Spectra of sounds and images
– lecture notes –

Pedro M. Q. Aguiar
Institute for Systems and Robotics / IST

February 2008

Abstract

These notes overview the spectral representation, emphasizing its compactness for describing acoustic signals, *e.g.*, music, and visual patterns, *e.g.*, textures and contours.

1 Sinusoids in Nature

We start by looking at the general class of sinusoidal signals. One of the reasons why this class is relevant in practice is that many physical systems produce signals that are well approximated by sinusoids. As an example, suggested by [1], consider the tuning fork in Figure 1.

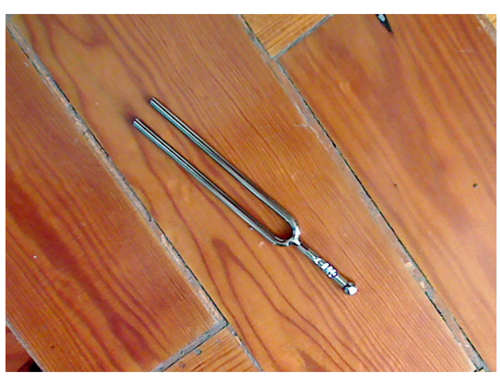

Figure 1: Tuning fork.

When stroked, the tuning fork produces a sound, a persistent "hum". Figures 2 and 3 plot this sound, which is usually called a "pure tone". Is it a coincidence that the (stationary part of the) signal looks like a sinusoid ?

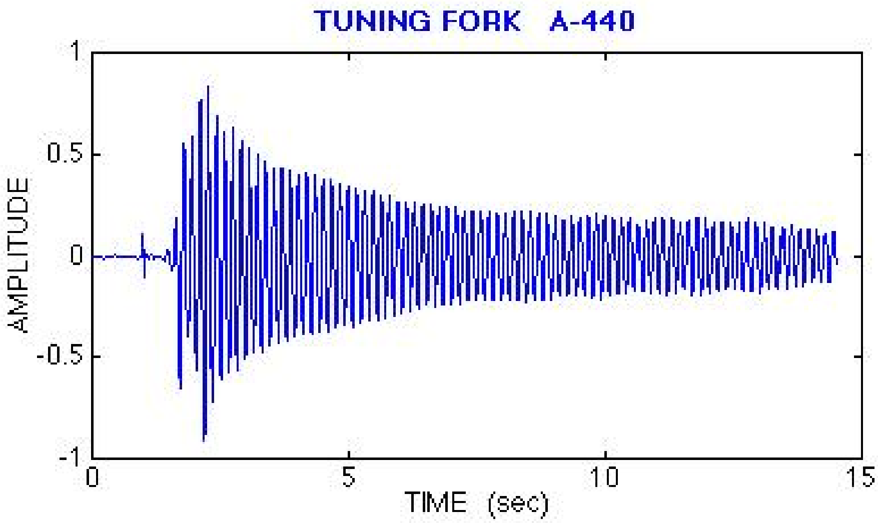

Figure 2: Acoustic signal produced by a tuning fork. (From [1].)

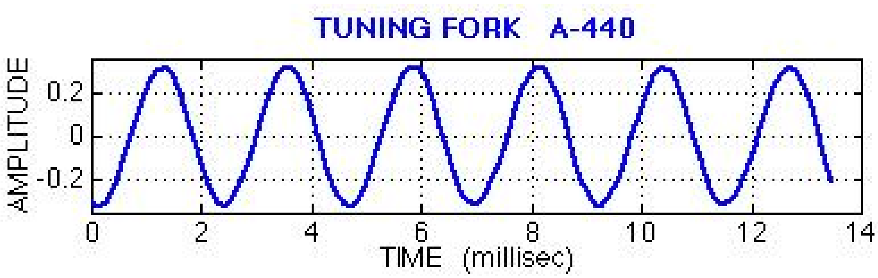

Figure 3: Stationary part of the signal in Figure 2, shown at a different scale.

Naturally, when stroked, the tuning fork tines vibrate. This vibration, transmitted through the air particles surrounding the fork, produces the audible sound. To derive the way the tuning fork tines (and thus the air particles) vibrate, consider the schematic diagram in Figure 4. It represents the fork after being stroked, *i.e.*, one of the tines is deformed by a given displacement from its equilibrium position.

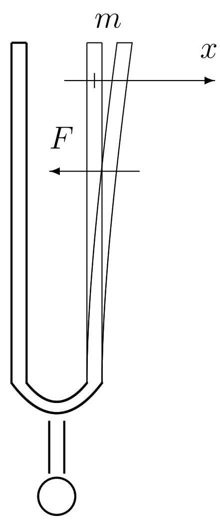

Figure 4: Vibrating tuning fork model.

Unless the initial deformation is so large that the tine bends (or breaks!), the metal can be thought as an elastic material. Thus, there is a (time-varying) restoring force $F(t)$ that is proportional to the deformation $x(t)$ (Hooke's law), both represented in Figure 4:

$$F(t) = kx(t)\,, \tag{1}$$

where k is an elastic constant measuring the stiffness of the material.

Naturally, the restoring force produces a proportional acceleration $a(t)$ on the tine (Newton's second law), *i.e.*,

$$a(t) = \frac{d^2x(t)}{dt^2} = -\frac{1}{m}F(t)\,, \tag{2}$$

where m is the mass of the tine. The minus signal is due to the fact that, obviously, the deformation $x(t)$ and the restoring force $F(t)$ have opposite directions (see Figure 4).

