Toward a Probabilistic Fourier Analysis on Audio Signals

Hugo Tremonte de Carvalho
Federal University of Rio de Janeiro (UFRJ)
Institute of Mathematics (IM)
Department of Statistical Methods (DME)
hugo@dme.ufrj.br
Orcid: 0000-0003-0776-0400

DOI: 10.46926/musmat.2021v5n2.91-102

Abstract: *In audio signal processing there are several tools employed to perform time-frequency analysis, being the spectrogram one of the most widely used. In a nutshell, it can be understood as a visual representation of the frequency content of an audio signal as it varies with time. Here we propose a probabilistic alternative to the spectrogram, which can be roughly interpreted as the most likely frequency to be present within an audio signal, also along time. This will be achieved by computing a specific posterior distribution in a Bayesian context. Preliminary experiments indicating the suitability of this object are presented, and potential applications to audio signal processing are outlined.*

Keywords*: Audio Signal Processing. Statistical Signal Processing. Bayesian Inference. Fourier Analysis. Spectrogram.*

I. Introduction

The main goal of this short note is to provide an answer to the following question: "how likely is some particular frequency to be present in some audio signal, regardless of its amplitude and phase, and the variance of any existing superimposed noise?". In fact, the answer here is not new, being firstly proposed in [2] for more general discrete-time signals. The novelty in this paper is the application of this general framework to audio signals, allowing then for probabilistic counterparts of well-known objects in audio signal processing, such as the spectrogram and the chromagram, and opening new possibilities to time-frequency related tasks, such as fundamental frequency estimation [6]. The paper is organized as follows: Section II introduces some basic notation to be used throughout the exposition, recalls the theory proposed in [2, Chap. 2] with more explained details in the analytical computations, and also extends the discussion on interpretations and suitability of the obtained quantities; in Section III, based on the previous discussions, we propose the probabilistic spectrogram and qualitatevely compares it with its classical counterpart; and finally, in Section IV we draw some conclusions and point out future works in this direction.

Received: November 29th, 2021
Approved: December 21st, 2021

II. Bayesian spectrum estimation

Recall the formerly presented question: "how likely is some particular frequency to be present in the signal $x(t)$, regardless of its amplitude and phase, and the variance of any existing superimposed noise?". In this Section, we follow the discussion proposed on [2, Chap. 2], with more detail on some analytical computations and interpretations. It is important to remark that the exposed answer is not at its greatest generality, which can be found on [2, Chap. 3]; but we choose to maintain a simpler discussion for the sake of clarity. Moreover, preliminary computational experiments demonstrated only a negligible improvement by applying the more general framework.

i. The model

Let $x[n]$, for $n = 1, \ldots, N$, be a digital audio signal, derived from an analogical audio signal, denoted by $x(t)$, via uniform sampling with known frequency f_s, usually 44,100 Hz for an audio signal with the typical CD quality. More specifically, $x[n]$ is the n-th time sample of $x(t)$ and is given by $x(n/f_s)$. For answering the question in the beginning of Section II, we will assume the following model for $x(t)$:

$$x(t) = f(t) + \varepsilon(t), \tag{1}$$

where $f(t) = B_1 \cos(2\pi\omega t) + B_2 \sin(2\pi\omega t)$, $\varepsilon(t) \sim \mathcal{N}(0, \sigma^2)$ are observations of independent white noise, and $\omega > 0$ is measured in Hz.

Notice that $f(t)$, the systematic part of the signal $x(t)$, is equivalent to the more intuitive function $B\cos(2\pi(\omega t + \varphi))$, for adequate choices of B and φ. However, we will keep the former choice, since the latter will imply in more complicated computations. Still, two points are still unclear: 1) why only to use a single frequency to model the signal, since a musical signal will be much more complex than a simple sinusoid? and 2) why is the hypothesis of Gaussian white noise reasonable?

Essentially, we are looking for a compromise between simplicity and accuracy of our modeling. If we try to superimpose more and more frequencies in function $f(t)$, we will be approaching a perfect representation of the musical signal $x(t)$ by means of the Fourier Transform; and if we try to perfectly model how the observed signal $x(t)$ deviates from a single frequency and then incorporate it in the noise $\varepsilon(t)$, it will become extremely complex and thus intractable. The gaussianity of $\varepsilon(t)$ still could be questioned, but we justify it intuitively via the Maximum Entropy Principle [5]: the additive noise comprises several distinct deviations of the observed signal $x(t)$ from the model $f(t)$, therefore, the "most non-informative" probability distribution with finite mean and variance and supported in $\mathbb{R}$ is the Gaussian distribution. Moreover, it is still possible to argue that our goal is not to achieve a perfect substitute for the observed signal $x(t)$ by means of the model $f(t)$, but to answer probabilistic questions about the presence of a particular frequency on the observed signal $x(t)$.

ii. Formulating the problem

Given the remarks at the end of the last section, we may proceed with a few steps of the inference procedure. After the time sampling procedure, we have a sequence of observations given by

$$x[n] = f[n] + \varepsilon[n], \text{ for } 1 \leq n \leq N, \tag{2}$$

where $f[n] = f(n/f_s)$, and the $\varepsilon[n] \sim \mathcal{N}(0, \sigma^2)$ are independent. This hypothesis of white Gaussian noise implies the following distribution for the random vector $\boldsymbol{\varepsilon} = [\varepsilon[1], \ldots, \varepsilon[N]]^T$:

$$p(\boldsymbol{\varepsilon}|\sigma) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\varepsilon[n]^2}{2\sigma^2}\right\} \propto \sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N} \varepsilon[n]^2\right\}. \tag{3}$$

