Quantifying the rationality of rhythmic signals

Alexandre Guillet, Alain Arneodo, Pierre Argoul, Françoise Argoul

"Theoretical physics, wavelets, analysis, genomics
An indisciplinary tribute to Alex Grossmann"

edited by P. Flandrin, S. Jaffard, T. Paul and B. Torresani to appear in the collection
"Applied and Numerical Harmonic Analysis", Springer.

Quantifying the rationality of rhythmic signals

A. Guillet, A. Arneodo, P. Argoul, and F. Argoul

Abstract Rhythms and vibrations represent the quintessence of life, they are ubiquitous (systemic) in all living systems. Recognising, unfolding these rhythms is paramount in medicine, for example in the physiology of the heart, lung, hearing, speech, brain, the cellular and molecular processes involved in biological clocks. The importance of the commensurability of the frequencies in different rhythms has been thoroughly studied in music. We define a log-frequency correlation measure on spectral densities that gives the distribution of frequency ratios (rational or irrational) in between two signals, and this measure is generalized to a time-log-frequency correlation measure using analytic wavelets. We illustrate these concepts on numerical signals (sums of sine functions) and voice recordings from the Voice-Icar-Federico II database. Finally, with a second correlation operation from two of these log-frequency correlation measures we introduce another measure that we call *sonance*, from which we can estimate the pitch transposition that would produce "harmony" (ratios of their harmonics being rational numbers) of two voices sung together.

1 Introduction

Scientific approaches of natural systems have been revolutionized in the last part of the XXth century with the advent of miniaturized electronic and computer systems. Beyond their impressive beauty, it was offered to human beings to demonstrate that nature is constructed from multi-scale intertwinned networks, (in time and in space) and that these networks are the field of highly complex nonlinear dynamics

A. Guillet, A. Arneodo[†], F. Argoul
CNRS, UMR5787, Laboratoire Ondes et Matière d'Aquitaine, Université de Bordeaux, France
e-mail: name@email.address

P. Argoul
MAST-EMGCU, Univ Gustave Eiffel, IFSTTAR, F-77477 Marne-la-Vallée, France e-mail: pierre.argoul@email.address

(non linear and/or non stationary rhythms) [1, 2, 3]. Even though apparently distinct biological rhythms (endogenous and exogenous) have been recognised as universal features of all organisms (neural signals, heart, hormone secretion, metabolism, tidal, circadian, lunar, seasonal, annual clocks, life cycle, ...) [4], the variability of these rhythms and their spatio-temporal interplay is still considered as incidental or ignored. Despite the fact that we can concretely demonstrate that the frequencies of these rhythms pave more than 10 decades, still, time (and frequency) is considered as varying linearly in living systems. In particular the presence of strong nonlinearities can give us greater sensing resolution to less intense stimuli. These mechanisms are ubiquitous across animal species and across all sensory modalities. Interestingly, the mappings between an external stimuli and the internal perception (psychophysical) of scales and laws are rather logarithmic than linear. A simple and more commonly encountered example for the non-specialist is the perception and emission of acoustic vibrations (sounds) by living species, these processes occur in logarithmic scales in time and frequency domains [5]. It has also been demonstrated experimentally that the cochlear filters of the inner ear are not spaced at linear frequency intervals but that their spacing is approximately logarithmic [6].

The emission of sound (speech, songs) by human cord tract (larynx, pharynx, mouth) is a complex nonlinear process that combines both muscles and tissues with different temporal and spatial scales, and the entire autonomic and central nervous systems. In this study, we analyse human voice signals (a single note maintained for a few seconds) that characterise the physiology of the vocal organ (larynx-pharynx-mouth) in healthy and pathological situations. To compare different signals and their spectral composition, we define a log-frequency correlation measure on spectral densities that gives the distribution of frequency ratios (rational or irrational numbers) between two signals. Using the wavelet transform formalism we extend this measure to a time-frequency correlation measure, that offers the possibility to estimate the temporal variability of this log-frequency correlation. We introduce reference spectral expansions as sums of Dirac terms that resume the characteristic property of these voice signals (harmonics as integer multiples of a fundamental frequency). Finally, we define a new integral cross-correlation of the previously defined measure which quantifies the rationality of the rhythms of two compared signals. We call it *sonance*, by analogy with the term consonance (resp. dissonance) that counts the perceived affinity or agreement (resp. disagreement) between different sounds. We validate this method on numerical model and voice signals collected from different sources. The first section is this introduction. The second section describes the mathematical methodology for log-frequency correlations (or spectrum of frequency ratios) and its generalization to time-frequency expansions in terms of analytic wavelet transforms. The third section illustrates these concepts on numerical signals (sums of sine functions) and voice recordings from the Voice-Icar-Federico II database, introduces the *sonance* measure and illustrates it on the previously computed log-frequency correlation measures of voice signals. Finally, we leave the medical application of voice dysphonia diagnosis with the comparison of an untrained voice with a singer voice that have similar spectral envelopes.

2 Spectrum of frequency ratios. Formalism and time-frequency generalization

2.1 Correlation functions for signal comparison

Let us consider two signals x and y of finite total energy $L^2(\mathbb{R}) : \langle x, x\rangle < +\infty$ and $\langle y, y\rangle < +\infty$ where $\langle \cdot, \cdot\rangle$ is the ordinary inner product of $L^2(\mathbb{R})$ and $\overline{x}$ is the complex conjugate of x:

$$\langle x, y\rangle = \int_{-\infty}^{+\infty} \overline{x(u)} y(u) \mathrm{d}u. \tag{1}$$

The comparison of these two signals x and y is usually performed through a deterministic correlation function $R[x, y](\xi)$ constructed from a time shift (translation) operator $\mathbf{T}_\xi$:

$$R[x, y](\xi) = \langle x, \mathbf{T}_\xi y\rangle = \int_{-\infty}^{+\infty} \overline{x(u)} y(u + \xi) \mathrm{d}u. \tag{2}$$

This definition, given for energy signals or square-integrable functions, can be extended to power signals. Thus, for signals which can be described by sums of periodic functions (stochastic signals with finite power), the cross-correlation function reads:

$$C[x, y](\xi) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \overline{x(u)} y(u + \xi) \mathrm{d}u, \tag{3}$$

When $x = y$, we get the auto-correlation function $C[x, x](\xi)$, that characterises the similarity between observations of a same signal as a function of the lag ξ between them. The auto-correlation function is Hermitian: $C[x, x](-\xi) = \overline{C[x, x](\xi)}$. The absolute value of $C[x, x](\xi)$ is maximum at the origin, where the auto-correlation function is real, positive and equal to the power of the signal x. When the signal x is real, this implies that the auto-correlation function is real and even.

Note that when $u = t$, t being the time variable, the function $C[x, y](\xi)$ is the cross-correlation function commonly used for time signals but u could be replaced by any other type of variable, and in particular the frequency (or log-frequency) when comparing spectral signals, as will be discussed below.

Translation-based correlation functions are very important for physics. They turn functions of a relative quantity (such as time or space position whose value depends on a translation from an arbitrary origin) into a function of an absolute quantity (such as time or space interval). However, the value of absolute quantities that have a physical dimension still depends on its comparison with an arbitrary standard: the physical unit. Since a scaling is involved, the unit plays the role of an arbitrary origin for the logarithm of these quantities. That is the reason why dilation-based correlation functions can be of interest for physics, as long as they compare functions of an absolute quantity: a new variable made of the ratio of two absolute quantities with the same physical unit neither depends on an origin nor on the unit; it is a pure proportion.

To extend the concept of correlation functions to absolute physical quantities, we need first to revisit the definition of the inner product. We make use of the logarithm to change from the translation-invariant group $(\mathbb{R}, +)$ to the dilation-invariant one $(\mathbb{R}^+, \times)$. The change of variable $u = \log v$ applied on Eq. (1) yields:

$$\langle X, Y \rangle = \int_{-\infty}^{\infty} \overline{X(u)}\, Y(u) \mathrm{d}u = \int_{0}^{\infty} \overline{X(\log v)} Y(\log v) \mathrm{d} \log v. \tag{4}$$

The change from the function $X(u)$ and the measure $\mathrm{d}u$ to the function $X \circ \log(v)$ and the measure $\mathrm{d} \log v = \mathrm{d}v/v$ means, for numerical computations, that we replace linearly sampled functions by geometrically sampled ones (of positive variable). In the following, we choose to make explicit the composition with the logarithm in each function. The previous translation operator $\mathbf{T}_{\xi}$ is naturally replaced by a dilation operator $\mathbf{D}_q$:

$$\mathbf{T}_{\log q}[X](\log v) = \mathbf{D}_q[X \circ \log](v) = X(\log(qv)). \tag{5}$$

Combining Eqs (4) and (5), we obtain from Eq. (2) a similar correlation function adapted to geometrically sampled signals:

$$R[X, Y](\log q) = \int_{0}^{\infty} \overline{X(\log v)} Y(\log(qv)) \mathrm{d} \log v, \tag{6}$$

where q is positive. For functions X and Y, the finite energy condition for the validity of this integral takes the form $\langle X, X \rangle < +\infty$. It can also be reformulated for finite power signals in a similar way as in Eq. (3).

The dilation correlation function in Eq. (6) inherits the following symmetry and linearity properties from Eq. (2):

$$R[Y, X](\log q) = \overline{R[X, Y](-\log q)}, \tag{7}$$
$$R[X, Y + Z](\log q) = R[X, Y](\log q) + R[X, Z](\log q). \tag{8}$$

Note that the logarithm does not allow to study functions of a negative absolute quantity (for instance negative delays or frequencies), nor negative ratios $q < 0$.

