

# A method for continuously assessing the autonomic response to music-induced emotions through HRV analysis

Michele Orini · Raquel Bailón · Ronny Enk ·  
Stefan Koelsch · Luca Mainardi · Pablo Laguna

Received: 1 October 2009 / Accepted: 22 February 2010 / Published online: 19 March 2010  
© International Federation for Medical and Biological Engineering 2010

**Abstract** Interest in therapeutic applications of music has recently increased, as well as the effort to understand the relationship between music features and physiological patterns. In this study, we present a methodology for characterizing music-induced effects on the dynamics of the heart rate modulation. It consists of three steps: (i) the smoothed pseudo Wigner-Ville distribution is performed to obtain a time–frequency representation of HRV; (ii) a parametric decomposition is used to robustly estimate the time-course of spectral parameters; and (iii) statistical population analysis is used to continuously assess whether different acoustic stimuli provoke different dynamic responses. Seventy-five healthy subjects were repetitively exposed to pleasant music,

sequences of Shepard tones with the same tempo as the pleasant music and unpleasant sounds overlaid with the same sequences of Shepard tones. Results show that the modification of HRV parameters are characterized by an early fast transient phase (15–20 s), followed by an almost stationary period. All kinds of stimuli provoked significant changes compared to the resting condition, while during listening to pleasant music the heart and respiratory rates were higher (for more than 80% of the duration of the stimuli,  $p < 10^{-5}$ ) and the power of the HF modulation was lower (for more than 70% of the duration of the stimuli,  $p < 0.05$ ) than during listening to unpleasant stimuli.

**Keywords** Heart rate · Heart rate variability (HRV) · Music · Time–frequency analysis · Wigner-Ville distribution

---

M. Orini (✉) · R. Bailón · P. Laguna  
Communication Technology Group, Aragón Institute of  
Engineering Research, University of Zaragoza, M. de Luna 1,  
50018 Zaragoza, Spain  
e-mail: michele@unizar.es

M. Orini · R. Bailón · P. Laguna  
CIBER de Bioingeniería, Biomateriales y Nanomedicina  
(CIBER-BBN), Zaragoza, Spain

M. Orini · L. Mainardi  
Department of Bioengineering, Politecnico di Milano, via Golgi  
39, 20133 Milano, Italy

R. Enk · S. Koelsch  
Languages of Emotion, Cluster of Excellence der Freien  
Universität Berlin, 14195 Berlin, Germany

R. Enk  
Max Planck Institute for Human Cognitive and Brain Sciences,  
Leipzig, Germany

S. Koelsch  
Department of Psychology, University of Sussex, Pevensey  
Building, Falmer, Brighton BN1 9QH, UK

## 1 Introduction

Almost everyone loves music because, among other reasons, it evokes particular emotional states. These emotional states are related to brain and autonomic nervous system (ANS) activity, but the relationship between musical and autonomic features is far from being completely understood. The use of music for therapeutic purposes or, more generally, for improving our well-being is a matter of increasing interest [6, 7, 11, 12, 17, 27, 32], but only little is known about how music can modulate physiological parameters such as heart rate and breathing rate.

In order to continuously quantify and characterize the autonomic response to sound stimuli, an appropriate methodology is needed. In particular, this methodology should be able to track music-induced dynamics in the ANS, and it should include an appropriate statistical study

to assess whether different musical stimuli induce significantly different autonomic responses.

The purpose of this study is to present a comprehensive framework for characterizing the dynamics induced by different musical stimuli on the autonomic modulation of heart rate. In order to characterize the autonomic response to musical stimuli, heart rate variability (HRV) is studied by time–frequency (TF) analysis, which provides a representation of a signal in both time and frequency domains simultaneously. Spectral analysis of HRV is a well-established technique to assess autonomic activity [24, 31] and it was already proposed, in a stationary context, as a critical index for the assessment of autonomic effects elicited by music [17]. In non-stationary conditions, the tracking of the HRV spectral components (namely, the LF [0.04, 0.15] Hz and HF [0.15, 0.4] Hz components) provides the assessment of the autonomic dynamics, and different methods have been proposed for the TF analysis of HRV [21, 22]. In this study, we employ a TF approach based on the smoothed pseudo Wigner-Ville distribution (SPWVD) combined with a parametric decomposition which allows to obtain a denoised spectrum and a robust estimation of the HRV parameters. In order to characterize the dynamic autonomic response to a given musical stimulus, the individual response patterns for each subject are estimated, and a sample-by-sample statistical analysis is performed.

## 2 Methods

### 2.1 Time–frequency analysis of HRV

The Wigner-Ville distribution (WVD) provides an excellent localization of spectral components in TF domain, but its applicability is limited by the unavoidable presence of interference terms, which are not inherent to the signal and which oscillate in the TF plane [10, 15, 16]. The SPWVD is a filtered version of the WVD, proposed to reduce the interference terms while maintaining a good TF resolution. Its discrete version is defined as:

$$S(n, m; g(n), h(k)) = 2 \sum_{k=-K+1}^{K-1} s_{x_a}(n, k) e^{-j\frac{2\pi k}{M}m}, \tag{1}$$

$$s_{x_a}(n, k) = |h(k)|^2 \sum_{n'=-N+1}^{N-1} [g(n')x_a(n+n'-k)x_a^*(n+n'+k)], \tag{2}$$

where  $n$  and  $m \in [-M + 1:M]$  are the discrete time and frequency indexes;  $k \in [-K + 1:K - 1]$ , with  $2K - 1 < 2M$ , is the time-lag index of the autocorrelation function (ACF);  $s_{x_a}(n, k)$  is the smoothed instantaneous ACF of the analytic version of the HRV signal  $x_a(n)$ . In (2),

interferences are attenuated by using a kernel composed by the time smoothing function  $g(n)$  and by the frequency smoothing function  $|h(k)|^2$ . As also proposed in [23], the following smoothing functions were used:

$$g(n) = \begin{cases} \frac{1}{2N-1} & \text{if } n = (-N + 1), \dots, (N - 1) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and

$$|h(k)|^2 = \begin{cases} e^{-\gamma|k|} & \text{if } k = (-K + 1), \dots, (K - 1) \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

Note that (i) the degree of time and frequency smoothing depends on the length of the rectangular window ( $2N - 1$ ) and on the damping factor of the exponential window ( $\gamma$ ); (ii) using  $g(n)$  and  $|h(k)|^2$  as in (3) and (4), the energy of  $x_a(n)$  is conserved.

