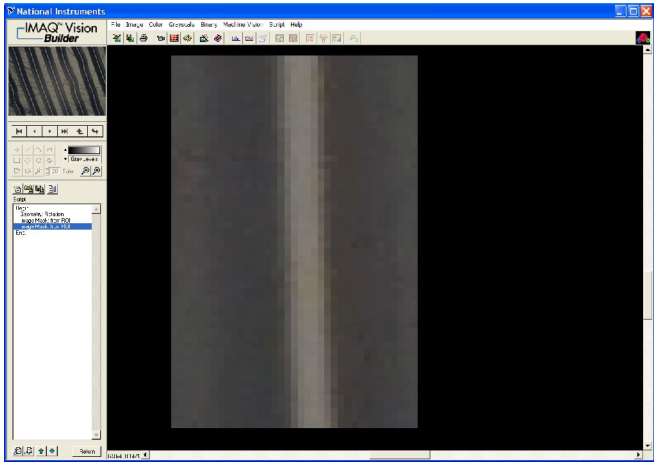
- Imaging
  - Requires sufficient resolution to measure the minimal undulations of the surface. The scale is ~0.2-1 μm (lateral) for the pre-vinyl era.
  - Lateral recordings can be imaged in 2D or 3D.
  - Cylinders, vertical disks, and complex damage structures require 3D
  - Method must be fast enough for efficient application.
- Processing
  - Effective algorithms to capture information content of grooves and reject noise.
- Analysis
  - Transform spatial pattern of groove position into audio response – physical model.
- Conversion to standard audio formats

Parameter	**78 rpm, 10 inch**	**Cylinder**
Revolutions per minute	78.26	80-160
Max/Min radii (inches)	4.75/1.875	2 – 5 fixed
Area containing audio data (mm^2)	38600	16200 (2”)
Total length of groove	152 meters	64-128 meters
Groove width at top inches (μm)	0.006 (160)	variable
Grooves/inch G_d	96-136	100-200
Groove spacing inches	0.007-0.01	0.005 – 0.01
Ref level peak velocity@1KHz	7 cm/sec	NA
Max groove amplitude (inches)	0.004-0.005	
Noise level below reference, S/N	17-37 dB	
Dynamic range	30-50 dB	
Groove max amplitude at noise level	1.6 - 0.16 μm	

Imaging Methods

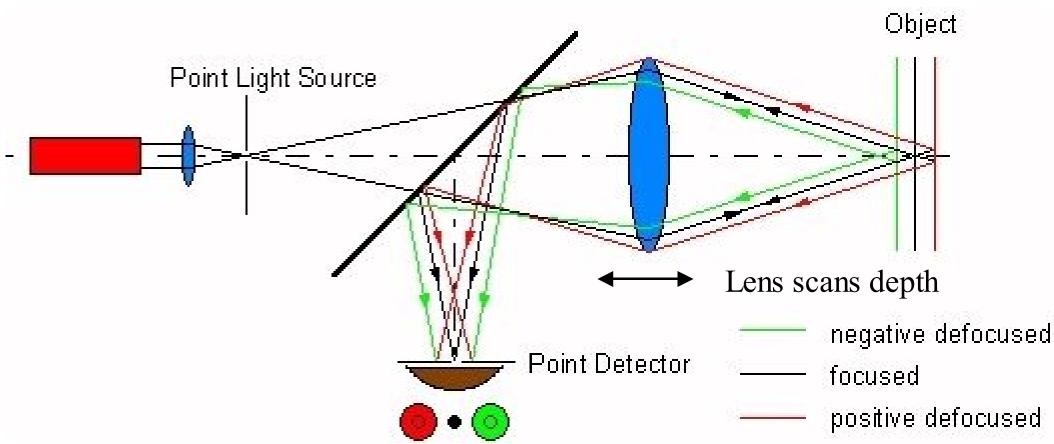
- Electronic Cameras – 2D or horizontal only view, frame based
- Confocal Scanning – 3D or vertical+horizontal view, point based
- Chromatic sensors – 3D, point based
- White Light Interferometry – 3D, frame based

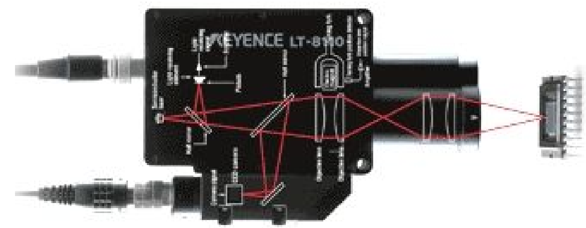
Electronic Camera

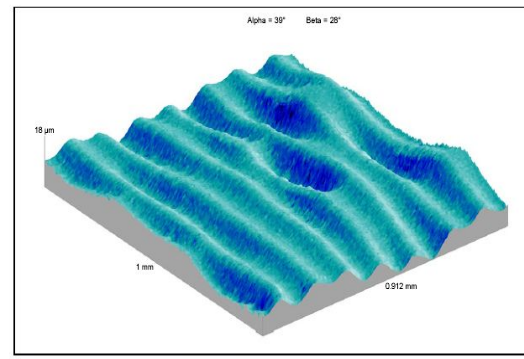


- CCD or CMOS image sensor
- Practical field of view is 0.7 x 0.54 mm
- Camera contains 768 x 494 pixels, up to few Mega-Pixels
- 1 pixel = 0.91 x 1.09 microns on the record surface
- Magnification and pixel size yield sufficient resolution for audio data measurement due to pixel interpolation
- Entire frames acquired at 30-1000 fps (up to 10,000 fps possible)

Laser Confocal Scanning Microscope







- Acquires image point by point
- Vertical resolution is ~0.1 micron
- Commercially available
- Point light source is reflected from measurement surface and detected at point detector. Only in-focus rays give signal in detector.
- Complete depth scan occurs 1400 times/sec for each point, averaging?
- Horizontal resolution set by point size 1-2 microns

Chromatic Confocal Sensor

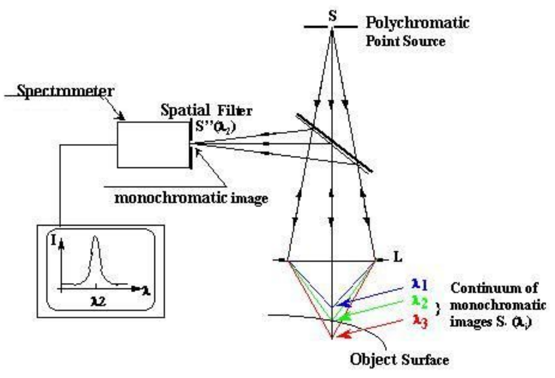
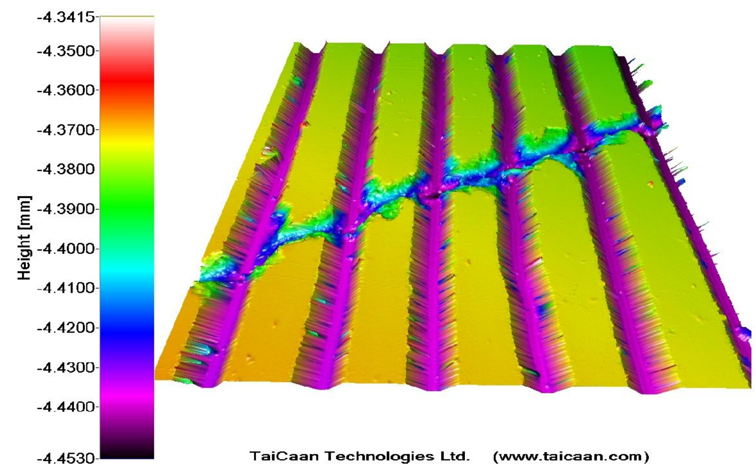
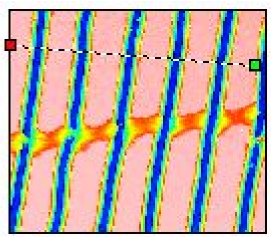


Figure 4 : Chromatic Confocal setup for 3D Surface metrology

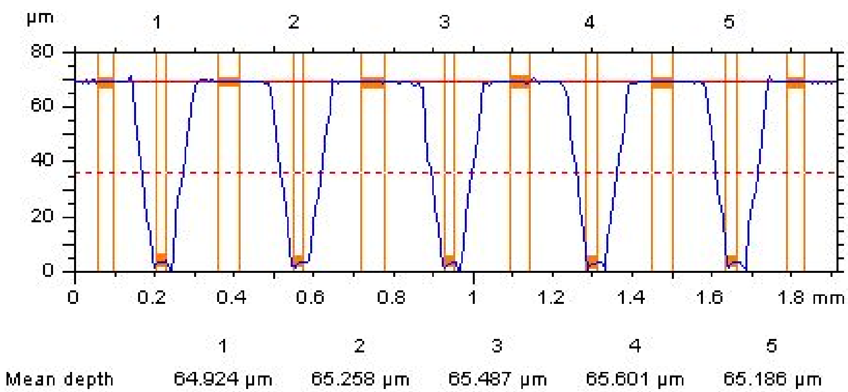


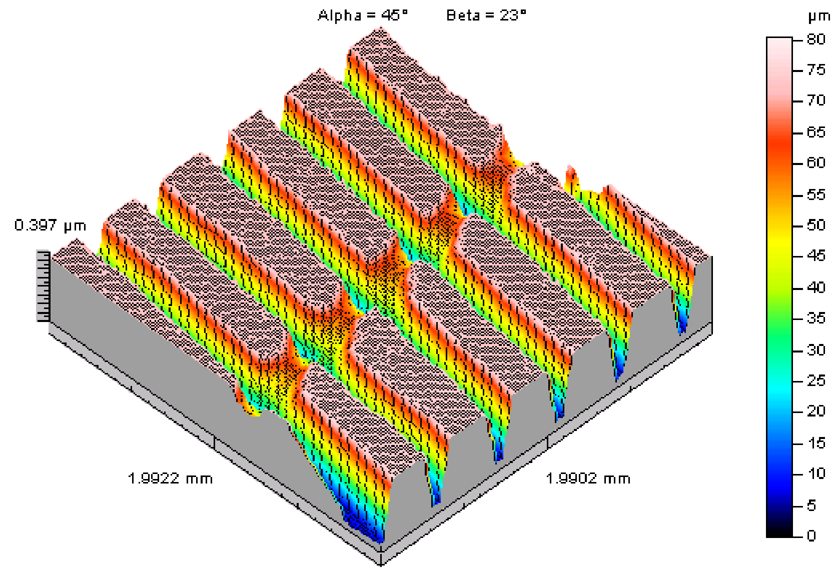
- Different colors image to different depths on the sample simultaneously – potential for faster scan, up to 4000 points/second may be possible, depending on surface
- Signal detected by spectrometer (color sensitive).
- Issue of reflection off sloped surfaces – data loss