Note that $\varepsilon[n] = x[n] - f[n]$, for $n = 1, \ldots, N$, is a change of variables from vector $\boldsymbol{\varepsilon}$ to $\mathbf{x} = [x[1], \ldots, x[n]]^T$ with unitary Jacobian, that leads to the following distribution for $\mathbf{x}$:

$$\begin{aligned} p(\mathbf{x}|B_1, B_2, \omega, \sigma) &\propto \sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x[n] - f[n])^2\right\} \\ &= \sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\{x[n] - [B_1\cos(2\pi\omega n/f_s) + B_2\sin(2\pi\omega n/f_s)]\}^2\right\}. \end{aligned} \tag{4}$$

When considered as a function of $\{B_1, B_2, \omega, \sigma\}$, the expression in Equation 4 is called the *likelihood* and it quantifies the probability of observing the signal $\mathbf{x}$ if the parameters are given by some particular value of $\{B_1, B_2, \omega, \sigma\}$. Maximizing this function in variables $\{B_1, B_2, \omega, \sigma\}$ leads to the *maximum likelihood estimator* (MLE), interpreted as as the values of $\{B_1, B_2, \omega, \sigma\}$ for which the observed data $\mathbf{x}$ is the most probable.

But notice that this estimate has a severe problem: it does not answer our question! In order to convince ourselves, let us compare the question with the interpretation given above:

- **QUESTION**: "How likely is some particular frequency (ω) to be present in the signal $x(t)$, regardless of its amplitude and phase (both encoded in B_1 and B_2), and the variance of any existing superimposed noise (σ^2)?"
- **INTERPRETATION OF THE MLE**: "The values of B_1, B_2 (encoding amplitude and phase), ω (frequency), σ (standard deviation – square root of the variance – of the superimposed noise) for which the observed data $\mathbf{x}$ is the most probable."

Notice that in the interpretation of the MLE the probability is with respect to $\mathbf{x}$, where in the question it should be considered over the frequency ω. Moreover, the question is only about the frequency disregarding all the other parameters, but the interpretation considers all of them in the optimization process. The key to address these issues is Bayesian inference, which will be briefly discussed in the next subsection, in a more general context.

iii. Interlude – A glimpse of Bayesian inference

Essentially, Bayesian inference is a method of statistical inference where the information contained in observed data can be used to update the knowledge about parameters of interest. More precisely, let Z be a random variable which probability function of probability density function is denoted by $p(z|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathbb{R}^k$ is a vector of parameters. We observe $\mathbf{z} = [z_1, \ldots, z_N]^T$ independent samples from Z and we want to estimate vector $\boldsymbol{\theta}$ from this data. The likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}) = p(\mathbf{z}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(z_n|\boldsymbol{\theta}), \tag{5}$$

and quantify the probability of observing data $\mathbf{z}$ if the parameter vector is $\boldsymbol{\theta}$. Maximizing this function with respect to $\boldsymbol{\theta}$ will lead to its *maximum likelihood estimation*, which can be interpreted as the value of $\boldsymbol{\theta}$ for which the observed data is the most probable.

If we have prior information about $\boldsymbol{\theta}$, encoded in the so-called *prior distribution* $p(\boldsymbol{\theta})$, we can use Bayes theorem and obtain the *posterior distribution* for the parameters:

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{\mathcal{L}(\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{z})} \propto \mathcal{L}(\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{6}$$

since $p(\mathbf{z})$ is a normalization term which in our application can be ignored. Therefore, maximizing the posterior distribution will lead to the *maximum a posteriori* (MAP) estimate for $\boldsymbol{\theta}$, representing the most probable parameters for the corresponding set of observation, a much easily interpretable quantity.

For more details on statistical and Bayesian inference, we refer the reader to [3, 4], respectively.

iv. The (Bayesian) inference procedure

By employing the Bayes' theorem, we are able to invert the conditional probability $p(\mathbf{x}|B_1, B_2, \omega, \sigma)$ and obtain

$$p(B_1, B_2, \omega, \sigma|\mathbf{x}) \propto p(\mathbf{x}|B_1, B_2, \omega, \sigma)p(B_1, B_2, \omega, \sigma). \tag{7}$$

The new term $p(B_1, B_2, \omega, \sigma)$ can encode previous knowledge about these variables, but we will adopt a conservative approach and use *non-informative priors*, that is, prior distributions that describe the most vague knowledge about them. For the parameters that can assume any real value (in this case B_1, B_2, ω)[1], we consider an improper prior distribution $p(B_1, B_2, \omega) \propto 1$, and for the scale parameter σ, which must be strictly positive, we impose the Jeffreys prior, given by $p(\sigma) \propto 1/\sigma$. By assuming prior independence between B_1, B_2, ω, σ, we have that

$$\begin{aligned} p(B_1, B_2, \omega, \sigma|\mathbf{x}) &\propto p(\mathbf{x}|B_1, B_2, \omega, \sigma)p(B_1, B_2, \omega, \sigma) \\ &\propto \sigma^{-(N+1)} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\{x[n] - [B_1\cos(2\pi\omega n/f_s) + B_2\sin(2\pi\omega n/f_s)]\}^2\right\}. \end{aligned} \tag{8}$$

