VisualAudio: an Optical Technique to Save the Sound of Phonographic Records

Sylvain Stotzer, Ottar Johnsen, Frédéric Bapst, Cédric Milan, Christoph Sudan, - Stefano S. Cavaglieri, Pio Pellizzari
University of Applied Sciences of Fribourg - Fonoteca Nazionale Svizzera

ABSTRACT

The optical retrieval and storage technique called VisualAudio provides a way to extract sound information from a phonographic record without any mechanical contact. The process is straightforward: we take a picture of each side of the disk using a dedicated analog camera, we store the film as our working copy, and when needed, we scan the film and process the image in order to extract the sound. This technique can be used to retrieve the sound of older records that are in such a bad shape that no regular turntable can be used. A working prototype has been built and was used to retrieve the sound from several records. A new prototype is currently being designed and built that will significantly improve the quality of the extracted sound. In this paper, we describe the principles and the characteristics of the different parts of the system and we analyze the performances. We hope that the new prototype will be used to save the sound from numerous old records.

1. INTRODUCTION

Cutting a disk was in practice the only way to preserve sounds until the introduction of magnetic tape in the early 50's. Therefore there are huge collections of phonographic records, for example in radio stations and national sound archives. Such archives include pressed disks produced by record companies as well as direct cut disks obtained by the direct recording of radio programs with often a great cultural value and available only as unique copies. Disks, and in particular acetates and shellacs, are fragile. All the records are deteriorating with time [1]. Worse, many records would be destroyed by the movement of the stylus from even the best turntables. Thus, we risk loosing an important cultural heritage. This is a big concern for the sound archivists.

We are proposing a technique called VisualAudio to preserve the sound of old records. In the proposed solution the records are read optically (without any contact) and the sound is extracted from the optical information. Based on this idea, a prototype has been built and has demonstrated that this technique can be used to extract the sound with a good quality. We are currently designing and building a new prototype that will significantly improve the quality of the extracted sound.

2. PHONOGRAPHIC FUNDAMENTALS

2.1. Cutting

The disk recorder transforms an electrical signal into a mechanical displacement of the cutting stylus. The cutting stylus cuts the groove into the surface of the lacquer.

2.2. Reproducing

The reproducing stylus reads the groove of the disk. The mechanical displacement of the stylus is transformed into an electrical signal. The signal flows through a number of electronic devices reaching a loudspeaker. The loudspeaker transforms the electrical signal into an acoustic vibration.

2.3. Disk characteristics

The amplitude of sound signal is stored either in the groove position for constant amplitude recording, or in the derivative of the radial groove position for constant velocity recording.

Most of the disks are recorded with constant velocity. This implies that low frequencies need a wide space to be recorded and that high frequencies generate too small displacements and are lost in noise. The solution is to filter the sound with an equalization curve at recording to reduce the low frequencies and boost the high frequencies. The preamplifier does the opposite to restore the original signal at playback.

	78 rpm records	**33 rpm records**
Groove width	31-187 µm	25.4 µm
Groove deviation	75 µm	28 µm
Distance between grooves	200-300 µm	85-125 µm
Bandwidth	100-12000 Hz	30-16000 Hz
Signal to noise ratio	32-40 dB	45-60 dB
Groove shape	round	triangular

Table 1. Disk characteristics [2].

Several such curves have been used (NAB, FFRR, AES, …) but the RIAA was standardized and broadly used since 1955 [2]. The RIAA recording preemphasis has a low-turnover of 500 Hz, and the frequencies below are attenuated by 6 dB/octave. The high-turnover is 2122 Hz, and the frequencies above are boosted by 6 dB/octave. RIAA records may be considered as constant velocity between 500 Hz and

2122 Hz, and constant amplitude for lower and higher frequencies.

Typical disk characteristics are given in table 1. It should be noted that for most interesting historical records, direct cut disks, the bandwidth is rarely greater than 6 kHz.