Chromatic Confocal Sensor

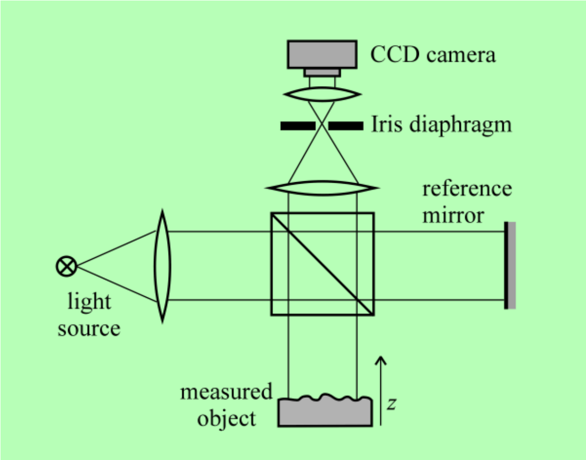


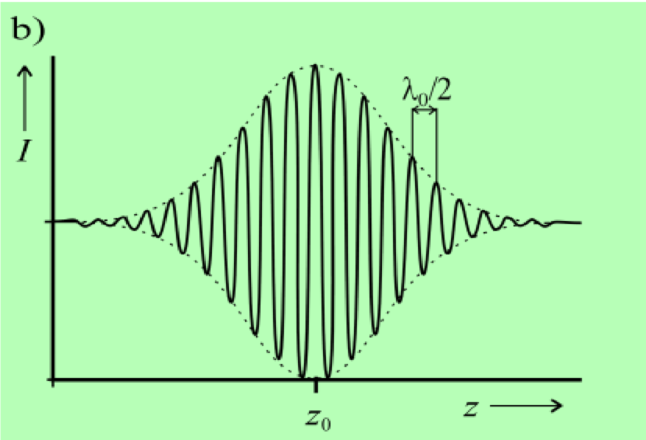
Extracted profile





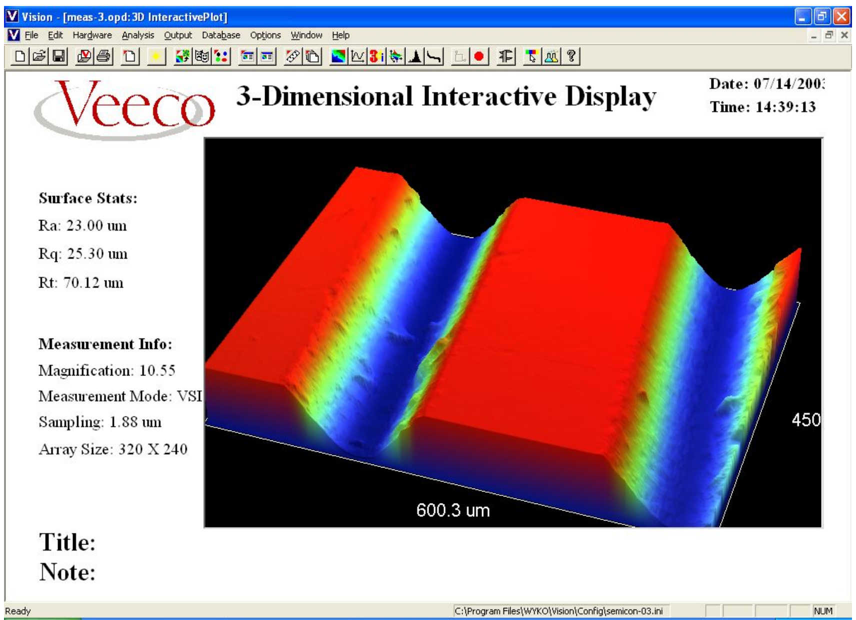
White Light Interferometry





- Interference principle – waves combine constructively for equal distances traveled.
- Sample is scanned in depth and imaged by frames, at each depth a different interference pattern is found. Frame size is typically 0.6 x 0.4 mm.
- Horizontal resolution is like 2D electronic camera but takes many vertical slices.
- Current systems run 60 fps and require 1-20 seconds per view for vertical scan.
- Scan time depends upon surface angle and reflectivity.
- Potential for faster systems with high fps cameras?

White Light Interferometry



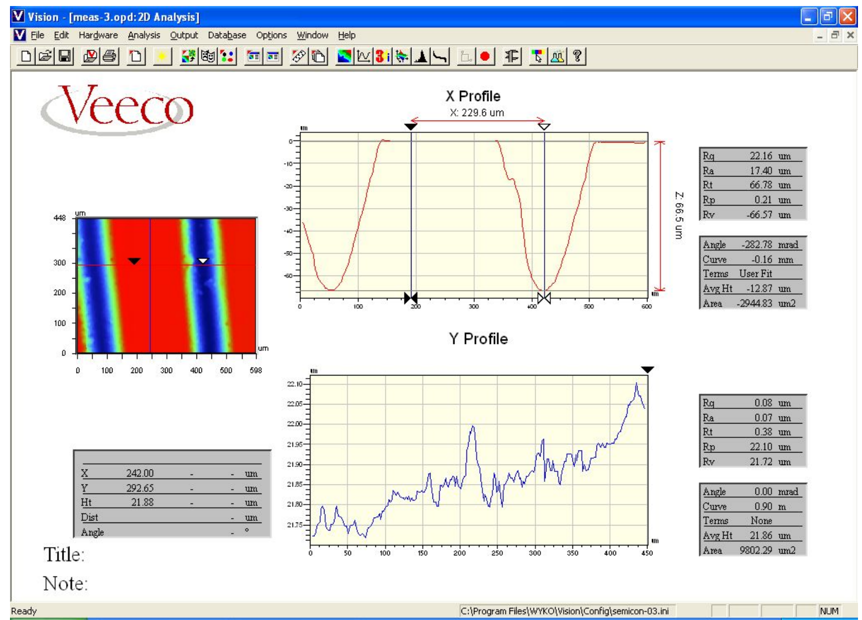


Image from 78 rpm record surface

Issue of data loss from groove sides – angle effect, time required to measure

White Light Interferometry: Scratch

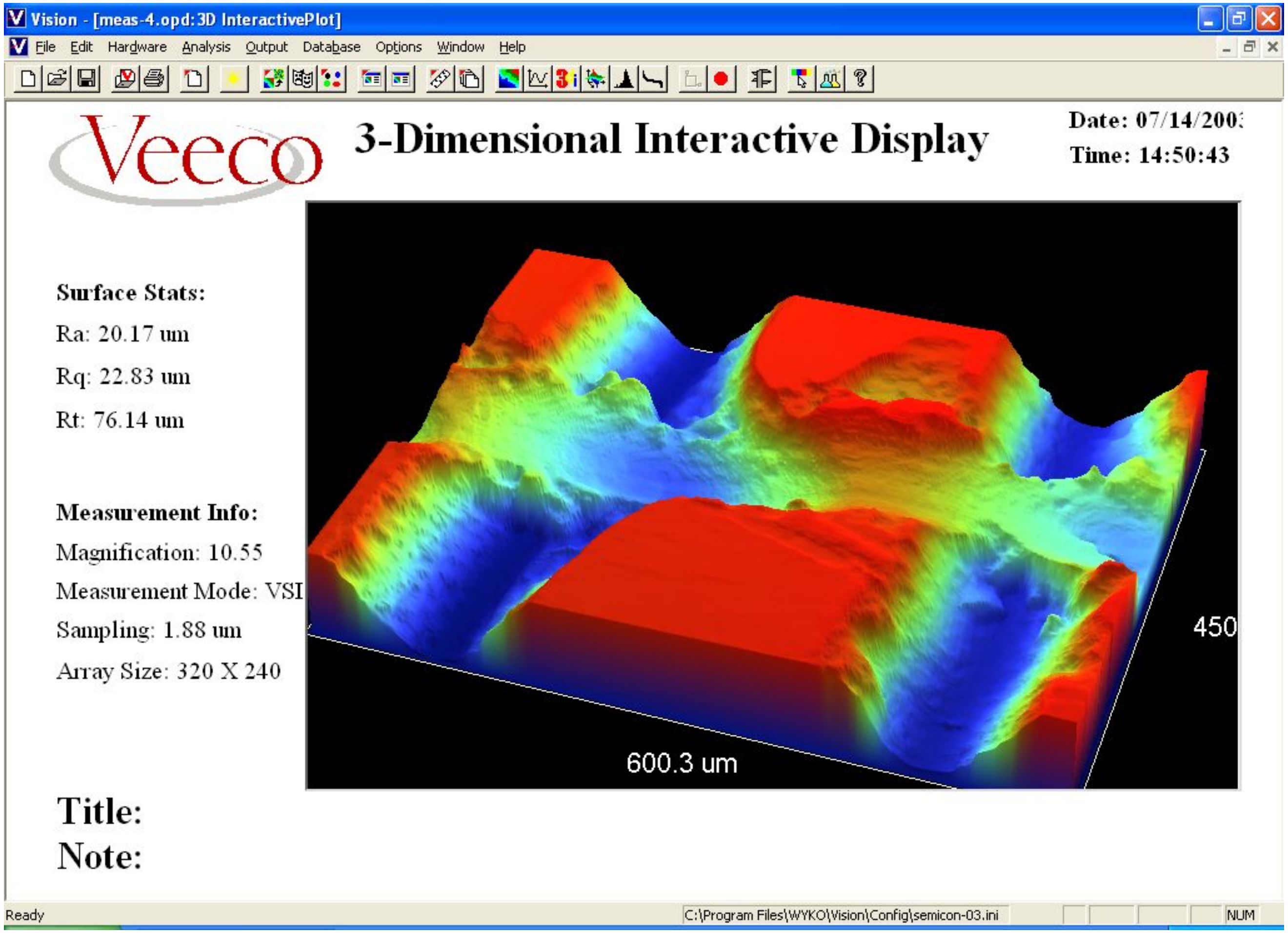
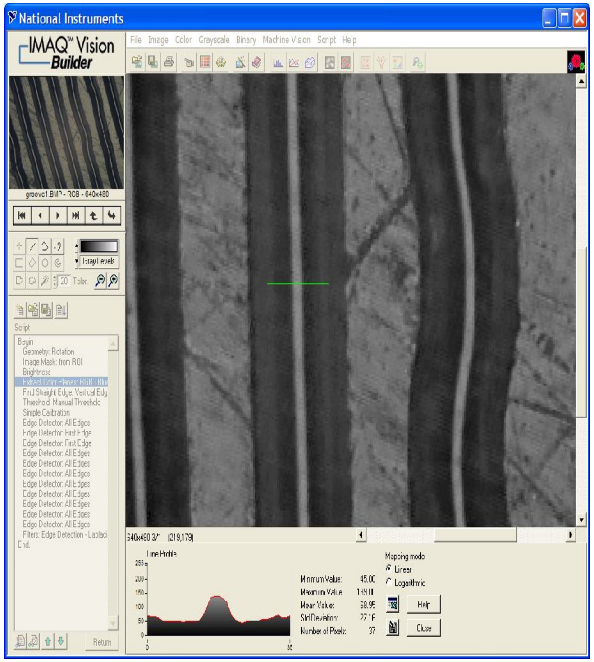
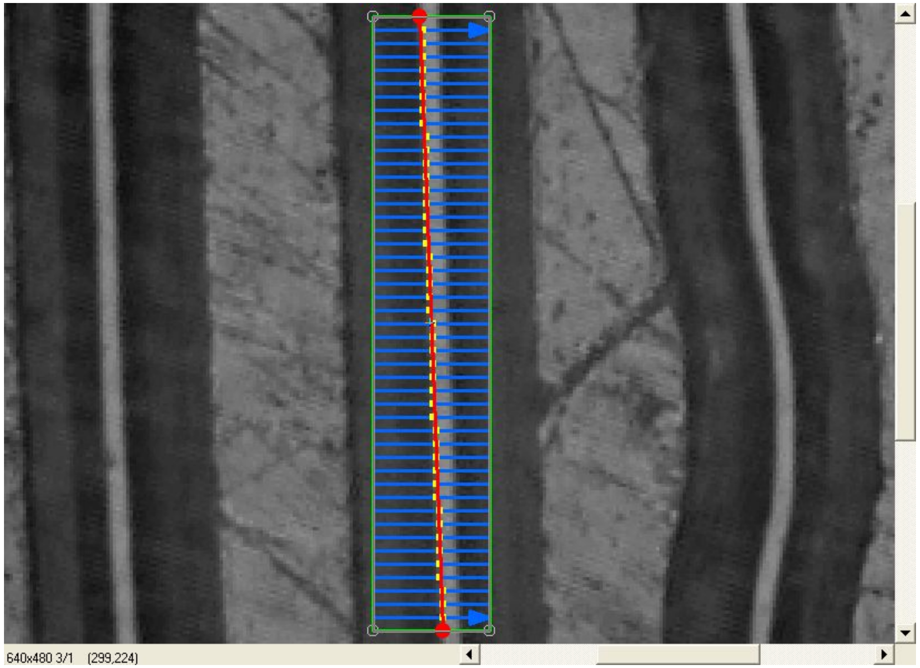