Firstly, let us expand the summation of squares inside the exponential to obtain:

$$\begin{aligned} &\sum_{n=1}^{N}\{x[n] - [B_1\cos(2\pi\omega n/f_s) + B_2\sin(2\pi\omega n/f_s)]\}^2 = \\ &\quad = \sum_{n=1}^{N}x[n]^2 - 2B_1\sum_{n=1}^{N}x[n]\cos(2\pi\omega n/f_s) - 2B_2\sum_{n=1}^{N}x[n]\sin(2\pi\omega n/f_s) \\ &\quad + B_1^2\sum_{n=1}^{N}\cos^2(2\pi\omega n/f_s) + B_2^2\sum_{n=1}^{N}\sin^2(2\pi\omega n/f_s) \\ &\quad + 2B_1B_2\sum_{n=1}^{N}\cos(2\pi\omega n/f_s)\sin(2\pi\omega n/f_s). \end{aligned} \tag{9}$$

Equation 9 can be simplified (exactly, without approximations) by using trigonometric identities, but this leads to tedious computations to prove such identities, or to evocations of some aspects from Linear Algebra ([2, Chap. 3] presents a discussion on this direction). We will not follow this path here and opt to employ some approximations, since we want to keep this discussion simple and easy to follow. Moreover, preliminary computational tests indicate that the gain in the

[1] In the beginning of Section II we constrained ω to be strictly positive for obvious physical reasons, but regarding the periodicity of trigonometric functions negative values of ω will also be allowed, from the mathematical viewpoint, to simplify the computations.

accuracy of the frequency estimation with the exact procedure is mostly negligible. In the light of these aspects we choose to approximate some of the terms in Equation 9 as follows:

- The term $\sum_{n=1}^{N} x[n]^2$ is constant (as a function of $\{B_1, B_2, \omega, \sigma\}$), and it will be denoted as $N\overline{x^2}$, where $\overline{x^2} = \frac{1}{N}\sum_{n=1}^{N} x[n]^2$.
- Both terms $\sum_{n=1}^{N} x[n]\cos(2\pi\omega n/f_s)$ and $\sum_{n=1}^{N} x[n]\sin(2\pi\omega n/f_s)$ are related to the real and imaginary parts of the Discrete Fourier Transform, and will be denoted, respectively, by $R(\omega)$ and $I(\omega)$.
- Both expressions $\sum_{n=1}^{N} \cos^2(2\pi\omega n/f_s)$ and $\sum_{n=1}^{N} \sin^2(2\pi\omega n/f_s)$ can be rewritten by using the simple trigonometric identities[2] $\sin^2(\alpha) = \frac{1-\cos(2\alpha)}{2}$ and $\cos^2(\alpha) = \frac{1+\cos(2\alpha)}{2}$ to obtain:

$$\sum_{n=1}^{N} \cos^2(2\pi\omega n/f_s) = \frac{N}{2} + \frac{1}{2}\sum_{n=1}^{N} \cos(4\pi\omega n/f_s), \tag{10}$$

$$\sum_{n=1}^{N} \sin^2(2\pi\omega n/f_s) = \frac{N}{2} - \frac{1}{2}\sum_{n=1}^{N} \cos(4\pi\omega n/f_s). \tag{11}$$

  Analogously, the term $\sum_{n=1}^{N} \cos(2\pi\omega n/f_s)\sin(2\pi\omega n/f_s)$ can be rewritten, using the formula for the sine of a sum of angles, to obtain:

$$\sum_{n=1}^{N} \cos(2\pi\omega n/f_s)\sin(2\pi\omega n/f_s) = \frac{1}{2}\sum_{n=1}^{N} \sin(4\pi\omega n/f_s). \tag{12}$$

  In [2, p. 17] it is claimed that these three trigonometric sums can be neglected, when compared to $N/2$, the dominant term appearing in the last two lines of Equation 9, assuming that the rather vague conditions $N \gg 1$ and $\omega N/f_s \ll 1$ are valid. Indeed, computational tests verified that for the human hearing frequency range (about 20 Hz to 20,000 Hz), the usual sampling rate of a CD-quality audio signal ($f_s = 44{,}100$ Hz) and the window length here employed ($N = 4{,}096$ – see Sec. III for more details), these neglected terms are two to three orders of magnitude smaller than $N/2$, validating the claims above in our context.

Given the discussion above, we can rewrite Equation 9 as[3]:

$$\begin{aligned}\sum_{n=1}^{N} \{x[n] - [B_1\cos(2\pi\omega n/f_s) + B_2\sin(2\pi\omega n/f_s)]\}^2 = \\ &= N\overline{x^2} - 2B_1R(\omega) - 2B_2I(\omega) + B_1^2\frac{N}{2} + B_2^2\frac{N}{2} \\ &= N\left[\overline{x^2} - \frac{2}{N}[B_1R(\omega) + B_2I(\omega)] + \frac{1}{2}(B_1^2 + B_2^2)\right].\end{aligned} \tag{13}$$

[2] Both these identities follow from the formulas for the sine and cosine of a sum of angles and from the fundamental trigonometric relation $\sin^2(\alpha) + \cos^2(\alpha) = 1$.

[3] The first symbol "=" is not accurate in Equation 13, being the "≈" preferable. However, we will go ahead with the slight abuse of notation, in order to simplify the notation.