(a)

(b)

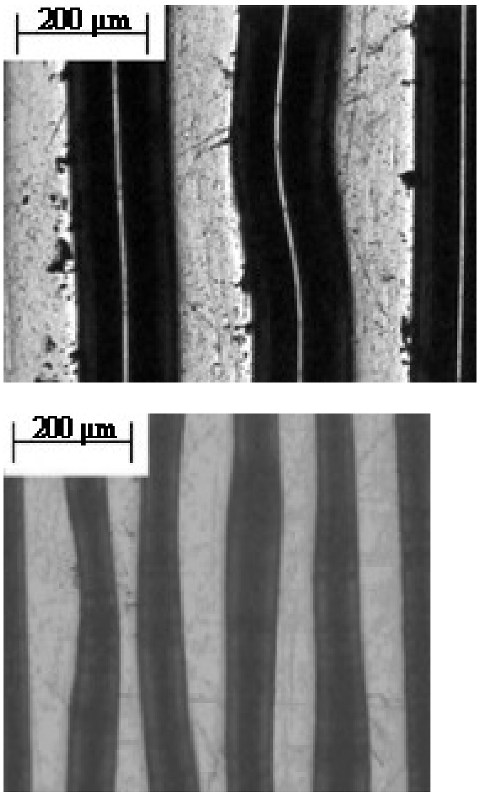


Fig 1. Microscope top view of (a) 78 rpm record and (b) 33 1/3 rpm record.

Fig 1 shows pictures of the surface of records as seen on a microscope. It should be observed that on the above picture, the sound information is contained in the horizontal position (it is the radial stylus displacement on a turntable). The dark lines are the grooves. Notice also that the 78 rpm (rotation per minute) record is in bad shape. Fig 2 shows the shape of the groove. The groove's shape depends mainly on the kind of disk. 78 rpm records have round grooves. Microgroove 33 1/3 rpm records have triangular grooves; stereophonic sounds can be stored by using the horizontal position of the groove for the monophonic part (left+right) and the depth for the stereophonic part (left-right). This means also that the right and left channels are stored on the position of the left and right walls.

The groove's shape has an important impact on the disk picture. Triangular grooves do not reflect light and therefore appear as a unique trace on the negative film. While the walls of the round grooves don't reflect light, the bottoms of the grooves do. Therefore there are two traces for each round groove.

(a)

(b)

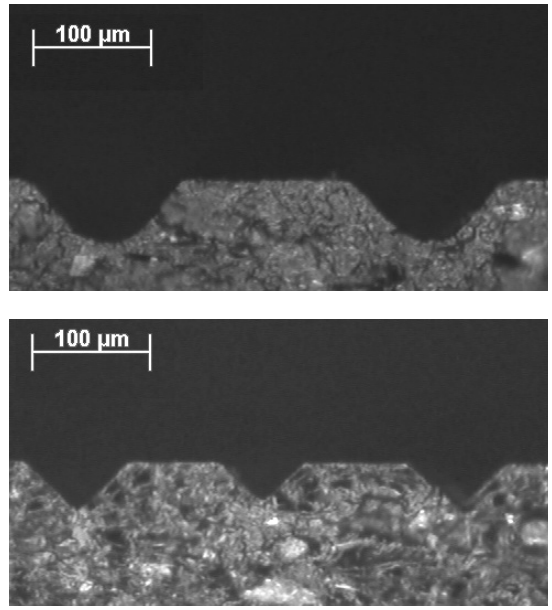


Fig 2. Profile of the groove of (a) a 78 rpm record and (b) a 33 1/3 rpm record.

3. THE VISUALAUDIO CONCEPT

This concept is based on the observation of the disk surface, using a microscope as shown on Fig 1. It is the radial displacement of the groove that contains the sound (which is also contained in the depth, for stereophonic and vertical cut records). As we see the groove's displacement, we can "see" the sound. It means that the sound information is contained in the image of the record. This concept leads to a 3 steps concept (which is shown on Fig 3) [3] [4]:

1) An analog picture of each side of a disk (either 33 1/3 or 78 rpm) is shot. The film must have a high spatial resolution and be relatively large, since we wish to catch the finest details of the groove. This process can be done quickly. The film is cheap, and can be stored for a long time (more than 100 years). That way, the sound information is preserved in case the original disk deteriorates.
2) When anyone wants to recover (i.e. to listen to) the sound, the film, with the picture of the disk, is scanned using a specially designed rotating scanner, and digitized. At this point a digital image of the record is stored.
3) The sound must then be extracted from the digital image. This requires image processing techniques in order to extract the radial displacement of the groove, to detect cuts and to correct other defects. Digital signal processing must be applied to the groove signal to extract the sound.