### 2.2 Parameter extraction

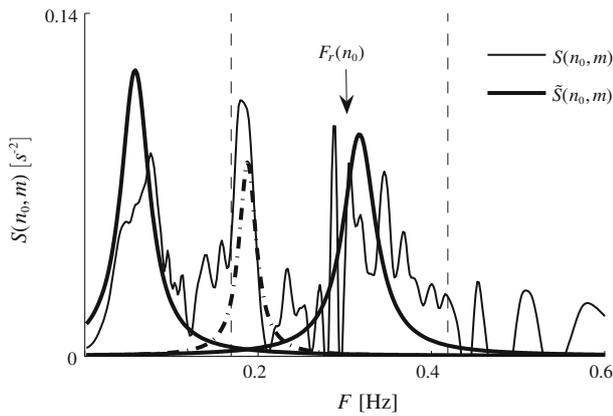
The time course of the HRV spectral parameters (such as instantaneous central frequencies  $F_{LF}(n)$ ,  $F_{HF}(n)$  and the instantaneous powers  $P_{LF}(n)$ ,  $P_{HF}(n)$  of LF and HF bands) was obtained performing a parametric decomposition of  $s_{x_a}(n, k)$  [23] and using the respiratory rate to estimate the HF component [1]. The method of decomposition is an extension of the Prony modeling, designed to estimate the parameters of damped sinusoids embedded in noise. In brief, the instantaneous ACF  $s_{x_a}(n, k)$  is modeled, for every time-instant  $n_0$ , as the sum of  $q$  damped sinusoids which are in strict relation with the instantaneous LF and HF components:

$$s_{x_a}(n_0, k) = \sum_{i=1}^q C_i(n_0) e^{j2\pi f_i(n_0)k - d_i(n_0)|k|} + \xi(n_0, k) \tag{5}$$

for  $k \in [-K + 1:K - 1]$ . In (5)  $\xi(n_0, k)$  is a white gaussian noise, which takes into account both background noise and model inaccuracies,  $C_i(n_0)$ ,  $d_i(n_0)$  and  $f_i(n_0)$  are the amplitude, damping factor and normalized frequency of the  $i$ th damped complex sinusoid. The parameters  $C_i(n_0)$ ,  $d_i(n_0)$  and  $f_i(n_0)$  are obtained from  $s_{x_a}(n_0, k)$  with  $k \in [0:K - 1]$ , using the least square method described in [20], which provides accurate estimation of exponentially damped sinusoidal signals in noise. The method combines linear backward prediction and singular value decomposition to separate signal components from noise. If we take the Fourier transform of the instantaneous ACF described in (5), ignoring the noise term, we get a denoised approximation of  $S(n, m)$ :

$$\tilde{S}(n, m) = \sum_{i=1}^q \tilde{S}_i(n, m), \tag{6}$$

where  $\tilde{S}_i(n, m)$  denotes the discrete Fourier transform of each damped sinusoid in (5). In Appendix, it is shown that

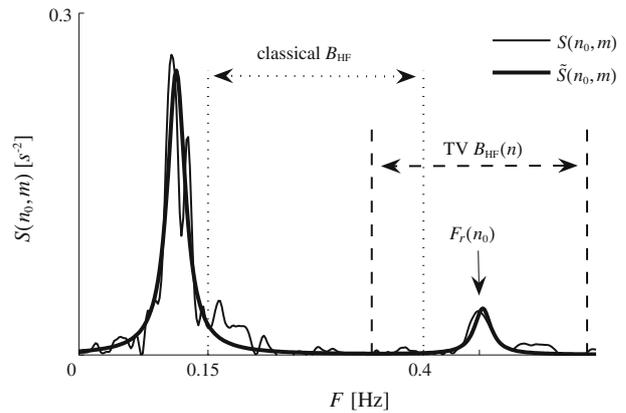


**Fig. 1** Continuous line The SPWVD,  $S(n_0, m)$ , computed at time  $n_0$ . Bold line The spectrum components,  $S_{LF}(n_0, m)$  and  $S_{HF}(n_0, m)$ , reconstructed using the parametric decomposition. Dashed bold line The interference term

the power associated to a signal component whose instantaneous ACF is a complex damped sinusoid can be analytically obtained from coefficients  $C_i(n)$ . The LF component is selected as the sinusoid with highest power in LF range, while the HF component is selected as the sinusoid whose central frequency lies in the HF range and is closer to the respiratory rate. The use of specific selection criteria for LF and HF components is particularly useful when the TF filtering is not sufficient to completely remove the interference terms. Indeed, in such a situation, the sinusoid which corresponds to the interference term, whose instantaneous central frequency lies in between the LF and the HF components [15], is automatically discarded from the instantaneous power and frequency estimation. This is shown in Fig. 1, where we observe that the parametric decomposition of a highly noisy instantaneous spectrum  $S(n_0, m)$  (in continuous line) allows to estimate LF and HF component (in bold line) and to discard the interference term (in dashed bold line).

### 2.3 Dynamic definition of HF band

It is generally accepted that the HF component of the HRV mainly reflects respiratory sinus arrhythmia [14] and that  $F_{HF}(n)$  is related to the respiratory frequency  $F_r(n)$ . The knowledge of  $F_r(n)$  can therefore be used to adjust the range of HF band and to improve the estimation of both LF and HF contributions. In traditional spectral analysis, the HF range is usually fixed [0.15, 0.4] Hz, but there are situations in which this range will lead to misestimation of the HF component, i.e. when respiratory rate decreases below 0.15 Hz (9 breaths per minute) or increases above 0.4 Hz (24 breaths per minute). In order to avoid these errors, the HF band is made time-varying and respiratory dependent [3]:



**Fig. 2** Continuous line The SPWVD,  $S(n_0, m)$ , computed at time  $n_0$ . Bold line The spectrum components,  $S_{LF}(n_0, m)$  and  $S_{HF}(n_0, m)$ , reconstructed using the parametric decomposition. Dotted vertical lines Traditional HF frequency band. Dashed vertical lines TV respiration-dependent HF band

$$B_{LF} : [0.04, 0.15] \text{ Hz} \tag{7}$$

$$B_{HF}(n) : F_r(n) \pm [-0.125, 0.125] \text{ Hz},$$

where  $B_{LF}$  and  $B_{HF}(n)$  identify the LF and HF bands, respectively. In such a way, the HF band is dynamically adjusted around  $F_r(n)$ , and thus a correct quantification of HF component is obtained even when the respiratory rate lies outside the traditional HF range. An example is shown in Fig. 2, where  $F_r = 0.465$  Hz. In this case, the use of a respiratory-dependent range (plotted in dashed vertical lines) allows to include the HF component, while the use of traditional HF range (dotted vertical lines) misses the HF component. When using the above rule, one has to pay attention to the fact that  $B_{LF}$  and  $B_{HF}(n)$  partially overlap when  $F_r(n) < 0.275$  Hz (16.5 breaths per minute). In this case, if in the spectrum there are two peaks, and at least one is in  $B_{LF}$  and the other in  $B_{HF}(n)$ , the one centered around the lower frequencies is considered the LF component and the one centered around the higher frequencies is considered the HF component. In those cases where  $F_r(n)$  is much lower, the overlap between the two ranges increases and just one peak is usually observed. In such a situation, the separation of LF and HF components is not straightforward due to non-linear interactions between sympathetic and parasympathetic modulations [5].