By substituting Equation 13 in Equation 8, we have that:

$$\begin{aligned} p(B_1, B_2, \omega, \sigma|\mathbf{x}) &\propto p(\mathbf{x}|B_1, B_2, \omega, \sigma) p(B_1, B_2, \omega, \sigma) \\ &\propto \sigma^{-(N+1)} \exp\left\{ -\frac{N}{2\sigma^2} \left[\overline{x^2} - \frac{2}{N}[B_1 R(\omega) + B_2 I(\omega)] + \frac{1}{2}(B_1^2 + B_2^2) \right] \right\}. \end{aligned} \tag{14}$$

v. Marginalization of nuisance parameters

Now, with the simplified expression for the posterior distribution, given in Equation 14, we are able to marginalize the undesired parameters and obtain only $p(\omega|\mathbf{x})$, given by

$$p(\omega|\mathbf{x}) \propto \int_0^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(B_1, B_2, \omega, \sigma|\mathbf{x}) \; dB_1 dB_2 d\sigma. \tag{15}$$

To perform the integrals in B_1 and B_2 (and then obtaining $p(\omega, \sigma|\mathbf{x})$) we must resort to the following Lemma:

Lemma 1. *Let $a \in \mathbb{R}$, $\mathbf{b} \in \mathbb{R}^m$, and $\mathbf{C} \in \mathbb{R}^{m \times m}$ be numerical, vectorial, and matricial constants, respectively. Assume also that matrix $\mathbf{C}$ is invertible. Then, the following equality holds:*

$$\int_{\mathbb{R}^m} \exp\left\{ -\frac{1}{2}(a + \mathbf{b}^T\mathbf{y} + \mathbf{y}^T\mathbf{C}\mathbf{y}) \right\} \, d\mathbf{y} = \frac{(2\pi)^{m/2}}{|\det(\mathbf{C})|^{1/2}} \exp\left\{ -\frac{1}{2} \left[a - \frac{\mathbf{b}^T\mathbf{C}^{-1}\mathbf{b}}{4} \right] \right\}. \tag{16}$$

The proof of this lemma, although simple, is quite tedious, since it involves completing the squares in the argument of the left exponential, identifying the kernel of a multivariate normal distribution, and rearranging the terms to obtain the desired result. Here we simply employ it and refer the interested reader to [1] for the proof.

As can be verified by a simple computation, by defining $a = \frac{N\overline{x^2}}{\sigma^2}$, $\mathbf{b} = -\frac{2}{\sigma^2}\begin{bmatrix} R(\omega) \\ I(\omega) \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} N/2\sigma^2 & 0 \\ 0 & N/2\sigma^2 \end{bmatrix}$, and substituting these values in the left-hand side of Equation 16 we recover exactly Equation 14. Therefore, by the direct application of Lemma 1, we have that

$$p(\omega, \sigma|\mathbf{x}) \propto \sigma^{-(N-1)} \exp\left\{ -\frac{N}{2\sigma^2} \left[\overline{x^2} - \frac{2}{N} C(\omega) \right] \right\}, \tag{17}$$

where $C(\omega)$ is defined as $[R(\omega)^2 + I(\omega)^2]/N$. This quantity is called the *periodogram* in [2, p. 7], and is a natural appearance in this probabilistic context of the squared-magnitude of the Discrete Fourier Transform.

Finally, in order to integrate Equation 17 with respect to σ, we can make a change of variables and use the kernel of an Inverse-Gamma distribution, a more classical approach – perhaps the one employed by Bretthorst in the computations omitted in [2, pp. 18–20]. To avoid extensive computations, we refer to the more recent work [7], where the *Inverse-Nakagami distribution* is proposed, which probability density function, in the variable $\sigma > 0$, is given by

$$\frac{2}{\Gamma(\lambda)} \left(\frac{\lambda}{\xi} \right)^{\lambda} \sigma^{-2\lambda-1} \exp\left(-\frac{\lambda}{\xi\sigma^2} \right), \tag{18}$$

being λ and ξ strictly positive real parameters. By using the fact that every probability density function must integrate 1 over its support, we can use the normalizing constant of this distribution to obtain the result

$$\int_0^{+\infty} \sigma^{-2\lambda-1} \exp\left(-\frac{\lambda}{\xi\sigma^2} \right) \, d\sigma = \frac{\Gamma(\lambda)}{2} \left(\frac{\xi}{\lambda} \right)^{\lambda}. \tag{19}$$

By comparing the term being integrated in Equation 19 with the joint density of $\{\omega, \sigma\}$ in Equation 17, we obtain that the suitable values of λ and ξ are given by

$$\lambda = \frac{N-2}{2}, \tag{20}$$

$$\xi = \frac{N-2}{2}\left[\overline{x^2} - \frac{2}{N}C(\omega)\right]^{-1}. \tag{21}$$

Therefore, by using the result of Equation 19 and discarding the multiplicative terms that does not depend on ω, we have that that

$$p(\omega|\mathbf{x}) \propto \left[1 - \frac{2C(\omega)}{N\overline{x^2}}\right]^{\frac{2-N}{2}}, \tag{22}$$

an expression for the distribution we are looking for.

Preliminary computational experiments indicated that trying to compute directly this quantity may cause a numerical overflow. Therefore, we opt to compute its logarithm[4], given by

$$\log_{10} p(\omega|\mathbf{x}) = \frac{2-N}{2}\log_{10}\left[1 - \frac{2C(\omega)}{N\overline{x^2}}\right] + \text{constant terms}, \tag{23}$$

where the constant terms are the logarithm of the multiplicative constants unconsidered along the computations. Recall that considering a strictly positive function or its logarithm is equivalent for finding its local maxima, because the logarithm is a strictly crescent function.

We now present an application of this distribution to audio signal processing.