Image Processing

- Each image consists of a set of "pixels"
- A pixel is an intensity or height measurement at a horizontal position (x,y)
- Image processing is a collection of mathematical operations performed on the pixels to extract information from the image, including:
  - Measure profiles and distances
  - Find transitions (edge detection)
  - Shape detection (morphology) and transformation
  - Alter values based upon neighboring pixels (filtering)

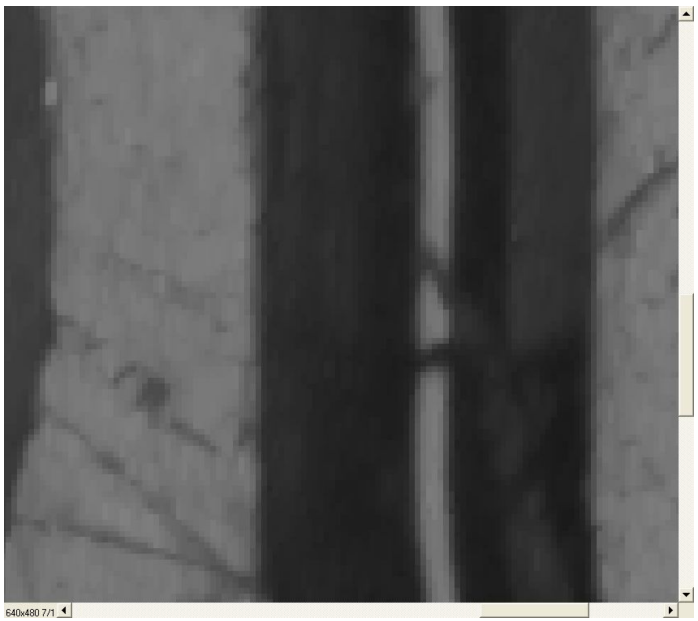
Image Processing





- Brightness profile in grayscale image across a feature
- Edge detection along a series of lines

Image Processing



Effect of iterated "dilation" operator on 1x3 pixel clusters
Dust particle is removed from the image

Signal Analysis

- Once the groove pattern is properly imaged and acquired an analysis is performed to extract the audio data.
- Based upon the physics of the recording process.
- Groove data is already in digital form so analysis methods are numerical.

Physics of Mechanical Sound Recording

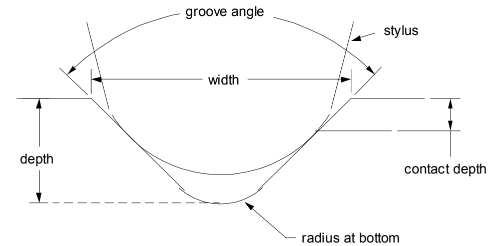
- Playback stylus rides in groove
- For magnetic recording and playback stylii, signal is proportional to stylus velocity

$$A_p = \frac{v_p}{2\pi f}$$

"constant velocity recording"

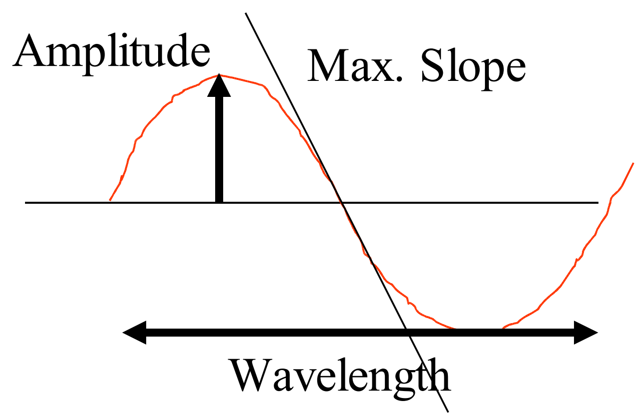
- Mediated by equalization scheme to attenuate low frequencies and boost high frequencies
- Levels are compared by amplitude

$$dB = 20\log\left(\frac{v}{v_{ref}}\right)$$



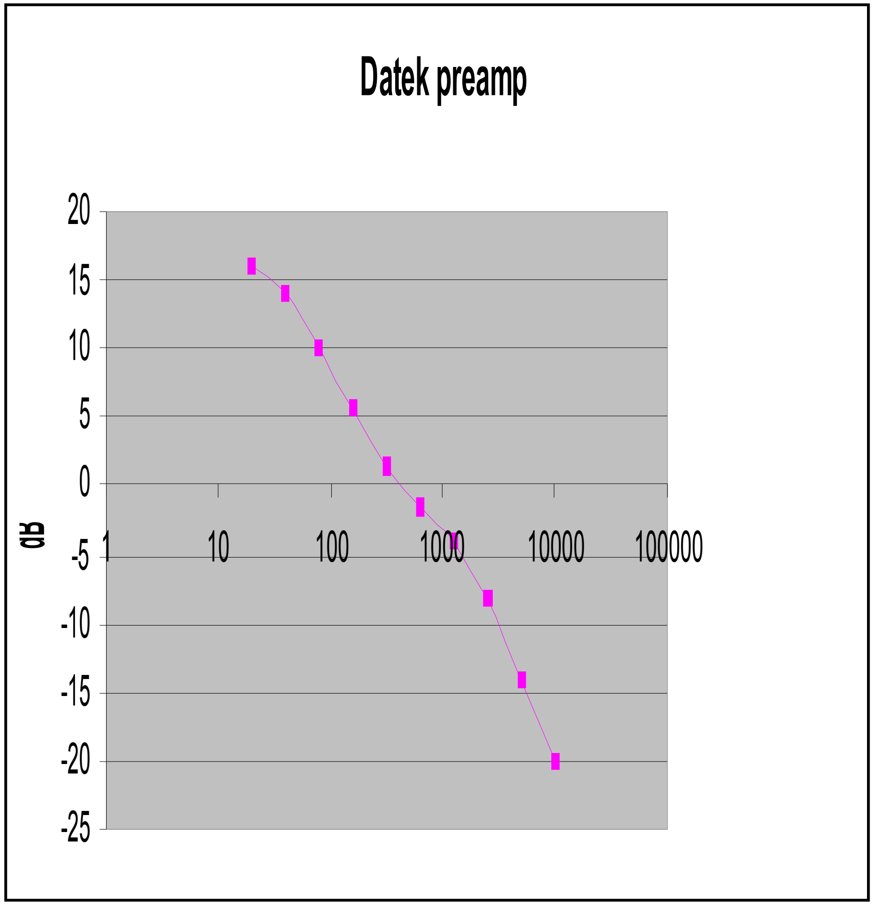
parameter	78 rpm coarse	33 1/3 ultrafine
width	0.006 - 0.008	0.001
depth	~0.0029	~0.0006
contact depth	0.0008	0.0004
radius	0.0015-0.0023	0.00015
angle	82 - 98	87 - 92

units are inches or degrees



Equalization and Reconstruction

- The constant velocity recording characteristic is modified as follows:
  - Low frequencies are attenuated to avoid excessive groove excursion
  - High frequencies are boosted above surface noise floor
- Playback is equalized to compensate for this.
- Optical reading of groove displacement is differentiated numerically to determine stylus velocity and then equalized



Speed of Method (1)

- Frame based methods
  - Overlapping images to enable stitching – 20%
  - Assume 0.7 x 0.54 mm = 0.378 mm^2 frame
  - 78 rpm disk: 38600 mm^2 = 123,000 frames, @30 fps: 1.2 hours/scan (realtime: 680 fps)
  - Cylinder: 16200 mm^2 = 51,000 frames, but vertical scan requires 1-10 seconds per frame: 14 -140 hours per scan (with 60 fps camera)

Speed of Method (2)

- Point scan based methods (confocal microscopes)
  - Very sensitive to number of points required for reconstruction
  - High density example: 4x4 μm grid:
    - 78 rpm disk: 152 meters x 160 μm: 1.5 billion points, @1000 points/second: 400 hours per scan
    - Cylinder: 16200 mm^2 : 1 billion points, @4000 points/second: 70 hours per scan (better surface angle)
  - Low density example: sample groove with 3 points across, at 8 μm intervals along length. Identify defects and return with high density selectively.
    - 78 rpm disk: 3 x 152 meters / 8 μm = 57 million points, @1000 points/second: 16 hours per scan
    - Cylinder ~3 hours per scan

Speed of Method - Comments

- Discussion only covered image acquisition, but data transmission, real time processing, and storage requirements are also significant.
  - Raw 2D images of 78 rpm disk: 570 Mbytes per second of audio data (88 Kbytes/sec in WAV file)
  - Immediate pre-processing (DSPs) could provide reduction.
- Only 2D camera is reasonably efficient for mass digitization (?). Slow scans OK for special reconstructions. But 3D required for cylinders.

