

PHONOGRAPHIC SOUND EXTRACTION USING IMAGE AND SIGNAL PROCESSING

Sylvain Stotzer, Ottar Johnsen, Frédéric Bapst, Christoph Sudan, – Rolf Ingold

University of Applied Sciences of Fribourg – University of Fribourg, Switzerland
sylvain.stotzer@eif.ch, ottar.johnsen@eif.ch, frederic.bapst@eif.ch, rolf.ingold@unifr.ch

ABSTRACT

The optical retrieval and storage technique called VisualAudio provides a way to extract sound information from an analog disk 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 old records that are in such a bad shape that no regular turntable can be used. A working prototype has been built and has retrieved the sound from several records. In this paper, we present signal processing issues of this system and in particular a signal to noise ratio analysis.

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 a single copy. 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 losing an important cultural heritage.

2. THE VISUALAUDIO CONCEPT

In record players, a needle follows the position of the groove and converts it into an electrical signal corresponding to the sound. It means that it is the radial displacement (and the depth for stereophonic and vertical cut records) of the groove that contains the sound information. Then the sound information is contained in the image of the record. This observation has led to a technique called VisualAudio, a 3 step concept which is shown on figure 1 [2] [3]:

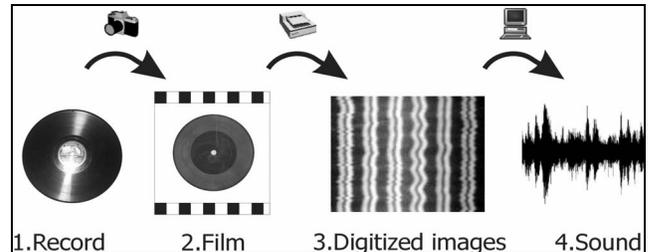


Fig. 1: The VisualAudio concept.

1. An analog picture of each side of a disk (either 33 or 78 rpm) is shot. The film must have a high spatial resolution and be as large as possible (about 1:1), 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 disks deteriorate.
2. When one wants to recover the sound, the film is digitized using a specially designed rotating scanner.
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 (which contains the sound), to detect cuts and to correct other defects. Digital signal processing must be applied to the groove signal to extract the sound.

Based on this idea, a prototype has been built and we have demonstrated that this technique can be used to extract the good quality sound. We hope that this technique can help the sound archives to preserve our cultural heritage. Recovered sound samples can be found on www.eif.ch/visualaudio.

2.1. 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 the 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.

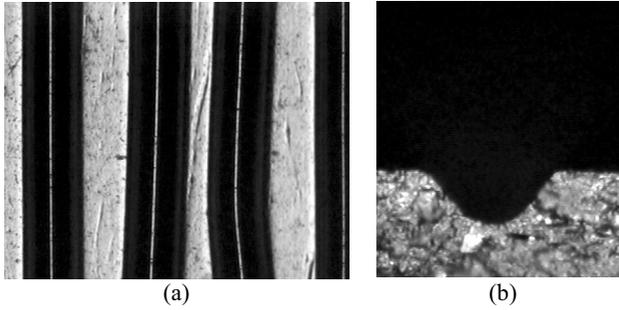


Fig. 2: Microscope top view (a) and profile (b) of 78 rpm.

Several such curves have been used (NAB, FFRR, AES, ...) but the RIAA was standardized and broadly used since 1955 [4]. 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.

The main disk characteristics that will be useful for image processing are listed in the table below [4]:

	78 rpm	33 rpm
Groove width:	31-187 μm	25.4 μm
Groove deviation	75 μm	28 μm
Groove spacing	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

It should be noted that for the most historical records, direct cut disks, the bandwidth is rarely greater than 6 kHz. Figure 2 shows a picture of a record as seen on a microscope (a) and the shape of the groove (b). On the figure 2 (a), the sound information is contained in the horizontal position (it is the radial stylus displacement on a turntable), while the vertical axis corresponds to the time. The dark lines are the grooves, and the thin white lines the bottom of the grooves.