III. The probabilistic spectrogram

The *spectrogram* is a fundamental tool in time-frequency analysis of audio signals, and it can be interpreted as the frequency content of an audio signal along time. In order to understand it more precisely, recall that [8, Sec. 2.5] defines the Short-Time Fourier Transform as

$$\mathcal{Y}[m,k] = \sum_{n=0}^{N-1} y[n + mH]w[n]e^{-2\pi ikn/N}, \tag{24}$$

for $m \in \mathbb{Z}$[5] and $k = 0, \ldots, \lfloor N/2 \rfloor$. The quantity H is the hop size, and it is related to the overlap between two consecutive windows; a common value to adopt is $H = \lfloor N/2 \rfloor$, indicating 50% of overlap.

Essentially, Equation 24 computes the Discrete Fourier Transform of the signal

$$x_m[n] = y[n + mH]w[n], \text{ for } n = 0, \ldots, N-1, \tag{25}$$

that is, an excerpt of length N from an audio signal y, properly smoothed on the edges by some window function w, with support of size N. The spectrogram is then defined as the squared-magnitude of each value of the Short-Time Fourier Transform, that is, $|\mathcal{Y}[m,k]|^2$. For more details on the spectrogram and on the choice of windowing functions, see [8, Sec. 2.5].

[4] The base of the logarithm is of minor importance here, but we will adopt 10, since it is the standard base used to transform magnitude to dB, an usual unit in Signal Processing.

[5] Note that m is in practice restricted to a finite set $\{0, \ldots, M\}$ such that these successive hops of size H cover the entire signal y. Since this specific range will not be directly used here, we avoided its explicit introduction, for the sake of clarity.

Since the probability density function on Equation 22 is related to the presence of frequencies ω in signal $\mathbf{x}$, instead of computing the Discrete Fourier Transform of the windowed signal $\mathbf{x}_m$, we can compute the probability distribution $p(\omega|\mathbf{x}_m)$, for each $m \in \mathbb{Z}$. The collection of all these probability distributions for $m \in \mathbb{Z}$ is called the *probabilistic spectrogram* of the signal y. This representation is expected to be more sparse than the spectrogram, since it computes the most likely frequencies at each time-frame and may disregard high-order harmonics of the instrument. Moreover, for each $m \in \mathbb{Z}$, we can compute the maximum a posteriori estimate for ω, that is, argmax $p(\omega|\mathbf{x}_m)$. This sequence of values may provide an interesting low-dimensional feature useful in Music Information Retrieval[6].

In order to illustrate this concept, we consider the subject of *Ricercar a 6*, the six-voice fugue that is commonly regarded as the high point of *The Musical Offering* (BWV 1079), by Johann Sebastian Bach, interpreted in two instruments: the fortepiano and the harpsichord[7]. Both audio signals were downloaded from YouTube in MP3 format with 320 Kbps, manually trimmed to contain only the subject, exported in WAV format and reduced to a monophonic signal. The sampling frequency is 44,100 Hz, the length of the analysis window was taken to be $N = 4{,}096$ time-samples with hop size of $H = 2{,}048$, implying in 50% of overlap between adjacent windows.

Figures 1 and 2 display the spectrogram of the subject of *Ricercar a 6* played on the fortepiano and the harpsichord, respectively. Note that the harpsichord signal contains much more harmonics than the fortepiano one, as it is naturally expected for both instruments.

In Figures 3 and 4 are displayed the logarithm of the probabilistic spectrogram of the subject of *Ricercar a 6*, played on the pianoforte and harpsichord, respectively. Blue color indicates negligible probability, and colors close to yellow indicate more likely ones. The y axis is restricted from 0 Hz to 2,500 Hz, since there are no likely frequencies above this value. Notice that these representations also capture higher harmonics of the harpsichord, but are cleaner than the spectrogram.

Indeed, Figures 5 and 6 display both the probabilistic spectrograms as in Figures 3 and 4, but the red dots indicate the most likely frequencies at each time frame. Notice that, although visually distinct, they look more similar than the spectrograms in Figures 1 and 2. This fact indicates that these sequences of frequencies can be regarded as an interesting audio feature, and a refinement of this quantity may lead to advances in algorithms of fundamental frequency detection.

IV. Conclusion and future works

In this paper we recalled and detailed the computations done in [2, Cap. 2], in order to study the posterior distribution of frequencies present within an audio signal, disregarding its amplitude and phase, and also the variance of any existing superimposed noise. This probability distribution was used to propose a new concept in time-frequency analysis of audio signals, the *probabilistic spectrogram*, illustrated by two musical examples.

Althoguh not quantitatively analyzed, the proposed quantity seems promising to be further investigated, specially its usage in Music Information Retrieval tasks, mainly because of its sparseness and high interpretability. This is addressed as a future work, together with the development of the *probabilistic chroma features*, corresponding to the probability allocated in the frequencies bands related to the twelve pitch classes[8].

[6] In [8, Sec. 7.1.2] it is described that peaks of the spectrogram provides an useful fingerprint for the taks of content-based audio retrieval.

[7] Both interpretations are available on YouTube. The fortepiano version is played by Leo van Doeselaar (`https://www.youtube.com/watch?v=fjxKy3pP41w`) and the harpsichord one is interpreted by Iain Simcock (`https://www.youtube.com/watch?v=AYw2E5F930M`).

[8] Assuming an equal emperament tuning with fundamental frequency of A_4 defined as 440 Hz.