### 2.4 Experimental procedure

Seventy-five subjects without any formal musical education (age range: 18–35 years, mean age:  $24.5 \pm 3.2$  years, 36 female) participated in an experiment designed to characterize the effects of acoustic stimuli with different emotional valence. All subjects were right-handed with an handedness quotient  $> 90$  according to the ‘Edinburgh

Handedness Inventory 9' [28] and reported to have normal hearing. During the experiment, four conditions were employed:

1. *Pleasant condition (P)* Six excerpts of joyful instrumental dance-tunes from the last four centuries were used as pleasant stimuli (all major–minor tonal music, each had a duration of 90 s). The tempo of the excerpts, measured in beats per minute, was 73, 105, 117, 124, 128, 169, mean =  $119.33 \pm 31.37$ .
2. *Unpleasant condition (U)* Six unpleasant stimuli were electronically manipulated counterparts of six musical pieces from the last four centuries (all major–minor tonal music with rather minor scales and slow tempo). For any of these stimuli, a new soundfile was created in which the original excerpt was recorded simultaneously with two pitch-shifted versions (one being one semitone above and the other a tritone below the original pitch), and subsequently recorded backwards in order to introduce many dissonant structures. In order to match the metre of these stimuli with the metre of the original pleasant ones, series of Shepard tones [33] were overlaid over the manipulated (unpleasant) musical excerpts. The Shepard tone had a stationary pitch and a duration of 100 ms, and each Shepard tone was one semitone higher than the previous Shepard tone (i.e. tones were perceived as a rising chromatic scale). Shepard tones were used to guarantee that the frequency spectrum of the stimuli is comparable, e.g. between beginning and end of the stimuli. The time interval between Shepard tones was chosen to match the tempo of the pleasant excerpts.

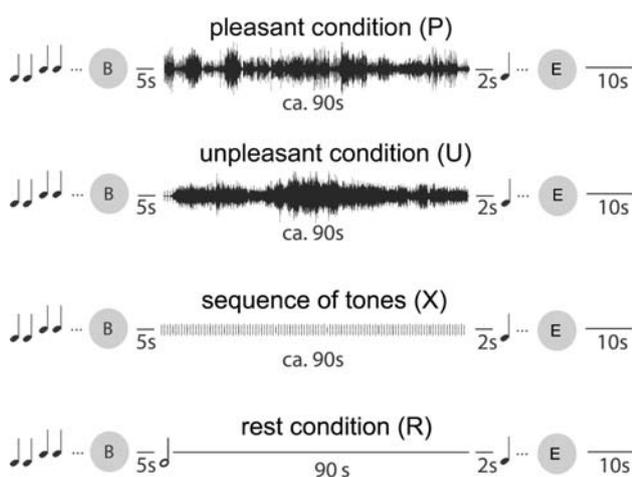
3. *Sequence of Shepard tones (X)* Shepard tones were presented separately, i.e. without music, as control for the unpleasant stimuli. The purpose was to assess whether Shepard tones alone provoked the same effect as the more complex and dissonant unpleasant excerpts. The particular structure of the sequence creates the auditory illusion of a tone that continually ascends or descends in pitch, yet which ultimately seems to get no higher or lower. The time interval between tones was chosen to match the tempo of pleasant and unpleasant excerpts and six sequences were used.
4. *Resting condition (R)* In addition to these three stimulus categories, there were also six resting intervals of 90 s duration in which no acoustic stimulus was presented.

All stimuli (pleasant, unpleasant and Shepard tones) were matched for volume (average root mean square power =  $-21.12 \pm 0.94$  dB) and were presented as illustrated in Fig. 3. Each stimulus began with a start-signal-tone (four short ascending sine wave tones) and ended with an end-signal-tone (a high, short single sine wave tone). For the resting condition, after the start-signal-tone, trials were indicated by a new 100 ms, 400 Hz sine wave tone. Stimuli were presented to every subject in the same pseudo-randomized sequence, designed so that each experimental condition followed all other conditions with equal probability.<sup>1</sup> After the end-signal-tone of each trial, participants had to indicate how pleasant or unpleasant they felt at the end of the trial by pressing response buttons. The rating task was followed by a 10-s pause until the next start-signal-tone appeared. Participants were instructed to listen carefully to the auditory stimuli with eyes closed and to tap the metre of the stimuli with their right index finger. This task was employed to control whether listeners paid attention not only to the pleasant but also to the unpleasant music and to the single tones sequence. No tapping was required during the resting condition.

## 2.5 Signal acquisition and preprocessing

Standard 12 lead electrocardiograms were measured using a 32 MREFA amplifier (Twente Medical Systems, Enschede, Netherlands) and digitized with a sampling rate of 1000 Hz. The recorded ECG signal was processed to derive the HRV signal and to estimate respiratory frequency. After detection of QRS complexes, the heart rate  $d_{HR}(n)$  was derived by integral pulse frequency modulation (IPFM) model, which also accounts for the presence of ectopic beats [25], and then evenly resampled at  $F_s = 4$  Hz,

<sup>1</sup> More in detail, the entire sequence was: X,U,P,X,U,R,X,P,R,U,X,P,U,X,R,P,U,R,P,X,R,U,P,R.



**Fig. 3** Experimental design—the trial begins (B) after a start-signal-tone. Musical pieces and silent period were followed by short signal tones that prompted participants to rate (E) their current emotional state. A 10-s pause was added before presentation of the next auditory cue

using spline interpolation. The HRV signal  $x(n)$  was then obtained by filtering  $d_{HR}(n)$  with a high-pass filter (0.03 Hz cut-off frequency). Using the methodology explained in Sect. 2.2, the HRV signal was processed to estimate the HRV spectral parameters, i.e. the instantaneous central frequencies  $F_{LF}(n)$  and  $F_{HF}(n)$ , powers  $P_{LF}(n)$  and  $P_{HF}(n)$  as well as the total instantaneous power,  $P_{tot}(n) = P_{LF}(n) + P_{HF}(n)$ , and a measure of the sympatho-vagal balance  $\alpha_{LF}(n) = \frac{P_{LF}(n)}{P_{tot}(n)}$  [24]. Finally,  $F_r(n)$ , necessary to make the HF band respiratory dependent as in (7), was indirectly estimated from the ECG-derived respiratory signal using the method presented in [2].