2.2. Scanning

The film lies on a glass turntable and is digitized by a 2048 sensors linear camera. On one rotation of the film, we scan a ring of the film. By a radial displacement of the camera, adjacent rings are scanned in order to digitize the whole record. The sampling rate of the camera combined with the turntable speed defines the image sampling frequency, ranging from 25 to 200 ksamples per ring. This frequency combined with the disk speed (33 rpm, 78 rpm, ...) defines the audio signal sampling frequency, ranging from 13.75 to 110 kHz for 33 rpm, and from 32.5 to 260 kHz for 78 rpm records.

As the camera is mounted on a 10x magnification optic, each of the 2048 sensors corresponds to a $1 \times 1 \mu\text{m}$ portion of the static image. As the glass tray rotates, each CCD-sensor integrates the light on a rotating length of the film. This integration time is constant but corresponds to different lengths on the outer or inner part of the disk. The disk area averaged by one pixel may

vary from $1 \times 2.66 \mu\text{m}$ up to $1 \times 37.44 \mu\text{m}$ depending on the sampling rate and the radial position on the disk.

2.3. Image characteristics

Due to the shape of the groove, the digitized images show one trace for 33 rpm grooves and two traces for 78 rpm grooves, meaning that we have 2 edges for 33 rpm and 4 edges for mono 78 rpm grooves. The left and right edges contain different information on stereo grooves, while on mono records, the groove width is constant and all the edges are supposed to be equidistant. These correlated signals may be used in different ways to improve the quality of the VisualAudio processing:

- If an edge is damaged by a scratch or hidden by a dust, the real position of the groove can be estimated by using only the other edges.
- If the edges are all of the same quality, we get a better accuracy by averaging the position of all the edges. This improves the sound quality with a gain of 3 dB for two edges and 6 dB for four edges, since the noise on the different edges can be considered as statistically independent.

3. NOISE ANALYSIS AND MODELLING

Several kinds of noises and distortions could be identified on the recovered audio signal. Mainly:

- Broadband noise: random noise. After sampling, this noise has a uniform spectral density up to half the sampling frequency. It appears as a hiss on the audio signal.
- Clicks or impulsive noise: non-linear effects, or physical local damages.
- Wow and flutter: pitch variation effects. Frequency modulation of the signal. Wow is due to a very low frequency (up to 6 Hz) and flutter describes higher frequencies perturbation.
- Parasite frequencies: additional frequencies caused by external sources.

Extracting the sound from the record is an image and signal processing problem, considering that we have a signal that is degraded by noise and distortions produced by the bad quality of the record, the picture taking, the scanning and the groove extraction algorithm.

3.1. Quality of the record

The surface of the disk produces two kinds of noises. The grain of the disk material (shellac, PVC, acetate) is really small and appears as white noise on the image.

Dust particles, scratches or damage to the groove's edge produce local spots on the image corresponding to impulse noise on the edge and therefore audio signal. This kind of damages can be corrected by interpolation, but we can't avoid some local signal distortion in some cases on the extracted sound, as the original information is no more entirely recoverable.

3.2. Picture taking

The exact size of the film grain as it visually appears on the film is hard to define because the emulsion on the film is made of a

~10 μm depth layer of gelatin containing randomly distributed silver halide grains, with a size of 0.2 to 2 μm . The exact shape (kind of filamentary structure) of these grains depends on the film, the exposure time and the development process. So the observable grain structure is more an overlap of grain particles at different depths, and this may produce white noise on the edges of the grooves of the acquired image. Taking into account that the image is scanned (sampled), it means that the edges have a constant noise spectrum up to half of the sampling frequency.

The disk illumination should be uniform, otherwise the traces width and intensity may vary on the film, therefore creating wow effect and complicating the image processing step.