Replacing (1) in (2), we obtain an equation that describes the dynamics of the fork, *i.e.*, the time evolution of the tine deformation $x(t)$:

$$\frac{d^2x(t)}{dt^2} = -\frac{k}{m}x(t)\,. \tag{3}$$

The solution of the second-order linear differential equation (3) is given by a sinusoidal signal,

$$x(t) = A\cos(\omega_0 t + \phi) \ . \tag{4}$$

In fact, by replacing (4) into (3), we get

$$-\omega_0^2 A\cos(\omega_0 t + \phi) = -\frac{k}{m} A\cos(\omega_0 t + \phi) \ , \tag{5}$$

which holds for all t if the frequency of the sinusoidal signal is

$$\omega_0 = \sqrt{\frac{k}{m}} \, , \tag{6}$$

regardless of the values of the parameters A and ϕ.

We conclude that the fork tines (and thus the air particles) vibrate sinusoidally (in agreement with the observation in the plot of Figure 3), with a frequency ω_0 that is uniquely determined by the tuning fork characteristics: it increases with the metal stiffness and decreases with the tine mass. Parameters A and ϕ, respectively, amplitude and phase shift, are not determined by the dynamics of the fork, but rather by its initial deformation, *i.e.*, by two auxiliary conditions required to make unique the solution of the second-order differential equation (3), *e.g.*, $x(0) = x_0$, $\left.\frac{dx(t)}{dt}\right|_{t=0} = v_0$.

The fact that forks like this vibrate at a fixed frequency makes them useful to tune musical instruments. For example, the fork in Figure 1 is a A-440 tuning fork. It's natural frequency is 440Hz ($\omega_0 = 2\pi \times 440$ rad s^{-1}), the frequency of the note A ("Lá") above middle C ("Dó") on a Western musical scale. This corresponds to a period $T = \frac{1}{440} \simeq$2.27ms, which is observed in the plot of Figure 3 (notice that there are five periods in approximately 11.5s).

The parameters ω_0, A, and ϕ in (4) are an efficient (*i.e.*, compact) representation of the (stationary part of the) acoustic signal produced by the tuning fork.

2 Spectral harmonics

Although other physical devices produce sounds that are approximately sinusoidal, *e.g.*, whistles, the sinusoidal model is far from being general, even

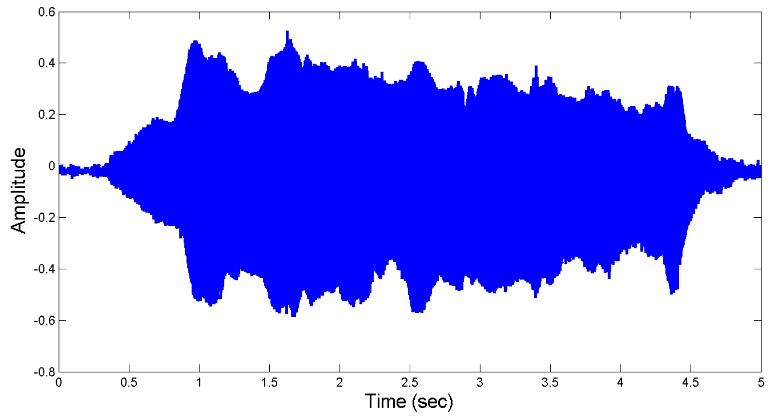

Figure 5: Acoustic signal produced by a violin.

to describe single musical notes. As an example, see in Figures 5 and 6 how the sound of an A above middle C looks like when played by a violin.

From the plot in Figure 6, we see that, although not sinusoidal, the (stationary part of the) acoustic signal produced by the violin is periodic. Furthermore, the period is the same as the one of the sound produced by the A-440 tuning fork (see Figure 3), giving intuition on why the fork can be used to tune the violin.

Naturally, the acoustic signal in Figure 6 does not admit such a compact representation as a sinusoidal signal. However, the time evolution of a vibrating string, such as the one producing the sound in Figure 6, can also be predicted from the fundamental laws of Physics, leading to a partial differential equation (because space derivatives are also involved), whose solution is a superposition of sinusoids. Rather than studying the dynamics of the vibrating string, we invoke the more general result wonderfully captured by Jean Baptiste Fourier in 1807, which states that the generality of periodic signals [1] can be obtained by summing sinusoids of frequencies multiples of a fundamental frequency (usually called spectral harmonics, or partials).

To clarify, consider a generic periodic signal $x(t)$, *e.g.*, the signal in Figure 6. The period T of the signal defines its fundamental frequency:

$$\omega_0 = \frac{2\pi}{T} \,. \tag{7}$$

[1] Exceptions are irrelevant in practice, *e.g.*, signals with infinite number of discontinuities in a finite interval.

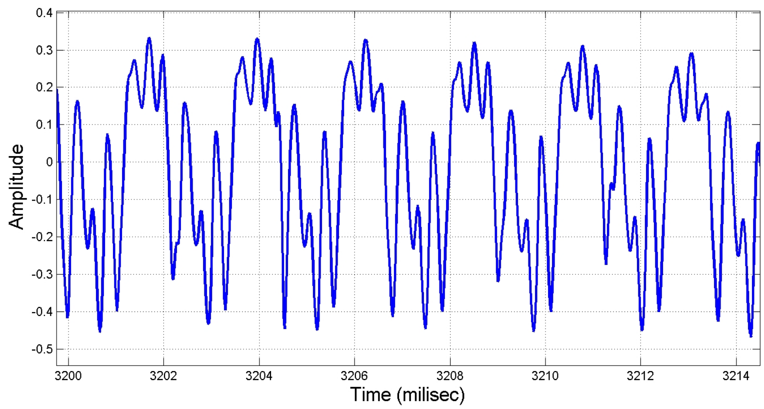

Figure 6: Stationary part of the signal in Figure 5, displayed at a different scale. Compare to the plot in Figure 3.

