# Gaussian Processes for Music Audio Modelling and Content Analysis

Alvarado Duran, PA; STOWELL, DF; IEEE International Workshop on Machine Learning for Signal Processing (MLSP)

http://arxiv.org/abs/1606.01039

# GAUSSIAN PROCESSES FOR MUSIC AUDIO MODELLING AND CONTENT ANALYSIS

*Pablo A. Alvarado, Dan Stowell*

Queen Mary University of London

## ABSTRACT

Real music signals are highly variable, yet they have strong statistical structure. Prior information about the underlying physical mechanisms by which sounds are generated and rules by which complex sound structure is constructed (notes, chords, a complete musical score), can be naturally unified using Bayesian modelling techniques. Typically algorithms for Automatic Music Transcription independently carry out individual tasks such as multiple-F0 detection and beat tracking. The challenge remains to perform joint estimation of all parameters. We present a Bayesian approach for modelling music audio and content analysis. The proposed methodology based on Gaussian processes seeks joint estimation of multiple music concepts by incorporating into the kernel prior information about non-stationary behaviour, dynamics, and rich spectral content present in the modelled music signal. We illustrate the benefits of this approach via two tasks: pitch estimation and inferring missing segments in a polyphonic audio recording.

***Index Terms***— Gaussian processes, kernel design, music signals, content analysis, audio restoration.

## 1. INTRODUCTION

In music information research, the aim of audio content analysis is to estimate musical concepts which are present but hidden in the audio data [1]. With this purpose, different signal processing techniques are applied to music signals for extracting useful information and descriptors related to the musical concepts. Here, musical concepts refers to parameters related to written music, such as pitch, melody, chords, onset, beat, tempo and rhythm. Then, perhaps the most general application is one which involves the prediction of several musical dimensions, that of recovering the score of a music track given only the audio signal [2]. This is known as automatic music transcription (AMT) [3].

AMT refers to extraction of a human readable and interpretable description from a recording of a music performance. We refer to polyphonic AMT in cases where more than a single musical pitch plays at a given time instant. The general task of interest is to infer automatically a musical notation, such as the traditional western music notation, listing the pitch values of notes, corresponding timestamps and other expressive information in a given audio signal of a performance [4]. Transcribing polyphonic music is a nontrivial task, especially in its more unconstrained form when the task is performed on an arbitrary acoustical input, and music transcription remains a very challenging problem [5].

Real music signals are highly variable, but nevertheless they have strong statistical structure. Prior information about the underlying structures, such as knowledge of the physical mechanisms by which sounds are generated, and knowledge about the rules by which complex sound structure is compiled (notes, chords, a complete musical score), can be naturally unified using Bayesian hierarchical modelling techniques. This allows the formulation of highly structured probabilistic models [4]. On the other hand, typically, algorithms for AMT are developed independently to carry out individual tasks such as multiple-F0 detection, beat tracking and instrument recognition. The challenge remains to combine these algorithms, to perform joint estimation of all parameters [3].

We present the design, implementation, and results of experiments of an alternative Bayesian approach for audio content analysis on monophonic, and polyphonic music signals with the possibility of being used for AMT. We use Gaussian process (GP) models for jointly uncovering music concepts from audio, by introducing a direct connection between the music concepts and the model hyper-parameters. The proposed methodology allows to incorporate in the model prior information about physical or mechanistic behaviour, nonstationarity, time dynamics (local periodicity, and non constant amplitude envelope), spectral harmonic content, and musical structure, latent in the modelled music signal. Specifically in the context of music informatics, we present kernels that embody a probabilistic model of music notes as time-limited harmonic signals with onsets and offsets. The presented approach can describe polyphonic signals, by encouraging partial or complete overlapping between the latent processes that represent each sound event or music note. A comparison with related work is provided in section 3.4. We illustrate the benefits of this approach via two tasks: pitch estimation in monophonic music and inferring missing segments in a polyphonic audio recording.

P.A.A thanks to Colciencias for funding.