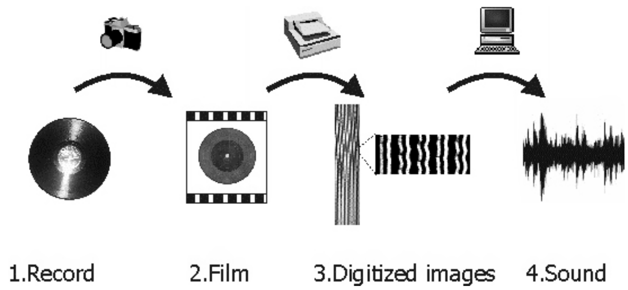


Fig 3. The VisualAudio concept.

3.1. Why use the intermediate photographic step?

The groove position must be evaluated very accurately, requiring high image resolution. This can be done with a microscope. Unfortunately many disks have a warping of up to 1 mm. This exceeds the depth of field of a microscope. Using the intermediate photography step allows us to work with a larger depth of field while imaging the disk, but ensures that the image to be digitized (the film) fits in the reduced depth of field required by the microscope's optics.

Time is a critical issue in such an archiving system, and the challenge is to save a large amount of records quickly, before their complete physical destruction. Taking a picture of the disks is a quick way to store an (almost) analog copy of the sound content in its current stage of conservation. The sound extraction could then be done on demand.

Fig 4. Photographic system used for the picture taking.

4. PICTURE TAKING

With such tiny groove deviations, 75 µm (78 rpm) or 28 µm (33 1/3 rpm), particular care is a must in the picture-taking phase. Maximum resolution obtainable with this system is limited both by the camera lens and by the film. High-resolution films have a resolution of about 500 lines, corresponding to 1000 dots/mm. This resolution is due to the film grain and can be considered as noise. The optical system limits also the resolution, in particular due to the diffraction of the lens and the depth of field. Here the resolution is even much lower. Happily, these produce a low-pass effect that is much less annoying than noise. The film must have about the same size as the record. Our pictures were made in a photo laboratory. Fig 4 shows the photographic system. Notice that a new greatly improved photographic system is being built. The calculations below refer to the new system.

4.1. Optical system

The resolution of an optical system is limited by lens aberrations, diffraction and depth of field. Lens aberrations are mainly chromatic and geometric. Both are negligible when using good quality lenses with small apertures. Chromatic aberrations are eliminated with monochromatic light.

The resolution of our optics is limited by: a) the circle of confusion and b) Fraunhofer diffraction [5]. The circle of confusion is the blur caused by the depth of field (*DOF*). The diameter of the circle of confusion C can be approximated using the focal length f and the opening diameter of the lens D at a 1:1 enlargement ratio:

$$C \simeq DOF \frac{D}{4f}$$

The Fraunhofer diffraction produces Airy patterns: the image of a point through a lens is a spot (blur), called Airy disk. The diameter of the Airy disk is defined using the light wavelength λ:

$$d_{airy} = 2.44\lambda \frac{f}{D}$$

The total blur B_{photo} of the picture taking is:

$$B_{photo} = \sqrt{C^2 + d_{airy}^2}$$

For $DOF = 1\,mm$ and $\lambda = 0.41\,\mu m$, the blur is minimal when: $f/D = 15.8$. The total blur is then $B_{photo} = 22.3\,\mu m$.

We have selected a focal length of 420 mm, corresponding to a total system length of 1.7 m for a 1:1 enlargement. We therefore choose $D = 27\,mm$ to satisfy the optimum relative aperture $f/D = 15.8$. Larger opening should be used for flatter disks in order to increase the resolution.

4.2. Illumination

The main purpose of the lighting is to provide a homogeneous illumination of the disk as seen from the lens.

The illumination depends mainly on the material reflectivity. Most disks are black, and the reflectivity factor of a black surface is very low, as most of the light is absorbed. But since the disks are bright, their reflectivity is mainly specular, meaning that most of the reflected light has a reflective angle equal to the incidence angle. Therefore a strong and directional lighting ensures sufficient contrast, even if the luminance of the black is low.

The best solution is to have a directional light that illuminates the disk uniformly from the lens point of view. This directionality also lowers the parasite illumination that could decrease the contrast by being reflected by the groove walls.

A monochromatic blue light is an interesting solution, as its short wavelength improves the sharpness and fits to the spectral response of the film.

4.3. Film

Sharpness, contrast and graininess are of high importance in the choice of the film. The sharpness of black and white films is much better than color films due to their emulsion. Thus only black and white films are considered for VisualAudio, as the main information in disk pictures is perfectly rendered by gray levels.