In the case of the signal in Figure 6, the fundamental frequency is the same frequency of the sinusoidal signal produced by the A-440 tuning fork. Fourier proposed to represent $x(t)$ by what is now called its spectral decomposition, *i.e.*, a linear combination of sinusoids of frequencies $k\omega_0$:

$$x(t) = A_0 + \sum_{k=1}^{\infty} A_k \cos\left(k\omega_0 t + \phi_k\right) . \tag{8}$$

Naturally, the parameters $\{A_0, A_k, \phi_k, k = 1, \ldots\}$ determine the "shape" of a period of the signal.

For the spectral representation (8) of the violin acoustic signal in Figure 6, the first component, *i.e.*, the one corresponding to the fundamental frequency, is precisely the (possibly scaled and delayed version of the) fork sinusoidal acoustic signal in Figure 3. In practice, many signals, *e.g.*, the violin sound in Figure 6, are well approximated by truncating expression (8) to a finite (and small) set of, say, K harmonics, thus the parameters $\{A_0, A_k, \phi_k, k = 1, \ldots, K\}$ are a compact representation for those signals.

For simplicity, the spectral representation in (8) is often written in terms of complex exponentials, leading to what is usually called the Fourier series of the signal:

$$x(t) = \sum_{k=-\infty}^{\infty} X_k e^{jk\omega_0 t} , \tag{9}$$

where we introduced the "complex amplitudes" $\{X_k, -\infty < k < \infty\}$, related to the parameters $\{A_0, A_k, \phi_k, k = 0, 1, \ldots\}$ in (8) by

$$X_0 = A_0, \qquad X_k = \frac{A_k}{2} e^{j\phi_k}, \qquad X_{-k} = X_k^* = \frac{A_k}{2} e^{-j\phi_k}, \qquad k \geq 1\,. \tag{10}$$

These "complex amplitudes", *i.e.*, the coefficients of the spectral representation, or the Fourier series (9), are uniquely determined by the signal $x(t)$, through

$$X_k = \frac{1}{T} \int_T x(t) e^{-jk\omega_0 t}\, dt\,, \tag{11}$$

see, *e.g.*, [2] for details.

As another example of the usefulness of the spectral representation for compactly representing periodic signals, Figure 7 plots the coefficients $\{X_k\}$ for an approximation of an acoustic signal produced by an human voiced sound. The majority of these coefficients are zero and only five sinusoidal components are enough to approximate the real sinal (Figure 8 illustrates the synthesis by using these components).

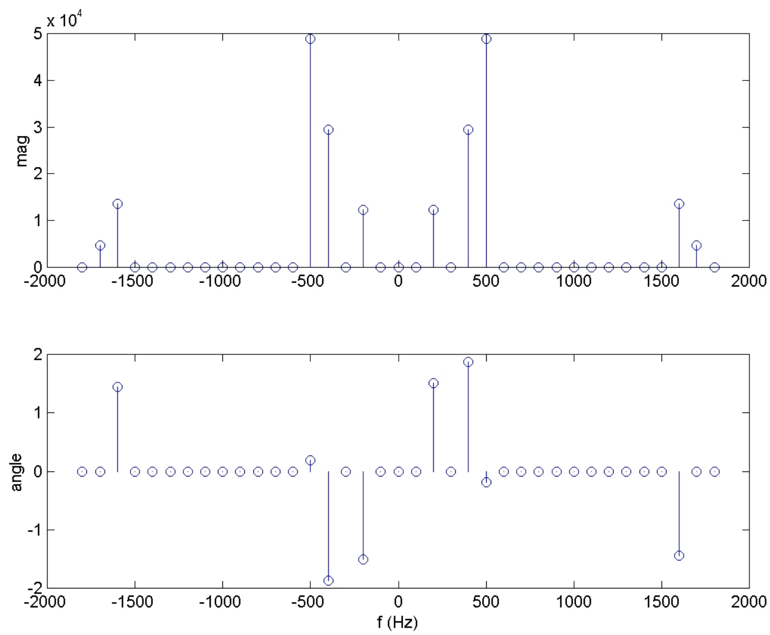

Figure 7: Spectral analysis of a voiced sound.