2.2 Spectrum of frequency ratios: a frequency ratio distribution

For the application of interest here, the unfolding of rhythms from real signals (their spectral “timbre”), we concentrate on “geometric” spectral densities that we define as real and positive functions $\mathcal{S}(\log f) \geq 0$ of the logarithm of the frequency. The log-frequency correlation function between two such densities

$$R[\mathcal{S}_1,\mathcal{S}_2](\log q) = \int_0^\infty \mathcal{S}_1(\log f)\mathcal{S}_2(\log(qf))\mathrm{d}\log f, \tag{9}$$

captures all the spectral relations between frequency modes of $\mathcal{S}_1(\log f)$ and $\mathcal{S}_2(\log f)$. $R[\mathcal{S}_1,\mathcal{S}_2](\log q)$ is positive and gives the distribution of frequency ratios q of $\mathcal{S}_1(\log f)$ and $\mathcal{S}_2(\log f)$, hence the notation R for ratio distribution. Similarly to standard correlation function of linearly sampled variables, the existence of this integral $R[\mathcal{S}_1,\mathcal{S}_2](\log q)$ requires that both distributions $\mathcal{S}_1(\log f)$ and $\mathcal{S}_2(\log f)$ be square integrable with the geometric measure of f (linear measure for $\log f$). Both the log-frequency distribution $\mathcal{S}(\log f)$ and the frequency ratio distribution $R[\mathcal{S}_1,\mathcal{S}_2](\log q)$ can be normalised as probability density functions:

$$\int_0^\infty R[\mathcal{S}_1,\mathcal{S}_2](\log q)\mathrm{d}\log q = \int_0^\infty \mathcal{S}_1(\log f)\mathrm{d}\log f \int_0^\infty \mathcal{S}_2(\log f)\mathrm{d}\log f = 1\ . \tag{10}$$

Frequency ratio distributions can be written in analytic form from spectral densities defined as isolated or sum of Dirac δ functions. For example, the two spectral densities $\mathcal{S}_j(\log f) = \delta(\log \frac{f}{f_j})$, $j = 1,2$ have a single frequency ratio $\frac{f_2}{f_1}$, and give a frequency ratio distribution $R[\mathcal{S}_1,\mathcal{S}_2](\log q) = \delta(\log \frac{qf_1}{f_2})$. If we define $\mathcal{S}(\log f)$ as a doublet of Dirac deltas $\mathcal{S}(\log f) = \mathcal{S}_1(\log f) + \mathcal{S}_2(\log f)$, from the linearity property Eq. (8), we can write the ratio distribution $R[\mathcal{S},\mathcal{S}](\log q) = \delta(\log \frac{qf_1}{f_2}) + 2\delta(\log q) + \delta(\log \frac{qf_2}{f_1})$. This simple analytic case is illustrated in Fig. 1, where we distinguish from $R[\mathcal{S},\mathcal{S}](\log q)$ three peaks, corresponding to the frequency pairs: (4:4) and (8:8) for $\log q = 0$, (4:8) for $\log q = \log 2$, and (8:4) for $\log q = -\log 2$.

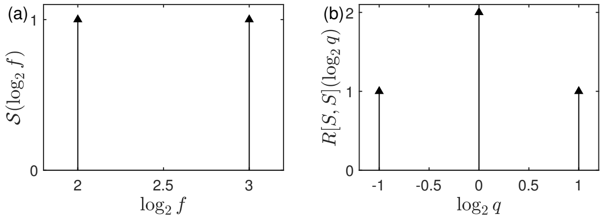

Fig. 1 (a) Ideal distribution $\mathcal{S}(\log f)$ in log-frequencies of a doublet of Dirac deltas such that the highest frequency is twice the lowest. (b) Representation of the spectrum of self-relations $R[\mathcal{S},\mathcal{S}](\log q)$ in logarithmic scale (base 2). The peak of ratio $\log_2 q = 0$ represents the self-relation of each frequency peak, whereas the ratios $\log_2 q = -1, 1$ represent their cross-relations.

The log-frequency spectral distributions $\mathcal{S}(\log f)$ cannot be assimilated to lin-frequency spectral densities defined the from Fourier transform of s: $\hat{s}(f) = \int_{-\infty}^{+\infty} s(t)e^{-2\pi i f t}\mathrm{d}t$, because their computation from linear measures in time and frequency faces some difficulties. The main one is practical, power spectral densities estimated with Fast Fourier Transform (FFT) algorithms are sampled linearly, whereas the integral of Eq. (9) requires a geometric frequency sampling. Re-sampling strategies of the Fourier spectra have been proposed in the literature [7], and could be used for stationary signals, however they require greater memory size and are computer time-consuming. Importantly, in the context of physiological signals which are often non-stationary, the extension of time-averaged spectral quantities to time-frequency distributions is mandatory. The wavelet transform answers to both issues, it provides not only a time-frequency representation of the spectral quantities, but also allows a geometric sampling in frequency. Using time-frequency decompositions we can straightforwardly extend our definition of log-frequency ratio distributions Eq. (9) to time-log-frequency ratio distributions for the analysis of non-stationary signals.

2.3 Wavelet transform formalism

Time-frequency analysing tools based on the wavelet transform have been introduced in the second half of the twentieth century and applied to many scientific domains for characterising and modelling non-stationary processes [8, 9, 10, 11, 12, 13, 14]. The wavelet transform of a finite energy signal $s(t) \in L^2(\mathbb{R})$ is defined as its inner product with the shifted copies of an analysing absolute integrable and finite energy wavelet $\psi(t) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ [9, 14, 15, 16]:

$$\mathcal{W}_{\psi}^{(p)}[s](a,b) = \langle s, \psi_{a,b} \rangle = a^{-\frac{1}{p}} \int_{-\infty}^{+\infty} s(t) \overline{\psi\left(\frac{t-b}{a}\right)} \mathrm{d}t, \tag{11}$$

$b \in \mathbb{R}$ and $a \in \mathbb{R}^+$ are the shift and scaling parameters. $\overline{\psi}$ is the complex conjugate of the analysing wavelet ψ, p is a parameter which defines the normalisation of the wavelet.
Two values of p are usually in the more common definitions found in the literature: $p = 1$, corresponding to the $L^1(\mathbb{R})$ norm and $p = 2$, corresponding to the $L^2(\mathbb{R})$ norm, respectively.
$p = 1$ often used for time-localized signals with different amplitudes, is appropriate when the magnitude of the modulus wavelet transform is wished to reflect the amplitude of the analysed signal $s(t)$. $p = 2$ is appropriate when the modulus-squared wavelet transform is wished to reflect the energy of the analysed signal $s(t)$.

In the frequency domain, the expression of the wavelet transform reads:

$$\mathcal{W}_{\psi}^{(p)}[s](a,b) = a^{1-\frac{1}{p}} \int_{-\infty}^{+\infty} \hat{s}(f)\, \overline{\hat{\psi}(af)} e^{2i\pi f b} \mathrm{d}f\ , \tag{12}$$

where $\hat{s}, \hat{\psi}$ denote the Fourier transforms of the signal and the wavelet.
This time-scale representation is quite suited for non-stationary signals since it localizes the analysis around time b and operates a band-pass filtering scaled by the parameter a. Importantly, a can be sampled arbitrarily, in our case we will sample it geometrically. It is common practice to consider the scale a as proportional to an inverse frequency $\frac{1}{f_a}$:

$$a = \frac{f_\psi}{f_a}, \tag{13}$$

where f_ψ is a characteristic frequency of the mother wavelet ψ. Three meaningful frequencies are classically used for f_ψ [17]: the peak frequency f_ψ^0 where the frequency domain mother wavelet magnitude $\left|\hat{\psi}(f)\right|$ is maximum, the energy (norm 2) frequency f_ψ^* which is the mean of $\left|\hat{\psi}(f)\right|^2$ and the norm 1 frequency $\check{f}_\psi$, that can be interpreted as an instantaneous frequency for progressive wavelets. An asymmetry in the frequency domain of the mother wavelet leads to distinct values for the previous frequencies f_ψ.

For the computation of the log-frequency correlation functions, the expression $\mathcal{W}_\psi[s](f_\psi/f_a, b)$ for the wavelet transform given in Eq. (12) can be turned to a time-frequency analysis by using Eq. (13) for a given characteristic frequency f_ψ:

$$\mathcal{W}_\psi^{(p)}[s]\left(\frac{f_\psi}{f_a}, b\right) = a^{1-\frac{1}{p}} \int_{-\infty}^{+\infty} \hat{s}(f)\, \overline{\hat{\psi}\left(\frac{f_\psi}{f_a} f\right)} e^{2i\pi f b} \mathrm{d}f\ . \tag{14}$$

For our applications to physiological signals, the Banach space $L^1(\mathbb{R}, \mathrm{d}t)$ norm corresponding to $p = 1$ will be preferred for the wavelet transform definition. The main reason is due to the fact that when rescaling time in the input signal as $s\left(\frac{t}{\rho}\right)$, with $\rho > 0$, both the time and the scale of the wavelet transform are rescaled, but without changing its magnitude. Thus as the Fourier transform of $s\left(\frac{t}{\rho}\right)$ is: $\rho\,\widehat{s}(\rho f)$, Eq. (12) when $p = 1$ leads to $\mathcal{W}_\psi^{(1)}[s]\,(\frac{a}{\rho}, \frac{b}{\rho})$. The (1) is dropped in the following. The peak frequency f_ψ^0 will be then adopted for the characteristic frequency f_ψ in Eqs (13), (14).

The admissibility condition for an analysing wavelet $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ establishes that the number

$$c_\psi = \int_0^{+\infty} |\hat{\psi}(u)|^2\, \frac{\mathrm{d}u}{u} \tag{15}$$

must be finite, nonzero and independent of $f \in \mathbb{R}^+$. If this admissibility condition is fulfilled, then every $s \in L^2(\mathbb{R})$ can be reconstructed from the convergent integral:

$$s(t) = \frac{1}{c_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{\infty} \mathcal{W}_\psi[s]\,(a, b)\, \psi\left(\frac{t-b}{a}\right) \frac{\mathrm{d}a}{|a|} \mathrm{d}b\ . \tag{16}$$

2.3.1 Time and frequency window for the analysing wavelet

The time-frequency window can be computed from the expression of the analysing wavelet ψ, assuming that ψ and $\hat{\psi}$ verify $t\psi(t) \in L^2$ and $f\hat{\psi}(f) \in L^2(\mathbb{R})$ [18]. If the center and the radius (with the norm 2) of the window function ψ are respectively t_ψ^* and Δ_ψ, $\psi((t-b)/a)$ is a window function with center $b + at_\psi^*$ and radius equal to $a\Delta_\psi$:

$$[b + at^*_\psi - a\Delta_\psi, b + at^*_\psi + a\Delta_\psi] \; . \tag{17}$$

This windows narrows (respectively widens) for small (resp. large) values of a. In the frequency domain, the window of $\hat{\psi}$ is defined similarly, assuming that the center and width of $\hat{\psi}$ are f^*_ψ and $\Delta_{\hat{\psi}}$, $\psi(af)$ is centered around f^*_ψ/a and has a radius $\Delta_{\hat{\psi}}/a$:

$$\left[\frac{f^*_\psi}{a} - \frac{1}{a}\Delta_{\hat{\psi}}, \frac{f^*_\psi}{a} + \frac{1}{a}\Delta_{\hat{\psi}}\right] \; . \tag{18}$$

In the following discussion, the center f^*_ψ of $\hat{\psi}$ is assumed to be positive. There are different ways of defining the wavelet resolution, called the quality factor of the wavelet. A first definition, given in [19], uses the bandwidth and the norm 2 frequency as follows:

$$Q^* = \frac{f^*_\psi/a}{2\Delta_{\hat{\psi}}/a} = \frac{f^*_\psi}{2\Delta_{\hat{\psi}}} \; , \tag{19}$$

which is independent of the scale parameter a. Alternatively, we could also use the full width at half maximum height of $|\hat{\psi}(f)|^2$ instead of $\Delta_{\hat{\psi}}$. We thus define another quality factor, $\tilde{Q}$, such as

$$\tilde{Q} = \frac{f^0_\psi}{|f_2 - f_1|} \tag{20}$$

where $|\hat{\psi}(f_1)|^2 = |\hat{\psi}(f_2)|^2 = |\hat{\psi}(f^0_\psi)|^2/2$ and $f_1 < f^0_\psi < f_2$. This factor is usually computed to characterise the qualitative damping behavior of simple damped oscillators [20].