978-1-5090-0746-2/16/$31.00 ©2016 IEEE

## 2. GP REGRESSION FOR MUSIC SIGNALS

GP-based machine learning is a powerful Bayesian paradigm for nonparametric nonlinear regression and classification [6]. GPs can be defined as distributions over functions such that any finite number of function evaluations $\mathbf{f} = [f(t_1), \cdots, f(t_N)]$, have a jointly normal distribution [7]. A GP is completely specified by its mean function $\mu(t) = \mathbb{E}[f(t)]$ (in this work it is assumed to be $\mu(t) = 0$), and its kernel or covariance function

$$k(t, t') = \mathbb{E}\left[(f(t) - \mu(t))(f(t') - \mu(t'))\right], \qquad (1)$$

where $k(t, t')$ has hyper-parameters $\boldsymbol{\theta}$. We write the GP as $f(t) \sim \mathcal{GP}(\mu(t), k(t, t'))$. The regression problem concerns the prediction of a continuous quantity [7], here a function $f(t)$, given a data set $\mathcal{D} = \{(t_i, y_i)\}_{i=1}^{N}$, where the audio samples $y_i$ are assumed as noisy measurements of $f(t)$ at typically regularly-spaced time instants $t_i$ (though GP regression framework allows for irregular sampling or missing data), i.e. $y_i = f(t_i) + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma^2_{\text{noise}})$. In GP regression for mono channel audio signals, instead of estimating parameters $\boldsymbol{\eta}$ of fixed-form functions $f(t, \boldsymbol{\eta}) : \mathbb{R} \mapsto \mathbb{R}$ where the time input variable $t \in \mathbb{R}$, we model the whole function $f(t)$ as a GP. That is, instead of putting a prior over the function parameters $\boldsymbol{\eta}$, we introduce a prior over the function $f(t)$ itself [8, 9].

The underlying idea in GP regression is that the kernel introduces dependencies between function $f(t)$ values at different inputs. Thus, the function values at the observed points give information also of the unobserved points [6]. The structure of the kernel (1) captures high-level properties of the unknown function $f(t)$, which in turn determines how the model generalizes or extrapolates to new test time instants [10]. In this way, prior knowledge about proprieties of music signals can be introduced by choosing a proper kernel that reflects those characteristics. In section 2.2 we study in more detail the design of kernels.

### 2.1. Model definition

Under a non-parametric Bayesian regression approach using GPs we are interested in calculating the posterior distribution over a stochastic function evaluated at test points $\mathbf{t}_*$, that is, the joint distribution of the vector $\mathbf{f}$ observed only via noisy measurements $\mathbf{y}$, then

$$p(\mathbf{f}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f}) \times p(\mathbf{f}|\boldsymbol{\theta})}{p(\mathbf{y})}, \qquad (2)$$

where $p(\mathbf{y}|\mathbf{f})$ corresponds to the likelihood, $p(\mathbf{f}|\boldsymbol{\theta})$ to the prior, $\boldsymbol{\theta}$ are the model hyper-parameters (prior parameters), $p(\mathbf{y})$ is the evidence or marginal-likelihood, and $p(\mathbf{f}|\mathbf{y})$ is the posterior or conditional predictive distribution. Assuming that conditioned on $f(t_i)$ the signal observations $y_i$ are i.i.d. (independent and identically distributed), the joint probability distribution of all the observations $\mathbf{y}$ follows a Gaussian distribution corresponding to $p(\mathbf{y}|\mathbf{f}, \sigma^2_{\text{noise}}) = \mathcal{N}(\mathbf{y}|\mathbf{f}, \sigma^2_{\text{noise}}\mathbf{I}_N)$, where $f_i = f(t_i)$ and $\mathbf{I}_N$ is an identity matrix of size $N$. The prior $p(\mathbf{f}|\boldsymbol{\theta})$ is obtained using the definition of GPs introduced at the beginning of this section. Given a finite set of corrupted observations $\mathbf{y}$, then the finite set of GP function evaluation values $\mathbf{f}$ follows a normal marginal distribution $p(\mathbf{f}|\boldsymbol{\theta})$ conditioned on the hyper-parameters $\boldsymbol{\theta}$, whose mean is zero and whose covariance is defined by a Gram matrix $\mathbf{K}_f$, this is $p(\mathbf{f}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}_f)$, where the covariance matrix is calculated using (1), i.e. $[\mathbf{K}_f]_{i,j} = k(t_i, t_j)$ [11]. The *marginal-likelihood* (or evidence) $p(\mathbf{y})$ mentioned before in (2) is the integral of the likelihood times the prior [7]