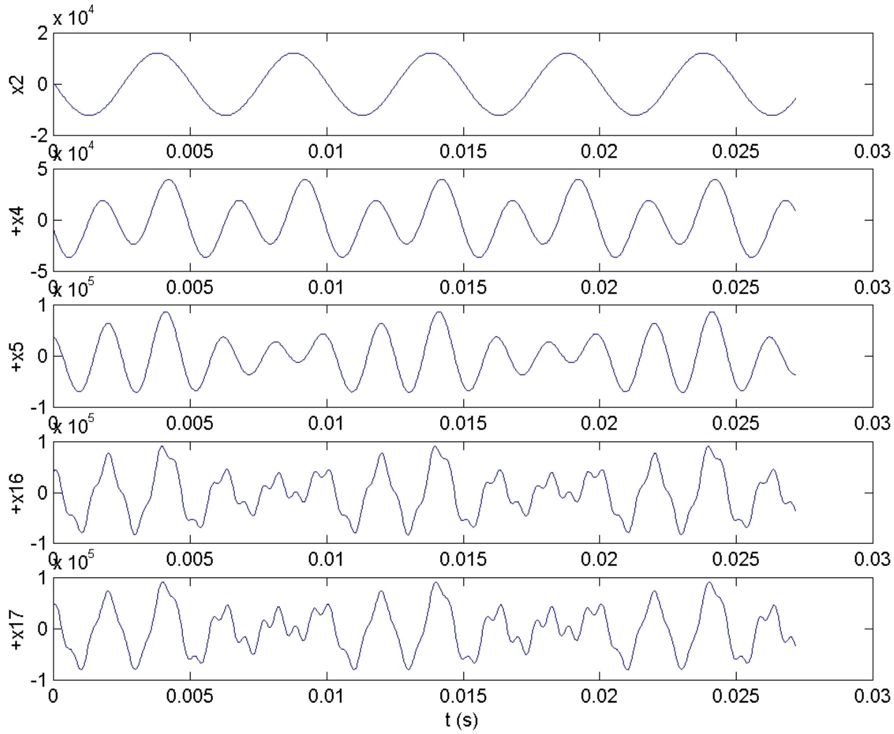

Figure 8: Spectral synthesis of a voiced sound.

3 Images, textures, and contours

The analysis in the previous section is easily extended to two-dimensional (2D) signals, *i.e.*, images, where independent variable time t is now replaced by the two spatial independent variables x and y. In particular, the sinusoidal building block is an image $I(x, y)$ given by

$$I(x, y) = A\cos\left(\omega_x x + \omega_y y + \phi\right) , \tag{12}$$

where ω_x stands for the spatial frequency along the x axis and ω_y for the frequency along y.

Figures 9 to 11 show three examples of sinusoidal images obtained by representing, in a gray level scale, the result of evaluating expression (12) for different pairs of values for ω_x, ω_y (x is represented on the horizontal axis and y on the vertical one). The resulting images are planar waves, whose direction

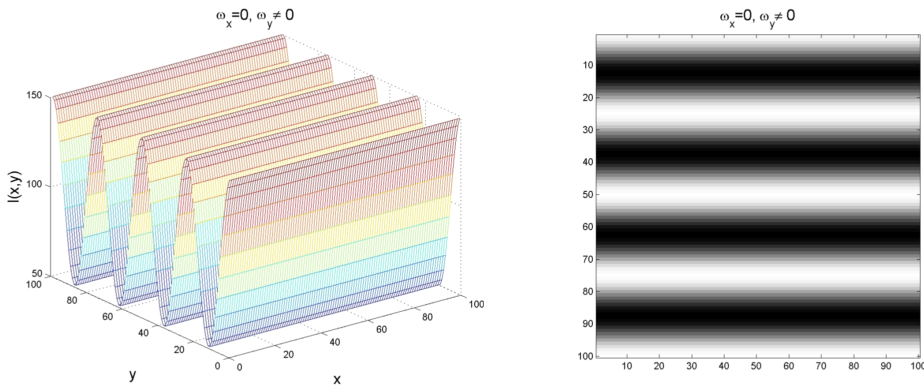

Figure 9: Sinusoidal image. Left: image intensity as a function of x and y. Right: corresponding gray level image.

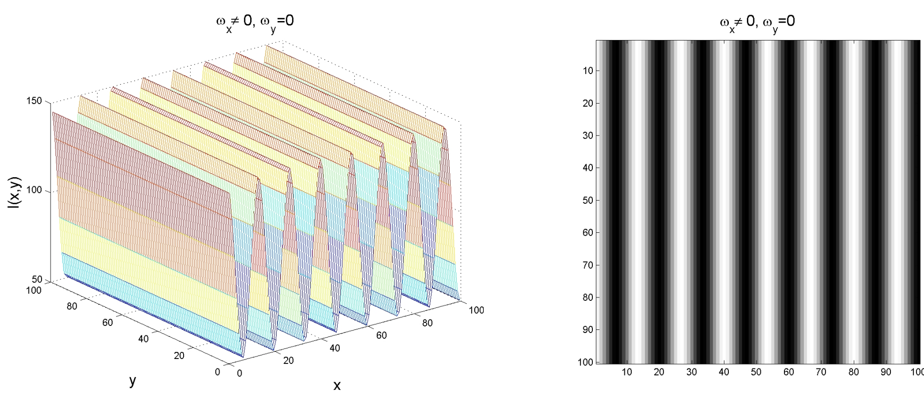

Figure 10: Sinusoidal image.

and frequency is determined by ω_x and ω_y. For example, in Figure 10, for $\omega_x \neq 0$ and $\omega_y = 0$, we get an horizontal wave, while in Figure 9, for $\omega_y \neq 0$ and $\omega_x = 0$, we get a vertical one. In Figure 11, for $\omega_x \neq 0, \omega_y \neq 0$, we get a diagonal wave, whose direction is the one of the vector (T_x, T_y) and whose period is given by $\sqrt{T_x^2 + T_y^2}$, where $T_x = \frac{2\pi}{\omega_x}$ and $T_y = \frac{2\pi}{\omega_y}$ (prove this, as an exercise).