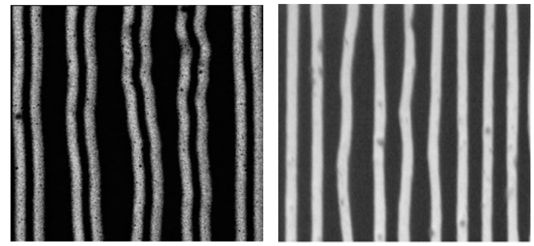
The sensitive part of a film is made of a 5 to 10 µm layer of gelatin in which silver halide crystals are randomly distributed. The size of these grains is 0.2 to 2 µm, depending on the film and on its development. The apparent graininess of a film is due to grain clumps created by the overlap of individual grains at different depths of the emulsion. These grains and graininess produce noise on the edges of the groove image.

Both the emulsion and the development define the contrast, represented by the γ of a film. Increasing γ means increases in contrast or tone separation:

- When γ is high, the shades are saturated: there are only a few different gray levels in the picture, and most of the light and dark tones are then transparent or opaque.
- When γ is low, all the different light tones of the image are rendered on the film, but the film grain is coarser and thus the edge has more noise.

The film speed may be of importance as the picture taking duration is the time critical phase of the VisualAudio process; but fast emulsion films have much coarser grain and thus lower sharpness and are therefore not acceptable.

Notice that these films are negatives. The black parts correspond therefore to the regions with the greatest illumination, i.e. the flat parts of the disk. The white parts correspond to the grooves that are regions with less illumination. We observe that for the 33 1/3 rpm record, we see a single line for each groove, while for the 78 rpm record we see 2 parallel lines corresponding to the 2 sides of the groove, while the middle of the groove is flat and therefore black. Fig 5 shows enlargements of the films.



(a) (b)

Fig 5. Enlarged extracts from the films of (a) a 78 rpm record and (b) a 33 1/3 rpm record.

5. FILM SCANNING

The specially built circular film scanner is shown on Fig 6.

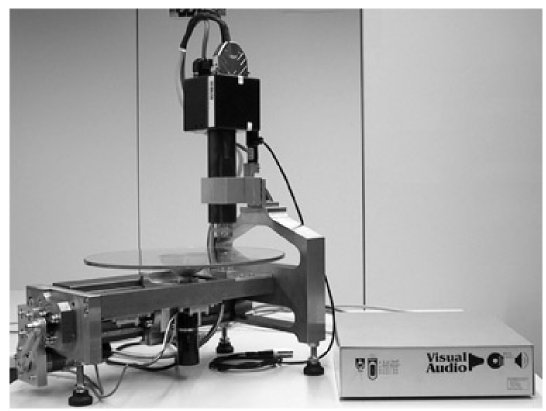


Fig 6. Rotating scanner for the films.

During the scanning, the film lies on a glass turntable. A 2048-sensor CCD-linear camera mounted on microscope optics is fixed above the glass and a light source located below the tray lightens the film by transparency. During each rotation of the glass turntable, we scan a ring of the film, whose width depends on the optics magnification. By radially displacing the tray, adjacent rings are scanned in order to digitize the whole record. The sampling frequency ranges from 25 to 200 ksamples per ring. This rate combined with the disk speed (33 1/3 rpm or 78 rpm) defines the audio signal sampling frequency, ranging from 13.75 to 110 kHz for 33 1/3 rpm, and from 32.5 to 260 kHz for 78 rpm records. The circular scanner has the advantage of transforming the circular disk picture into a rectangular image, thus avoiding a coordinate transformation.

5.1. Optics

The equations for the resolution of a microscope optics are similar to those presented in section 4.1, except that they are based on the numerical aperture (*NA*) that defines the light gathering ability and resolving power of a lens:

$$C = DOF \frac{NA}{\sqrt{1 - NA^2}} \quad \text{and} \quad d_{airy} = \frac{0.61\lambda}{NA}$$

$DOF = 50\ \mu m$ is sufficient to encompass the photosensitive emulsion depth and the warping of the rotating glass tray. With $NA = 0.25$ and $\lambda = 0.5\ \mu m$, the total scanning blur is:

$$B_{scanning} = \sqrt{C^2 + d_{airy}^2} = 13\ \mu m$$

5.2. Illumination

The illumination should be homogeneous on the plane captured by the CCD camera. This is achieved by using a lens condenser on the light source.

Parasite light should be minimized, especially reflections in the optical tube, to avoid blur and light spots.