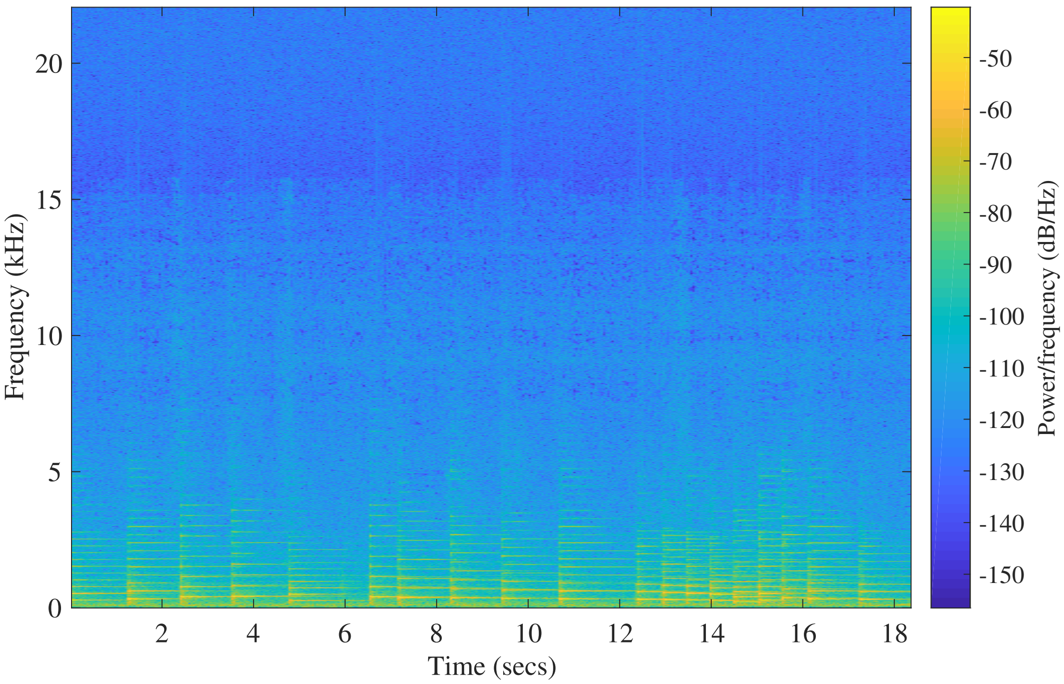

Figure 1: *Spectrogram of the subject of* Ricercar a 6, *played on the fortepiano.*

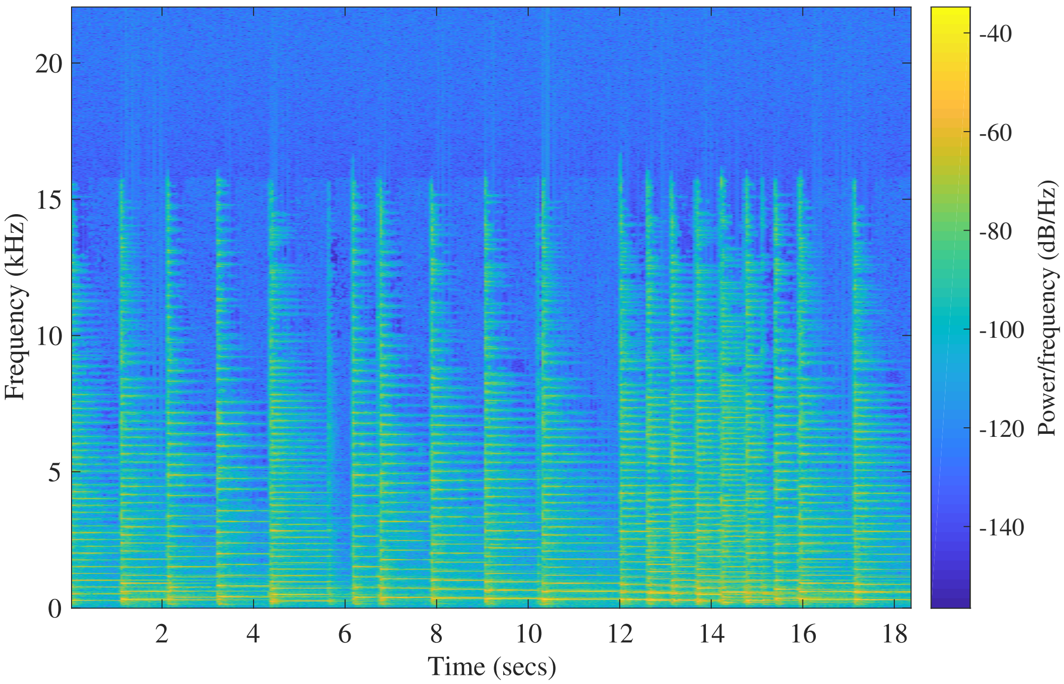

Figure 2: *Spectrogram of the subject of* Ricercar a 6, *played on the harpsichord.*