Naturally, the theory outlined in the previous section for representing one-dimensional (1D) periodic signals in terms of sinusoids, extends to 2D in a straightforward way, see *e.g.*, [3]. Rather than dive into the corresponding equations, which are simple extensions of their 1D version above, let us

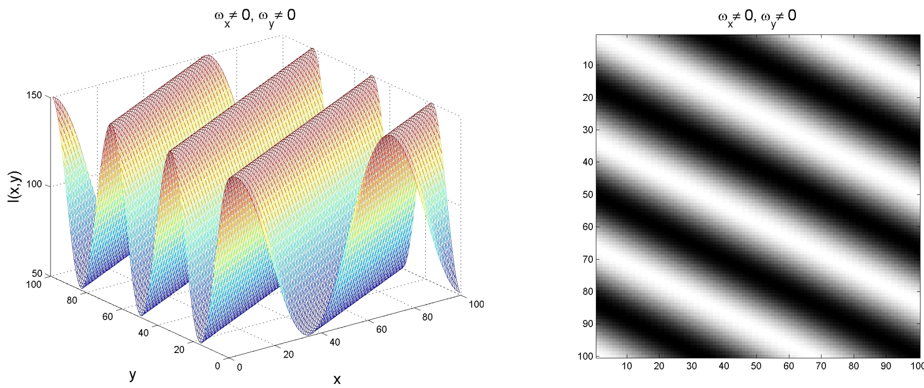

Figure 11: Sinusoidal image.

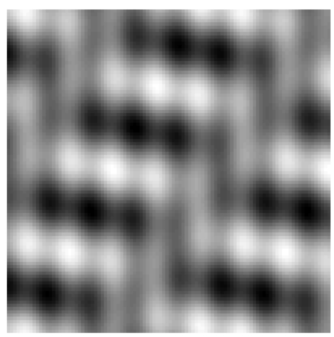

Figure 12: A textured pattern obtained by linear combination of the images on the right side of Figures 9, 10, and 11.

illustrate, with a simple example, that sinusoidal images also lead to useful compact representations. See the textured pattern image shown in Figure 12. This image, looking somehow complex and resembling a natural tissue, was obtained by a simple weighted sum of the sinusoidal images represented in Figures 9 to 11. In fact, in a similar way as for 1D signals, periodic images, *i.e.*, 2D patterns that repeat in space, are also compactly represented as linear combination of sinusoids.

We now discuss how Fourier-like representations are also useful to represent other types of visual content, namely, 2D contours. In fact, identifying the visual field with the complex plane, a closed contour can be seen as a periodic (complex) signal, see the example in Figure 13. On the left, we see an arbitrary contour. Considering a parametric representation in terms of a parameter $t \in [0,1]$, the contour corresponds to the geometric locus of the points $\{x(t), y(t)\}$. This way, an equivalent representation for the contour is the complex periodic signal $z(t) = x(t) + jy(t)$, of which one period is represented (both real and imaginary parts) in the right plot of Figure 13.

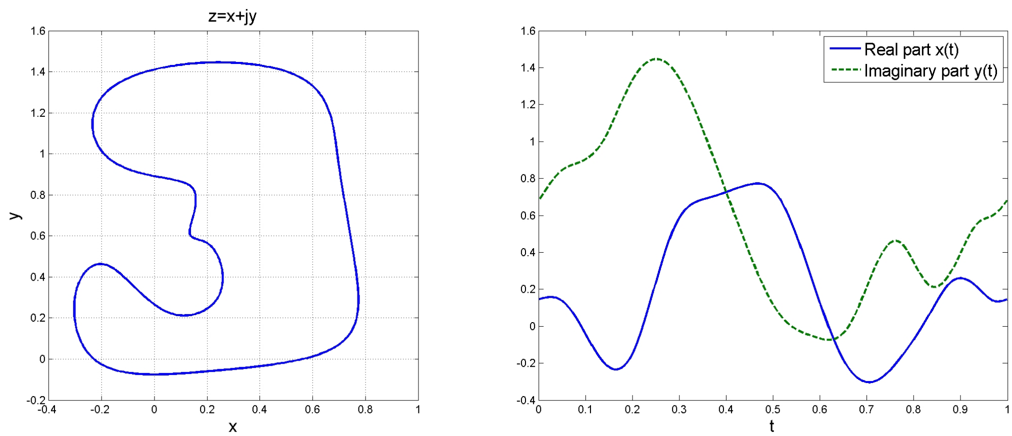

Figure 13: A contour as a periodic signal in the complex plane.

The periodic signal $z(t)$ admits, in turn, a spectral representation in terms of the expressions (9) and (11). In Figure 14, we represent the coefficients $\{X_k\}$ of the Fourier series of $z(t)$ [2]. We see that the magnitude of the coefficients is only relevant for small $|k|$, meaning that a compact representation for the contour is then obtained by simply storing those coefficients.

Figure 15 shows some of the successive approximations of the contour, obtained by using the synthesis expression (9), with increasing number of spectral components. Notice how the bottom right contour of Figure 15, obtained with only 10 spectral components, is almost visually indistinguishable from the original contour in Figure 13.

[2]Note that the coefficients of the Fourier series do not verify $X_{-k} = X_k^*$ because the signal is complex, *i.e.*, there is no equivalent representation such as (8) or the relations (10).