The choice of the quality factor is essential to obtain an adapted time-frequency resolution and consequently a “good” analysis of the processed signals. The authors in [21] propose three bounds to obtain a range of acceptable values. When the signal is composed of several frequency components, the proximity of their characteristic frequencies provides a lower bound. The exponential decay rate of the amplitude imposes another upper bound. Eventually, the length of the signal determines yet another upper bound.

2.3.2 Choice of the analysing wavelet: the Grossmann wavelet

In the absence of a suitable unifying theory for wavelet behaviors, the choice of a particular wavelet for a particular problem may often appear arbitrary. For rhythmic signals, complex analytic analysing wavelets are preferred, leading to: $\hat{\psi}(f) = 0$, $\forall f \leq 0$. In that case, the measure appears naturally in these integrals (Eq. (16)), as in Eqs (4) and (6), because the analysing wavelet is scale invariant (under dilations).

In the following, we choose a single-parameter progressive wavelet, introduced for the decomposition of Hardy functions by Grossmann and Morlet [8]:

$$\hat{\psi}_Q(f) = \begin{cases} \psi_0 e^{-\frac{1}{2}\left(Q \log \frac{f}{f_0}\right)^2} & \forall f > 0\,; \\ 0 & \forall f \leq 0\,, \end{cases} \tag{21}$$

of peak frequency $f_\psi^0 = f_0$, for which the maximum value is ψ_0. This wavelet is symmetric in log-frequencies about $\log f_0$. The other characteristic frequencies are $f_\psi^* = f_0 e^{\frac{3}{4Q^2}}$ and $\check{f}_\psi = f_0 e^{\frac{3}{2Q^2}}$. Both previously defined quality factors depend on Q only:

$$Q^* = \frac{1}{2}\left(e^{\frac{1}{2Q^2}} - 1\right)^{-\frac{1}{2}}, \tag{22}$$

$$\tilde{Q} = \left(2 \sinh \frac{\sqrt{\log 2}}{Q}\right)^{-1}. \tag{23}$$

When Q is large enough, the leading term in the expansions gives $Q^* \simeq Q/\sqrt{2}$ and $\tilde{Q} \simeq Q/\sqrt{\log 2}$ respectively, followed by a term of order $\frac{1}{Q}$. Consequently, we will refer to the parameter Q as the quality factor for this wavelet.

When choosing the value $\psi_0^2 = \frac{Q}{\sqrt{\pi}}$, the admissibility constant is one and $|\hat{\psi}(f)|^2$ can be considered as a probability density function in log-frequencies:

$$c_\psi = \int_0^\infty |\hat{\psi}_Q(f)|^2 \mathrm{d} \log f = 1\ . \tag{24}$$

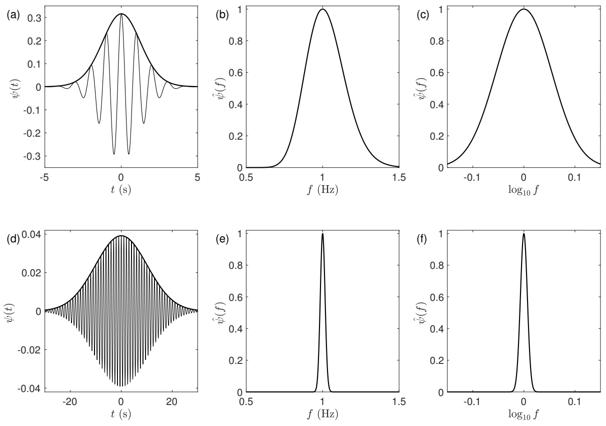

Fig. 2 Grossmann analysing wavelet (log-normal in frequency). $\psi(t)$ is computed by inverse Fourier transform of $\hat{\psi}_Q(f)$ (Eq. (21)). (a) $\psi_Q(t)$. (b) $\hat{\psi}_Q(f)$ in linear-frequency scale. (c) $\hat{\psi}_Q(f)$ in a (base 10) logarithmic frequency scale. (a,b,c) are computed for $Q = 8$. (d,e,f) Same as (a,b,c) for $Q = 64$. In (a) and (d) $|\psi_Q(t)|$ (respectively $\Re\left\{\psi_Q(t)\right\}$) are plotted in thick (resp. light) black lines.

In Fig. 2, we plot the Grossmann wavelet for two values of Q, respectively $Q = 8$ (top plots) and $Q = 64$ (bottom plots). For larger Q values, the number of oscillations of $\Re\left\{\psi_Q(t)\right\}$ and its width increases whereas $\hat{\psi}_Q(f)$ narrows. We can observe in Fig. 2(c,f) that the wavelet in the log-frequency domain is symmetric around $\log f_0 = 0$ whereas it is asymmetric around $f_0 = 1$ in linear frequencies (Fig. 2(b,e)). An important aspect of the oscillating progressive wavelets is how many oscillations are fitting inside their time-window [19, 17]. This number of oscillations determines the acuteness of the local frequency detection of a given rhythm and is of order Q. If this number is too large, the wavelet averages over too much oscillations and cannot provide a correct estimation. Conversely, if the number of oscillations is insufficient (less that ~ 3) the detection of a local rhythm will not be possible. The choice of this parameter is particularly important if the signal presents sharp transitions or close frequencies, as will be illustrated in the following figures.

The authors in [17] showed that the Grossmann wavelet can be seen as a scaling limit of a general family of progressive wavelets with two parameters, the Morse wavelet [22, 23, 24, 25, 26, 27, 28]. The Cauchy-Paul wavelet, intensively used in quantum mechanics and in the context of analytic functions [29], as well as the analytic version of the derivative of Gaussian wavelet or the Airy wavelet all belong to the Morse family.

2.4 Extension of frequency ratio distributions to time-frequency ratio distributions

From the Grossmann progressive wavelet transform defined in the previous section, we define a time-frequency distribution for non-stationary signals:

$$\mathcal{S}^{(Q)}(\log(f_a), b) = \left|\mathcal{W}_{\psi_Q}[s]\left(\frac{f_0}{f_a}, b\right)\right|^2 . \tag{25}$$

Note that the integral of the wavelet transform definition in Eq. (12) is sampled linearly in f, but that the values of the frequencies f_a (or scale a) can be chosen arbitrarily, for our purpose we will select them geometrically distributed. In the following, $b = t$ and $f_a = f$ are considered as time and frequency parameters, which simplifies the notation of $S^{(Q)}(\log f, t)$. This distribution is computed for strictly positive values of f and we can extend the definition of the cross-correlation function to time-frequency distributions:

$$\begin{aligned} R[\mathcal{S}_1^{(Q)}, \mathcal{S}_2^{(Q)}](\log q, t) &= \int_0^\infty \mathcal{S}_1^{(Q)}(\log f, t)\mathcal{S}_2^{(Q)}(\log(qf), t)\mathrm{d}\log f & (26)\\ &= \int_0^\infty \left|\mathcal{W}_{\psi_Q}[s_1]\left(\frac{f_0}{f}, t\right)\right|^2 \left|\mathcal{W}_{\psi_Q}[s_2]\left(\frac{qf_0}{f}, t\right)\right|^2 \mathrm{d}\log f . & (27) \end{aligned}$$

The log-frequency autocorrelation function is defined as $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$. The temporal mean of $\mathcal{S}^{(Q)}(\log f, t)$: $\langle \mathcal{S}^{(Q)} \rangle_t(\log f)$ that can be seen as a power spectral density based on the wavelet transform $\mathcal{W}_{\psi_Q}[s]$.

2.4.1 Computation of the log-frequency correlation function

Using the convolution theorem, $R[\mathcal{S}_1^{(Q)}, \mathcal{S}_2^{(Q)}]$ can be computed quite efficiently using the fast Fourier transform (FFT) several times (that discretizes the Fourier transform here denoted $\mathcal{F}$): on a first step with respect to the time variable the signal (noted $\mathcal{F}$), and on a second step with respect to the log-frequency variable (noted $\mathcal{F}_{\log f}$) and the computation step is an inverse FFT in log-frequency space (noted $\mathcal{F}_{\log f}^{-1}$).

$$R[\mathcal{S}_1^{(Q)}, \mathcal{S}_2^{(Q)}](\log q, t) = \mathcal{F}_{\log f}^{-1}\left[\overline{\mathcal{F}_{\log f}\left[|W_{\psi_Q}[s_1](.,t)|^2\right]}\mathcal{F}_{\log f}\left[|W_{\psi_Q}[s_2](.,t)|^2\right]\right](\log q) \tag{28}$$

$$\text{where } W_{\psi_Q}[s](f,t) = \mathcal{F}^{-1}\left[\overline{\hat{\psi}}_Q\left(\frac{f'}{f}\right)\mathcal{F}[s](f')\right](t)\ . \tag{29}$$

This supposes that the frequency f (or scale a) parameter of the CWT is sampled geometrically. The slowest operations consist in matrix multiplications. The fact that the second step requires Fourier transforms of the distributions $\mathcal{S}$ on log-frequency scale implies that the computed range of log-frequency values is enlarged, and padded with zeros to avoid extra-ratios arising from the FFT computation by an artificial periodisation of the $\mathcal{S}$ distribution.