5.3. Digitizing

Since the camera is mounted on a 10x magnification optics, each 10x10 µm sensor integrates the light of a 1x1 µm portion of the static image. As the glass tray rotates, each sensor acquires a rotating area of the film, and quantizes it with 256 levels. The integration time is constant, corresponding to constant audio signal durations, but the acquired sectors on the outer or inner parts of the disk have different lengths. Thus the film area averaged by one pixel may vary from 1x3 µm up to 1x38 µm depending on the sampling frequency and the radial position on the disk. This integration is a weighted averaging, as the film is moving during the sensor exposure time. Some external variations may also influence the weights parameters; for instance the motor vibrations that cause sampling time shifts and thus random modulations of the extracted sound.

5.4. Total system resolution

Disregarding noise, the worst case resolution of the system R_{system} and blur B_{system} can be approximated using the blur of each component:

$$\frac{1}{R_{system}^2} = B_{system}^2 = B_{photo}^2 + B_{film}^2 + B_{scanning}^2$$

With a film resolution of $600\ lp/mm$ (line pairs per mm):

$$B_{system} = \sqrt{22.36^2 + 1.7^2 + 13^2} = 25.92\mu m$$

$$R_{system} = 38.57\ \ lp/mm$$

The picture-taking phase is obviously the limiting factor of the whole process. However if the disk is flatter than assumed, *DOF* decreases and the resolution increases. This shows that it is possible to retrieve the groove edges positions by image processing as the worst case blur is smaller than the groove width (Table 1).

6. IMAGE PROCESSING

The purpose of the image processing is to estimate the displacement of the groove edges, which contain the sound information.

6.1. Image characteristics

Due to the rotational scanning, a ring of the disk is represented by a rectangular image (Fig 7). Each ring contains 5 to 20 complete rotations of the groove, depending on the optical magnification, the disk type (78 rpm, 33 1/3 rpm) and the radial displacement of the groove due to the spiral and the centering of the film on the glass turntable. Acquisitions made on a negative film display the grooves as light traces and the intergrooves as dark areas.

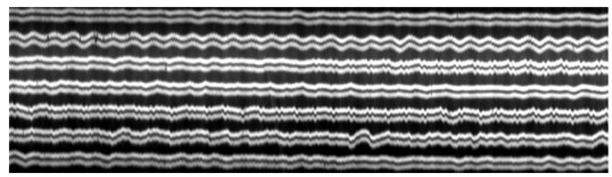


Fig 7. Section of a ring image acquired on a 78 rpm record.

6.2. Image degradation

The VisualAudio process degrades the image due to a combination of blurring, noise and saturation.

The blur produces low-pass filtering and transforms the step edges of the groove into ramp edges. It spreads the edge position information over several pixels, called transitions. Optical blur on an image can be modeled as a convolution with a gaussian function. Therefore the combination of several gaussian blurs during the imaging chain can also be modeled as a gaussian blur.

An edge could cross several pixels at the same sampling period, thus increasing the number of pixels of the transition.

Noise is produced at every step of the VisualAudio process: dust, grain, quantizing noise, CCD noise, etc… The low-pass filters caused by the blur eliminate the higher frequencies of this noise, except for the CCD-noise produced at the end of the imaging chain. Most of these noises are additive except the CCD-noise that is multiplicative, meaning that it has higher amplitudes in the light shades.

The tone transfer function caused by the saturation means loses of information: it enhances the dynamic (number of gray levels used) in part of the transition area, but flattens the dynamic in the dark and light areas.

6.3. Sound correction by image processing

The high frequency noise is filtered by the integration and by the blur caused by the acquisition process. An additional median filtering can be applied on the radial direction on each intensity profile to lower the remaining noise prior to the edge detection.

Scratches, spots and other big image defects might cause severe damage to the extracted sound. Therefore they must be detected and replaced by interpolation at the end of the image processing phase. If the blur low-pass filters are too strong, the noise produced by these defects is completely spread in the blur and it produces a shift of the edge position.

A blurred image still contains the necessary information for the edge detection. The blur by itself does not need to be removed for the image analysis. It just forces to work with a larger number of pixels to reach the same accuracy as would be attainable with a non-blurred image. By forcing to work with several pixels, this blur provides the advantage of lowering the influence of the CCD-noise on the extracted sound.

Saturation destroys part of the necessary information in the intensity profile, and there is no way to recover them with sufficient accuracy. Saturation would only be acceptable with a constant blur width, or on a step edge, not on a ramp edge. Otherwise a random shift is added to the position of the detected edge and that produces noise on the extracted audio signal.