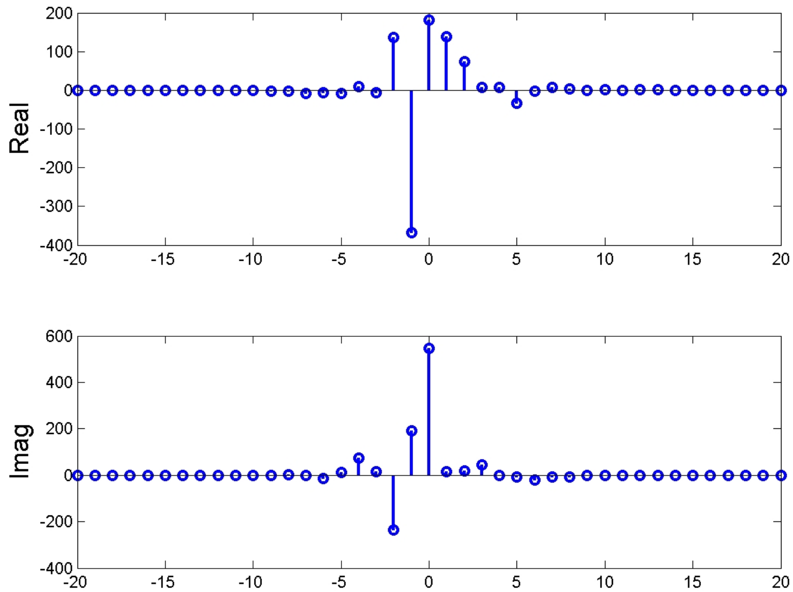

Figure 14: Spectral analysis of the contour in Figure 13.

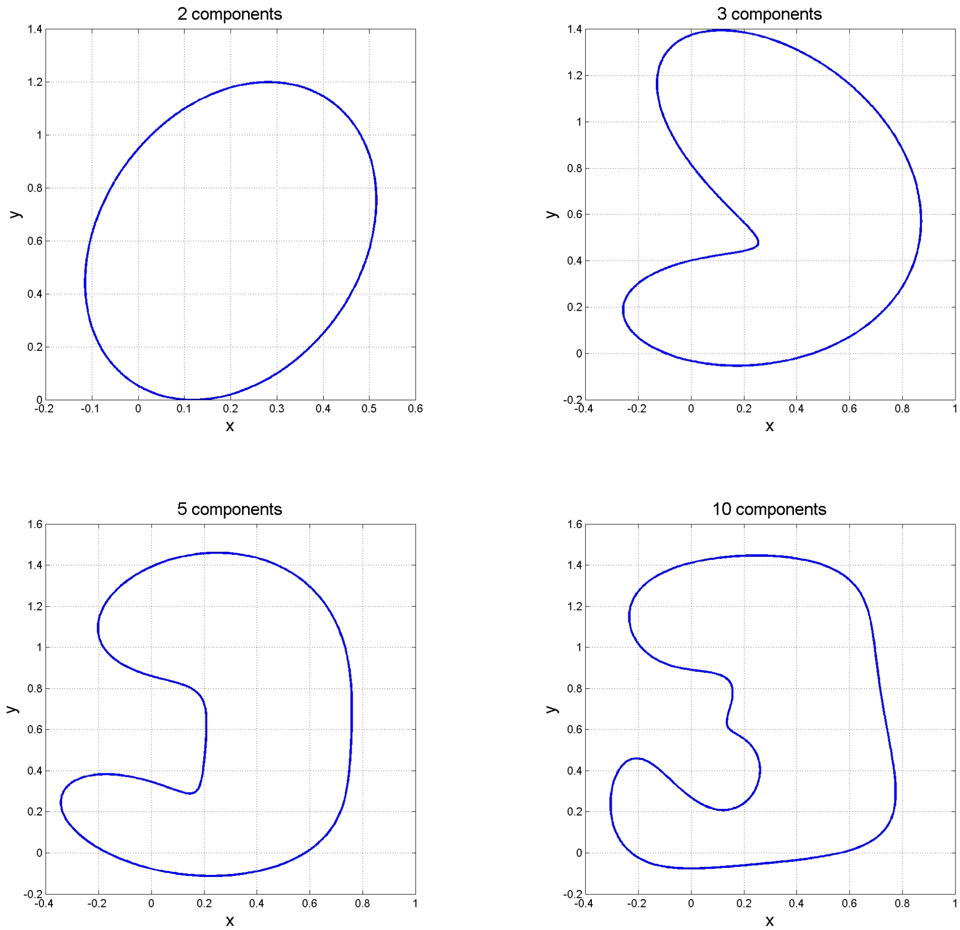

Figure 15: Spectral approximations of the contour in Figure 13.

4 Time-frequency representation and musical notation

After having illustrated that the spectral analysis provides a compact representation for repetitive patterns, we must now refer that it is not so for signals in general. In fact, the generality of signals do not repeat in time or space: just think of a general image, *e.g.*, Figure 16, a pronounced phrase, or an entire musical piece (see extract in Figure 17).

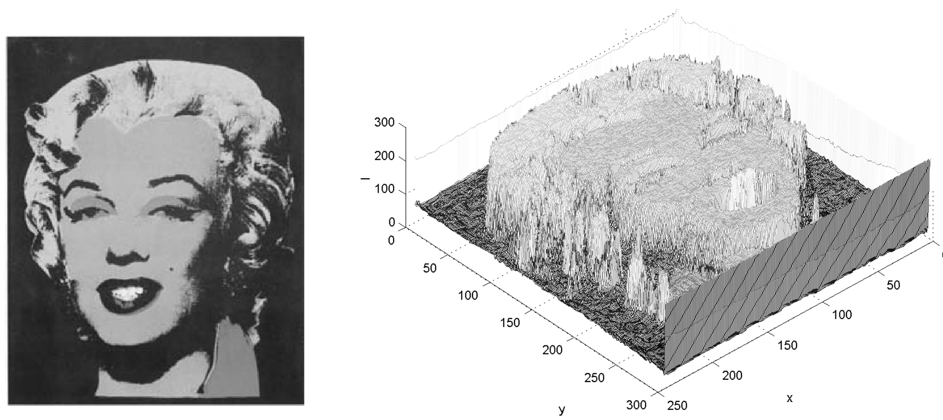

Figure 16: Example of a non-stationary image.

