AN APPROXIMATE BAYESIAN FUNDAMENTAL FREQUENCY ESTIMATOR

Jesper Kjær Nielsen[†], *Mads Græsbøll Christensen*[‡], *Søren Holdt Jensen*[†]

[†]Aalborg University
Dept. of Electronic Systems
Niels Jernes Vej 12, DK-9220 Aalborg
{jkn,shj}@es.aau.dk

[‡]Aalborg University
Dept. of Architecture, Design & Media Technology
Niels Jernes Vej 14, DK-9220 Aalborg
mgc@create.aau.dk

ABSTRACT

Joint fundamental frequency and model order estimation is an important problem in several applications such as speech and music processing. In this paper, we develop an approximate estimation algorithm of these quantities using Bayesian inference. The inference about the fundamental frequency and the model order is based on a probability model which corresponds to a minimum of prior information. From this probability model, we give the exact posterior distributions on the fundamental frequency and the model order, and we also present analytical approximations of these distributions which lower the computational load of the algorithm. By use of simulations on both a synthetic signal and a speech signal, the algorithm is demonstrated to be more accurate than a state-of-the-art maximum likelihood-based method.

Index Terms— Fundamental frequency, Bayesian inference and model comparison, Zellner's g-prior.

1. INTRODUCTION

Fundamental frequency estimation is an important problem in several applications such as music processing [1], speech processing [2], and electrocardiography [3]. In these applications, an approximately periodic signal is often assumed to be observed in additive noise, and the problem is to estimate the fundamental frequency from the observations. A periodic signal consists of $l-1$ overtone partials $\{\omega_i\}_{i=2}^{l}$ which are related to the fundamental frequency ω by $\omega_i = i\omega$. For a periodic signal, the fundamental partial and the overtone partials are referred to as harmonic components or harmonics, and the number of harmonics l is usually unknown. Numerous fundamental frequency estimation algorithms have been suggested in the literature. The simplest algorithms are the non-parametric methods based on, for example, the auto-correlation function [4] or the cepstrum [5]. The more advanced algorithms are based on parametric models of the observed signal, and inference about the fundamental frequency is typically based on maximum likelihood (ML) methods [6], subspace-based methods [7], filtering methods [8], or Bayesian methods [9,10]. We refer the interested reader to [11] for a review of many of the non-Bayesian methods. Only a few of the suggested methods assume that the number of harmonics is unknown. In order to perform model selection, these methods typically add an order dependent penalty term to the log-likelihood function or compare the angle between subspaces [12]. These methods can usually not be used for neither model comparison nor model averaging.

Mathematically speaking, the problem of estimating the fundamental frequency is typically defined in the following way. We observe a discrete-time signal of the form

$$x(n) = s(n) + e(n), \quad n = 0, 1, \cdots, N-1 \qquad (1)$$

where $\{x(n)\}_{n=0}^{N-1}$, $\{s(n)\}_{n=0}^{N-1}$, and $\{e(n)\}_{n=0}^{N-1}$ are the observed signal, the systematic part of the signal, and the stochastic part of the signal, respectively. The systematic part consists of l harmonics and is at time index n given by

$$s(n) = \sum_{i=1}^{l} \alpha_i \exp(ji\omega n) \qquad (2)$$

where $j = \sqrt{-1}$ is the imaginary unit. For the i'th harmonic, the complex amplitude α_i, the in-phase component a_i, the quadrature components b_i, the amplitude A_i, and the phase ϕ_i are related by $\alpha_i = a_i + jb_i = A_i \exp(j\phi_i)$. Note that a real-valued signal can be cast into the form of a complex-valued signal by computing its down-sampled analytic signal [13]. Provided that the frequencies of the first and last harmonics are not too close to zero and π (relative to N), respectively, the solution to the estimation problem using the down-sampled analytic signal yields the same result as for the real-valued signal [11]. In this paper, we use the complex-valued signal model since it leads to simpler notation and faster algorithms [7, 11].

The primary aim of this paper is to develop a default estimation scheme for estimating the fundamental frequency and the number of harmonics. By the word *default* we mean that we develop a nearly user-parameter free algorithm which automatically follows from a minimum of prior information and a few minor approximations. The approximations are made so that we obtain closed-form expressions and a computational load comparable to the methods suggested in [11]. Moreover, we compare the proposed method to the algorithm proposed in [11, Sec. 2.6], and we indicate by use of simulation examples that our proposed method is superior to this. The paper is organised as follows. The inference method and the probability model is described in Sec. 2, and inference about the fundamental frequency and the number of harmonics is made in Sec. 3. In Sec. 4, we develop an approximation which decreases the computational load, and we evaluate the approximation in Sec. 5.