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\boldsymbol{\theta})\mathrm{d}\mathbf{f}. \qquad (3)$$

Since the likelihood $p(\mathbf{y}|\mathbf{f})$ and the prior $p(\mathbf{f}|\boldsymbol{\theta})$ are multivariate Gaussian distributions, we can directly calculate the integral in (3). Using the properties of the normal distribution [11] for marginal and conditional normal distributions, we obtain $p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_y)$, where the values in the matrix $\mathbf{K}_y = \mathbf{K}_f + \sigma^2_{\text{noise}}\mathbf{I}$ depend on the hyper-parameters $\boldsymbol{\theta}$ (we have included $\sigma^2_{\text{noise}}$ in the hyper-parameters vector). The reason it is called the marginal likelihood, rather than just likelihood, is because we have marginalized out the latent Gaussian vector $\mathbf{f}$ [12]. The computation of the posterior distribution of the GP conditioned on the set of measurements $\mathbf{y}$ and estimation of the parameters $\boldsymbol{\theta}$ of the covariance function of the process correspond to learning in this non-parametric model [6]. Using the properties of Gaussian distribution [11, 7], the posterior has the form $p(\mathbf{f}|\mathbf{y}) = \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}_{\text{pos}}, \mathbf{K}_{\text{pos}})$, where the posterior mean is $\boldsymbol{\mu}_{\text{pos}} = \mathbf{K}_f\mathbf{K}_y^{-1}\mathbf{f}$, and the posterior covariance matrix is $\mathbf{K}_{\text{pos}} = \mathbf{K}_f - \mathbf{K}_f^{\top}\mathbf{K}_y\mathbf{K}_f$.

### 2.2. Kernel design

Some of the broad properties of audio signals are non-stationarity, rich spectral content, dynamics (locally periodic, non constant amplitude envelope), mechanistic patterns, and music structure. Therefore we seek covariance functions that can describe or reflect these properties. One powerful technique for constructing new kernels is to build them out of simpler kernels as building blocks [13, 11]. We use some properties showed in [11] for building non-stationary covariance functions. To construct non-stationary kernels we combine basic stationary covariance functions. We use *change-windows* in order to be able to model notes or sound events which are not continuously active but have a beginning and an ending in the music signal. As in [10] we define a change-window by multiplying two sigmoid functions. The parameters of the change-windows are directly related with the location, onset and offset of the sound events. In the present work we will use manually-specified onset/offset locations. We assume the complete process $f(t)$ is a linear combination of $M$ random process, representing each one a note or sound event. In this way

$$f(t) = \sum_{m=1}^{M} \phi_m(t) f_m(t), \tag{4}$$

where each GP $[f_1, f_2, \cdots, f_M]$ is independent with respect to each other. It is important to highlight that $M$ is directly related with the number of notes or sound events in the signal. On the other hand, $\phi_m(t)$ are the respectively change-windows that allow a specific GP $f_m(t)$ to appear or vanish in certain parts of the input space (time). In this sense the proposed approach can handle polyphonic signals, by encouraging partial or complete overlapping between change-windows. It can be shown the general expression for the covariance function $k_f(t, t')$ is given by