### 2.6 Statistical analysis

Statistical analysis was designed and performed to assess whether significant differences exist between the time-course of each physiological parameter in the four experimental conditions. Particular attention is paid to quantify differences on dynamic responses provoked by each musical stimulus, and thus the statistical differences are quantified on a sample-by-sample basis. The following notation is used: subjects are indexed with  $i \in [1:L]$ , conditions with  $E \in [P, U, X, R]$ , stimuli which belong to the same condition with  $j \in [1:6]$ . A general physiological parameter is indicated with  $\phi(n)$ . According to this notation,  $\phi_{i,E,j}(n)$  represents the time-course of a physiological parameter during the  $j$ th repetition of condition  $E$  for subject  $i$ . We first assume that, for each subject, the median time-course over the six repetitions of the same musical condition is representative of the individual autonomic response pattern to that stimulus. This is defined as:

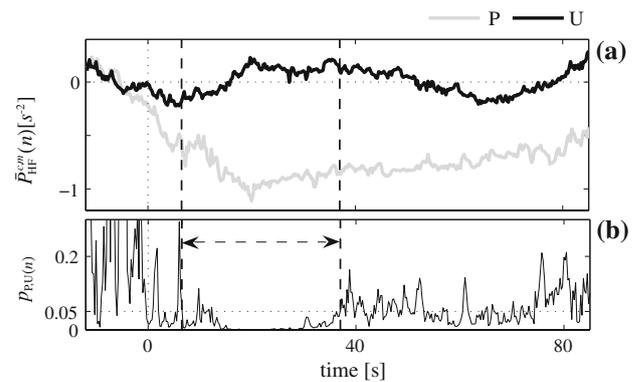
$$\phi_{i,E}^{c,m}(n) = \text{median} \left\{ \phi_{i,E,j}^c(n) \right\}_{j=1}^6 \quad (8)$$

where  $\phi_{i,E,j}^c(n)$  is the baseline-corrected parameter computed as:

$$\phi_{i,E,j}^c(n) = \phi_{i,E,j}(n) - \frac{1}{\tilde{T}} \sum_{n=-\tilde{T}}^{-1} \phi_{i,E,j}(n) \quad (9)$$

where  $\tilde{T} = (TF_s)$  and  $T = 12$  s is the duration of an interval preceding the onset of each  $j$ th trial. This particular normalization subtracts the parameter mean values (computed before the onset of each stimulus) from each parameter and it is used to highlight only the transient phenomena produced by the actual condition and not by the past history or by the different subject reference.

At every instant  $n_0$ , the value of  $\phi_{i,E}^{c,m}(n_0)$  for the subject  $i$  represents the  $i$ th realization of a statistical population  $\Omega_{E,n_0} = \{ \phi_{1,E}^{c,m}(n_0), \phi_{2,E}^{c,m}(n_0), \dots, \phi_{L,E}^{c,m}(n_0) \}$ , obtained by collecting, from all the subjects, the same parameter, at the same time  $n_0$ . By iteratively doing a pairwise comparison



**Fig. 4** **a** Mean trend of  $P_{HF}^{c,m}(n)$  during pleasant ( $P$ ) and unpleasant ( $U$ ) conditions. **b** Time-course of the  $p$ -value quantifying the differences between  $P_{HF,P}^{c,m}(n)$  and  $P_{HF,U}^{c,m}(n)$ . Vertical dashed lines mark the epoch (from about 6 to 38 s) in which response patterns are significantly different ( $p < 0.05$ )

of the four statistical populations  $\Omega_{P,n}$ ,  $\Omega_{U,n}$ ,  $\Omega_{X,n}$  and  $\Omega_{R,n}$  (i.e. six comparisons) and repeating the test for different  $n$ , the time-course of six  $p$ -values  $p_{E_k, E_l}(n), (E_k, E_l) \in [P, U, X, R]$  with  $E_k \neq E_l$ , is estimated. Given that the  $\Omega_{E,n}$  are not normally distributed, the Mann–Whitney test is used. The continuous estimation of  $p_{E_k, E_l}(n)$  allows to assess the time after which the autonomic response to the different conditions differs, and for how long these differences are significant.

An example of this procedure is shown in Fig. 4, where the mean time-course of  $P_{HF,i}^{c,m}(n)$ , denoted  $\bar{P}_{HF}^{c,m}(n)$ , during pleasant and unpleasant conditions are reported. The resulting time-course of the  $p$ -value  $p_{P,U}(n)$ , shown in Fig. 4b, allows to identify the time epochs in which physiological responses are different.

### 3 Results

Results were obtained considering  $L = 75$  participants and the TF analysis parameters  $(2N - 1)$ ,  $K$  and  $\gamma$  were set equal to 101 samples (25 s), 1024 samples and  $1/64$  samples<sup>-1</sup>, respectively, (see Sect. 2.1).

A representative HRV signal segment and its TF representations are plotted in Fig. 1. The SPWVD  $S(n, m)$  estimated as in (1) and the TF distribution  $\tilde{S}(n, m)$ , obtained through the parametric decomposition (6), are reported in Fig. 1b and c, respectively. The parametric decomposition allows to discard the interference terms which were still present in  $S(n, m)$  and it improves the localization in the TF plane of the LF and HF components. From this representation, it is possible to qualitatively characterize the trends related to every particular condition. Low frequency component modulation, centered around 0.1 Hz, increased during the short intervals in

between two conditions. As expected, the HF component followed the quick variation of the respiratory rate (dotted line) and its trend appears to be highly condition dependent: the instantaneous frequency of HF was higher in *P* and lower in *R*. Differences in the dynamic response to the musical stimuli were continuously quantified through the statistical analysis explained in Sect. 2.6. An example is shown in Fig. 4, where the induced response patterns in  $P_{\text{HF}}(n)$  during *P* and *U* are shown. During pleasant condition,  $P_{\text{HF}}^{c,m}(n)$  first decreased and, after about 20 s, it slowly increased toward baseline values, while during unpleasant condition, it first slightly decreased for about 6 s, then it increased and it maintained higher values from 20 to 50 s from the beginning of the excerpt. The *p*-value, reported in Fig. 4b, shows that the response patterns resulted significantly different in the temporal window 6–38 s after the onset of the stimuli (interval marked by vertical dashed lines). It is worth noting that, even if between *P* and *U* mean trends were separated, differences were significant only in the first part of the response. This shows the importance of the *p*-value tracking in the result evaluation.

The mean time course of the physiological parameters are shown in Fig. 6 and illustrated in Sect. 3.1. The results of the statistical analysis are summarized in Fig. 7. In this graphic, the response patterns of the physiological parameters to the experimental conditions are compared pairwise. Every column represent a physiological parameter  $\phi(n)$  and the column length represents the duration (% of stimuli duration) in which  $\phi_{E_k}(n) > \phi_{E_l}(n)$  for a significance level  $p < 0.05$  with  $E \in [P, U, X, R]$ . A negative value for the column length indicates inverse relationship (i.e.  $\phi_{E_k}(n) < \phi_{E_l}(n)$ ). For example, in the sixth column is reported that  $P_{\text{HF}}(n)$  was significantly lower during *P* than during *U*, and this difference was maintained for 71% of the duration of the stimulus. Notably, some stimuli provoked parameter changes which were different for the entire duration of the stimulus. The strength of these differences is quantified by the median *p*-values estimated in the intervals of statistical significance. Symbols \* and ° indicate that median  $p(n) < 0.005$  and median  $p(n) < 5 \cdot 10^{-5}$ , respectively.

### 3.1 Time course of physiological parameters

Significant differences in the autonomic response patterns to pleasant music, unpleasant stimuli, sequence of Shepard tones and resting condition were observed.