3 Computation of log-frequency distributions from numerical and real signals

3.1 Model signals constructed from sine functions

In Fig. 3, we construct an artificial non-stationary signal from the sum of two sine functions: $s(t) = \sin(\phi_1(t)) + \sin(\phi_2(t))$, with $\phi_2(t) = 4\pi t$ linear in time, and $\phi_1(t) = 2\pi t H(-t) + 3\pi t H(t)$ with the Heaviside step function H, and we compare the wavelet transform analysis for the two quality factors $Q = 8$ and $Q = 64$. With this signal we estimate a lower bound of Q that is suitable for a frequency discrimination according to [21]: for $t < 0, Q \gtrsim 10$ and for $t > 0, Q \gtrsim 14$. Moreover, the signal length gives the constraint $Q \lesssim 285$. For $t < 0$, the signal possesses two frequencies, highlighted on the colour-coded image of $\mathcal{S}^{(Q)}(\log f, t)$ by two horizontal bands ($f_1 = 1$ and $f_2 = 2$), the width of which depends on the quality factor Q, ($Q = 8$ near the lower acceptable Q bound in Fig. 3(b) and $Q = 64$ in

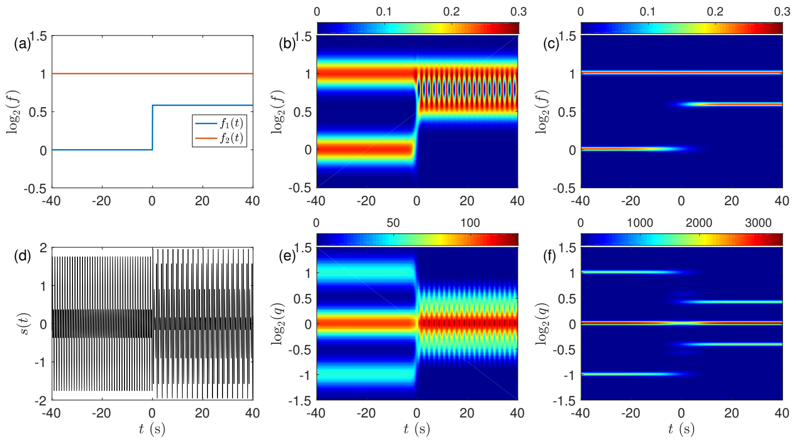

Fig. 3 Analysis of a model signal defined as the sum of two sine functions $s(t) = \sin(2\pi f_1(t)t) + \sin(2\pi f_2 t)$, with $f_2 = 2$ constant, and $f_1(t) = H(-t) + \frac{3}{2}H(t)$ with the Heaviside step function. (a) Plot of the frequencies $f_1(t)$ and $f_2(t)$ in (base 2) logarithmic scale. (b) $\mathcal{S}^{(8)}(\log f, t)$, computed for $Q = 8$. (c) $\mathcal{S}^{(64)}(\log f, t)$, computed for $Q = 64$. (d) Temporal signal $s(t)$ in the time window [-40s, 40s]. (e) $R[\mathcal{S}^{(8)}, \mathcal{S}^{(8)}](\log q, t)$. (f) $R[\mathcal{S}^{(64)}, \mathcal{S}^{(64)}](\log q, t)$. $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$ is defined in Eq. (26).

Fig. 3(c)). For $t > 0$, we can again recognise the two bands f_1 and f_2, and, as for $t < 0$, their narrowing for the larger Q values. The transition zone of this two bands, below and above $t = 0$, needs to be discussed. Fig. 4 highlights this transition with sections of $\mathcal{S}^{(Q)}(\log f, t)$ performed for remarkable values of f; 1, 3/2 and 2. From the sections of Fig. 4(a), we estimate the width of this transition ~ 4.8s for $Q = 8$, and ~ 39s for $Q = 64$. Another interesting phenomenon emerges in the $t > 0$ regime, where the two frequency bands become closer. A low-frequency modulation of the wavelet transform squared modulus in the intermediate frequency range $[f_1, f_2]$ with period 2s appears, corresponding to frequency $f_m = f_2 - f_1$ (0.5Hz in this example). The matrix of the wavelet transform modulus is not simply the superimposition of the wavelet transform squared moduli of the sine alone, $|\mathcal{W}_{\psi_Q}[s_1 + s_2](a, b)|^2 \neq |\mathcal{W}_{\psi_Q}[s_1](a, b)|^2 + |\mathcal{W}_{\psi_Q}[s_2](a, b)|^2$, but extra terms such as $2\hat{\psi}(af_1)\hat{\psi}(af_2)\cos(2\pi(f_2 - f_1)b)$ are also involved and are not negligible when f_1 and f_2 become too close (which is the case in Fig. 3(b)). This effect disappears quite completely for larger Q values because the product $\hat{\psi}(af_1)\hat{\psi}(af_2)$ vanishes. We conclude that the choice of Q is a compromise between two objectives, (i) discriminating close frequencies (in which case larger Q values will be preferred), (ii) affording a correct temporal resolution for the detection of steep frequency changes (in which case smaller Q values will be more efficient).

Fig. 3(e,f) shows the corresponding colour-coded maps $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$ for the same signal and the same values of Q (8 and 64). We recognise for $t < 0$ three horizontal bands of constant q, corresponding respectively to frequency ratios $q = 1/2, 1, 2$. The intensity of the middle band ($q = 1$) is more contrasted (×2) because it corresponds to the sum of self-relations (f_1:f_1) and (f_2:f_2). The two

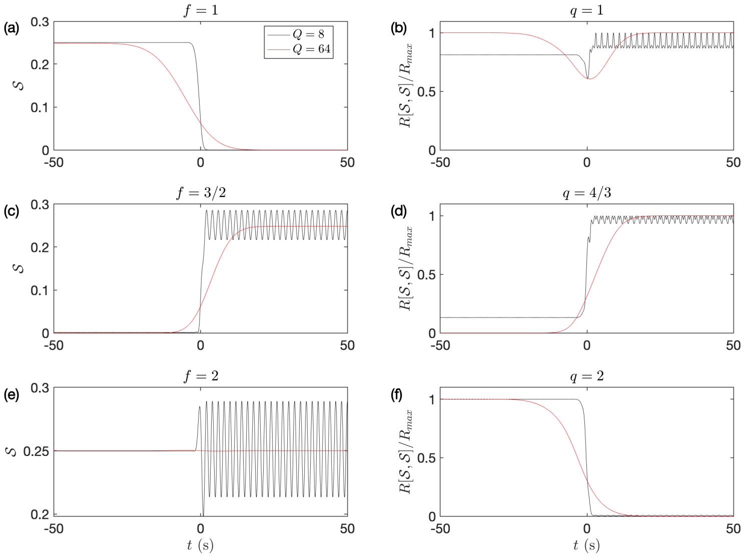

Fig. 4 (a, c, e) Sections of the log-frequency spectral distributions $\mathcal{S}^{(8,64)}(\log f_i, t)$ selected from Fig. 3(b,c) for the three frequencies $f_1 = 1$, $f_2 = 3/2$ and $f_3 = 2$ Hz, and the same values of the quality factor Q (8 and 64). (b, d, f) Sections of $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q_i, t)$ selected from Fig. 3(e,f) for three values of q: $q_1 = 1$, $q_2 = 4/3$, $q_3 = 2$, corresponding to local maxima of $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}]$. For each selected q value, $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}]$ was scaled by its maximum in the time interval.

symmetric weaker bands correspond to cross-frequency ratios $(f_1{:}f_2)$ and $(f_2{:}f_1)$. For $t > 0$, the three bands become closer, and similarly to the maps of $\mathcal{S}^{(Q)}$ in Fig. 3(b), a slow temporal modulation of $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$ superimposes to the bands, due to coupling terms in the wavelet transform modulus. As expected, and similarly to what was observed on $\mathcal{S}^{(Q)}(\log f, t)$ maps, increasing Q from 8 to 64 produces a strong narrowing of the bands and a strong reduction of the low-frequency modulation. The sections at fixed q of these $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$ maps are shown in Fig. 4 to highlight similarly the transition zone around $t = 0$, its widening for larger Q values, and the slow temporal modulations observed for $Q = 8$. Due to the use of the Grossmann wavelet, sections at fixed t of both $\mathcal{S}^{(Q)}(\log f, t)$ and $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$ are Gaussian of widths $(\sqrt{2}Q)^{-1}$ and Q^{-1} respectively when the bands are not interfering (independent of t).

Another family of model signals (Fig. 5(e)), particularly interesting with respect to the applications to voice signals, is defined as the sum of sine functions $\sum_{i=1}^{n} \sin(2\pi f_i t)$ with $f_i = i f_1$, i positive integer. In Fig. 5, we take $n = 6$ and perform the same time-frequency analysis with a Grossmann analysing wavelet with two values of Q, respectively $Q = 8$ (b,c) and $Q = 128$ (f,g). We note again that the larger Q, the finer and distinguishable the peaks of both $\mathcal{S}^{(Q)}(\log f, t)$ and $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$. The already noticed low-frequency modulations in the previous example again appear in this example (Fig. 5(b,f)) for $Q = 8$. Amazingly the frequency of this slow mode is precisely the fundamental frequency of this signal, and this modulation is the most intense for the highest harmonic ($f_6 = 6f_1$), this effect is due to the ordering of these 6 frequencies as integer multiples of f_1, giving a constant frequency step between successive harmonics $f_{i+1} - f_i = f_1$. This $f_1 = 1$ Hz slow modulation mode appears when two frequencies of the list are too close (in log-scale) for being separated properly by the analysing wavelet.

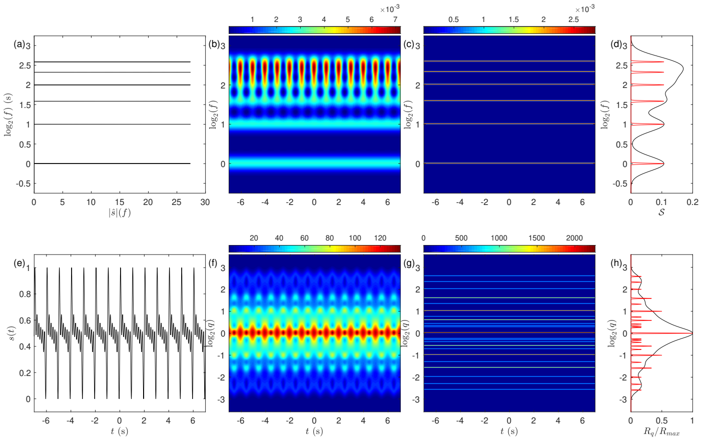

Fig. 5 Analysis of a model signal defined as the sum of 6 sine functions $s(t) = \sum_1^6 \sin(2\pi f_i t)$, with $f_i = i f_1$ constant. (a) Plot of $\log(f(t))$. (b) $\mathcal{S}^{(8)}(\log f, t)$, computed for $Q = 8$. (c) $\mathcal{S}^{(128)}(\log f, t)$, computed for $Q = 128$. (d) $\mathcal{S}^{(Q)}(\log f, t = 0)$. (e) Temporal signal $s(t)$ in the time window [-6.5s, 6.5s]. (f) $R[\mathcal{S}^{(8)}, \mathcal{S}^{(8)}](\log q, t)$. (g) $R[\mathcal{S}^{(128)}, \mathcal{S}^{(128)}](\log q, t)$. (h) $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t = 0)$. In (d) and (h) the plots for $Q = 8$ (resp. 128) are coloured in black (resp. red). $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$ is defined in Eq. (26).