Speed of Method - Comments

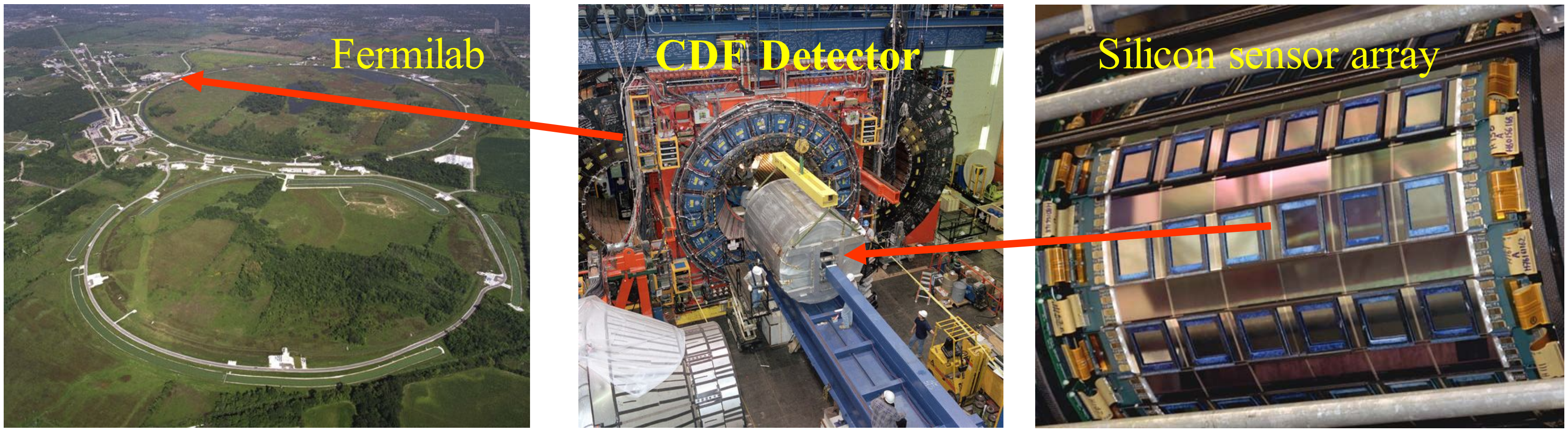
- 3D methods have excessive vertical resolution, perhaps some time could be recovered with relaxed vertical scan protocol?
- Speed of 3D surface profilers could increase with new technologies (faster cameras, higher frequency drivers…) recall – none of this was possible 10 years ago.
- Topic is ripe for collaboration with 3D surface profiling industry.

Context

- The methods and techniques described here may be unfamiliar in an audio engineering context.
- They are however widely used in other fields.
  - Automated inspection
  - Microbiology
  - Security and military
- From an experimental particle physics context this approach seemed natural.

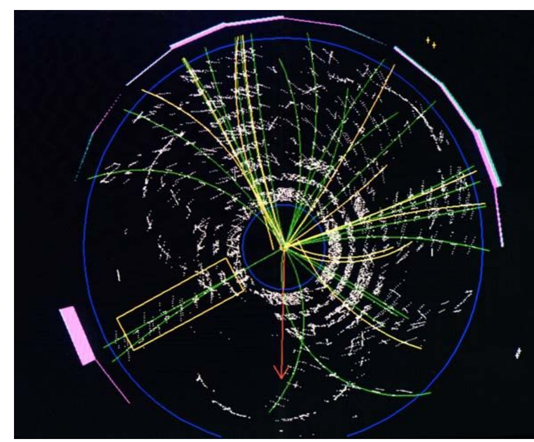
Particle Physics Methods

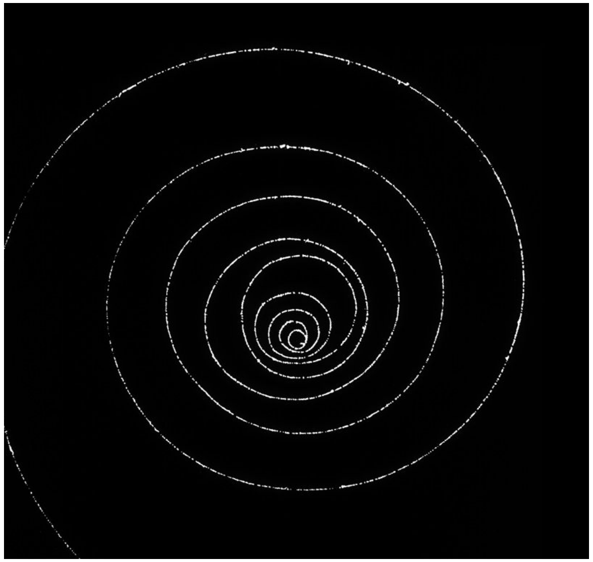
High energy accelerators are used to study basic nature of matter and energy and to re-create conditions of the early universe.



>Precision mechanical survey methods are required to build sensor array
>Massive data collection and analysis
>Pattern recognition and image processing to analyze signals and noise observed in detectors

Computer event display





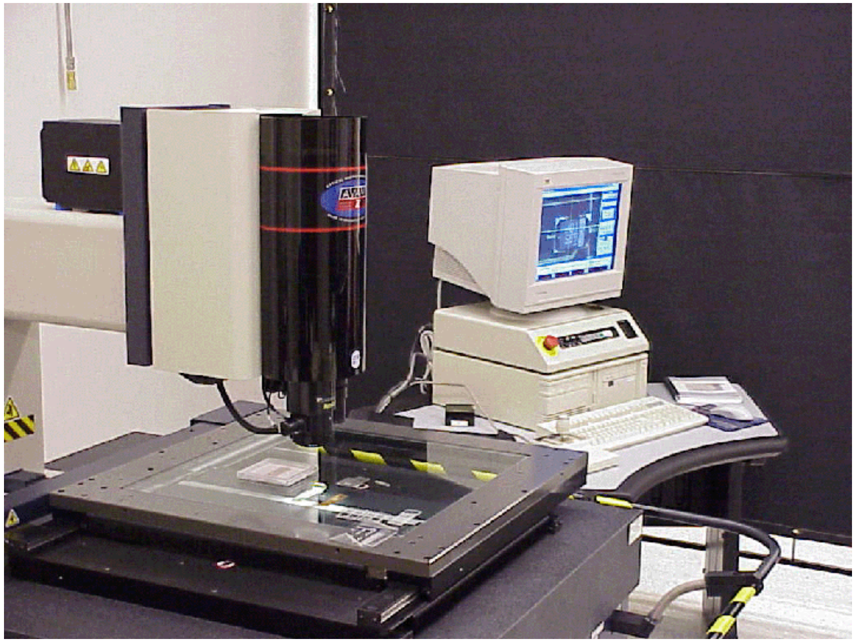
Photograph of bubbles formed along the trajectory of an electron as it loses energy in a “Bubble Chamber”.

Measurement of tracks in a particle detector is similar to following the groove in a mechanical recording – pattern recognition, noise reduction issues are familiar.

Measurement methods and precision required to build detectors are similar to that required for audio reconstruction

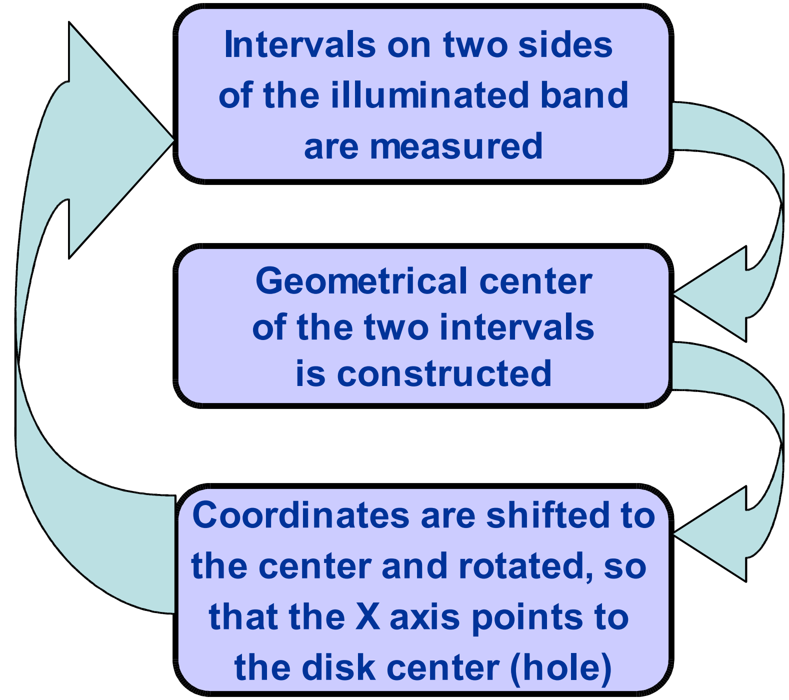
Test of Concept using 2D Imaging

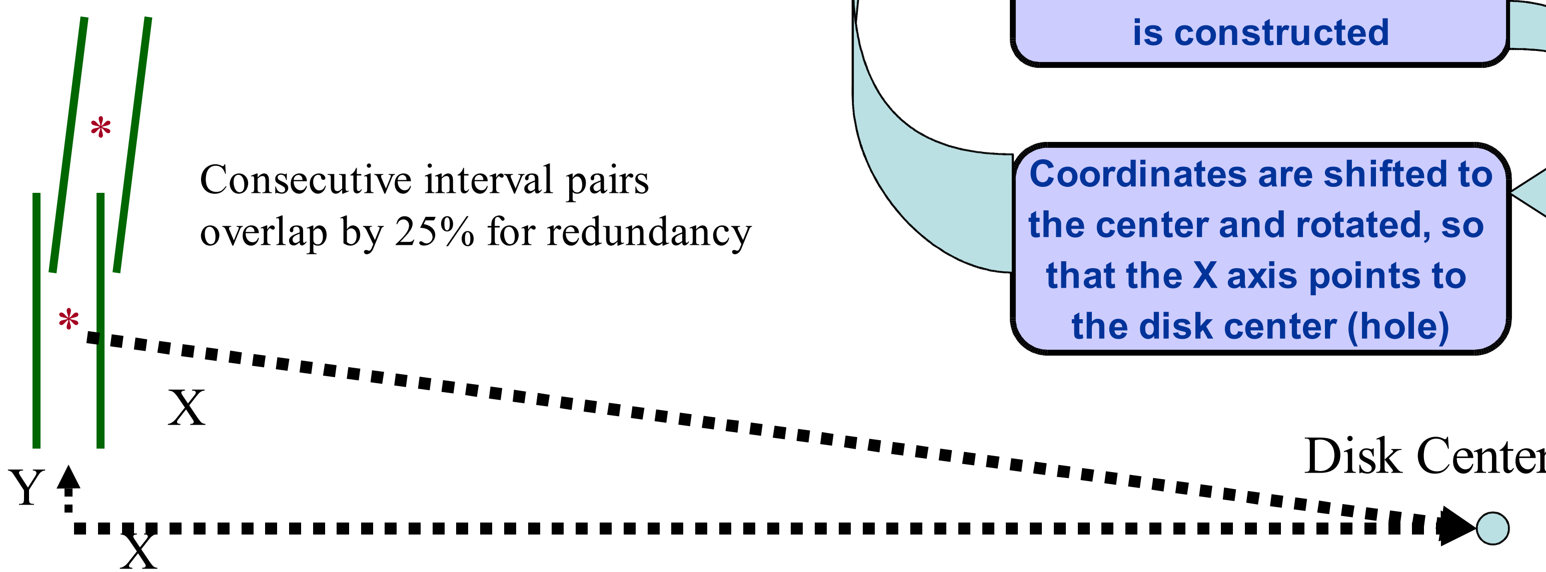
- Precision optical metrology system already in use for Particle Physics detector construction at LBL. Non-contact.
- “SmartScope” manufactured by Optical Gauging Products.
- System features zoom microscope with electronic camera and precision stage motion in x-y-z.
- Includes image acquistion with pattern recognition and analysis & reporting software
- Wrote program to scan grooves, report, and process data (offline).