- $d_{\text{HR}}(n)$ : As shown in Fig. 6a, during the first 10 s, the heart rate,  $d_{\text{HR}}(n)$ , decreased for all the conditions, and after a transient of about 20 s it stabilized around values  $d_{\text{HR},R}(n) < d_{\text{HR},X}(n) \approx d_{\text{HR},U}(n) < d_{\text{HR},P}(n)$ . All

the differences between the conditions were significant (at least during more than 80% of their duration, as reported in Fig. 7), except between *U* and *X*. Between *P* and *U*, the significance was reached after about 16 s, while between *P* and *R* and between *U* and *R*, differences were significant from the first instants until the end of the conditions.

- $F_{\text{LF}}(n)$ : The mean trend of  $F_{\text{LF}}(n)$  (Fig. 6b) slightly increased from baseline for about 20–30 s, regardless of the kind of stimulus and no relevant difference was detected between the conditions.
- $F_{\text{HF}}(n)$ : This parameter and the related  $F_r(n)$  experienced the most marked differences among different musical stimuli (see Fig. 6c, d); their mean trend highly differed from a condition to another with little variability. The response of  $F_{\text{HF}}(n)$  was fast, and the difference between  $F_{\text{HF},P}(n)$  and  $F_{\text{HF},U}(n)$  was significant after only 8 s. After 10–15 s,  $F_{\text{HF}}(n)$  reached a stable state where  $F_{\text{HF},R}(n) < F_{\text{HF},X}(n) \approx F_{\text{HF},U}(n) < F_{\text{HF},P}(n)$ . Between *P* and *R* and between *U* and *R*, differences were significant during all the 90 s (median *p*-value  $< 10^{-10}$ ), mainly due to the fact that while acoustic stimuli provoked an increase in  $F_{\text{HF}}(n)$ , the silence of resting condition was observed to provoke the opposite effect.
- $P_{\text{LF}}(n)$ : After the beginning of the trials,  $P_{\text{LF}}(n)$  was observed to decrease for any condition for about 15–20 s, before reaching a quasi-stationary state (Fig. 6e). Its decrease, which was statistically significant from the first seconds of the trials onset ( $p < 0.001$ ), had approximately the same rate for every condition. After the transients, it was observed that  $P_{\text{LF},P}(n) < P_{\text{LF},U}(n) < P_{\text{LF},X}(n) \approx P_{\text{LF},R}(n)$ . Even if the mean trend during *P* was lower than during *U*, their great variability made this difference not significant.
- $P_{\text{HF}}(n)$ : Fig. 6f shows that  $P_{\text{HF},X}(n)$  and  $P_{\text{HF},U}(n)$  fluctuated around baseline values, while  $P_{\text{HF},R}(n)$  and  $P_{\text{HF},P}(n)$  evidenced opposite trends, by increasing ( $P_{\text{HF},R}(n)$ ) and decreasing ( $P_{\text{HF},P}(n)$ ) in the first 20 s from the beginning of the stimulus. Almost immediately, we may observe that  $P_{\text{HF},P}(n) < P_{\text{HF},X}(n) \approx P_{\text{HF},U}(n) < P_{\text{HF},R}(n)$ . The differences between *P* and *R* and between *U* and *R* were almost instantaneously significant, while between *P* and *U* significance was reached after about 6 s and it was maintained for about 30 s (see also Fig. 4).
- $P_{\text{tot}}(n)$ : Total power was also estimated (Fig. 6g). During all the conditions,  $P_{\text{tot}}(n)$  tended to decrease from baseline values. During pleasant condition, both  $P_{\text{LF}}(n)$  and  $P_{\text{HF}}(n)$  decreased and this is reflected on the strong reduction of  $P_{\text{tot},P}(n)$ , while the increasing of  $P_{\text{HF}}(n)$  during *R* made  $P_{\text{tot},R}(n)$  higher than during all the other conditions. The significance of the differences

observed in  $P_{\text{tot}}(n)$  was larger than for  $P_{\text{LF}}(n)$  but smaller than for  $P_{\text{HF}}(n)$ .

- $\alpha_{\text{LF}}(n)$ : In Fig. 6h, the mean baseline-corrected time-course of the sympatho-vagal balance  $\alpha_{\text{LF}}(n)$  is shown. It was observed that  $\alpha_{\text{LF}}(n)$  decreased during about 15 s, regardless of the specific condition and no significant differences were detected between its response patterns.

## 4 Discussion

In this study, we presented a methodology to continuously quantify music-induced changes in HRV. This approach consists of three steps. First, the SPWVD is used to obtain a TF representation of the HRV signal. Second, the time course of physiological parameters is estimated through a parametric decomposition which makes use of dynamical adjustment of HF range based on respiratory frequency. Third, we performed a statistical population analysis to continuously assess whether the dynamics involved in the autonomic response to different musical stimuli are significant or not.

### 4.1 The time–frequency framework

Response patterns of the ANS are dynamic, and there is the need for an appropriate methodology capable of tracking these changes. Compared to other possibilities, such as Continuous Wavelet Transform, Spectrogram and time-varying autoregressive methods [21], the SPWVD has the advantage of an excellent joint TF resolution [15]. In addition, the SPWVD was used, in a recent study [7], to estimate the changes in the instantaneous power spectrum of the respiratory signal induced by the listening to classical music excerpts.

In the TF analysis of HRV based on the WVD, the simultaneous presence of time-varying LF and HF components gives rise to interference terms which should be reduced using an appropriate kernel [16]. In simulation studies, the smoothing function used in (2) was shown to provide a reliable estimation of LF and HF dynamics, also in high non-stationary conditions [29]. Other kinds of kernels, capable of achieving a wide range of different shapes [9] or of automatically adjusting to the signal structure [4], have been proposed in the literature. However, a fixed kernel is used in this study to maintain the same TF resolution for all subjects.

The parametric decomposition used to estimate the time-course of the spectral parameters allows to separate the information which is considered inherent to the signal from both noise and interference terms. This methodology was first proposed in [23] and here it was combined with

the respiratory rate to improve the tracking of the HF component [1].