We observe in Fig. 5(d) that for $Q = 8$ (black curve), the higher frequency components cannot be discriminated. This phenomena is even more visible on the ratio distribution $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$ in Fig. 5(h). $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$ presents an odd number of peaks, it is symmetric around the central peak ($q = 1$). In that example, each of the sixth frequency components contributes to this central peak. 11 lateral peaks emerge for $q > 1$, and accumulate closer to the central peak. The positions of these peaks correspond to all the possible distinguishable frequency ratios of the signal, and the amplitude of these peaks is proportional to the number of combinations of frequencies that produces a given ratio. To distinguish all the peaks in Fig. 5(g,h) it was necessary to increase Q to 128. The total number of frequency ratios (for $q > 1$) is $\sum_{i=1}^{n-1} i = n(n-1)/2$ if $n > 2$, in this example it is equal to 15. When there is no redundancy in the frequency ratios, for instance if harmonic frequencies are prime multiples of the fundamental frequency, each frequency ratio occurs once in $R[\mathcal{S}^{(Q)}, \mathcal{S}^{(Q)}](\log q, t)$.

3.2 Physiological signals: voice recordings

The voice signals reported in this manuscript were selected from the **VO**ice-**IC**ar-f**ED**erico II (VOICED) database [30] recorded by the “Institute of High Performance Computing and Networking of the National Research Council of Italy (ICAR-CNR)” and the Hospital University of Naples “Federico II” during 2016 and 2017. This database can be downloaded from the PhysioNet website [31]. It has been proposed lately as a new element in research on automatic voice disorder detection and classification. Together with medical phonetic examinations of a set of 208 individuals, among which 73 male and 135 female, voice signals, proportional to a local sound emission intensity, were acquired for about 4-5 s and sampled at 8000 Hz at 32 bit, vocal folds were examined by laryngoscopy and two medical questionnaires were collected at the ambulatories of Phoniatrics and Videolaryngoscopy of the “Federico II” Hospital of the University of Naples or at the medical room of the ICAR-CNR. The protocol description is reported in [30]. Dysphonia is a quite common voice disorder (1/3 of adults will suffer from it once in their lifetime), it may originate from a functional or organic alteration of the vocal apparatus and its mechanics and may not systematically be considered as pathologic [32, 33]. On the one side, laryngoscopy is an invasive technique that gives a direct view of the physical alterations of the vocal tract [34]. On the other side, the analysis of the voice acoustic signal is not intrusive and, thanks to the improvement of signal analysis methods, it can nowadays be used to guide or assist the recognition of the origin of a suspected dysphonia. Voice classification methods from voice recordings by the recognition and quantification of the voice timbre (or tone color) has rapidly attracted the interest of electronic and computer science engineers. Globally, one can classify these methods in three groups [35], (i) the time-domain methods which use autocorrelation functions or their variants [36, 37] to search for repeatability between a temporal waveform and its time lagged version, (ii) frequency domain methods which locate characteristic frequencies and conclude to a spectral “coloraturas” for the voice, these methods meet rapidly their limitations if the signal is not stationary, (iii) time-frequency domain techniques [38, 39, 40], that we have chosen for this study.

The voice signal $s(t)$, numbered #008, is that of a female of 51 years without deep vocal impairment at the time of the test, ranked in the group of reflux laryngitis (Fig. 6). This example was chosen because it has marked peaks which can be detected by thresholding the signal (this is quite rare, because it requires both a particular shape of the signal and a global stationarity of its amplitude). The Fourier spectra of this signal (reported in log-log and log-lin scales in Fig. 6(b) and (c) respectively) weight the power (in log-scale) of its spectral components; a fundamental mode with frequency $f_1 \sim 188$ Hz and higher modes (harmonics), ranked as integer multiples of f_1: if_1 with $i = 2, 3, 4, 5, \dots$ with different powers. This simple frequency decomposition was observed in most of the signals provided in the VOICED database, this is a conspicuous characteristic of the human voice. These voice signals appear as the alternance of quite regular large and sharp peaks (which give the fundamental mode) and smaller oscillations which may be very irregular. In some cases these smaller oscillations may be difficult to discriminate from the noise produced by some friction of the vocal tracts. Even though this type of signal can be compared to the sum of sine functions introduced in Fig. 5(e), the higher number of harmonics of this signal and their different power means that it could be reproduced by a nonlinear dynamical system (ruled by nonlinear ordinary differential equations) where the different frequency components follow nonlinear rules [42]. Our purpose in this paper is not to discuss the physical and biological mechanisms or the modelling of voice signals, we have selected these examples as illustrations for our log-frequency correlation method because their spectral decomposition is very rich in harmonics (overtones) of the fundamental frequency.

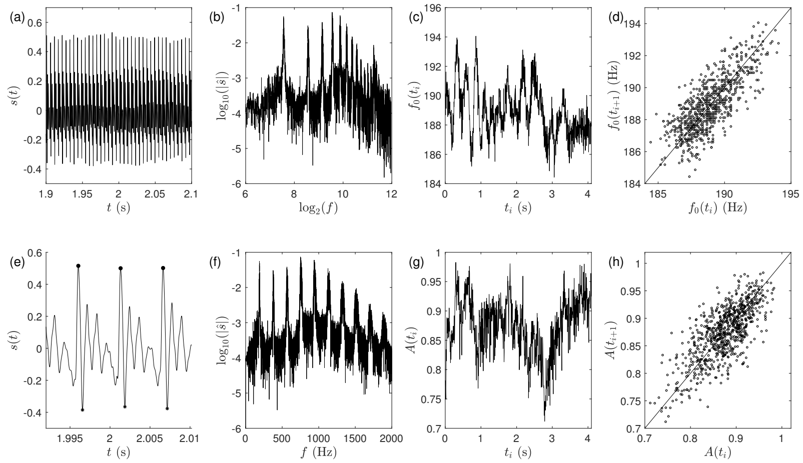

Fig. 6 Analysis of the #008 voice signal $s(t)$ from the voice-icar-federico-ii database [30]. (a) Zoom of the signal during 0.2 s. (b) $|\hat{s}|$ plotted versus f in logarithmic scales (base 10 and base 2 respectively). (c) Local frequency $f_1(t)$ computed from the detection of the extrema of larger prevalence from the signal (see (e)). (d) First return inverse of interpeak intervals $f_1(t_i) = 1/\Delta T_i$ scatter plot (these large amplitude peaks are marked with black dots in (e)). (e) Zoom of the signal on the short interval (20 ms) showing the local maxima $s_P(t_i)$ (black dots) and minima $s_p(t_i)$ (black stars) which are used to compute both the local frequency: $f_1(t_i) = 1/\Delta T_i$ and the amplitude $A(t_i)$ of each larger amplitude peak: $A(t_i) = s_P(t_i) - s_p(t_i)$. (f) $|\hat{s}|$ (in base 10 log-scale) plotted versus f in Hz (linear scale). (g) Amplitude of the largest peaks $A(t_i)$ versus time (see (e) for their detection). (h) First return peak amplitude ($A(t_{i+1})$ vs $A(t_i)$) scatter plot.

The temporal change of the fundamental mode frequency $f_1(t_i)$ and the largest peak amplitude $A(t_i)$ can be extracted from the #008 voice signal by thresholding its largest amplitude peaks (maxima: $s_P(t_i)$ and minima $s_p(t_i)$) as depicted in Fig. 6(e). Fig. 6(c) shows that $f_1(t)$ is modulated in time, suggesting an irregularity of the rhythm coming from some difficulty of the patient to maintain a constant value of f_1. In this example, a similar temporal modulation is also visible on the largest peak amplitude $A(t_i)$ (Fig. 6g). If these temporal variations were solely produced by instrumental noise, the first return scatter plots of f_1 and A at successive peaks would give a symmetric cloud of points around the diagonal. In Fig. 6(d) for the fundamental frequency modulation and in Fig. 6(h) for the amplitude modulation these first return scatter plots are anisotropic, meaning that the dispersion of these values extends beyond instrumental noise. This conclusion is also confirmed by the temporal evolution of $f_1(t_i)$ (Fig. 6(c)) and $A(t_i)$ (Fig. 6(g)), we notice that, in the first second, the modulations of $f_1(t_i)$ have the largest amplitude and are quasi-periodic, this first regime can also be recognised from the modulations of $A(t_i)$. This patient has a rather mild dysphonia (classified as produced by reflux laryngitis) which can be recognised by an important set of harmonics and a rather low vocal fold noise.

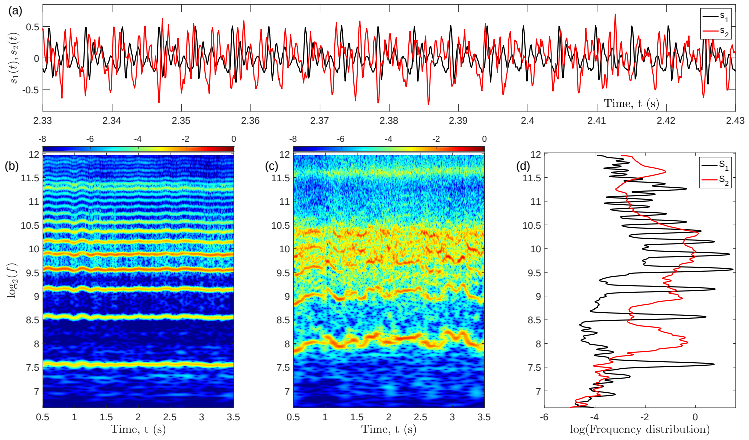

Fig. 7 Comparison of the time-frequency analysis of voice signals #008 (s_1) and #169 (s_2). (a) Zooms of s_1 and s_2 in a 0.1s window. (b,c) Associated time-frequency distributions (Eq. (25)) $\mathcal{S}^{(64)}[s_1](\log f, t)$, and $\mathcal{S}^{(64)}[s_2](\log f, t)$ computed with a Grossmann analysing wavelet and a quality factor $Q = 64$. The horizontal bands highlight the fundamental and harmonic frequencies. (d) Corresponding temporal averages of the frequency distributions reported in panels (b) black line and (c) red line. The ordinate of (d) (here the horizontal axes) is arbitrary and the frequency distributions are normalised.