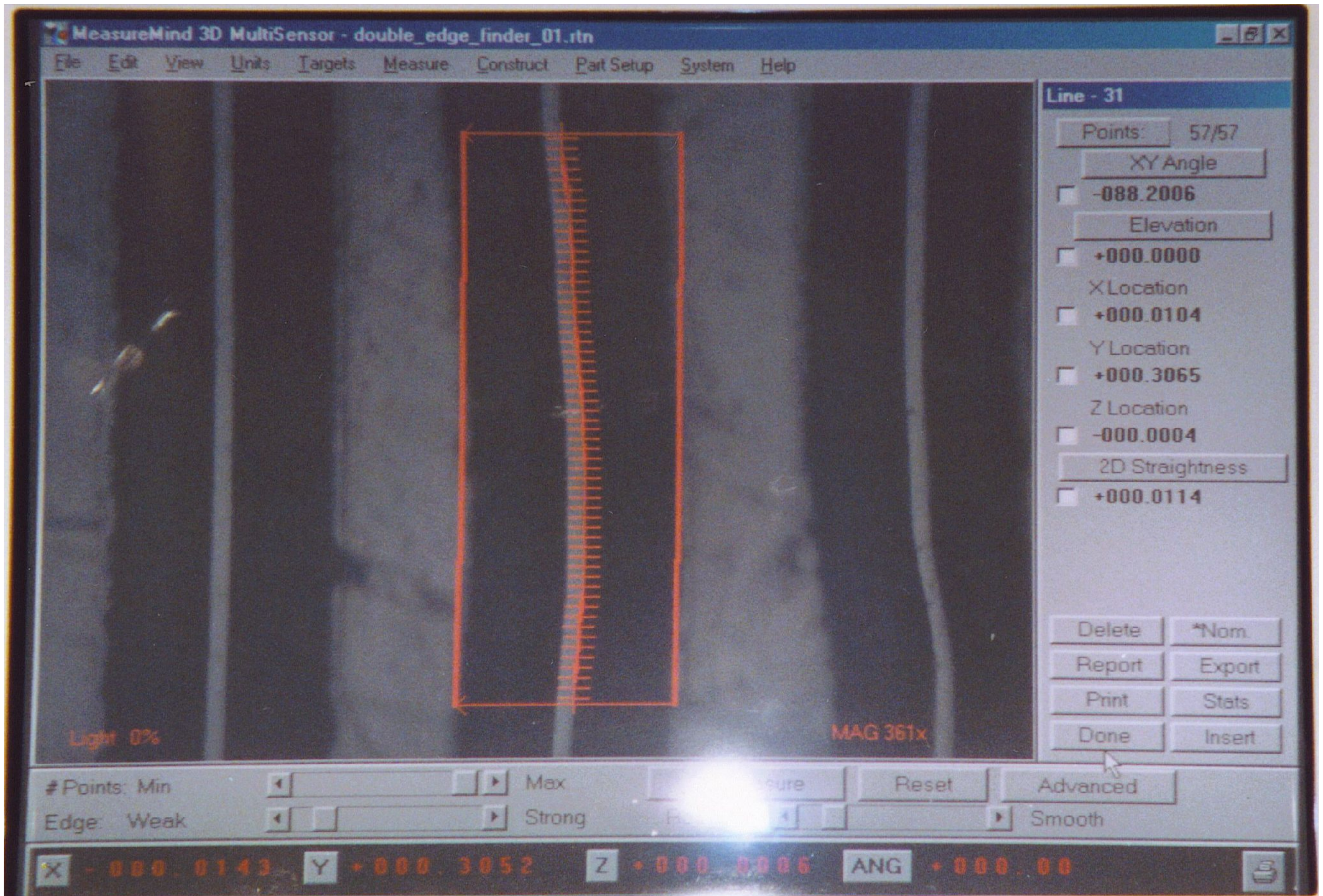
The SmartScope Program

An algorithm was devised to follow the groove approximately spiraling to the disk center:





Example of Smart Scope Edge Finding on Groove Bottom

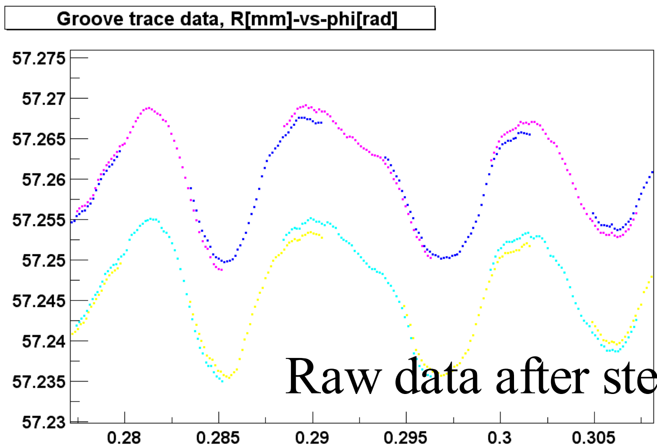


Technical Issues

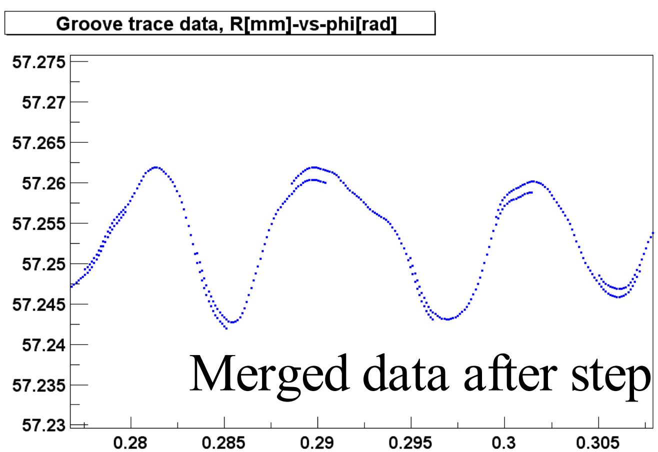
- System used was general purpose, not optimized for this application. It required ~40 minutes to scan 1 second of audio.
- No image recording (for offline use), system performed processing up front with built in software. Output data were found points along features.
- Automatic noise reduction due to rejection of dust or scratches in edge finding process. Also used bottom width measurement as a noise rejection tool offline.
- Number of points along groove, could be increased.
  - 8 μm steps = 66 KHz sampling
- Merging of adjacent frames
- Groove pattern must be differentiated numerically- algorithm selection

Offline Data Processing

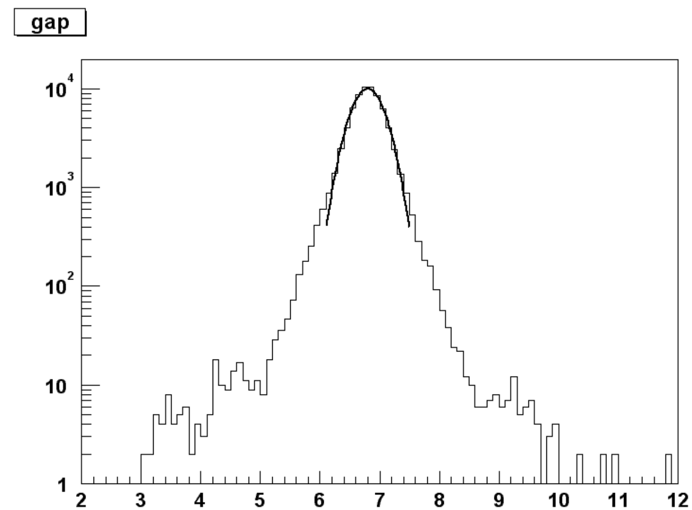
1. Reformatting data in one global coordinate system
2. Removal of big outliers
3. Filtering by selecting on the distance between interval pair; merging two sides into one.
4. Matching the adjacent frames
5. Fit the groove shape $R' = R0 + C*f' + A*\sin 2(f' + f0)$
6. Differentiation
7. Resampling; multiple runs addition; conversion to WAV format



Raw data after step (1)



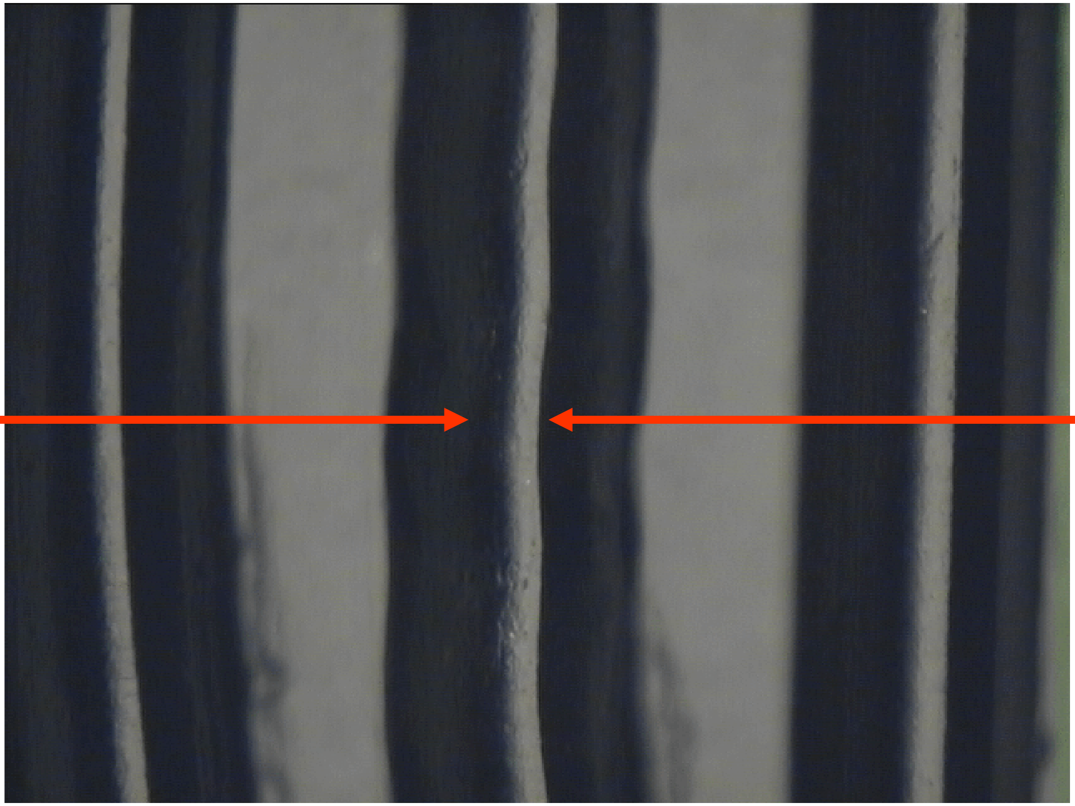
Merged data after step (3)



Gap between two edges [μm]
We impose 2.5 sigma cut to reduce noise.