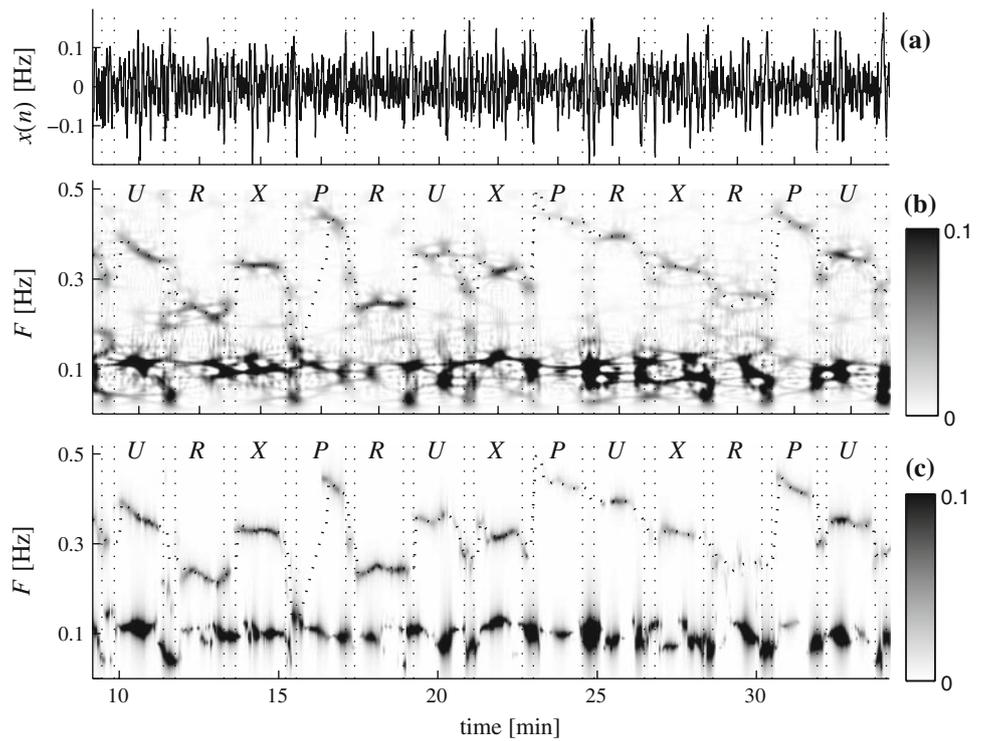
### 4.2 Dynamic HF band

The use of time-varying bands for a reliable estimation of HRV spectral components has been proposed in different applications. Most of these techniques make use of respiratory information to fix spectral boundaries [3, 18]. The inclusion of  $F_r(n)$  in the definition of the HF range is important especially in TF analysis. In our results, we observed marked variations in  $F_r(n)$  due to the musical stimuli, which bring  $F_r(n)$  to frequencies higher than 0.4 Hz (see  $P$  in Fig. 5). Thus, a dynamic adjustment is needed to correctly estimate these changes. The respiratory rate was obtained from the ECG using [2]. This method exploits the oscillatory pattern of the rotation angles of the heart's electrical axis as induced by respiration, and it is particularly suitable when dealing with highly non-stationary and noisy ECG recordings. In order to corroborate the use of [2] in this kind of studies, we compared our estimation of  $F_r(n)$  with the respiratory frequency obtained from a direct measurement of respiration (performed through respiratory belt), only available in 58 subjects. The mean error was low ( $0.0001 \pm 0.0045$  Hz, equal to  $-0.4 \pm 2.1\%$ ) in agreement with results reported in [2] for simulated signals. In our study,  $F_r(n)$  was always available, but in those cases where respiratory frequency is not available (neither from respiratory signal nor from the ECG),  $B_{\text{HF}}(n)$  may be automatically adjusted to the instantaneous spectral properties of the HRV signal [13].

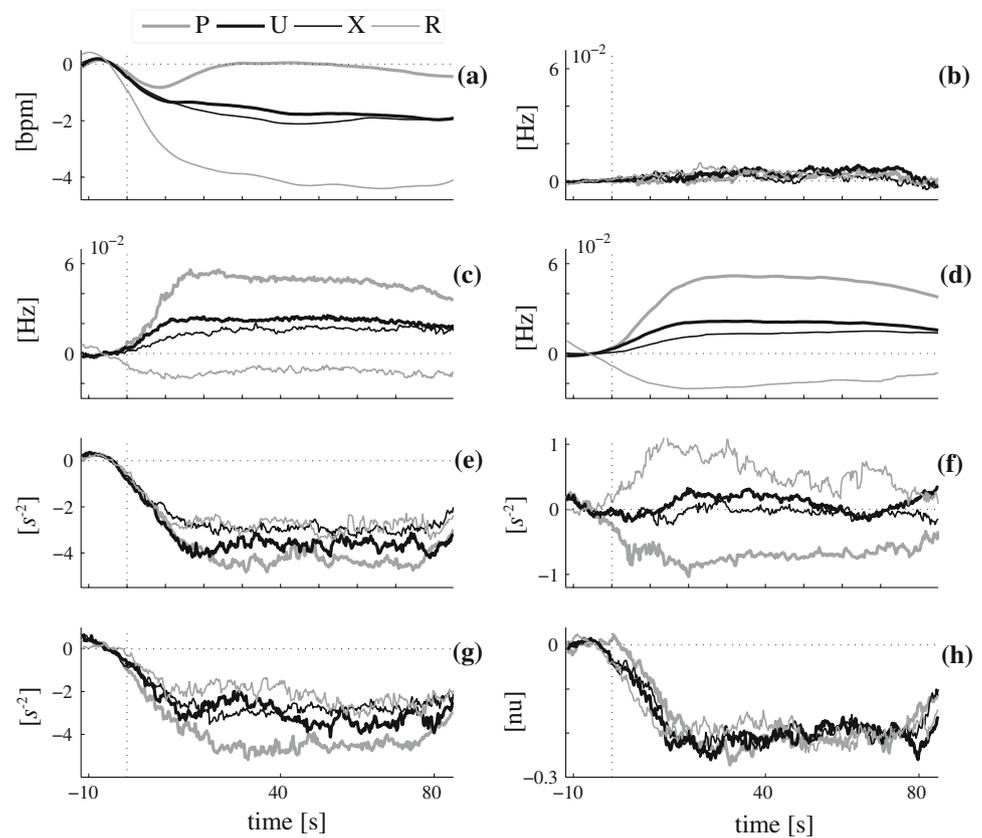
### 4.3 Physiological parameter changes during music stimuli

The purpose of the experimental procedure was to characterize the effect of stimuli with different emotional valence on the heart rate modulation. There was agreement between the assumed valence of the stimuli and the emotional state as reported by the participants. All conditions were rated significantly differently, being  $P$  the more pleasant and  $U$  the more unpleasant. The results showed significant differences in the time-course of HRV parameters during the different conditions. The mean trend of most of these parameters (see Fig. 6) was characterized by two different phases: a first early short phase of about 15–20 s during which parameters changed quickly and a second longer phase during which parameters approximately maintained the same values or varied slowly. These considerations should be taken into account in the design of analogous experiments and should warn against the use of time-invariant methods in this kind of studies. Given that

**Fig. 5** **a** Example of the HRV signal  $x(n)$ , **b**  $S(n, m)$ : smoothed pseudo Wigner-Ville distribution, **c**  $\tilde{S}(n, m)$ : TF distribution reconstructed using the parametric decomposition illustrated in Sect. 2.2. Dotted line The respiratory rate.  $P$  pleasant,  $U$  unpleasant,  $X$  sequence of Shepard tones,  $R$  rest