A mean or median filtering in the tangential direction would smoothen the signal and enhance the harmonic distortions. Therefore tangential corrections must be done in the frequency domain by audio filters.

6.4. Edge detection

The groove is almost perpendicular to the linear camera, therefore we assume that the edge direction is perpendicular to the intensity profile and that edge detection has to be processed only in radial direction to define every edge position at each sampling period. In order to get the most accurate estimation of the edge position, a priori knowledge of the groove picture must be considered:

- The groove has physical properties (Table 1).
- The groove's width is constant for mono records, meaning that the displacements of all the groove edges are equal.
- There are no standards for the intergroove's width.
- The sound properties define boundaries for the groove displacement between consecutive samples.

Due to the degradation, the edge is represented by a blurred area that can be described by four parameters:

- Location
- Transition width
- Dark intensity I_d
- Light intensity I_l

We need to determine the first parameter by using the other three, which could vary in time due to local conditions. As long as the blur is homogeneous (not saturated) in the radial direction for each sample line, the detection algorithm has to detect the middle of the blurred area. This could be determined from the width of the first derivative of the intensity profile, or from the zero-crossings of its third derivative. But derivatives are sensitive to noise; therefore we propose another method to detect the middle of the transition.

The derivative of the intensity profile is approximated by convolving the intensity profile with a box filter. The size of this box must be large enough to filter the noise and guarantee the uniqueness of each solution, but it is limited by two factors: the size of the transition and the intergroove spacing. This derivative is used to define the regions of interest for each groove, and the most relevant points for the transition fitting.

The ideal groove intensity profile can be modeled as a box for a 33 1/3 rpm record (or two boxes for a 78 rpm record). The blurred groove can then be modeled by a trapezoid, where the top and bottom are defined respectively by the lowest grey level I_d of the local intergroove and the highest grey level I_l of the current groove. Sides are approximated by least squares fitting of the points having the higher dynamic, based on the derivative (Fig 8). The edge is then located at the middle of the segment delimited by the crossings between the sides and I_d and I_l.

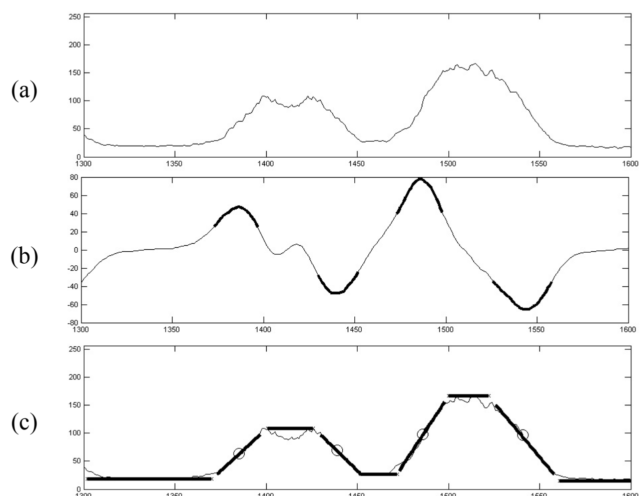


Fig 8. (a) Intensity profile from a 78 rpm groove. (b) First derivative of the profile. (c) Groove modelling by least squares fitting of the bold parts from the derivative. Edges are located by the circles at the middle of the transitions.

To evaluate the accuracy of the edge detection process, the differences were measured between consecutive detected edges positions on unmodulated grooves. Even if the blur is spread over several pixels, the standard deviation of these measures reaches 0.9 μm for 33 1/3 rpm records and 0.7 μm for 78 rpm records.

6.5. Extracting the sound from the groove position

The sound signal is obtained by extracting the radial position of the groove. Often, the sound is oversampled (the sampling frequency depends on the tangential distance between pixels). The extracted sound may be sampled up to 260 kHz. As the bandwidth for 78 rpm records is limited to 12 kHz it allows a great noise reduction, since most of it is above 15 kHz.