Issue of Edge Quality in 2D Image

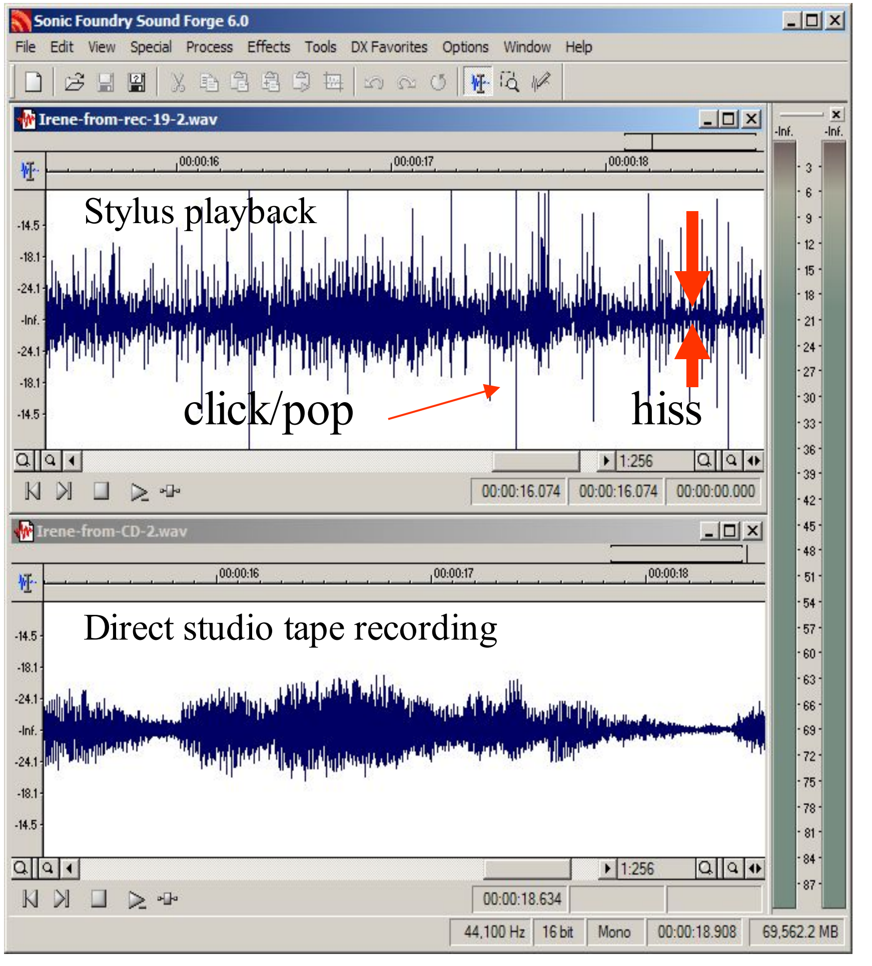
Poor edge definition



Good edge definition

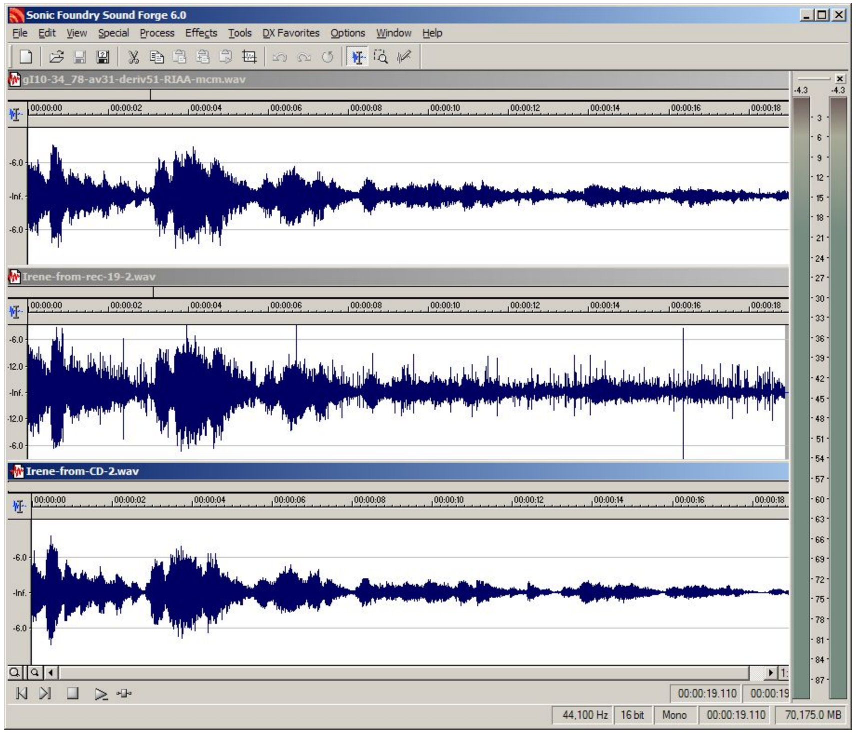
Noise Sources

- Surface noise or hiss
  - High frequency due to continuous imperfections in groove surface
- Transient impulse noise or “clicks and pops”
  - Due to discrete imperfections such as scratches, random and isolated
- Wow and flutter
  - Not really noise, systematic distortions such as motor speed, off axis rotation



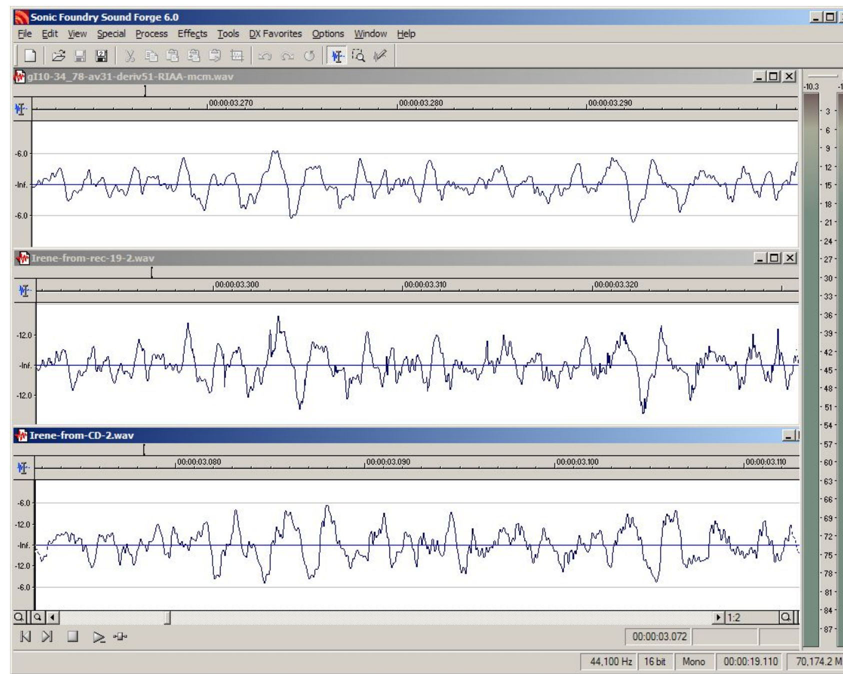
Results: 1st Sample

- Sample is 19.1 seconds
- From ~1950 78 rpm disk
- Top: Imaging method
- Middle: Played by stylus
- Bottom: Professionally re-mastered CD version
- Record was not cleaned.
- No "standard" digital noise reduction software used.



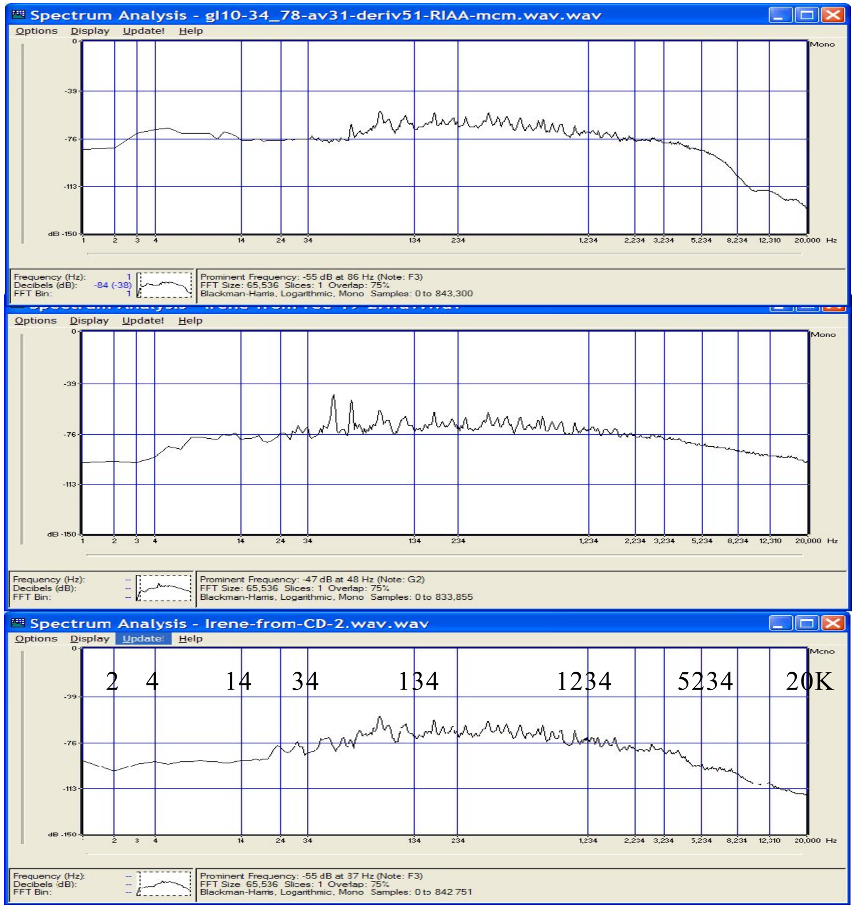
Zoom In for Detailed Comparison

- 40 ms portion shown
- Striking similarity between optical and stylus reconstruction
- Optical lacks clips and pops, certain noise features, high frequency structures (~10 kHz)
- Qualitative match to CD/tape version



Frequency Spectra

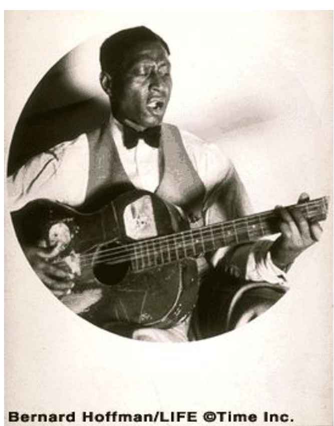
- FFT spectra of optical (top), stylus (middle), and CD/tape (bottom)
- Audio content in range 100 - 4000 Hz very similar
- More high frequency content in stylus version
- Effects of equalization and differentiation?
- Low frequency structure in optical sample (audible).

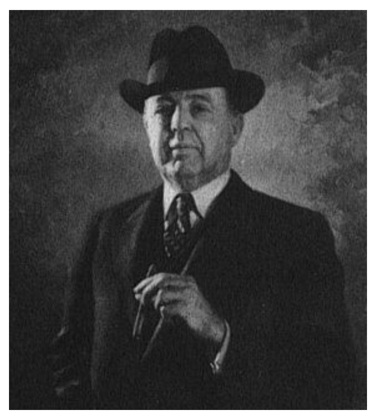


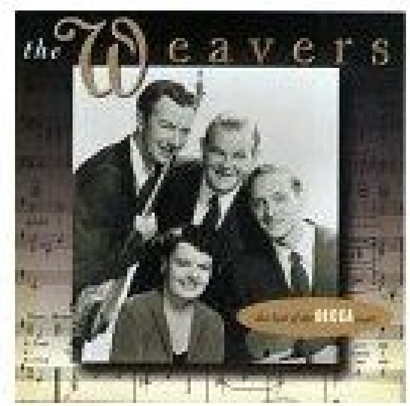
Sound Comparison

"Goodnight Irene" by H. Ledbetter (Leadbelly) and J.Lomax, performed by The Weavers with Gordon Jenkins and His Orchestra ~1950

- Sound from the CD of ***re-mastered tape***.
- Sound from the ***mechanical (stylus)*** readout.
- Sound from the ***optical*** readout.