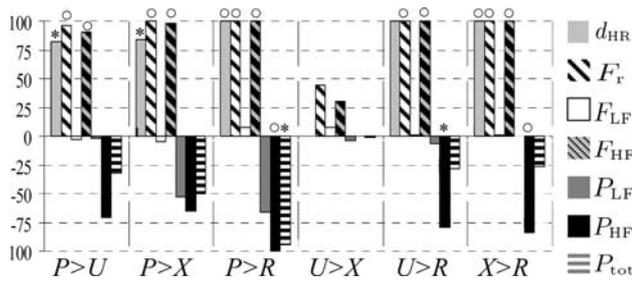


**Fig. 6** Mean trend, averaged over the entire population, of the changes provoked by pleasant ( $P$ ), unpleasant ( $U$ ), sequence of tones ( $X$ ) and rest ( $R$ ) conditions. Physiological parameters from (a) to (h), as detailed in Sect. 2.5, are: **a** the mean heart rate  $\bar{d}_{HR}^{c,m}(n)$ ; **b, c** the instantaneous frequency of LF and HF,  $\bar{F}_{LF}^{c,m}(n)$  and  $\bar{F}_{HF}^{c,m}(n)$ ; **d** the respiratory rate  $\bar{F}_r^{c,m}(n)$ ; **e, f** the instantaneous power of LF and HF,  $\bar{P}_{LF}^{c,m}(n)$  and  $\bar{P}_{HF}^{c,m}(n)$ ; **g** instantaneous total power  $\bar{P}_{tot}^{c,m}(n)$ ; **h** the sympatho-vagal balance  $\bar{\alpha}_{LF}^{c,m}(n)$ . Results are detailed in Sect. 3.1



the autonomic response to music-induced emotions was observed to be transient and not instantaneous, too short stimuli could be insufficient to elicit significant differences.

In addition, a pause between the end of a trial and the beginning of the next one is strongly recommended to assure the restoration of appropriate baseline values.



**Fig. 7** Results of the pairwise comparison between the conditions (see Sect. 2.6). Bar length represents the time duration (% of stimuli duration) in which  $P > U$ ,  $P > X$ ,  $P > R$ , etc. ( $p(n) < 0.05$ ). Negative values denote inverse relationship. \* and °: median  $p$ -values  $< 0.005$  and  $< 5 \cdot 10^{-5}$ , respectively

The largest differences in autonomic responses were observed between pleasant and rest conditions: significantly different patterns were observed in all parameters but  $F_{LF}(n)$  (see column  $P > R$  in Fig. 7). More importantly, even the effects induced by pleasant and unpleasant excerpts, that is by two acoustic stimuli comparable in loudness and tempo, could be discriminated by tracking  $d_{HR}(n)$ ,  $F_r(n)$ ,  $F_{HF}(n)$ ,  $P_{HF}(n)$  and, partially, by  $P_{tot}(n)$ . This difference is likely to be due to the different emotional valence of the two types of stimuli (pleasant vs. unpleasant). In contrast, unpleasant stimuli and the sequence of Shepard tones were observed to elicit rather comparable effects, and no physiological parameter could clearly discriminate between them. This similarity may be due to the absence of harmonic and melodic structures in both kinds of stimuli, or due to a predominant effect of the Shepard tones, present in both  $U$  and  $X$ , on the dissonant structures of the unpleasant excerpts.

Respiratory rate and  $F_{HF}(n)$  showed the greatest sensitivity to music-induced stimuli: any kind of acoustic stimulus provoked almost instantaneous differences with respect to resting condition, and differences between pleasant and unpleasant conditions were reached in only 4–8 s after the stimulus onset. A fast response in HF component is in agreement with previous studies on ANS stimulation, such as response to tilt test [26].

While the respiratory rate has been shown to be strongly affected by music [6, 7, 11, 12, 27], the role of heart rate is more controversial. In several works,  $d_{HR}(n)$  was observed to be significantly affected by acoustic stimuli with different valence [6, 27, 32], while in others this influence was not observed [11, 17]. Here, unpleasant stimuli were observed to induce a stronger decrease in  $d_{HR}(n)$  than pleasant music (see Fig. 6a). The triphasic waveform described in [8] and in [32] was also observed, even if here it appeared much smoother and longer: during pleasant condition  $d_{HR}(n)$  initially decreased for about 10 s, then it slightly increased for about 10 s, before decreasing again.

Heart rate variability was already proposed as a critical index for the assessment of the music effect in [17], where the physiological response to exciting and sedative music was compared in a stationary context. That study reported that exciting music decreased the activation of the parasympathetic nervous system. Here, both LF and HF modulations, assessed by  $P_{LF}(n)$  and  $P_{HF}(n)$ , were observed to be higher during resting condition than during acoustic stimuli listening, regardless of the internal structure of the excerpts. The power content in HF was higher during unpleasant than during pleasant condition, suggesting that the listening of pleasant music provoked a reduction of parasympathetic modulation. The pleasant excerpts elicited the opposite effect than resting condition, during which the vagal activity increased. The simultaneous decrease of the heart and respiratory rates and the increase of the power of HF modulation highlight the strong relaxing effect of silence, as already observed in [6].

As shown in Fig. 6e, when a condition began, the activation in LF band dramatically decreased within about 15 s with respect to baseline values. The difference between  $P_{LF}(n)$  values before and after each start-signal-tone are significant already after the first seconds of the trials. This pattern may be related to the expectation of a new coming stimulus (participants did not receive any previous information about the excerpts) and to the task of evaluation assigned to the participants, which may increase the sympathetic activity.

The decrease observed in the sympatho-vagal balance, which passed from a first LF predominance ( $\alpha_{LF}(n) \approx 0.6$ ) to an equilibrium in which the modulation in HF was slightly higher ( $\alpha_{LF}(n) \approx 0.4$ ), took about 20 s. This change was common to all the different conditions, and no statistical differences were observed, among the subjects, in  $\alpha_{LF}(n)$ , suggesting that the sympatho-vagal balance is higher during attention.

In this study, we proposed a comprehensive methodology for the assessment of physiological parameter changes related to musical stimuli and we focused our attention to HRV modification. Nevertheless, the listening to music provokes changes also in other signals related to the ANS [6, 7, 11, 12, 27, 32], and further studies should be therefore considered to achieve a comprehensive characterization of music-induced effects on the ANS. Additionally, the statistical analysis performed here may also be applied in bivariate TF analysis [19, 30] to assess whether different conditions provoke different dynamic interactions in cardiovascular or cardiorespiratory control, or to explore the relationship between physiological rhythm and musical profile as in [7]. The experimental results revealed the transient nature of music-related patterns and suggest the need for further studies on the dynamic relationship

between musical and autonomic features to improve the potential use of music in therapeutic applications.

**Acknowledgements** This work was supported in part by the Ministerio de Ciencia y Tecnología, FEDER, under Project TEC2007-68076-C02-02/TCM and by the Centro de Investigación Biomédica en Red (CIBER) de Bioingeniería, Biomateriales y Nanomedicina, Feder, through Instituto de Salud Carlos III (ISCIII).