6.6. The VisualAudio algorithm

The image processing algorithm can be decomposed in the same way as the acquisition: ring by ring and line by line. The VisualAudio sound extraction algorithm is displayed in Fig 9.

```
For each ring image
    For each line
            Optional median filtering
            Define region of interest
            Edge detection on the line
    Groove reconstruction:
      1) Reconstruct the edges belonging to the same
         rotation of the groove.
      2) Combine the extracted data from consecutive
         rotations.
    Non-linear corrections to the groove displacement
Combine the sound extracted from all rings
Low-pass filtering to remove out of band noise
Audio filter to remove the clicks, buzzes, hisses, ...
```

Fig 9. Pseudocode for the VisualAudio algorithm

7. SIGNAL TO NOISE CONSIDERATIONS

The easiest way to evaluate the sound extraction performances is by perceptual comparisons. But for sounds, that are natural speech or music signals, it's impossible to measure the noise and the distortions. Using test records having soundtracks containing a unique frequency, it is possible. We digitize a portion of a sine wave from the test record, and from the FFT of the reconstructed signal, we can calculate the signal to noise ratio. By analyzing this spectrum, we can also observe distortions, which help us finding their causes, correct them and improve the signal to noise ratio. As an example on Fig 10, we have the spectrum of a reconstructed 300 Hz sine wave. We observe noise, as well as peaks due to the scanner's motor vibration (intermodulations with the 110 Hz motor vibration).

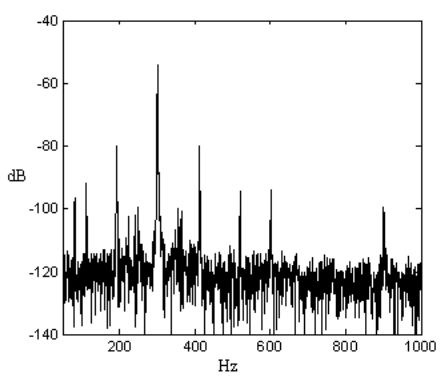


Fig 10. Spectrum of a reconstructed 300 Hz signal extracted from the HFN 001 33 1/3 rpm test record.

The aim of VisualAudio is to be as good as a turntable for records in good condition. It means to reach a signal to noise ratio of up to 60 dB for 33 1/3 rpm and 40 dB for 78 rpm records.

To analyze the signal to noise ratio (SNR), we assume that the noise produced by the film and by the acquisition has a uniform frequency distribution and that the original signal cut on the disk $s(t)$ is a sine wave: $s(t) = A_{f_0} \sin(2\pi f_0 t)$, where A_{f_0} is the maximum amplitude for the frequency f_0 . In the next paragraphs, we will estimate the maximum noise standard deviation on one edge σ_n allowed to reach satisfying signal to noise ratio using the VisualAudio process. This will be done for a 78 rpm disk and for the different recording modes, using the following parameters: B_n the noise bandwidth, B_s the sound bandwidth and N the number of groove edges used to retrieve the signal. The typical values for a 78 rpm are: $B_n = 130\,kHz,\quad B_s = 12\,kHz,\quad N = 4$.

7.1. SNR for constant amplitude records

In constant amplitude mode, the peak-to-peak groove deviation A is constant, defining the signal variance σ_x^2 :

$$\sigma_x^2 = \frac{A^2}{8}$$

The power of the noise is:

$$P_n = \frac{\sigma_n^2}{NB_n / B_s}$$

The signal to noise ratio is:

$$SNR = \frac{\sigma_x^2}{P_n} = \frac{A^2 NB_n}{8\sigma_n^2 B_s}$$

With a maximal groove deviation of $A = 75\,\mu m$, σ_n must be smaller than 1.75 μm to reach a 40 dB signal to noise ratio.

7.2. SNR for constant velocity records

In constant velocity mode, the derivative of the position represents the amplitude of the sound signal:

$$v(t) = \frac{ds}{dt} = 2\pi f_0 A_{f_0} \cos(2\pi f_0 t)$$

The SNR can then be calculated as follows [6]:

$$SNR = \frac{3 f_0^2 A_{f_0}^2 NB_n}{2\sigma_n^2 B_s^3}$$

In the case where $f_0 = 500\,Hz$ and $A_{f_0} = 75\,\mu m$, σ_n must be smaller than 0.25 μm to reach a 40 dB SNR.

7.3. SNR for equalized records

The signal and the noise of an equalized record are filtered by the preamplifier. With $f_0 = 500\,Hz$ and $A_{f_0} = 75\,\mu m$, σ_n must be smaller than 1.28 μm to reach a 40 dB SNR [6].

7.4. Constraints on the resolution

In the previous sections, we have defined the accuracy necessary in the estimation of the groove edges position. We get standard deviations that are near $1\mu m$. This seems to indicate that the task is impossible as the system blur is about $25\mu m$. But as most of the system blur is not noise but low-pass, an edge position can be estimated much more accurately than the blur, in particular since we are interested in the relative position. The groove standard deviation measures mentioned in section 6.4 satisfy the resolution constraints exposed in the SNR calculations.