$$k_f(t, t') = \sum_{m=1}^{M} \phi_m(t) k_m(t, t') \phi_m(t'). \tag{5}$$

We see that the overall process has a kernel consisting of a linear combination of the corresponding covariance functions of every subprocess. We assume each GP $f_m(t)$ in (4) is stationary. A random process is stationary (wide sense stationary WSS) if its mean is constant, and its kernel is a covariance function of $\tau = t - t'$, then we can write $k(t, t') = k(\tau)$ [14, 7]. It can be shown that the *spectral density* or *power spectrum* $S(s)$ of a WSS process corresponds to the Fourier transform (FT) of the covariance function, that is

$$S(s) = \int_{-\infty}^{\infty} k(\tau) e^{-js\tau} \mathrm{d}\tau, \tag{6}$$

thus

$$k(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(s) e^{js\tau} \mathrm{d}s. \tag{7}$$

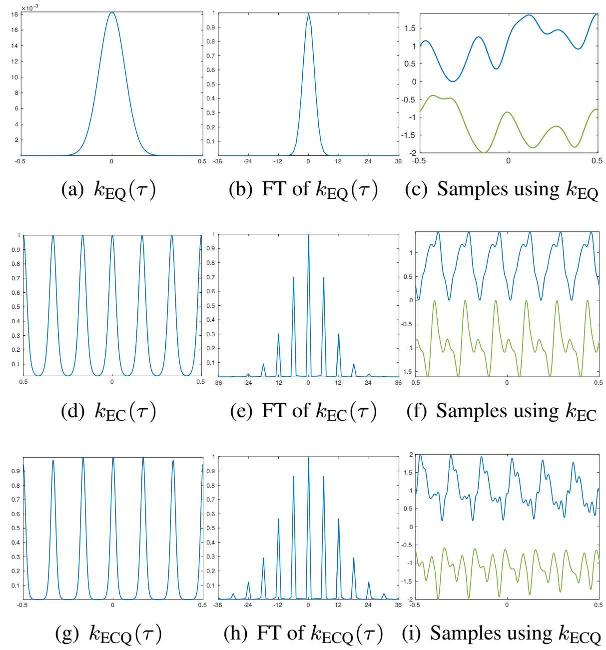
This is known as the Wiener-Khintchine theorem [7, 14]. Thus, we can do frequency-domain analysis for several covariance functions and decide which kernel is more appropriate for modelling the spectral content of music signals. The FT of the *exponentiated quadratic* covariance function

$$k_{\text{EQ}}(\tau) = \sigma^2 \exp\left(-\frac{\tau^2}{2l^2}\right), \tag{8}$$

as well as the FT of the *exponentiated cosine* kernel

$$k_{\text{EC}}(\tau) = \sigma^2 \exp\left[z \cos(\omega\tau)\right], \tag{9}$$

are shown in Fig 1(b)-1(e). Covariance function (8) is probably the most widely-used kernel within the kernel machines field, because the GP with a exponentiated-quadratic covariance function is very smooth [7]. In order to allow periodic kernels to describe functions where the amplitude envelope



(a) $k_{\text{EQ}}(\tau)$ (b) FT of $k_{\text{EQ}}(\tau)$ (c) Samples using $k_{\text{EQ}}$

(d) $k_{\text{EC}}(\tau)$ (e) FT of $k_{\text{EC}}(\tau)$ (f) Samples using $k_{\text{EC}}$

(g) $k_{\text{ECQ}}(\tau)$ (h) FT of $k_{\text{ECQ}}(\tau)$ (i) Samples using $k_{\text{ECQ}}$

**Fig. 1**. Frequency analysis of kernels (8), (9), (10) depicted in (a), (d), (g) respectively. (b), (e), (h) are their corresponding FT. (c), (f), (i) show sampled functions.