The second voice signal that we have selected is that of a female of 62 years (#169 in the VOICED database), with hyperkinetic dysphonia. In that case, the quasiperiodicity observed in signal #008 is so much disrupted that it is impossible to use the previous threshold method for extracting the largest signal peaks; the time-frequency analysis is required to check to which extent we can find a timbre for this voice, and how it changes with time. Fig. 7(b,c) reports the colour-coded images of the #008 (s_1) and the #169 (s_2) time-frequency distributions $\mathcal{S}^{(64)}(\log f, t)$. Averaging these frequency distributions, we get smoothed power spectrum distributions for these two examples (Fig. 7(d)). The fundamental band frequency of #169 is much broader than that of #008, and shifted to greater values $f_{1,1} \sim 188.8$Hz (voice #008) and $f_{1,2} \sim 268$Hz (voice #169), and we also note that it is quite impossible in #169 to discriminate more than one harmonics from the averaged frequency spectrum. The time frequency distributions in Fig. 7(b,c) highlight these differences. Whereas the fundamental mode band and its harmonics are weakly modulated in time for #008, that of #169 are very irregular, the third and fourth harmonics can be mixed up and indistinguishable, the harmonics above five are no longer visible. The vocal folds of #169 can no longer maintain their tight contact that is essentiel for a correct sound emission and the resulting effect, when hearing the voice, is that of a scratching noise which covers completely the expected tone. This person is quite unable to sing a melody.

3.3 Tuning voice pitches via the computation of correlation functions

3.3.1 Reference frequency distribution $\mathcal{S}_{0,j}$

An "ideal" frequency distribution $\mathcal{S}_{0,j}$ is first introduced in order to compare the real frequency distribution $\mathcal{S}_j^{(Q)}$ with a reference through the cross-correlation defined in Eq. (26). Let us consider the vibrating string model used to represent the sounds emitted by stringed instruments. When the string is plucked at its ends, its natural frequencies are integer multiples of the fundamental frequency depending on the square root of the force of tension of the string [41]. By analogy to this model, the reference frequency distribution $S_{0,j}(\log f)$ is a Dirac comb model in log scale defined as a sum of integer multiples of the fundamental frequency $f_{1,j}$ of the studied signal

$$\mathcal{S}_{0,j}(\log f) = \sum_n c_n \delta\left(\log \frac{f}{n f_{1,j}}\right), \tag{30}$$

with weights $c_n \geq 0$ and possible cut-off ($c_n = 0,\ \forall n > N$). We assume that the series of c_n is bounded: $\sum_n c_n < +\infty$.

We compare then the spectrum of the comb reference model to itself. We build an ideal ratio distribution by computing the auto-correlation of $\mathcal{S}_{0,j}$ derived from Eq. (27):

$$R_{00}(\log q) = R[\mathcal{S}_{0,j}, \mathcal{S}_{0,j}](\log q) = \sum_n \sum_m c_n c_m \delta\left(\log q \frac{n}{m}\right), \tag{31}$$

this ratio distribution depends neither on time, nor on the signal under study. When $\log q = 0$ ($q = 1$), $R_{00}(0) = \sum_n c_n^2 < +\infty$.

3.3.2 Cross-correlation $R[\mathcal{S}_{0,j}, \mathcal{S}_j^{(Q)}]$

The time distribution cross-correlation function between $\mathcal{S}_{0,j}$ and $\mathcal{S}_j^{(Q)}$ is deduced from Eq. (27):

$$R[\mathcal{S}_{0,j}, \mathcal{S}_j^{(Q)}](\log q, t) = \sum_n c_n \mathcal{S}_j^{(Q)}(\log(qnf_{1,j}), t). \tag{32}$$

Different degrees of “frequency matching” can also be captured by high peaks in $R[\mathcal{S}_0, \mathcal{S}_j^{(Q)}](\log q, t)$, especially when q is a simple frequency ratio of harmonics of $f_{1,j}$ and $f_{1,0}$, for example 1:2 - octave, 2:3 - fifth, 3:4 - fourth would give a perfect consonance, 3:5 - major sixth and 4:5 - major third would give a medial consonance, 5:6 - minor third and 5:8 - minor sixth would give imperfect consonance. “Unmatched” frequency configurations would be obtained if a couple of frequency ratio of harmonics belong to the dissonance list: 8:9 - major second, 8:15 - major seventh, 9:16 - minor seventh, 15:16 - minor second, 32:45 ($\sim 1/\sqrt{2}$) - tritone [43].

In the following, we will denote $R_{ij} = R[\mathcal{S}_i^{(Q)}, \mathcal{S}_j^{(Q)}]$, the time-frequency window is fixed ($Q = 64$).

3.3.3 Application to two voice signals from the VOICED data base

For each of the two voice signals #008 and #169, we construct a Dirac comb model as reference distribution. These distributions are such that their lowest frequency peak matches the signal fundamental frequency (for instance for the signal #008: $f_{1,1} = 188.8$Hz and for signal #169: $f_{1,2} = 268$Hz). The frequency of the highest harmonic of the comb model is limited by the sampling frequency F_s: $nf_{1,i} \lesssim F_s/2$ ($F_s = 8000$ Hz). We take $c_n = 1, \forall n \leq 15$:

$$\mathcal{S}_{0,j}(\log f) = \sum_{n=1}^{15} \delta\left(\log \frac{f}{nf_{1,j}}\right), \quad j = 1, 2. \tag{33}$$

For numerical computations, the frequency f is discretized in $f_k = f_{\min} \cdot \alpha^{k-1}$, with $k = 1, 2, 3, ...N$ and N the size of the frequency vector f, $f_{\min}$ its minimum value and α the geometric factor determined from N, $f_{\min} = 100$ Hz (fixed by the voice database), and $f_{\max} = F_s/2$. The correlation function $R_{0j}(\log q, t)$ is computed by combining this comb distribution with that of the voice signal as in Eq. (28), using the analytic expression of the $\log(f)$-Fourier transform ($\mathcal{F}_{\log f}$) of the comb distribution. We do not take its Fourier transform numerically because it is the source of numerical artefacts. For the comb model aligned to the fundamental frequency $f_{1,j}$ of signal s_j, it reads:

$$\mathcal{F}_{\log f}\left[\mathcal{S}_{0,j}\right](u) = \sum_{n=1}^{15} \exp(-i2\pi u \log(nf_{1,j}/f_{\min})). \tag{34}$$

u is the conjugated variable (through Fourier transformation) of $\log(f)$. In that space, we take the scalar product of $\mathcal{F}_{\log f}\left[\mathcal{S}_{0,j}\right]$ with the conjugate of $\mathcal{F}_{\log f}\left[\mathcal{S}_j^{(Q)}\right]$ (computed numerically from $W_{\psi_Q}[s_j](\log f, t)$) and compute its inverse Fourier transform to recover the correlation function $R[\mathcal{S}_{0,j}, \mathcal{S}_j^{(Q)}](\log q)$.

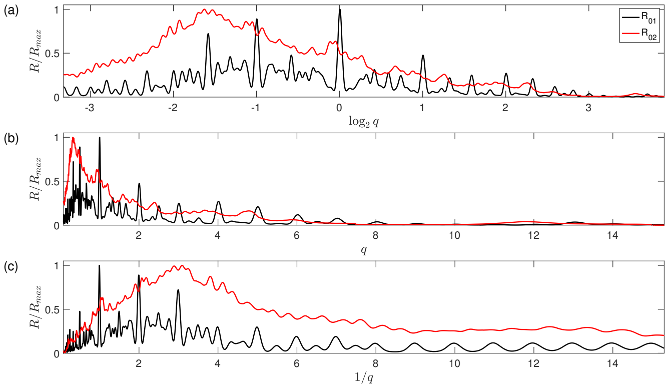

Fig. 8 Ratio distributions of voice signals ((1) #008 and (2) #169) and their frequency-matched Dirac comb models (R_{0j}). (a) Correlations of the frequency distribution $\langle R_{0i} \rangle = R[\mathcal{S}_{0,j}, \langle \mathcal{S}_j^{(Q)} \rangle_t](\log q)$ (j = 1, 2, Q = 64) with their reference comb frequency distributions (defined in the text). These correlations have been normalised to their maximum for ease of comparison. (b, c) Plot of $R[\mathcal{S}_{0,j}, \langle \mathcal{S}_j^{(Q)} \rangle_t]$ versus q for $q > 1$ (b) and $1/q$ for $q < 1$ (c).

The time-averaged frequency ratio distributions $R[\mathcal{S}_{0,j}, \langle \mathcal{S}_j^{(Q)} \rangle_t](\log q)$ of each voice signals #008 and #169 with its "best-fitted" Dirac comb model are presented in Fig. 8. If the signals were regular and quasi-stationary, these ratio distributions should pinpoint ratios corresponding to the multiples of the fundamental frequencies. The plot of these correlation functions in linear q and $1/q$ scales in Figs 8(b) and 8(c) highlights a strong asymmetry, it is due to different amplitudes of the fundamental mode and its harmonics compared to the constant coefficients in the comb model. Again, as for frequency distributions, we note a strong difference of the ratio distributions for signals #008 (s_1) and #169 (s_2). Confronting $R[\mathcal{S}_0, \mathcal{S}^{(Q)}](\log q, t)$ with $\mathcal{S}^{(Q)}(\log f, t)$ for voice signal #169 (Fig. 9) unveils important features which were not visible from the time averaged ratio distribution $R[\mathcal{S}_0, \langle \mathcal{S}^{(Q)} \rangle_t](\log q)$ (Fig. 8). Even if the fundamental mode frequency and its harmonics vary a lot during these 3s record, their ratios do not change dramatically, as a characteristic property of the mechanics of the vocal folds. In the middle of this signal ($1.4\text{s} < t < 1.55\text{s}$) (see Fig. 9(a) for a zoom in this interval), four flat ratio bands can be noticed, suggesting that this person put sufficient effort to recover for a short period of time a "mild sensation" of timbre. How this intermittent loss and recovery of the voice timbre occurs, the time range of these alternating sequences could be used as diagnosis criteria or aftercare follow-up (invasive intervention is necessary if soft or hard nodules are detected on the vocal cords (stage III), or voice exercises for earlier stages).