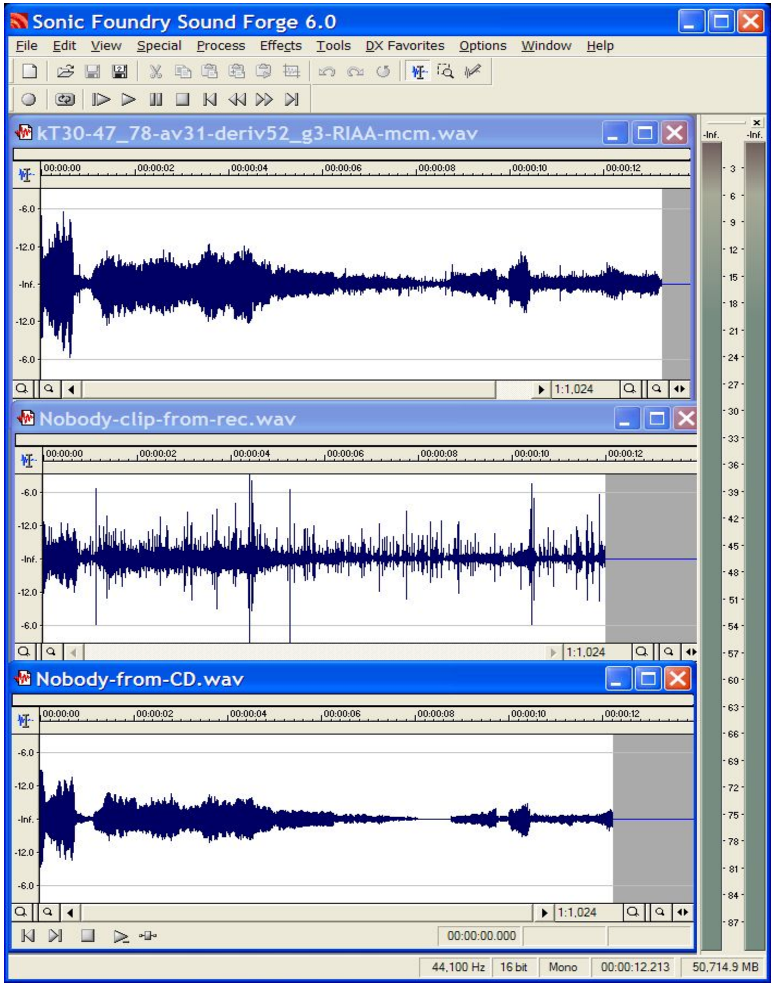






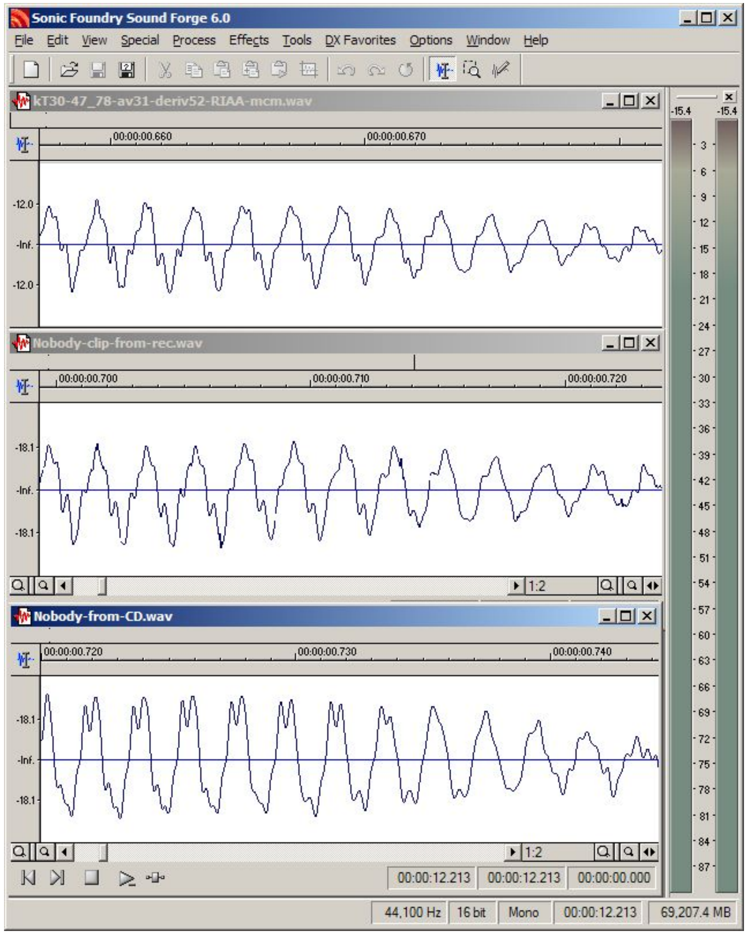
Results: 2nd Sample

- Sample is 13.2 seconds
- From 1947 78 rpm disk
- Top: Imaging method
- Middle: Played by stylus
- Bottom: Professionally re-mastered CD version
- Sample contains a pause at ~8 second point which can be used for noise studies



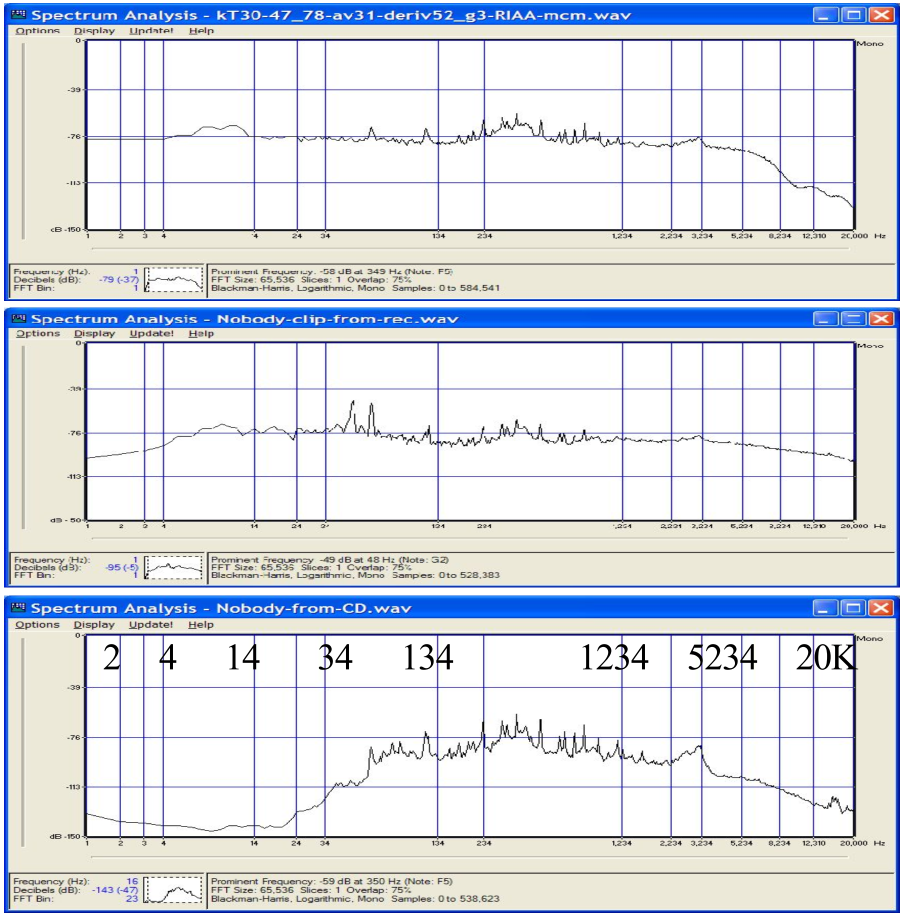
Zoom In for Detailed Comparison

- 25 ms portion shown
- Striking similarity between optical and stylus reconstruction
- Qualitative match to CD/tape version
- Strong structure around 500 Hz is typical of this sample – single voice and piano



Frequency Spectra

- Top (optical), middle (stylus), bottom (CD) versions
- Similar mid-range
- Low frequency difference



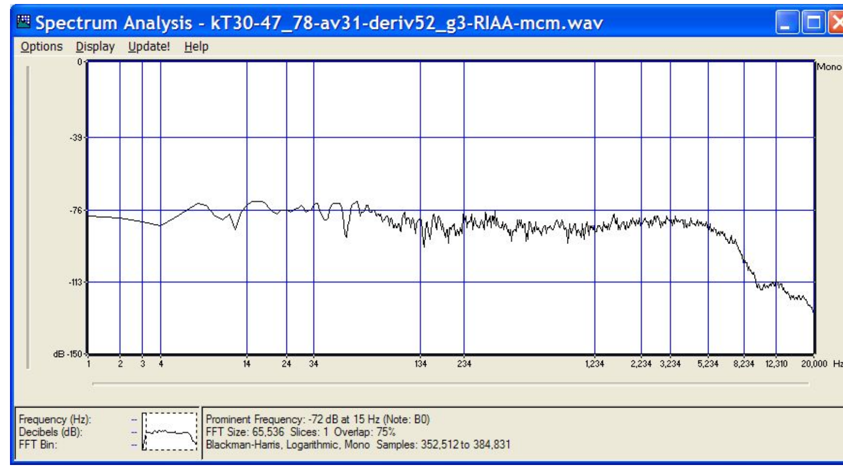
Sound Comparison

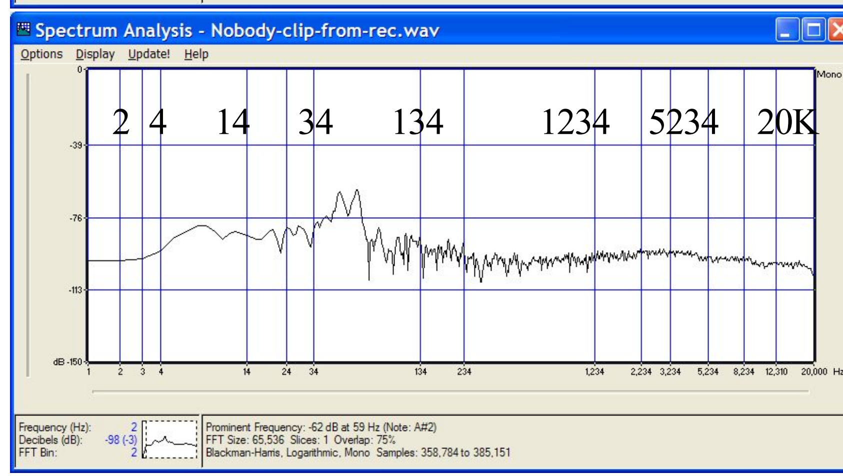
"Nobody Knows the Trouble I See" , traditional, performed by Marion Anderson, Matrix D7-RB-0814-2A, 1947

- Sound from the CD of ***re-mastered tape***.
- Sound from the ***mechanical (stylus)*** readout.
- Sound from the ***optical*** readout.

