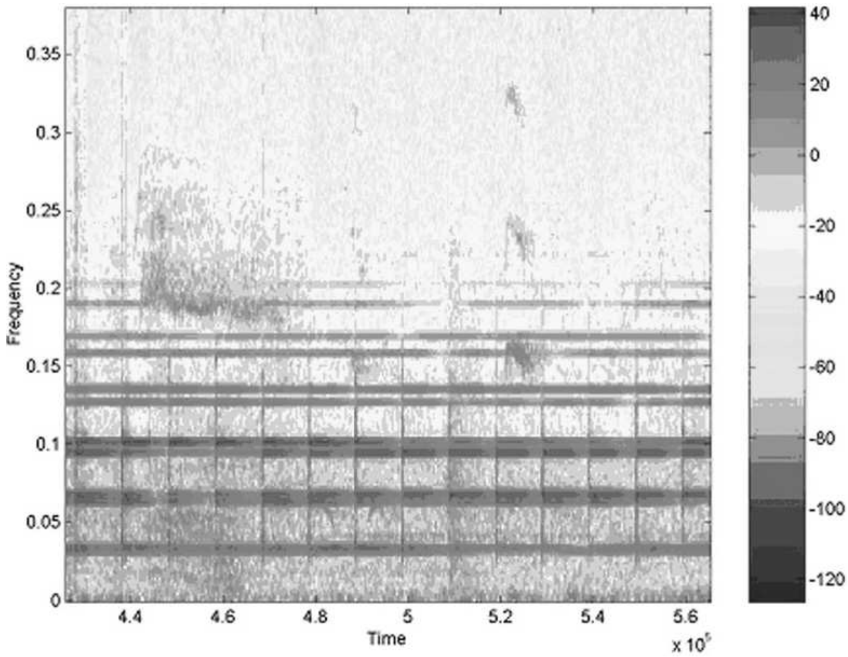
RE-INVENTING FOURIER

Jean-Julien Aucouturier, Temple University Japan, Tokyo, Japan.
E-mail: <aucouturier@gmail.com>.

Abstract

Artists asked to represent sound in a visual way mysteriously re-invented many of the concepts that preside over the mathematical signal transforms used in computer music. Their drawings adopted a systematic 2-dimensional structure, and sometimes resembled time-frequency representations such as the Fourier transform. This makes us ponder here over the different criteria of what makes a "good" computer representation for the scientist, and what makes a "good" visual work for the artist.

Fig. 1. Fourier spectrogram of a 1.5-second extract of a soundscape recorded at a train crossing in Tokyo, Japan. Regular vertical lines correspond to the ring of the signal. At time 5.2, one can see an isolated bird call. (© JJ. Aucouturier.)

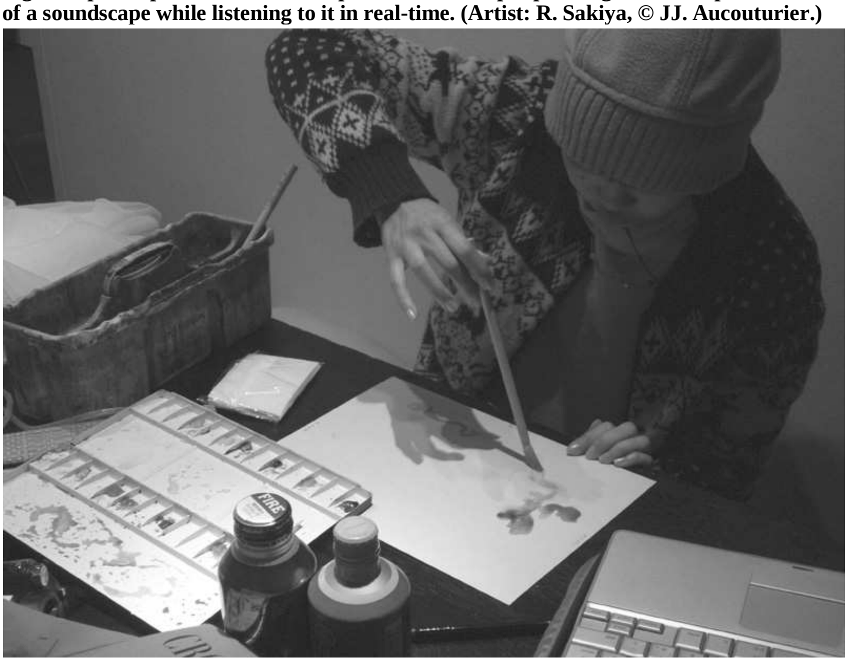


A sound waveform, when recorded in digital format, contains a lot of details: everything needed to play the sound with high-fidelity. Too much details in fact, as it is nearly impossible to understand what the sound is by looking at it. This is a bit like looking at a very high-definition image by zooming very close: we loose the big picture. If we want to see and make anything useful with sound, we need to find another representation, a "transform" that changes the sound file into something simpler, something of which we can make more sense. Computers can do many sound transforms, the most famous of which is the Fourier Transform.