3.3.4 Cross-correlation $R[\mathcal{S}_i^{(Q)}, \mathcal{S}_j^{(Q)}]$ of two voice signals

There are two interpretations for the log-frequency cross-correlation function $R[\mathcal{S}_i^{(Q)}, \mathcal{S}_j^{(Q)}](\log q, t)$, leading to different possible applications. Either we see it as a distribution of the ratios q between the frequencies of $\mathcal{S}_i^{(Q)}(\log f, t)$ and

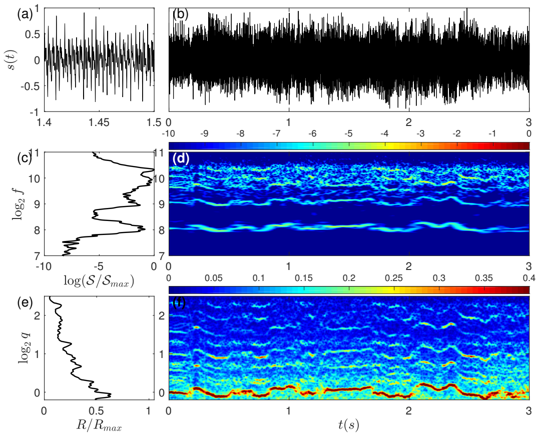

Fig. 9 Comparing $R[\mathcal{S}_0, \mathcal{S}^{(Q)}](\log q, t)$ with $\mathcal{S}^{(Q)}(\log f, t)$ for voice signal #169. (a) Zoom of $s(t)$ in the [1.4s, 1.5s] interval. (b) Plot of a middle selection of 3s out of the 4.5s recorded voice signal. (c) Temporal average of the frequency distribution $\langle \mathcal{S}^{(Q)} \rangle_t(\log f)$ computed with a Grossmann analysing wavelet with quality factor $Q = 64$. (d) Colour-corded map of the time-frequency distribution $\mathcal{S}^{(Q)}(\log f, t)$. (e) Ratio distribution of the averaged frequency distribution: $R[\mathcal{S}_0, \langle \mathcal{S}^{(Q)} \rangle_t](\log q)$. (f) Colour-coded map of the time-ratio distribution $R[\mathcal{S}_0, \mathcal{S}^{(Q)}](\log q, t)$.

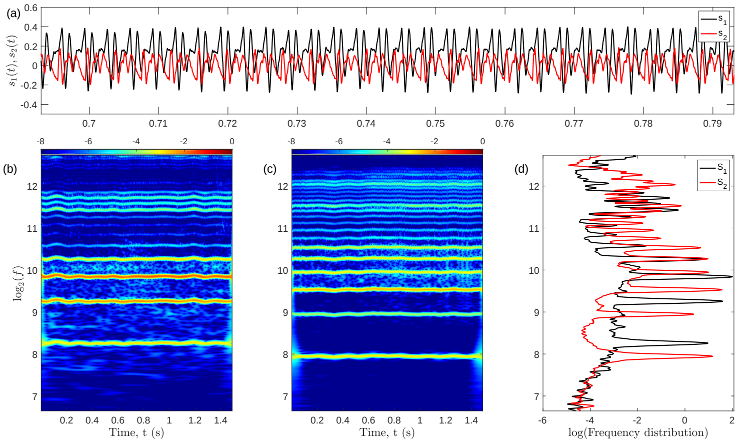

Fig. 10 Comparison of the time-frequency analysis for two "normal" voice signals: s_1 is a sung vowel, s_2 is simply a maintained vowel. (a) Zooms of s_1 and s_2 in a 0.1s window. (b,c) Associated time-frequency distributions (Eq. (25)) $\mathcal{S}_1^{(Q)}(\log f, t)$, and $\mathcal{S}_2^{(Q)}(\log f, t)$ computed with a Grossmann analysing wavelet with quality factor $Q = 64$. The horizontal bands highlight the fundamental and harmonic frequencies. (d) Corresponding temporal averages of the frequency distributions reported in panels (b) black line and (c) red line. The ordinate of (d) (here the horizontal axes) is arbitrary and the frequency distributions are normalised.

$\mathcal{S}_j^{(Q)}(\log f, t)$, or we view it for each q as a measure of how well $\mathcal{S}_i^{(Q)}(\log f, t)$ and $\mathcal{S}_j^{(Q)}(\log(qf), t)$ match. For instance, in the example reported here, $\mathcal{S}_1^{(Q)}, \mathcal{S}_2^{(Q)}$ are obtained from two different persons holding a pitch in their vocal range. The peaks in the correlation function $R[\mathcal{S}_1^{(Q)}, \mathcal{S}_2^{(Q)}](\log q)$ indicate the importance of the corresponding frequency ratios between the voices, in accordance with its first interpretation as a ratio distribution. When these ratios are close to simple rational numbers m/n, they indicate the presence of a $m : n$-synchronisation, that corresponds to the consonance of the voices simply sung together. In this way, we assess how rational the spectral relations of the voices are. The other interpretation is as follows: assuming the $S_2^{(Q)}(\log qf)$ models the second voice transposed by q to a different pitch, the peaks in $R[S_1^{(Q)}, S_2^{(Q)}](\log q)$ also indicate for which pitch transpositions the second voice would best match the first voice. This allows us to tune one voice with the other.

The possibility to match a real voice signal with reference model signals is very interesting because it can limit the maximum harmonics frequency for this cross-correlation, as an "intelligent" low pass filtering. This would not be possible by computing directly $R[\mathcal{S}_i^{(Q)}, \mathcal{S}_j^{(Q)}]$. We compare in Figs 10 and 11 two different "normal" voices ("a" vowel) from the clinic research in the speech therapy laboratory UNADREO in Toulouse (France). The frequency distributions $\mathcal{S}_i^{(Q)}(\log f, t)$ and $\mathcal{S}_j^{(Q)}(\log f, t)$ $(i \neq j)$ plotted in Fig 10, have the same characteristic frequency peaks structure as the voice #008, but we notice that the frequency distribution of the voice s_1 has greater energy in the harmonics around 1000 Hz, which is characteristic of the emission of trained singer voices.

The cross-correlation ratio distribution $R_{ij} = R[\mathcal{S}_i^{(Q)}, \mathcal{S}_j^{(Q)}]$ is quite different from the auto-correlation ratio distributions R_{00}, $R_{ii}^{(Q)}$ and $R_{jj}^{(Q)}$. A common reference Dirac comb is chosen for both s_1 and s_2 ($i = 1$ and $j = 2$) and is aligned to the fundamental frequency $f_{1,1} = 307.2$Hz. The highest central peak indicates the ratio of the fundamental frequencies, it is centered for R_{00}, R_{11} and R_{22} and shifted to $q = f_{1,2}/f_{1,1} = 247/307.2 \sim 2^{-0.315}$ for R_{12}. R_{00} shows very sharp and narrow peaks which line up symmetrically on either sides of $q = 1$. The amplitude of these peaks recapitulates the weighting of the frequency ratios q for simple comb models and gives us which distribution would be obtained if all the frequency components of the signals had exactly the same power. There is a very light asymmetry of R_{00} which comes from inescapable numerical limitations in the implementation of comb models from the Fourier transforms of Dirac δ-distributions. This small artefact could be diminished by increasing the frequency resolution and taking longer voice records.

3.4 Frequency distribution matching and *sonance*

To find the best match with the cross-correlation R_{12}, we propose then to compute

$$R[R_{00}, R_{12}](\log x, t) = \int_0^\infty R_{00}(\log q) R_{12}(\log xq, t) \mathrm{d} \log q \tag{35}$$

$$= \int_0^\infty R_{00}\left(\log \frac{q}{x}\right) R_{12}(\log q, t) \mathrm{d} \log q = \sum_n \sum_m R_{12}\left(\log \frac{m}{n} x, t\right). \tag{36}$$

This new quantity can be computed by two equivalent paths:

$$R[R_{00}, R_{12}](\log x, t) = R[R_{01}, R_{02}](\log x, t). \tag{37}$$

The new parameter x that we call pitch transposition can be understood in two ways. Either $R_{12}(\log xq, t)$ is seen as the distribution of ratios q between the first voice $\mathcal{S}_1^{(Q)}(\log f, t)$ and the second voice of transposed pitch $\mathcal{S}_2^{(Q)}(\log xf, t)$. Or the x in $R_{00}(\log(q/x))$ is seen as a varying ratio between the fundamental frequency of the ideal distribution $\mathcal{S}_0$.

Indeed, the best matching is expected when the pitch of the second voice is transposed to match the first one, thus when the voices are sung at unison, or equivalently when the fundamental frequency ratios are matched between the pairs of distributions: $x = f_{1,2}/f_{1,1}$.

As a result, for two voices of fundamental frequency $f_{1,1}$ and $f_{1,2}/x$, the quantity $R[R_{00}, R_{12}](\log x, t)$ as a function of the pitch transposition x has the following interpretation: it measures how "ideal" (similar to the model R_{00}) the spectral relations are between the voices. Extrema of this curve appear directly related to the musical property of consonance or dissonance of certain fundamental frequency ratios. For this reason, we call this quantity the *sonance* between the two voices, and we rewrite its definition (35) using the reference ratio distribution as the density of a measure $\mathrm{d}\text{𝄞}(\log q) = R_{00}(\log q)\mathrm{d}\log(q)$:

$$\text{𝄞}[R_{12}](\log x, t) = R[R_{00}, R_{12}](\log x, t) = \int_0^\infty R_{12}(\log xq, t)\mathrm{d}\text{𝄞}(\log q), \tag{38}$$

that we could denote equivalently $\text{𝄞}[\mathcal{S}_1^{(Q)}, \mathcal{S}_2^{(Q)}](\log x, t)$.

This *sonance* measure is a geometric function of a pitch transposition quantity x, its maxima indicate the optimum relative pitch transpositions for which the two voices sung together would match best. This term *sonance* bears some analogy with the concepts of consonance and dissonance which were first suggested by Pythagoras (sixth century BC), hence our choice of the symbol 𝄞, but this similarity of terms must be nuanced. Dissonance and consonance are not mathematical quantities since they have been used to describe an empirical sensation of human beings (combination of cochlea physiology and cognitive training) when hearing a mixture of sounds (two or more)[44, 45].