## Appendix

Consider a signal component  $x(t)$ , whose ACF is the damped complex sinusoid  $y(\tau)$ :

$$y(\tau) = Ce^{-d|\tau|+j2\pi v_0\tau} \quad (10)$$

with  $d > 0$ ,  $\tau$  and  $C \in \mathbb{R}$ . The Fourier transform of  $y(\tau)$  is equal to:

$$Y(v) = \int_{-\infty}^{\infty} Ce^{-d|\tau|+j2\pi(v_0-v)\tau} d\tau = \frac{2Cd}{4\pi^2(v-v_0)^2 + d^2}. \quad (11)$$

Note that  $y(\tau)$  is an hermitian function and  $Y(v)$  is real. The function  $Y(v)$  is the power spectral density of  $x(t)$  and has a Lorentian shape with a peak centered on frequency  $v_0$ . The power of the signal component  $x(t)$ ,  $P_x$ , is then closely related to coefficient  $C$  and it can be analytically obtained as the total area of  $Y(v)$ :

$$\begin{aligned} P_x &= \int_{-\infty}^{\infty} Y(v)dv = \int_{-\infty}^{\infty} \left[ \frac{2Cd}{4\pi^2(v-v_0)^2 + d^2} \right] dv \\ &= \frac{2Cd}{4\pi^2} \int_{-\infty}^{\infty} \left[ \frac{1}{(v-v_0)^2 + \left(\frac{d}{2\pi}\right)^2} \right] dv \\ &= \frac{2Cd}{4\pi^2} \frac{2\pi}{d} \left[ \arctan\left(\frac{(v-v_0)2\pi}{d}\right) \right]_{-\infty}^{\infty} = C \end{aligned} \quad (12)$$

## References

- Bailón R, Mainardi LT, Laguna P (2006) Time–frequency analysis of heart rate variability during stress testing using a priori information of respiratory frequency. In: Proceedings of computers in cardiology, pp 169–172
- Bailón R, Sörnmo L, Laguna P (2006) A robust method for ECG-based estimation of the respiratory frequency during stress testing. *IEEE Trans Biomed Eng* 53(7):1273–1285
- Bailón R, Laguna P, Mainardi L, Sörnmo L (2007) Analysis of heart rate variability using time-varying frequency bands based on respiratory frequency. In: Proceedings of the 29th international conference of the IEEE engineering in medicine and biology society. IEEE-EMBS Society, Lyon, pp 6674–6677
- Baraniuk RG, Jones DL (1993) A signal-dependent time–frequency representation: optimal kernel design. *IEEE Trans Signal Process* 41(4):1589–1602
- Baselli G, Porta A, Ferrari G (1995) Models for the analysis of cardiovascular variability signals. In: Malik M, Camm AJ (eds) Heart rate variability. Futura Publishing Company, Armonk, pp 135–145
- Bernardi L, Porta C, Sleight P (2006) Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart* 92(4):445–452
- Bernardi L, Porta C, Casucci G, Balsamo R, Bernardi NF, Fogari R, Sleight P (2009) Dynamic interactions between musical, cardiovascular, and cerebral rhythms in humans. *Circulation* 119(25):3171–3180
- Bradley MM, Lang PJ (2000) Affective reactions to acoustic stimuli. *Psychophysiology* 37(2):204–215
- Costa A, Boudreau-Bartels G (1995) Design of time–frequency representations using a multiform, tiltable exponential kernel. *IEEE Trans Signal Process* 43:2283–2301
- Flandrin P (1999) Time–frequency/time-scale analysis. Academic Press, San Diego
- Gomez P, Danuser B (2004) Affective and physiological responses to environmental noises and music. *Int J Psychophysiol* 53(2):91–103
- Gomez P, Danuser B (2007) Relationships between musical structure and psychophysiological measures of emotion. *Emotion* 7(2):377–387
- Goren Y, Davrath LR, Pinhas I, Toledo E, Akselrod S (2006) Individual time-dependent spectral boundaries for improved accuracy in time–frequency analysis of heart rate variability. *IEEE Trans Biomed Eng* 53(1):35–42
- Grossman P, Taylor EW (2007) Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions. *Biol Psychol* 74(2):263–285
- Hlawatsch F, Boudreaux-Bartels GF (1992) Linear and quadratic time–frequency signal representations. *IEEE Signal Process Mag* 9(2):21–67
- Hlawatsch F, Flandrin P (1997) The interference structure of the Wigner distribution and related time–frequency signal representations. In: The Wigner Distribution—theory and applications in signal processing. Elsevier, Amsterdam, pp 59–113
- Iwanaga M, Kobayashi A, Kawasaki C (2005) Heart rate variability with repetitive exposure to music. *Biol Psychol* 70(1):61–66
- Jasson S, Médigue C, Maison-Blanche P, Montano N, Meyer L, Vermeiren C, Mansier P, Coumel P, Malliani A, Swynghedauw B (1997) Instant power spectrum analysis of heart rate variability during orthostatic tilt using a time–frequency-domain method. *Circulation* 96(10):3521–3526
- Keissar K, Davrath LR, Akselrod S (2009) Coherence analysis between respiration and heart rate variability using continuous wavelet transform. *Philos Trans R Soc A* 367(1892):1393–1406
- Kumaresan R, Tufts D (1982) Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Trans Acoust Speech Signal Process* 30(6):833–840
- Mainardi LT (2009) On the quantification of heart rate variability spectral parameters using time–frequency and time-varying methods. *Philos Trans Ser A* 367(1887):255–275
- Mainardi L, Bianchi A, Cerutti S (2002) time–frequency and time-varying analysis for assessing the dynamic responses of cardiovascular control. *Critl Rev Biomed Eng* 30(1–2):181–223
- Mainardi L, Montano N, Cerutti S (2004) Automatic decomposition of Wigner distribution and its application to heart rate variability. *Methods Inf Med* 43:17–21
- Malik M, Bigger J, Camm A, Kleiger R, Malliani A, Moss A, Schwartz P (1996) Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Eur Heart J* 17(3):354

25. Mateo J, Laguna P (2003) Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Trans Biomed Eng* 50:334–343
26. Novak P, Novak V (1993) Time/frequency mapping of the heart rate, blood pressure and respiratory signals. *Med Biol Eng Comput* 31(2):103–110
27. Nyklíček I, Thayer J, Van Doornen L (1997) Cardiorespiratory differentiation of musically-induced emotions. *J Psychophysiol* 11(4):304–321
28. Oldfield RC (1971) The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9(1):97–113
29. Orini M, Bailón R, Laguna P, Mainardi LT (2007) Modeling and estimation of time-varying heart rate variability during stress test by parametric and non parametric analysis. In: *Proceedings of computers in cardiology*, pp 29–32
30. Orini M, Bailón R, Mainardi L, Mincholé A, Laguna P (2009) Continuous quantification of spectral coherence using quadratic time–frequency distributions: error analysis and application. In: *International conference on computers in cardiology*
31. Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim C, Suri J (2006) Heart rate variability: a review. *Med Biol Eng Comput* 44(12):1031–1051
32. Sammler D, Grigutsch M, Fritz T, Koelsch S (2007) Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* 44(2):293–304
33. Shepard RN (1964) Circularity in judgments of relative pitch. *J Acoust Soc Am* 36(12):2346–2353