Joseph Fourier is a 18th-century French physicist who invented a way to calculate the many simultaneous) speeds at which a phenomenon repeats itself. For sound, these frequencies notably tell us how high a given note sounds: the notes at the right end of a piano have a high frequency ("sharp"), those at the left end have a low frequency ("bass"). The Fourier transform (more precisely, a spectrogram of short-time fast Fourier transforms FFT) shows time horizontally and frequency vertically. A FFT contains a lot of information. It reveals things that are nearly impossible to hear: when our computer produced the FFT representation shown in Figure 1, we could *see* clearly (with a little practice) a bird call at time 5.2 (3 superimposed inverted-V shapes), which took us several minutes to *hear* even after we knew it was there.

We (humans) tend to describe sounds in many ways: their color, brightness, their timbre (piano or clarinet), their mood (happy or sad), most of which are impossible to calculate for a computer. Conversely, FFT shows a property of sound, "frequency", that humans cannot easily understand. We do not *hear* frequency (we hear *pitch*, which is quite a different construct in the case of everyday sounds [1]). The only reason computers do FFTs is because they can: the work of Fourier and others has yielded algorithms to let computers find/calculate these frequencies.

Fig. 2. A participant of the workshop sound/mindscapes painting a visual representation of a soundscape while listening to it in real-time. (Artist: R. Sakiya, © JJ. Aucouturier.)



The art students involved in the Sound/Mindscape workshop (held in Art Center Ongoing, Tokyo, Nov. 26~31, 2008 [2]) did not know about Fourier, and they didn't know about computers. When asked to represent a long sequence of sounds in a visual way, they reached for the paper and they painted. They explored different ways to trace sounds. After a while, they decided to try out some "rules" (like the rules of a game), i.e. time should go from left to right of the page, size of the brush should be proportional to the amplitude of the sound, etc. (Fig.3) After a long experimentation, they presented one way to draw sounds with which they were particularly happy. From left to right – they explained – is the *time*, and from bottom to top of the page, the *height* of the sound.

I gasped. They had re-inventing Fourier.

Before my eyes was a human rendering of a FFT spectrogram (Fig.4), splashes of paint drawn in synchronization with the sounds, representing them in a time-frequency space.

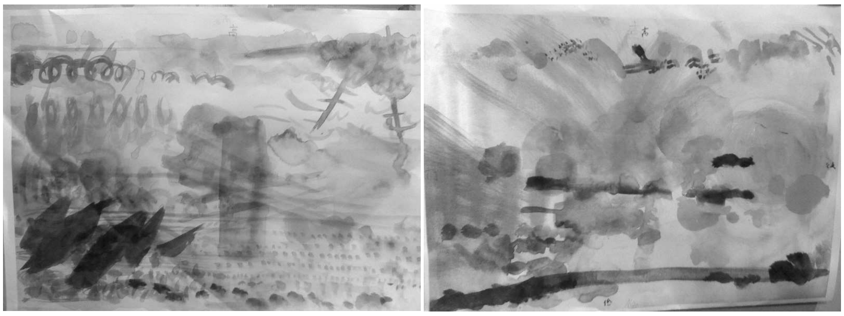
Then I realized my mistake. It wasn't frequency. When the artists talked about "sound height", they really meant this: height. How many meters in the air the sounds come from. Birds, HIGH in the sky; underground train, way down beLOW. What I was looking at was a *time-height* representation. Surprisingly close to Fourier frequency on a linguistic level, but different concept altogether.

Of course. I wasn't talking to computers.

Nevertheless, their painting had an intriguing appeal. It was well organized, the high/tall sounds balanced by a continuous stream of low sounds (which turned out to be footsteps), evenly distributed in time, yet not boring thanks to the occasional random occurrence of a stroke here and there. The rules brought some *things* out of the sound, made them come real and visual in a very unique way. I could feel why the artists were satisfied with this "transform". It was as satisfying for them as a Fourier transform is for me.

There, decades of computer research have favored the Fourier transform, for a purpose. The representation it provides is highly informative. It is compact, easy to understand, characterizes sounds precisely without drowning them in a haze of unnecessary details. With a FFT, one can identify bird songs, find the precise time of appearance of a plane, count how many trains in a day, or even transcribe the notes of a musical melody. It carries an optimal amount of information – something scientists since Claude Shannon can measure in a very precise way.