If the focal distance is too small, geometrical distortion appears on the outer part of the record. Therefore the extracted signal quality on the outer part won't be as good as on the inner part of the disk.

3.3. Scanning

The current rotating scanner is a prototype, and there are still several mechanical and optical aspects to improve. The speed is not perfectly constant. This is not annoying since the sampling is synchronized to the resolver of the motor. A motor producing 200 pulses per rotation drives the glass turntable. The scanner causes following distortions:

- Irregular sampling: the sampling is triggered by the resolver of the motor. Therefore, the angular scanning frequency should be constant (and therefore also the sampling of the sound). However, the resolver's uncertainty as well as the reducer's play produces a jitter of the sampling time, causing distortions (random modulation of the sound).
- Vibrations: the 200 pulses per cycle are mechanically propagated in the prototype and therefore appear on the extracted sound as 110 Hz (33 rpm) or 260 Hz (78 rpm) constant hiss with decreasing amplitudes at the harmonic frequencies. Intermodulations with the sound also appear.
- Optical distortions may also appear during scanning, including geometrical effects due to bad focusing, and non-constant illumination.

If the film is badly centered on the glass tray, the digitizing is not uniform during the rotation causing a wow effect on the extracted sound (same as with a regular record player with off-axis center). The bad centering can be corrected mechanically or by software, as far as it can be measured by image processing.

4. SIGNAL TO NOISE CONSIDERATIONS

It is not sufficient to characterize the sound extraction performances by perceptual comparisons. For sounds, that are natural speech or music signals, it's impossible to measure the noise and the distortions. Using test records, it is possible. We digitize a portion of a sine wave from the test record, and by 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 figure 3, we have the spectrum of a reconstructed 300 Hz sine wave. We observe noise, as well as peaks due to the motor vibration (intermodulations with the 110 Hz motor vibration).

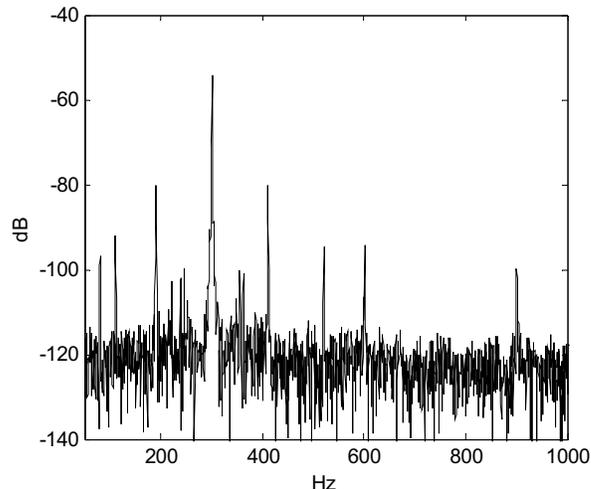


Fig. 3: Spectrum of a reconstructed 300 Hz signal extracted from the HFN 001 33 rpm Test Record.

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

To estimate the signal to noise ratio, 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:

$$B_n = 130 \text{ kHz}, \quad B_s = 12 \text{ kHz}, \quad N = 4.$$

4.1. 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 :

$$\sigma_x^2 = \frac{A^2}{8} \quad (1)$$

The power of the noise is:

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

The signal to noise ratio is:

$$\text{SNR} = \frac{\sigma_x^2}{P_n} = \frac{A^2 NB_n}{8\sigma_n^2 B_s} \quad (3)$$