Naturally, the analysis of the previous sections can be extended to cope with non-periodic signals, by letting $T \rightarrow \infty$, which originates the Fourier transform, see, *e.g.*, [2]. However, this approach lacks the nice compactness characteristic we observed for the Fourier series representation of periodic signals [3]. When dealing with acoustic musical signals, instead of attempting to describe general aperiodic signals with a single spectrum, we may take advantage of the fact that these signals are "locally periodic". In fact, although not "globally periodic", many musical signals are composed by shorter segments that exhibit periodicity, see the example in Figure 17. Thus, an useful

[3] Although not emphasized here, Fourier-like representations are also motivated by other characteristics, for example, its effectiveness in linear filtering, see *e.g.*, [1, 5], and its invariance with respect to certain groups of transformations in image analysis, see *e.g.*, [3, 4].

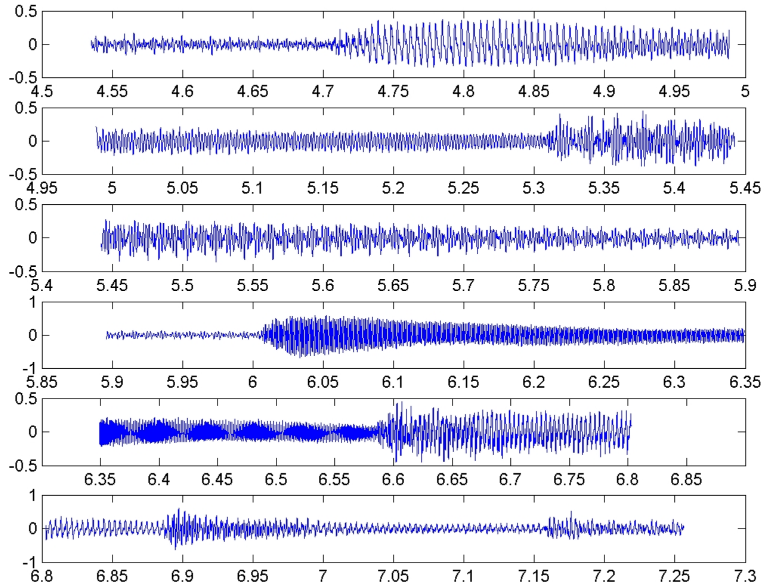

Figure 17: Example of a music segment. Note that, although non-periodic, it is largely "locally periodic".

description for this kind of signals is based on the set of spectral representations of the periodic segments. This description consists on a so-called time-frequency representations, because it describes the spectral content of the signal as a function of time, see *e.g.*, [5, 6].

The oldest time-frequency representation can be traced back to at least the 16th century, see an example in Figure 18. It was developed as a compact way to describe Western music. Time is on the horizontal axis. Each sound event, *i.e.*, each note (or pause), is denoted by a symbol, which is placed in an horizontal position that indicates its start time. The specific symbol used denotes the duration of the sound event (see, *e.g.*, [7], for a detailed description of the symbols). Frequency is on the vertical axis. The vertical position of the symbols indicate the fundamental frequency of the respective note, in a logarithmic scale.

The correspondence between the vertical position of the symbols indicating the notes, the corresponding (white) key in a piano keyboard, and the corresponding (fundamental) frequency is indicated in Figure 19. The piano

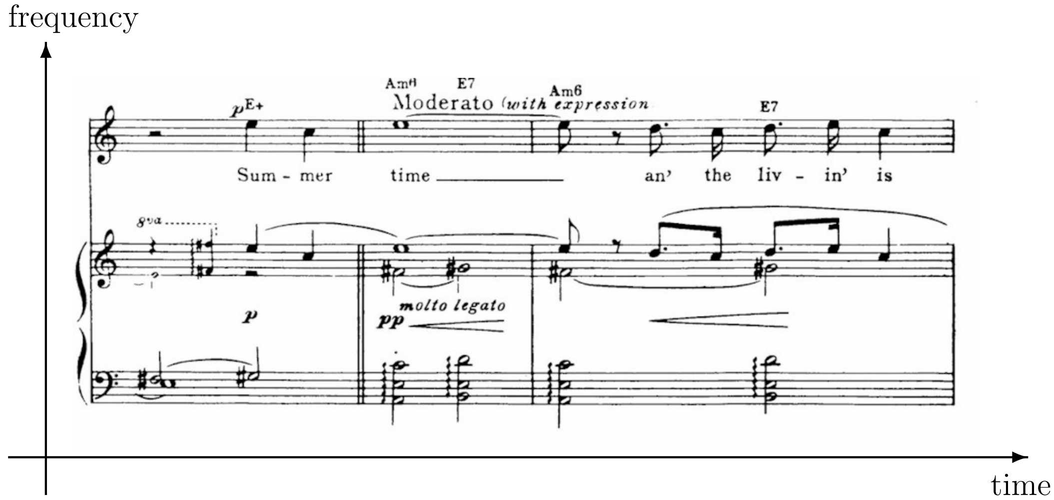

Figure 18: Western music notation: a time-frequency representation.