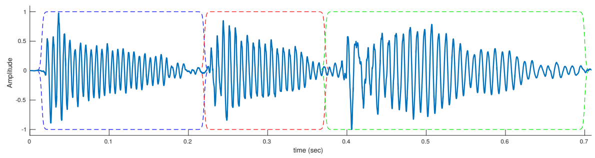
changes in time, we introduce a modification of expression (9). To do so, we multiply this kernel with (8). The resulting covariance function, called *exponentiated-cosine-quadratic* corresponds to

$$k_{\text{ECQ}}(\tau) = \sigma^2 \exp\left[z \cos(\omega\tau) - \frac{\tau^2}{2l^2}\right]. \tag{10}$$

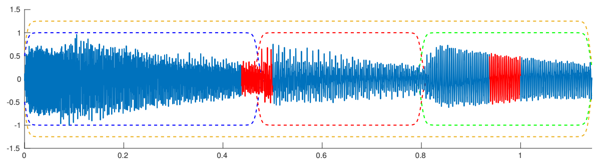
Fig. 1(h) depicts the FT of (10). We see that its spectral density keeps similar to the one obtained for (9) (see Fig. 1(e)). But the realizations sampled from a GP with this covariance function (Fig. 1(i)) show a smooth variation in the amplitude envelope, and also maintain the properties described by the previous kernel (9), i.e. a periodic structure with natural frequency and harmonics. This covariance function (10) seems to be more appropriate for modelling music signals in comparison with the two kernels presented previously ((8)-(9)). The hyper-parameter $\omega$ in (9)-(10) is directly related with the natural frequency or F0 of the modelled random processes.

## 3. RESULTS AND DISCUSSION

Experiments were done over real audio. We evaluated different kernel configurations on a pitch estimation task, and on a missing data imputation task. All experiments assume we previously know the number of change-windows and its locations. In the pitch estimation task all the parameters of the covariance function are known, except those related with the fundamental frequency of each sound event, i.e. the value of

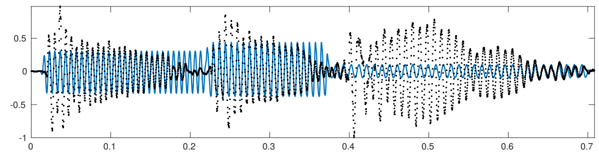


(a) Signal used for pitch estimation.

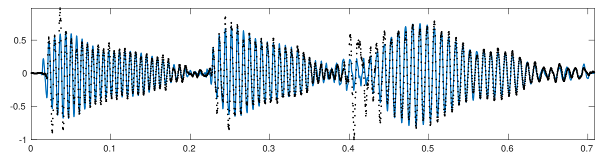


(b) Signal used for filling missing-data gaps.

**Fig. 2**. (a) analysed audio (blue line), change-windows (dashed lines). (b) observed data (blue line), missing-data gaps (red line), change-windows (dashed lines).



(a) Observations (dots), and posterior mean (continuous line) using (9).



(b) Observations (dots), and posterior mean (continuous line) using (10).

**Fig. 3**. Posterior mean for the pitch estimation experiments. (a) using $k_{\mathrm{EQ}}(\tau)$, and (b) using $k_{\mathrm{EQC}}(\tau)$.

$\omega_m$ in (9) and (10) when using these kernels in the general model (4). Thus, we focus on optimizing only these model hyperparameters from the data. In the missing data imputation task the score of the modelled piece of music audio is used for tuning manually the model hyperparameters.

### 3.1. Data

In this study we used two short audio excerpts, in order to explore the method, so that we can efficiently fit models and search in the hyperparameter space. The excerpt used for pitch estimation experiments corresponds to $0.7$ seconds of the song *Black Chicken 37* by Buena Vista Social Club. This segment of audio contains three notes of a bass melody (Fig. 2(a)). In the missing data imputation task we used polyphonic audio corresponding to $1.14$ seconds of Chopin's *Nocturne Op. 15 No. 1*, where more than one note occur at the same time. The segments of signal in red in Fig. 2(b) represent gaps of missing data. We reduced the sample frequency of both audio excerpts from $44.1$KHz to 8KHz. To infer hyperparameters we consider an empirical Bayes approach, which allow us to use continuous optimization methods. We maximize the marginal likelihood. This moves us up one level of the Bayesian hierarchy, and reduces the chances of overfitting [12]. Given an expression for the log marginal likelihood and its partial derivatives, we can estimate the kernel parameters using any standard gradient-based optimizer [12]. A gradient descent method was used for optimization.