Noise Study

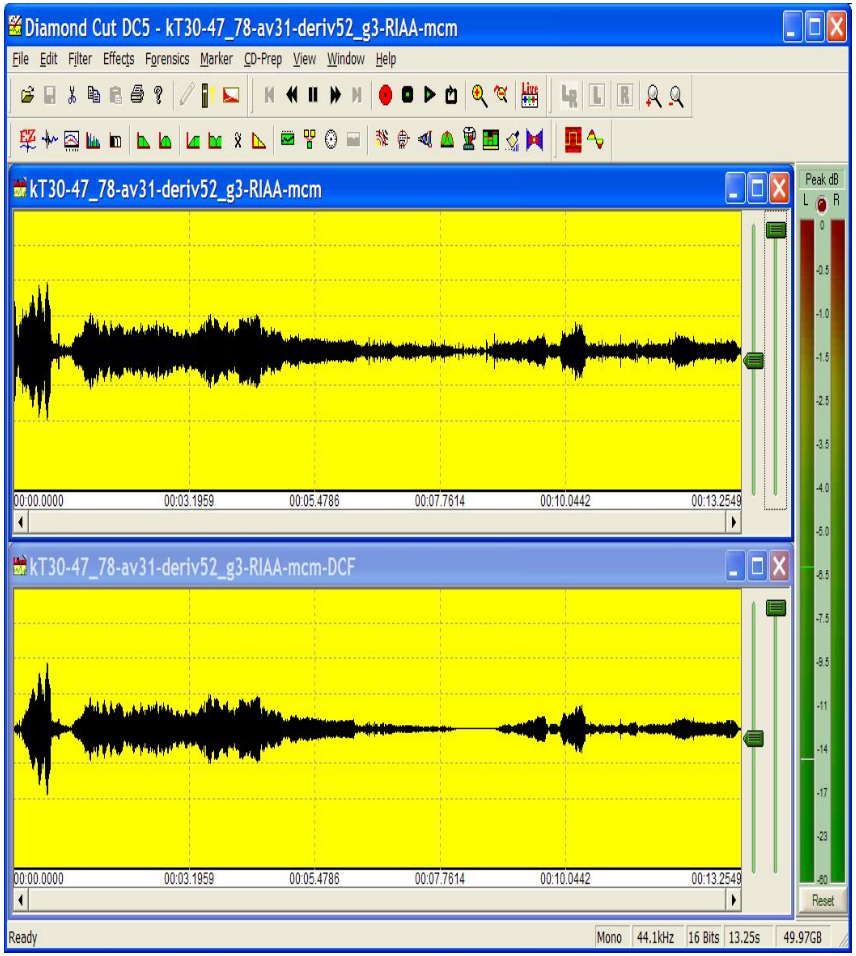
- Spectra of noise only segment at ~8 seconds
- Top (optical) bottom (stylus) versions
- Mid-range level is higher in optical sample





Effect of Classic Noise Reduction

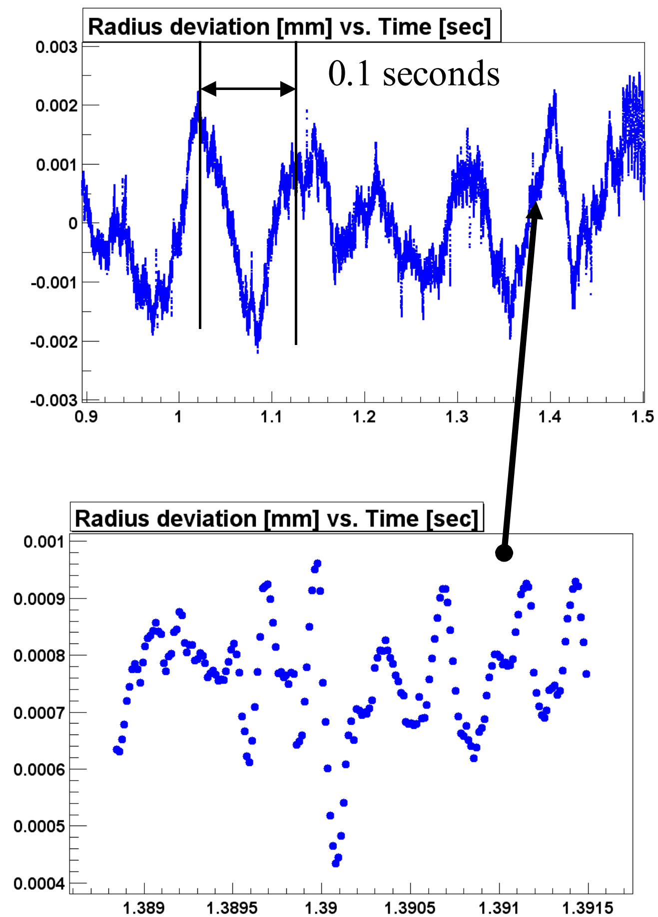
- Option to use commercial continuous noise filtering software on optical sample
- Result
  - Before
  - After



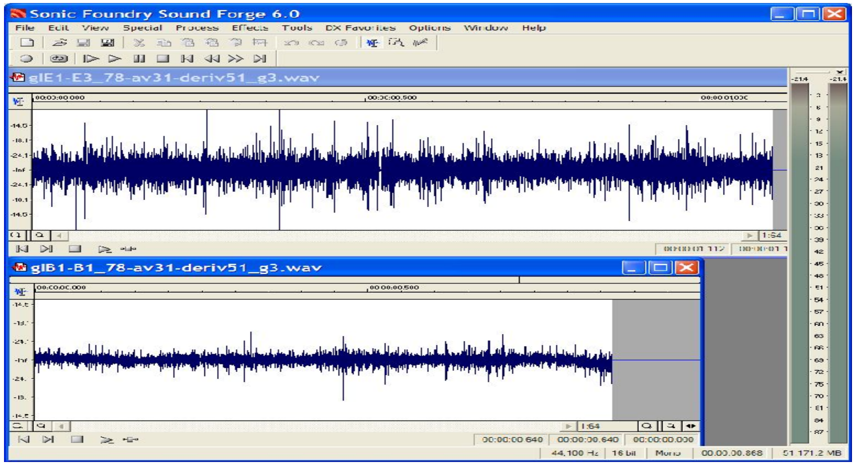
Physical Origin of Noise in Optical Reconstruction

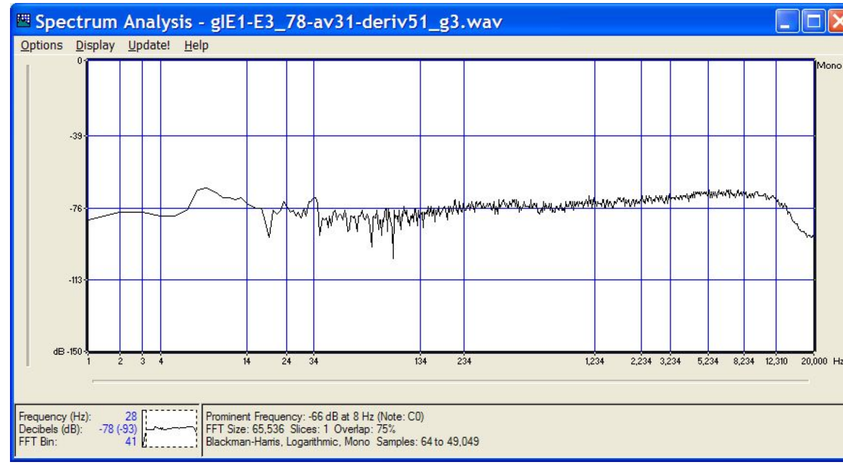
- View of raw groove shape data from region of pause, before differentiation into velocities.
- Upper plot is 0.6 second portion.
- Lower plot shows deviations about 10 Hz waveform.
- Each point is an independent edge detection across the groove bottom.
- Clear structures, spanning multiple points are resolved of typical scale:

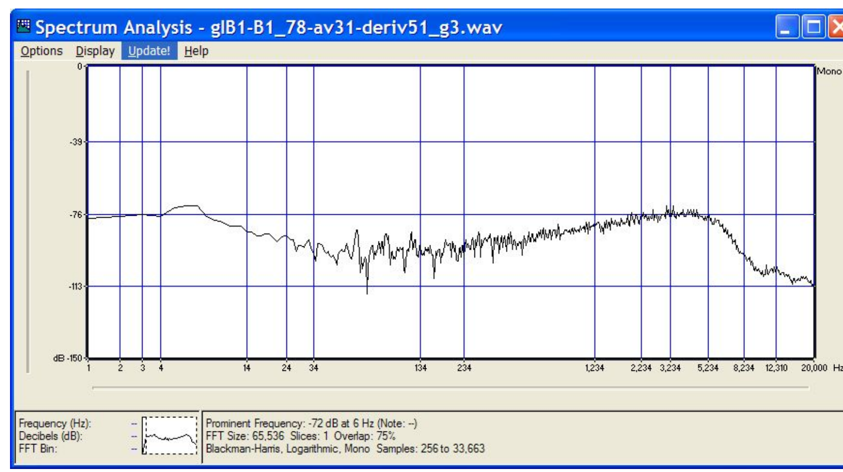
  100 microns (0.2 ms) x 0.2 microns !!!



Measurement of Noise at R_{min} & R_{max}







Optical readings
Upper sample is at outer radius
Lower sample is at inner radius
From "Goodnight Irene" disk
If noise is dominated by surface structures of constant size distribution the outer radius amplitude and frequency should be greater due to greater linear speed there

Conclusions

- Image based methods have sufficient resolution to reconstruct audio data from mechanical media.
- Method readily reduces impulse (click&pop) noise.
- Wide band "hiss" is present in 2D optical reconstruction.
  - Origin is not known definitively.
  - Insights into physical basis of noise. Observed structure may suggest ways to reduce further. More study required.

Conclusions

- Basic process is slow and/or data intensive compared to simple stylus playback.
  - 2D approach may be suitable for mass digitization.
  - How general is the 2D image quality?
  - At present 3D methods may be suitable for reconstruction of particular samples since they require ~hours per scan. Point scan is more flexible than frame based approach.
- Future potential in surface profiling field for full 3D reconstruction. This would be a good area for further research and collaboration. Optimizations.
- Report available LBNL-51983, submitted to JAES
- Info at our URL www-cdf.lbl.gov/~av

Areas for Further Work

- Studies of noise mechanisms in image data.
- Generality of 2D imaging across different samples.
- Effects of illumination and angle in 2D imaging.
- Development of hardware/software test bed for 2D imaging studies.
- Further study of 3D methods aimed at determining best candidate technology.
- Development of 3D profiling test bed hardware and software.
- Collaboration with industry.

Parameter	**78 r.p.m., 10 inch**	**33 1/3 r.p.m., 12 inch**
Groove width at top	0.006 inches	0.001-0.003 inches
Grooves/inch G_d	96-136	200-300
Groove spacing	0.007-0.01 inches	0.0033-.005 inches
Reference level peak velocity@1KHz	7 cm/sec	7 cm/sec (0.0011 cm)
Maximum groove amplitude	0.004-0.005 inches	0.0015-0.002 inches
Noise level below reference, S/N	17-37 dB	50 dB
Dynamic range	30-50 dB	56 dB
Groove max amplitude at noise level	1.6 - 0.16 μm	0.035 μm
Maximum/Minimum radii	4.75/1.875 inches	5.75/2.375 inches
Area containing audio data	38600 mm^2	55650 mm^2
Total length of groove	152 meters	437 meters