In Fig. 11, we compute the *sonance* of the two voice records shown in Fig. 10, and referenced to the same comb model defined from the first signal s_1. The comparison of the plots of the cross-correlation functions R_{01}, R_{02}, and R_{12} with the auto-correlation functions R_{11}, R_{12} and R_{00} of Fig. 11 draw our attention to important

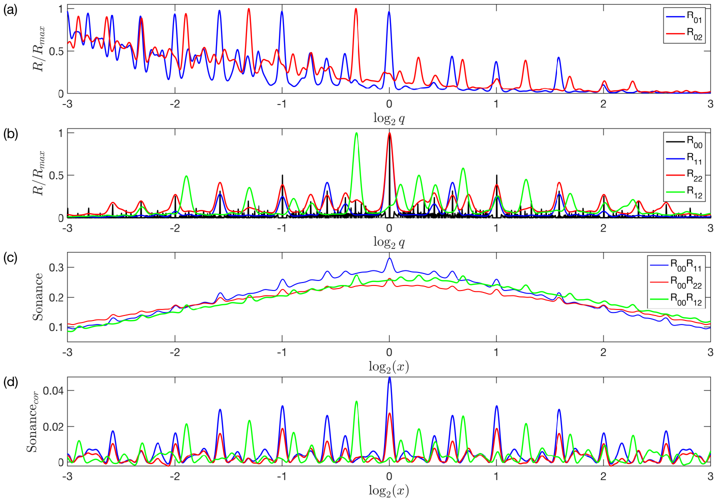

Fig. 11 Comparing ratio distributions and *sonance* of the two voice signals of Fig. 10. (a) Plots of the normalised ratio distributions $R_{0j} = R[\mathcal{S}_{0,1}, \langle\mathcal{S}_j^{(Q)}\rangle](\log q)$ with $j = 1, 2$. For both signals we use the same Dirac comb model with the fundamental mode frequency of voice s_1. (b) Plots of the normalised ratio distributions R_{11}, R_{22}, R_{12} with $R_{ij} = R[\langle\mathcal{S}_i^{(Q)}\rangle, \langle\mathcal{S}_j\rangle](\log q)$ computed from the two voices, and $R_{00} = R[\mathcal{S}_{0,1}, \mathcal{S}_{0,1}](\log q)$ computed from the comb Dirac model. (c) Plot of the *sonance* (cross- and auto-) of the two voices with respect to the reference comb model ratio distribution for an arbitrary pitch transposition $\log(x)$: 𝄞$[R_{ij}](\log x) = R[R_{00}, R_{ij}](\log x)$ with $i, j = 1, 2$ green line, $i, j = 1, 1$ blue line and $i, j = 2, 2$ red line. (d) Sonance curves of Fig. 11(c) are corrected by subtracting their lower envelope.

features. Whereas the auto-correlation functions R_{ii} are always symmetric, the cross-correlation functions are asymmetric, and their greater peaks are observed for q values below 1. We observe the same asymmetry in the cross-sonance profiles 𝄞$[R_{12}](\log x)$ shown in Fig. 11(d) (red line), the basal line of the *sonance* is larger for $x < 1$ than for $x > 1$. With the "auto-sonance" functions 𝄞$[R_{ii}](\log x)$, computed similarly as in Eq. (38), we recover the symmetry. The largest peaks of 𝄞$[R_{12}]$ point to the interval changes between the two voices that would lead to increased consonance, and its wells to losses of consonance of the voices sung simultaneously. The left-shift of 𝄞$[R_{12}]$ from the central ratio $x = 1$ corresponds to a change in pitch; voice s_2 has a lower pitch than voice s_1 (number 1). Finally, the almost inexistant peak of 𝄞$[R_{12}](0)$ indicates that these two voices sung together without any tone adjustement would be slightly dissonant because their fundamental and harmonic frequencies have few commensurability: the frequency ratios are not close to simple rational numbers m/n.

As a last remark, the *sonance* profile of the voices is directly influenced by two choices: first, the quality factor of the wavelet Q, which determines the distinguishability of the frequencies and their ratios, and second, the choice of the number of harmonics and possibly their amplitude in the reference comb model $\mathcal{S}_0$. We believe that, for a realistic *sonance* profile, Q should be related to the critical band of the ear [44] and is, together with the complexity of the reference ratio distribution R_{00}, representative of the musical training of the ear.

4 Conclusion

We have introduced time-log-frequency ratio distributions based on analytic wavelets that we have applied to model and physiological signals (voice records). A second correlation operation was defined to compare the matching of two of these ratio distributions, called *sonance* which estimates the pitch transposition that would produce "harmony" (small integer rational ratios of their harmonics) of two voices sung together. This work has shown that a geometric correlation function, in log-frequency is better suited to uncover characteristic frequency ratios between different signals. The application to voice records has been selected not only for its simplicity to perform and reproduce, but also because it gives credit to the origin of the concept of frequency ratios in voiced sounds. This method is presently generalized to other physiological signals (heart, brain, breath, muscles, vessels, . . .), and offers the possibility to compute cross-correlation distributions from signals of different nature, recorded from different organs or tissues.

Acknowledgements This work has been supported by ANR under contract: ANR-18-CE45-0012-01, EURLight S&T funding. We are very endebted to L. Delmarre, E. Harte, and S. Polizzi for fruitful discussions.

References

1. S.E. Jorgensen, B.D. Fath (eds.), *Encyclopedia of Ecology*, 1st edn. (Elsevier, Amsterdam, The Netherlands, 2008)
2. P.C. Ivanov, K.K.L. Liu, R.P. Bartsch, New Journal of Physics **18**(10), 100201 (2016)
3. A.L. Barabasi, N. Gulbahce, J. Loscalzo, Nature Reviews Genetics **12**(1), 56 (2011)
4. A. Goldbeter, *Au Coeur des Rythmes du Vivant. La Vie Oscillatoire* (Odile Jacob, Paris, 2018)
5. J.N. Oppenheim, M.O. Magnasco, Physical Review Letters **110**(4) (2013)
6. J. Schnupp, I. Nelken, A. King, *Auditory Neuroscience: Making Sense of Sound* (MIT Press, Cambridge, Mass, 2011)
7. G.V. Haines, A.G. Jones, Geophysical Journal International **92**(1), 171 (1988)
8. A. Grossmann, J. Morlet, SIAM Journal on Mathematical Analysis **15**(4), 723 (1984)
9. R. Kronland-Martinet, J. Morlet, A. Grossmann, International Journal of Pattern Recognition and Artificial Intelligence **1**(2), 273 (1987)
10. J.M. Combes, A. Grossmann, P. Tchamitchian, *Wavelets: Time-Frequency Methods and Phase Space* (Springer, Berlin, Heidelberg, 1989)

11. N. Delprat, B. Escudie, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, B. Torresani, IEEE Transactions on Information Theory **38**(2), 644 (1992)
12. R. Carmona, W.L. Hwang, B. Torresani, in *Wavelets and Statistics*, vol. 103, ed. by A. Antoniadis, G. Oppenheim (Springer New York, New York, NY, 1995), pp. 95–108
13. R. Carmona, W. Hwang, B. Torresani, IEEE Transactions on Signal Processing **45**(10), 2586 (1997)
14. R. Carmona, W.L. Hwang, B. Torresani, *Practical Time-frequency Analysis: Gabor and Wavelet Transforms with an Implementation in S*. Vol. 9 in Wavelet Analysis and its Applications (Academic Press, San Diego, 1998)
15. B. Torresani, *Analyse Continue par Ondelettes*. Collection Savoirs Actuels (InterEditions, Paris, 1995)
16. P. Flandrin, *Time-Frequency Time-Scale Analysis*, Vol. 10 in Wavelet Analysis and its Applications (Academic Press, San Diego, 1998)
17. J. Lilly, S. Olhede, IEEE Transactions on Signal Processing **57**(1), 146 (2009)
18. C.K. Chui, *An Introduction to Wavelets* (Academic Press, San Diego, 1992)
19. T.P. Le, P. Argoul, Journal of Sound and Vibration **277**(1-2), 73 (2004)
20. Y. Rocard, *Dynamique Générale des Vibrations* (Masson et cie, P aris 1943)
21. S. Erlicher, P. Argoul, Mechanical Systems and Signal Processing **21**(3), 1386 (2007)
22. I. Daubechies, T. Paul, Inverse Problems **4**(3), 661 (1988)
23. P.M. Morse, Physical Review **34**(1), 57 (1929)
24. I. Daubechies, J.R. Klauder, T. Paul, Journal of Mathematical Physics **28**(1), 85 (1987)
25. S. Olhede, A. Walden, IEEE Transactions on Signal Processing **50**(11), 2661 (2002)
26. J.M. Lilly, S.C. Olhede, IEEE Transactions on Information Theory **56**(8), 4135 (2010)
27. J.M. Lilly, S.C. Olhede, IEEE Transactions on Signal Processing **60**(11), 6036 (2012)
28. J.M. Lilly, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **473**(2200), 20160776 (2017)
29. T. Paul, K. Seip, in *Wavelets and Their Applications*, M.B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer and I. Raphael edn. (Jones and Bartlett, Boston, 1992), pp. 303–322
30. U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, L. Verde, Computers & Electrical Engineering **68**, 310 (2018)
31. A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley, Circulation **101**(23) (2000)
32. F. Le Huche, A. Allali, *La Voix, Tome 2: Pathologies Vocales d'Origine Fonctionnelle* (Elsevier Masson, Paris, 2010)
33. F. Le Huche, A. Allali, *La Voix, Tome 3: Pathologies Vocales d'Origine Organique* (Elsevier Masson, Paris, 2010)
34. G.S. Berke, B.R. Gerratt, Journal of Voice **7**(2), 123 (1993)
35. D. Jouvet, Y. Laprie, in *2017 25th European Signal Processing Conference (EUSIPCO)* (IEEE, Kos, Greece, 2017), pp. 1614–1618
36. M. Ross, H. Shaffer, A. Cohen, R. Freudberg, H. Manley, IEEE Transactions on Acoustics, Speech and Signal Processing **22**(5), 353 (1974)
37. S.Y. Lowell, R.H. Colton, R.T. Kelley, Y.C. Hahn, Journal of Voice **25**(5), e223 (2011)
38. S. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, San Diego, 1999)
39. L. Cohen, *Time-frequency Analysis* (Prentice Hall, Upper Saddle River, NJ, 1995)
40. N.E. Huang, S.S.P. Shen (eds.), *Hilbert-Huang Transform and its Applications*, 2nd edn. No. 16 in Interdisciplinary Mathematical Sciences (World Scientific Publ, Singapore, 2014)
41. M. Gérardin, D.J. Rixen, *Mechanical Vibrations. Theory and Application to Structural Dynamics*, 3rd edn. (John Willey & Sons, Ltd, Chichester, UK, 2015)
42. C.M. Travieso, J.B. Alonso, J. Orozco-Arroyave, J. Vargas-Bonilla, E. Noth, A.G. Ravelo-Garcia, Expert Systems with Applications **82**, 184 (2017).
43. H.V. Helmholtz, *Theorie Physiologique de la Musique Fondée sur l'Etude des Sensations Auditives* (Victor Masson et Fils, Paris, 1868)
44. R. Plomp, W.J.M. Levelt, The Journal of the Acoustical Society of America **38**(4), 548 (1965)
45. A. Kameoka, M. Kuriyagawa, The Journal of the Acoustical Society of America **45**(6), 1460 (1969)