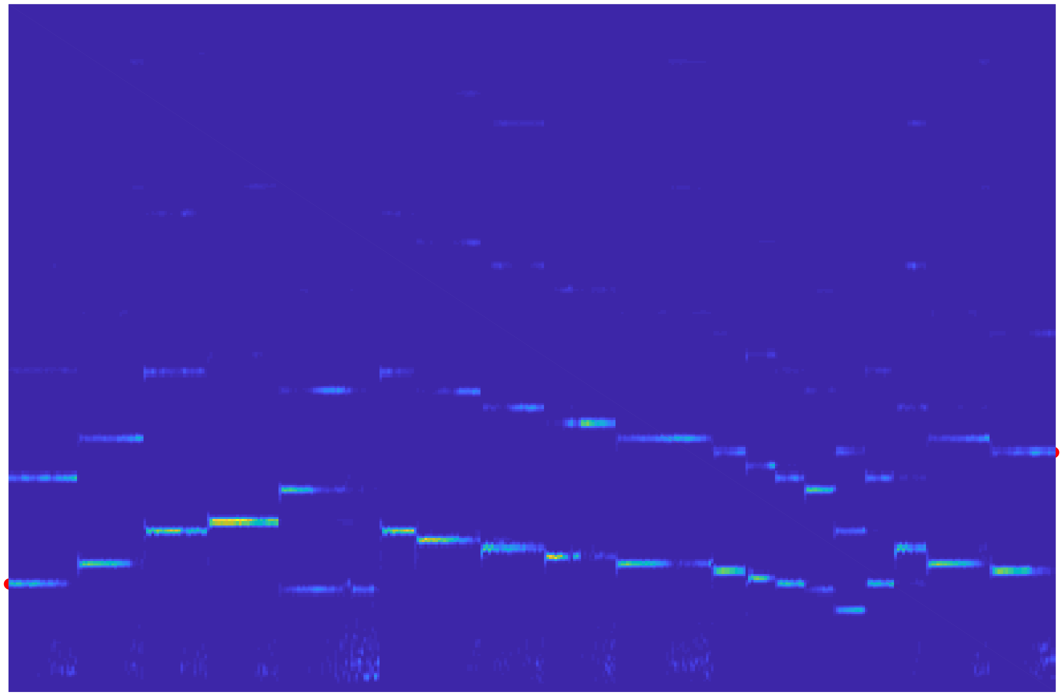

Figure 3: *Probabilistic spectrogram (logarithm; y axis restricted from 0 Hz to 2,500 Hz) of the subject of* Ricercar a 6, *played on the fortepiano.*

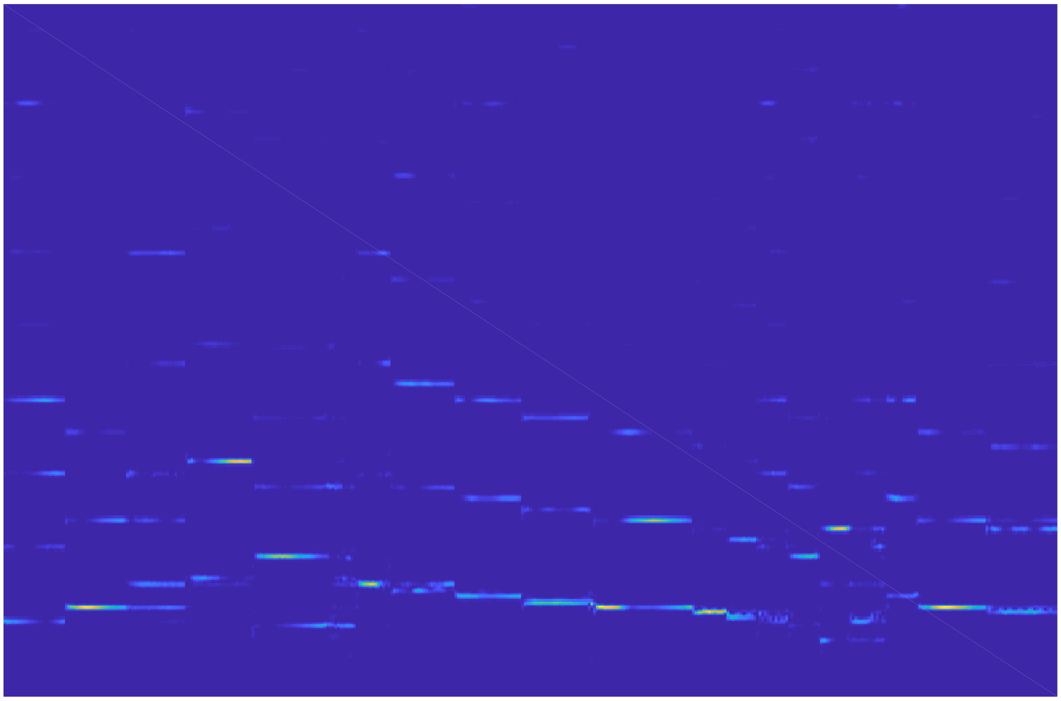

Figure 4: *Probabilistic spectrogram (logarithm; y axis restricted from 0 Hz to 2,500 Hz) of the subject of* Ricercar a 6, *played on the harpsichord.*