The limits of the VisualAudio system are caused by several noise and distortion sources that are difficult to estimate. For example, the record might be dusty, or the groove border is dented, the illumination over the whole disk might be non-uniform, the rotation speed of the film might not be constant and lead to irregular sampling corresponding to a frequency modulation of the sound, and vibrations might disturb the scanning.

8. RESULTS

Several prototypes have been built [7], [8] and [9] before the current one, each one leading to better results. Several music extracts from the latest prototype are available on the web site www.eif.ch/visualaudio. As you can hear, the sound is already of a reasonable quality. The processing time was at the beginning a limiting factor. Extraction of a few seconds of music took several hours. Several algorithm improvements have decreased the processing to nearly real time, meaning that a sound can be extracted from a film in about the same amount of time than when playing the record on a turntable.

The proposed system is especially useful for disks in bad condition, for example lacquers or broken records. A similar system, without the intermediate photographic step, but with an x-y scanning has recently been proposed, but the scanning time is much longer [10]. The image processing is more complicated, since the groove spiral must be unwound. Avoiding the intermediate step of the film might be an advantage for the signal to noise ratio.

We have obtained funding that will allow us to improve all the steps of the process. We are now in particular improving the optics of the picture taking. In addition, we are working on a new version of the scanner as well as on the algorithms for a better sound extraction.

9. CONCLUSION

A solution to extract the sound from records has been proposed and demonstrated with good sound quality. This technique has several advantages.

- The picture is shot without interfering with the surface of the disk.
- There is no need to manipulate the disk except for placing it on the photo stand.
- Disks in virtually all conditions (even delaminated, broken, deformed, etc.) can be read and the sound can be restored.
- Many disk formats (size, speed, cutting, etc.) can be read using the same equipment.
- Image processing is something very well established. It is relatively easy to make all kind of correction to the physical incoherencies of the disk.
- Film is a quite stable, small, and cheap carrier for storing sound information. This means that it might as well be used as a long-term storage medium.

We are confident that we will be able to transform this prototype into a system that can be used to save our sound heritage with high sound quality on a large scale.

ACKNOWLEDGEMENTS

The authors would like to thank Thierry Fumey for taking the pictures of the records, Olivier Chassot, Swen Ruegsegger, Pascal Bovet, Louis Schmitt, Sylvain Pruvot and Jacques Miauton for their help in this project. This project is funded by project HES-SO-01.01TI, by the Swiss National Science Foundation (project 21-64984.01) and the Gebert Rüf Foundation.

REFERENCES

[1] B. Fox, "Not fade away." New Scientist, 177 (2384), pp. 40-43, March 1st 2003.

[2] "Anthology of the Journal of Audio Engineering Society", "Disk Recording" vol. 1 & 2, 1953-1980.

[3] S. Cavaglieri, O. Johnsen, F. Bapst, "Optical Retrieval and Storage of Analog Sound Recordings", AES 20th International Conference, Budapest, 2001.

[4] O. Johnsen, F. Bapst, C. Sudan, S. Stotzer, S. Cavaglieri, P. Pellizzari, "VisualAudio: an Optical Technique to Save the Sound of Phonographic Records", IASA Journal, No 21, June 2003, pp38-47.

[5] F. L. Pedrotti, L. S. Pedrotti, "Introduction to Optics", 2nd ed., Prentice-Hall, 1992.

[6] S. Stotzer, O. Johnsen, F. Bapst, C. Sudan, R. Ingold, "Phonographic Sound Extraction Using Image and Signal Processing", ICASSP, Montreal, 2004

[7] O. Chassot, C. Miauton, "VisualAudio", Diploma project at Ecole d'ingénieurs et d'architectes de Fribourg, Switzerland, 2000.

[8] N. Bosi, "VisualAudio: acquisition de l'image", Diploma project at Ecole d'ingénieurs et d'architectes de Fribourg, Switzerland, 2000.

[9] Y. Pauchard, C. Sudan, "VisualAudio", Diploma project at Ecole d'ingénieurs et d'architectes de Fribourg, Switzerland, 2001.

[10] V. Fadeyev, C. Haber, "Reconstruction of Mechanically Recorded Sound by Image Processing", J. Audio Eng. Soc., vol. 51, no. 12, pp. 1172-1185, 2003.

Web site: www.eif.ch/visualaudio