### 3.2. Pitch estimation

For the pitch estimation task we tested two different models with kernels (9), and (10) respectively. We performed hyperparameters learning using all the observed signal shown in Fig. 2(a). This is because in this experiment rather than evaluating the prediction of the trained models, we were interested in the accuracy of pitch estimation. Covariance function (8) does not have any parameter we can link to the fundamental frequency of each sound event, that is why we omitted it here. We compared the GPs models results with the algorithm pYIN, a fundamental frequency estimator [15]. The trained model using $k_{\mathrm{EC}}(\tau)$ was able to estimate the pitch for each sound event with a RMS error of $0.6282$ semitones. On the other hand, the amplitude-envelope evolution of the signal is beyond the scope of the structure that this kernel can model (See Fig. 3(a)). This is because this covariance function can only describe constant amplitude-envelope, periodic signals, with a fundamental frequency and several harmonics (Fig. 1(f)). Results using (10) are shown in Fig. 3(b). We observe that although the posterior mean of the predictive distribution does not exactly fit the data, the model is able to learn the pitch of each of the three sound events with a smaller RMS error of $0.1075$ in comparison with the $0.1688$ RMS error obtained with pYIN. Variations in the amplitude envelope can also be described using (10).

### 3.3. Filling gaps of missing data in audio

We compared three different models predicting missing-data gaps. We studied kernels (8), (9), and (10). In Fig. 2(b) first gap (red segment) contains the transient (onset and attack [16]) of a sound event, whereas the second gap is located in a more stable segment of the data (smooth decay). Fig. 4(a)-4(b) depict the prediction using (8). These figures correspond to zoom in small sections of the signal where the gaps occur (Fig. 2(b)). We see that the model using this kernel overfits the data, i.e. the posterior mean (blue line) fits all the observed data (black dots) with high confidence (grey shaded area), but the confidence decreases and the prediction is quite poor in the input space zones where the data is not available (red dots). Also, we see that the model using (8) does not expect any periodic behaviour in the gaps. The RMS error for

**Table 1**. Filling gaps prediction RMS error.

| kernel | RMS transient gap | RMS decay gap |
|---|---|---|
| $k_{\mathrm{EQ}}(\tau)$ | 0.2265 | 0.3172 |
| $k_{\mathrm{EC}}(\tau)$ | 0.2143 | 0.0964 |
| $k_{\mathrm{EQC}}(\tau)$ | 0.0912 | 0.0355 |

both gaps is presented in Table 1.

Fig. 4(c)-4(d) show the prediction using covariance function (9). In the transient gap (Fig. 4(c)) the posterior mean (blue line) does not follows the data, this is because transients are short intervals during which the signal evolves in a nonstationary, nontrivial and unpredictable way [16]. opposite to this, the model using kernel (9) can only describe the behaviour of constant amplitude-envelope periodic stochastic functions. In the second gap (Fig. 4(d)) the posterior mean describes properly the periodic behaviour of the data, but it does not follow the amplitude-envelope of the observations. This is because this covariance function is able to describe periodic functions that have several harmonic components. The drawback of this kernel is that it assumes constant the amplitude of the periodic stochastic functions that describes. These different performance on the prediction is reflected on the RMS error obtained for each gap (Table 1).

Results using (10) are presented in Fig. 4(e)-4(f). We see that in Fig. 4(f) the posterior mean describes properly the periodic behaviour and amplitude envelope smooth evolution of the modelled signal. We observe that prediction on the decay gap using (10) is closer to the actual data (red dots) than the results obtained with (9) as well as (8). This is reflected in the smallest RMS error in table 1. This is because (10) allows to describe periodic functions that have several harmonic components and time-varying amplitude envelope. On the other hand, the prediction performance reduces for the transient gap (Fig. 4(e)). In order to model the onset, attack and decay of a sound event, covariance function (10) could be modified for modelling nonstationary amplitude envelope evolution.