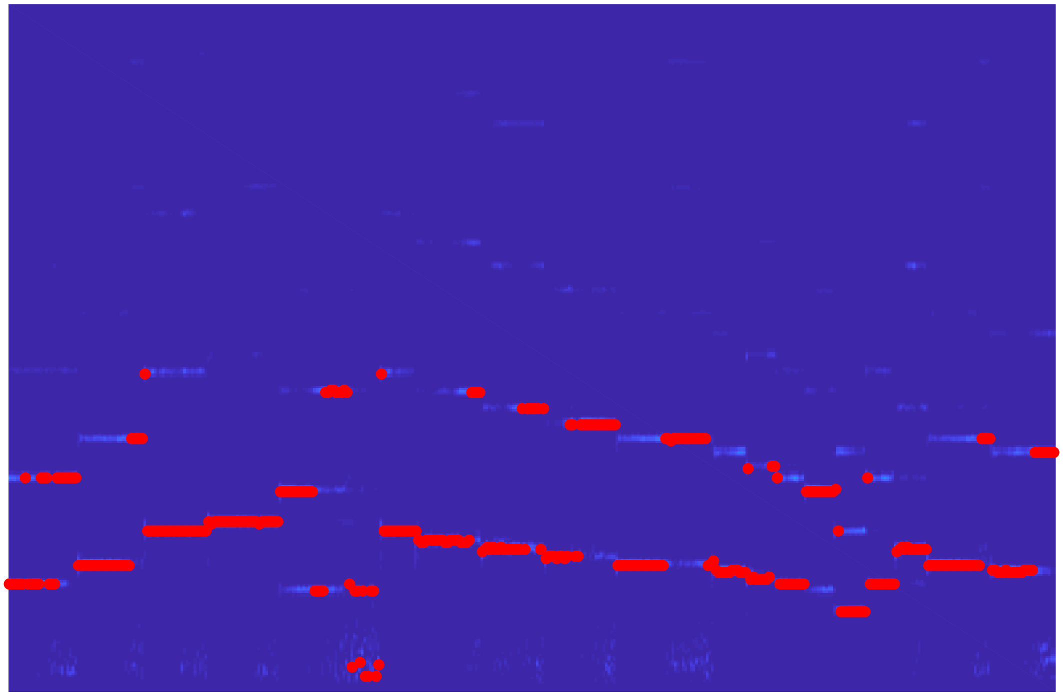

Figure 5: *Probabilistic spectrogram (logarithm; y axis restricted from 0 Hz to 2,500 Hz) of the subject of* Ricercar a 6, *played on the fortepiano. The red dots indicate most likely frequency at each time-frame.*

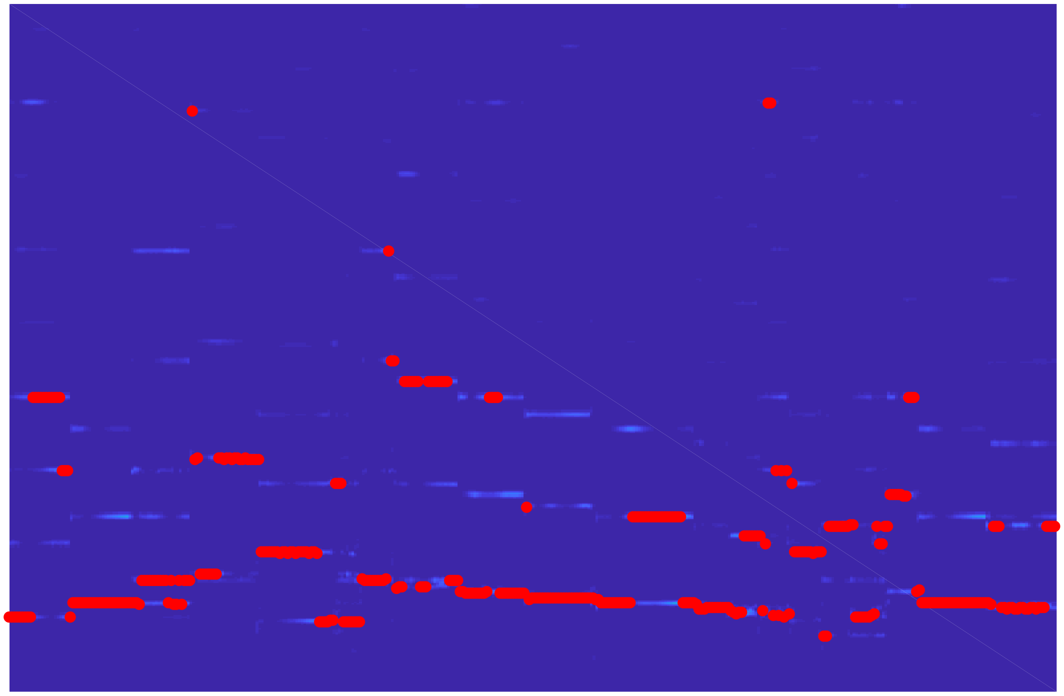

Figure 6: *Probabilistic spectrogram (logarithm; y axis restricted from 0 Hz to 2,500 Hz) of the subject of* Ricercar a 6, *played on the harpsichord. The red dots indicate most likely frequency at each time-frame.*

References

[1] Bishop, C. (2006) *Pattern Recognition and Machine Learning*. New York: Springer.

[2] Bretthorst, G. (1988) *Bayesian Spectrum Analysis and Parameter Estimation*. New York: Springer.

[3] Casella, G.; Berger, R. (2001) *Statistical Inference*. Boston: Cengage Learning.

[4] Gelman, A; Carlin, J.; Stern, H.; Dunson, D.; Vehtari, A.; Rubin, D. (2013) *Bayesian Data Analysis*. Boca Raton: CRC Press.

[5] Jaynes, E. (1982) On The Rationale of Maximum-Entropy Methods. *Proceedings of the IEEE*, 70/9, pp. 939–952.

[6] Klapuri, A. (ed.); Davy, M. (ed.) (2006) *Signal Processing Methods for Music Transcription*. New York: Springer.

[7] Louzada, F.; Ramos, P.; Nascimento, D. (2018) The Inverse Nakagami-m Distribution: A Novel Approach in Reliability. *IEEE Transactions on Reliability*, 67/3, pp. 1030–1042.

[8] Müller, M. (2015) *Fundamentals of Music Processing*: Audio, Analysis, Algorithms, Applications. New York: Springer.