Fig. 4. "Sound transforms" created by the workshop participants, painted in real-time while listening to 3-minute soundscape recordings. Paintings were constrained to have a precise 2- dimensional structure, agreed in advance between participants (water color on paper, © T. Murase, Y. Inagaki, 2008. Photographs © JJ. Aucouturier)

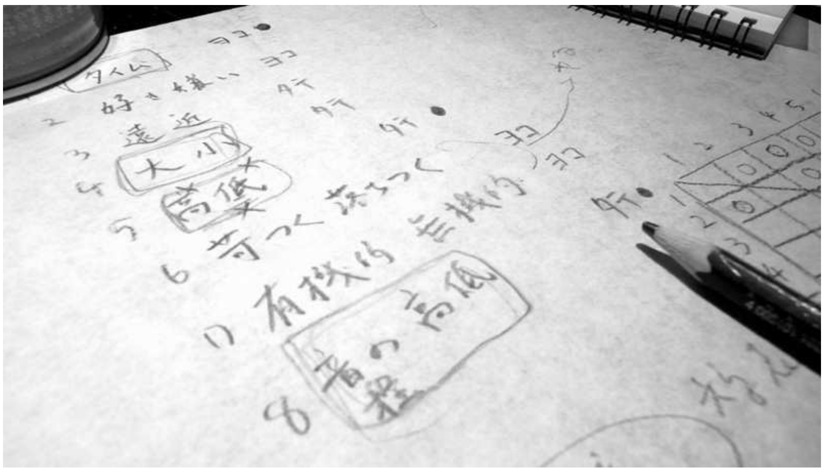


Here, hours of artistic research were eventually favoring the time-height transform. Why? What does this transform (which the artists called a "rule") do that other alternatives do not achieve? Why set down for this and not, say, loud sounds from left to right and bright sounds from bottom to top. The goal is not information – since the representation is not meant to be processed nor to be useful. Beauty? I suspect it's not just beauty – this would be a very limited vision of artistic research. Harmony? Interestingness? Artists are reluctant to define the goal of their practice, yet they talk a lot about "it". They talk about "cheap" representations (too obvious? Too objective?). They talk about honest piece of work (as in revealing some sincerely felt subjectivity?). What precisely? Why time and height? Provocatively, I try to write a mathematical formula characterizing the properties of their chosen representation, compared to their previous, less interesting attempts. Do you optimize the space coverage (think of maximum-entropy spectral estimation [3])? The distribution of color? Of stroke size (think of sparse transforms [4])? The ability to visualize both small events and global scenes (think of wavelet transforms [5])?

They giggle and resist what they probably view as my attempt to find a magic formula for art. I'm not. I'm sincerely trying to understand, not to reduce. Maths is probably not the right way, I'm willing to admit. But what else would we have? What other option? Precisely?

This is important, I thought, this is crucially important. It felt like a very rare instant of contact between my world and their world. Something that, if we can sustain for long enough, could change the way I do science, and change the way they do art.

Fig. 3. Working documents showing the rules used to paint visual representations of sound. Each line gives the extreme values of a dimension axis, with its favored orientation: 1- time (horizontal), 2- pleasing vs disagreeable (horiz.), 3- far vs close (vertical), 4- large vs small (vert.), 5- high vs low position (vert.), 6- unsettling vs calm (horiz.), 7- organic vs artificial (horiz.), 8- note height low vs high (vert.). On the right side, a matrix is used to combine the dimensions into 2-dimensional representations in a systematic way (© T. Murase, Y. Inagaki, 2008. Photographs © JJ. Aucouturier)



Workshop participants: Shoya Arai, Jean-Julien Aucouturier, Cathy Cox, Tadasuke Go, Takashi Hamada, Tomohiko Hayakawa, Lindee Hoshikawa, Hideaki Idetsuki, Yu Inagaki, Jess Mantell, Toshi Murase, Ai Nagashima, Rie Sakiya, Jaehwa Shin, Shintaro Soma, Nao Tokui.

References and Notes

1. Ives, D. T. and Patterson, R. D. **(**2008**).** "Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics," *J. Acoust. Soc. Am., 123(5)*

2. Sound/Mindscapes workshop and exhibition, Art Center Ongoing, Kichijoji, Tokyo, Japan, Nov. 26th~Dec.7th. Organized by JJ. Aucouturier and H. Idetsuki. <http://soundmindscapes.info>

3. Cover, T. and Thomas, J. (1991) Elements of Information Theory, John Wiley and Sons, Inc.

4. Daudet, L. Sparse and structured decompositions of signals with the Molecular Matching Pursuit, IEEE Transactions on Audio, Speech and Language Processing, Vol. 14(5), pp. 1808-1826, 2006

5. Tzanetakis, G., G.Essl, and P. R. Cook. "Audio Analysis Using the Discrete Wavelet Transform," In Proceedings of of WSES International Conference, Acoustics and Music (AMTA), Greece, 2001.