### 3.4. Related work

In [17] GPs are used for time-frequency analysis as probabilistic inference. Natural signals are assumed to be formed by the superposition of distinct time-frequency components, with the analytic goal being to infer these components by applying Bayes' rule [17]. GPs have also been used for audio source separation [18, 20]. In [18] the *mixture* signal is modelled as a linear combination of independent convolved versions of latent GPs or *sources*. The model splits the *mixture* signal in frames also considered independent, by using weight-functions. Thus each source is modelled as a series of concatenated locally stationary frames, each one with its corresponding covariance function. With this assumption the resulting signal is supposed to be non-stationary [18]. On the other hand, despite the approach we present also assumes the latent GPs $f_m$ in (4) as non-correlated, the observed signal is not framed into independent segments. Instead of using weight-functions that act over the observed data, we introduce change-windows $\phi_m$ influencing each latent GP ending up with latent processes representing specific sound events that happen at certain segments of time. Therefore the proposed model keeps the correlation between the observations throughout all the signal. That is what allows to make prediction in gaps of missing data (section 3.3). GPs have been used also for estimating spectral envelope and fundamental frequency of a speech signal [19]. Finally, GPs for music genre classification and emotion estimation were investigated in [21].

## 4. CONCLUSIONS

We discussed a GP regression framework for modelling music audio. We compared different models in pitch estimation as well as in prediction of missing data. We showed which kernels were more appropriate for describing properties of music signals, specifically: nonstationarity, dynamics, and spectral harmonic content. The advantage of this approach is that by designing a proper kernel we can introduce prior knowledge and beliefs about the properties of music signals, and use all that prior information to improve prediction. Computational complexity is an important limitation of GPs, therefore the presented work could be extended using efficient representations to model larger audio signals. Kernels as [22] could be studied for modelling harmonic content, and Latent Force models [23] for describing mechanistic characteristics.

## 5. REFERENCES


[1] X. Serra, M. Magas, E. Benetos, M. Chudy, S. Dixon, A. Flexer, E. Gómez, F. Gouyon, P. Herrera, S. Jorda, O. Paytuvi, G. Peeters, J. Schlüter, H. Vinet, and G. Widmer, *Roadmap for Music Information Research*, Creative Commons BY-NC-ND 3.0, 2013.

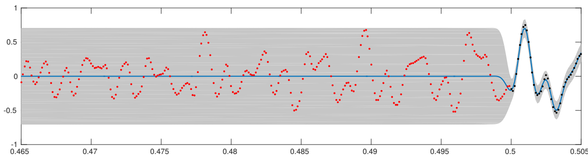
[2] M. Muller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1088–1110, 2011.

[3] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: Challenges and future directions," *J. Intell. Inf. Syst.*, vol. 41, no. 3, pp. 407–434, 2013.

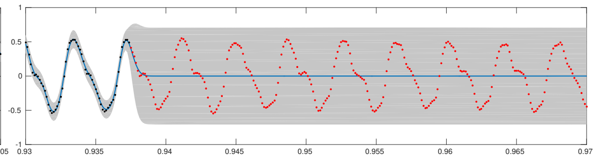
[4] A. T. Cemgil, S. J. Godsill, P. Peeling, and N. Whiteley, "Bayesian statistical methods for audio and music processing," *The Oxford Handbook of Applied Bayesian Analysis*, pp. 1–45, 2010.

[5] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: Breaking the glass ceiling," in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, 2012, pp. 379–384.