keyboard is organized in a repetitive pattern, where each segment doubles the frequency (*i.e.*, an octave). For example, since, as referred in the previous sections, the frequency of the A above middle C is 440Hz, the frequency of the A immediately above is 880Hz and the one of the A below is 220Hz (see Figure 19). In between (the majority of) the piano white keys, there are also black keys (whose note is represented in musical notation by using special symbols like $\sharp$ or $\flat$, see, *e.g.*, [7]). Overall, each octave is divided into 12 steps. In an equally tempered scale, each of these steps is equal, meaning that the ratio between the frequencies of consecutive keys is fixed, say r. Since 12 steps double the frequency,

$$r^{12} = 2 \quad \Leftrightarrow \quad r = 2^{1/12} = \sqrt[12]{2} \simeq 1.0595\,. \tag{13}$$

It is thus straightforward to compute the frequency of any note, knowing the frequency of another one. To make things clear, let us number the notes, starting by the lowest frequency key of the piano keyboard (see Figure 19): A-1, A$\sharp$($\Leftrightarrow$B$\flat$)-2, B-3, C-4, C$\sharp$($\Leftrightarrow$D$\flat$)-5, D-6, etc. To compute the frequency f_m of key number m, given the one of key n, f_n, we just have to multiply it (or divide), by the ratio r, the appropriate number os steps, *i.e.*, $m - n$ times:

$$f_m = f_n\, r^{m-n} = f_n\, 2^{(m-n)/12}\,. \tag{14}$$

As an example, let us compute the fundamental frequency of the middle C (key number 40). Since the fundamental frequency of the key 49 (A above

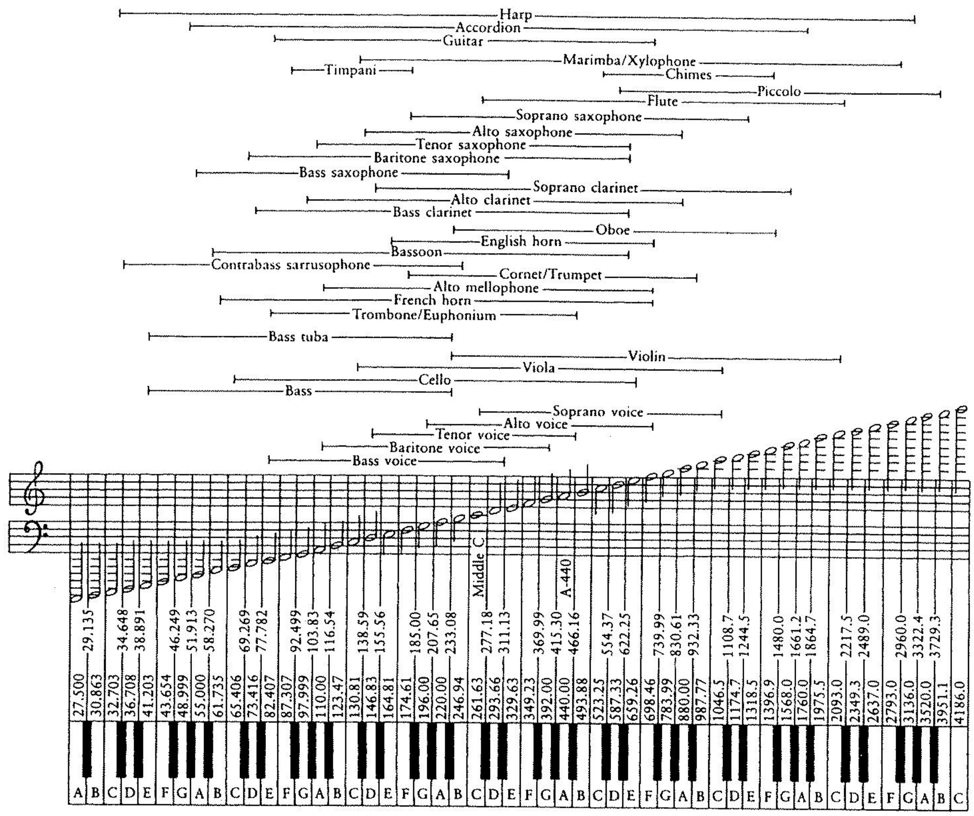

Figure 19: Piano keyboard and correspondent mapping to music sheet. See also the frequency ranges of other musical instruments and human voices. (From [8].)

middle C) is 440Hz, we get

$$f_{40} = f_{49}\, 2^{(-9/12)} \simeq 261.6256\text{Hz} \qquad (15)$$

for the middle C, according with what is indicated in Figure 19.

We end by emphasizing that the automatic inference of an highly structured representation for a music signal, such as the one based in the musical notation of Figure 18, is a tremendously difficult task. This is the ultimate goal of automatic music transcription, which, besides obvious applications in music analysis, is also required for structured audio coding, where very economic representations, such as MIDI [4] files, retain only an extremely compact description of the characteristics of a piece of music, see [6].

References

[1] James H. McClellan, Ronald W. Schaffer, and Mark A. Yoder. *DSP First – A Multimedia Approach*. Prentice Hall, 1999.

[2] Alan V. Oppenheim and Alan S. Willsky. *Signals and Systems*. Prentice Hall, 1996.

[3] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Prentice Hall, 1992.

[4] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2002.

[5] Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck. *Discrete-Time Signal Processing*. Prentice Hall, 1999.

[6] Anssi Klapuri and Manuel Davy, editors. *Signal Processing Methods for Music Transcription*. Springer, 2006.

[7] George T. Jones. *Music Theory*. Barnes & Noble, 1974.

[8] Jonh D. Barrow *The Artfull Universe Expanded*. Oxford University Press, 2005.

[4] Musical Instrument Digital Interface, a standard for exchanging data between electronic musical devices.