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

4.2. SNR for constant velocity records

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

$$v(t) = \frac{ds}{dt} = 2\pi f_0 A_{f_0} \cos(2\pi f_0 t) \quad (4)$$

The reconstructed signal $x(t)$ is equal to the original signal $s(t)$ with an additional noise $n(t)$, with variance σ_n^2 :

$$x(t) = A_{f_0} \sin(2\pi f_0 t) + n(t) \quad (5)$$

The velocity $y(t)$ of the reconstructed signal is then defined with $n_1(t) = \frac{dn}{dt}$:

$$y(t) = \frac{dx}{dt} = 2\pi f_0 A_{f_0} \cos(2\pi f_0 t) + n_1(t) \quad (6)$$

The power spectral densities of the signal $\phi_y(f)$ and of the noise $\phi_n(f)$ and $\phi_{n_1}(f)$ define the respective powers P_s and P_n and thus the SNR:

$$\phi_y(f) = \frac{4\pi^2 f_0^2 A_{f_0}^2}{4} (\delta(f + f_0) + \delta(f - f_0)) \quad (7)$$

$$\phi_n(f) = \frac{\sigma_n^2}{2NB_n} \quad (8)$$

As $n_1(t)$ is the derivative of $n(t)$:

$$\phi_{n_1}(f) = \frac{\sigma_n^2}{2NB_n} 4\pi^2 f^2 \quad (9)$$

$$P_s = 2\pi^2 f_0^2 A_{f_0}^2 \quad (10)$$

$$P_{n_1} = 2 \int_0^{B_s} \frac{\sigma_n^2}{2NB_n} 4\pi^2 f^2 df = \frac{4\sigma_n^2 \pi^2 B_s^3}{3NB_n} \quad (11)$$

$$SNR = \frac{3f_0^2 A_{f_0}^2 NB_n}{2\sigma_n^2 B_s^3} \quad (12)$$

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

4.3. SNR for equalized records

The signal and the noise of an equalized record are filtered by the preamplifier. If we consider the RIAA transfer function $H(f)$ as an example, the filtered noise and signal powers are then:

$$\phi_{n_2}(f) = \phi_{n_1}(f) |H(f)|^2 \quad (13)$$

$$P_{n_2} = \frac{4\sigma_n^2 \pi^2}{2NB_n} 2 \int_0^{B_s} |H(f)|^2 f^2 df \quad (14)$$

$$P_s = 2\pi^2 f_0^2 A_{f_0}^2 |H(f_0)|^2 \quad (15)$$

σ_n must then be smaller than $1.28 \mu m$ to reach a 40 dB SNR using (14) and (15), with $f_0 = 500$ Hz and $A_{f_0} = 75 \mu m$.

5. REMARKS

The proposed system is especially useful for disks in bad condition, for example lackered 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 [5]. 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.

6. CONCLUSION

Calculations have shown the maximum allowable noise on the groove edges in order to get good quality recovered sound. Due to the averaging sampling and the hard to define film grain structure, the above calculated standard deviations cannot be analytically related to the disk grain or to the film grain. However, the order of magnitude obtained by the calculations shows that it is possible to reach satisfying results using films to archive disk images and that the actual prototype has a sufficient resolution. It confirms the choices operated on the prototype. It also indicates the film quality and resolution necessary in the final system that is currently in the specification phase. Improvements are still needed at each step of the process: picture taking, scanning and image analysis to build a final system. After the first tests, 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

Thanks to Thierry Fumey, Yves Pauchard, Pascal Bovet and Jacques Miauton for their help in this project. This project was partly funded by project HES-SO-01.01TI and by the Swiss National Science Foundation (project 21-64984.01). We thank the Gebert R uf Foundation for funding the final prototype.

References

- [1] B. Fox, "Not fade away" New Scientist, 177 (2384), pp. 40-43, March 1st 2003.
- [2] S. Cavaglieri, O. Johnsen, F. Bapst, "Optical Retrieval and Storage of Analog Sound Recordings", AES 20th International Conference, Budapest, 2001.
- [3] O. Chassot, C. Miauton, N. Bosi, "VisualAudio", Diploma project at Ecole d'ing nieurs et d'architectes de Fribourg, Switzerland, 2000.
- [4] "Anthology of the Journal of Audio Engineering Society", "Disk Recording" vol. 1 & 2, 1953-1980.
- [5] 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