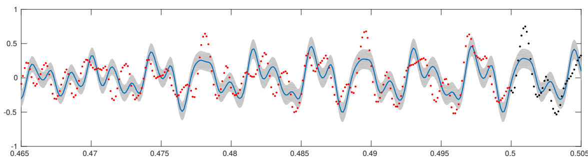
[6] S. Särkkä, A. Solin, and J. Hartikainen, "Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing:



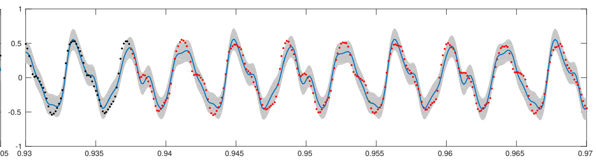
(a) Prediction on transient gap using $k_{\mathrm{EQ}(\tau)}$.



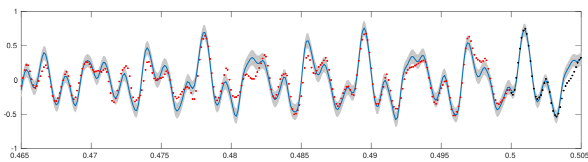
(b) Prediction on smooth decay gap using $k_{\mathrm{EQ}(\tau)}$.



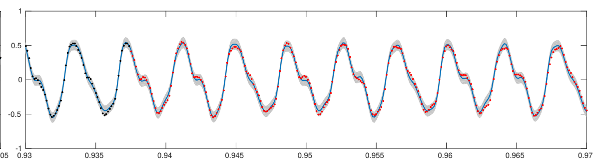
(c) Prediction on transient gap using $k_{\mathrm{EC}(\tau)}$.



(d) Prediction on smooth decay gap using $k_{\mathrm{EC}(\tau)}$.



(e) Prediction on transient gap using $k_{\mathrm{EQC}(\tau)}$.



(f) Prediction on smooth decay gap using $k_{\mathrm{EQC}(\tau)}$.

**Fig. 4**. Zoom in a portion of missing-data gaps. In each figure the continuous blue line represent the posterior mean, grey shaded areas correspond to the posterior variance, red dots are missing data, whereas black dots are observed data.


A look at gaussian process regression through kalman filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 51–61, July 2013.

[7] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2005.

[8] S. Särkkä and A. Solin, *Image Analysis: 18th Scandinavian Conference, SCIA 2013, Espoo, Finland, June 17-20, 2013. Proceedings*, pp. 172–181, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[9] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, 2012.

[10] J. R. Lloyd, D. K. Duvenaud, R. B. Grosse, J. B. Tenenbaum, and Z. Ghahramani, "Automatic construction and Natural-Language description of nonparametric regression models," in *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014, pp. 1242–1250.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag, New York, 2006.

[12] K. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

[13] J. Shawe and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, New York, 2004.

[14] K. Shanmugan and A. Breipohl, *Random Signals: Detection, Estimation and Data Analysis*, Wiley, 1988.

[15] M. Mauch and S. Dixon, "Pyin: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 659–663.

[16] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, 2005.

[17] R. E. Turner and M. Sahani, "Time-frequency analysis as probabilistic inference," *IEEE Transactions on Signal Processing*, vol. 62, no. 23, pp. 6171–6183, Dec 2014.

[18] A. Liutkus, R. Badeau, and G. Richard, "Gaussian processes for underdetermined source separation," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3155–3167, July 2011.

[19] K. Yoshii and M. Goto, "Infinite kernel linear prediction for joint estimation of spectral envelope and fundamental frequency," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013.

[20] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto, "Beyond nmf: time-domain audio source separation without phase reconstruction," in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, 2013.

[21] K. Markov and T. Matsui, "Music genre and emotion recognition using gaussian processes," *IEEE Access*, vol. 2, pp. 688–697, 2014.

[22] R. P. Adams A. G. Wilson, "Gaussian process kernels for pattern discovery and extrapolation," *International Conference on Machine Learning (ICML), 2013.*, 2013.

[23] M. Álvarez, D. Luengo, and N. D. Lawrence, "Linear latent force models using gaussian processes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 11, pp. 2693–2705, Nov 2013.