2. PROBLEM FORMULATION

Our primary aim is to make inference about the fundamental frequency ω and the model order l given our prior information I and the N data points, which we have collected in the vector $\boldsymbol{x}$. That is, we wish to find the posterior distributions

$$p(\omega, l|\boldsymbol{x}, I) = p(\omega|\boldsymbol{x}, l, I)p(l|\boldsymbol{x}, I) \qquad (3)$$

and some of their statistics such as the mode and the moments. The model order l labels a unique model $\mathcal{M}_l$ which is an hypothesis of how the data $\boldsymbol{x}$ have been generated. Each model $\mathcal{M}_l$ is parametrised by the model parameters $\boldsymbol{\theta}_l \in \Theta_l$ where Θ_l is the parameter space. For our problem, ω is one of these parameters, and

we refer to the remaining model parameters such as the noise variance and the complex amplitudes as the nuisance parameters. The relationship between the data $\boldsymbol{x}$ and the model $\mathcal{M}_l$ is given by the probability density function (pdf) $p(\boldsymbol{x}|\boldsymbol{\theta}_l, l, I)$ which is called the data model or the sampling distribution. When viewed as a function of the model parameters, the sampling distribution is referred to as the likelihood function, and it plays an important role in statistics where it is mainly used for parameter estimation. However, it cannot alone be used for model comparison or selection since it suffers from over-fitting. In a Bayesian framework, the model parameters and the model order are random variables with the prior distributions $p(\boldsymbol{\theta}_l|l, I)$ and $p(l|I)$, respectively, and these distributions contain our state of knowledge before any data are observed. After observing some data, we update our state of knowledge by transforming the prior distributions into the posterior distributions $p(\boldsymbol{\theta}_l|\boldsymbol{x}, l, I)$ and $p(l|\boldsymbol{x}, I)$. We obtain the desired posterior distributions in (3) by marginalising $p(\boldsymbol{\theta}_l|\boldsymbol{x}, l, I)$ over the nuisance parameters. The prior and posterior distributions for the model parameters and the model are connected by Bayes' theorem

$$p(\boldsymbol{\theta}_l|\boldsymbol{x}, l, I) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta}_l, l, I)p(\boldsymbol{\theta}_l|l, I)}{p(\boldsymbol{x}|l, I)} \tag{4}$$

$$p(l|\boldsymbol{x}, I) = \frac{p(\boldsymbol{x}|l, I)p(l|I)}{p(\boldsymbol{x}|I)} \tag{5}$$

where

$$p(\boldsymbol{x}|l, I) = \int_{\Theta_l} p(\boldsymbol{x}|\boldsymbol{\theta}_l, l, I)p(\boldsymbol{\theta}_l|l, I)d\boldsymbol{\theta}_l \tag{6}$$

is called the marginal likelihood or the evidence. For model comparison, the odds of two competing model orders k and i are often compared. In this connection, we define the posterior odds

$$\frac{p(k|\boldsymbol{x}, I)}{p(i|\boldsymbol{x}, I)} = \text{BF}[k, i]\frac{p(k|I)}{p(i|I)} \tag{7}$$

where the Bayes' factor is given by

$$\text{BF}[k, i] = \frac{p(\boldsymbol{x}|k, I)}{p(\boldsymbol{x}|i, I)} = \frac{m_k(\boldsymbol{x})}{m_i(\boldsymbol{x})} \tag{8}$$

where $m_k(\boldsymbol{x})$ is an unnormalised marginal likelihood whose normalising constant must be the same for both i and k. Working with $m_k(\boldsymbol{x})$ rather than the normalised marginal likelihood $p(\boldsymbol{x}|k, I)$ is usually much simpler. Moreover, $p(\boldsymbol{x}|k, I)$ does not even exist if we use improper priors. In this case, we must require that improper priors are only used for the common model parameters which have the same meaning across all models. Otherwise, the Bayes' factor is indeterminate [14]. Since the prior and posterior distributions on the model order are discrete, it is easy to find the posterior odds and the posterior distribution once the Bayes' factors are known. For example, the posterior probability mass function (pmf) on the model order is

$$p(l|\boldsymbol{x}, I) = \frac{\text{BF}[l; k]p(l|I)}{\sum_{i=1}^{L} \text{BF}[i; k]p(i|I)} \tag{9}$$

where the model order k labels some user selected base model which we compare all other models against. Therefore, the main challenge in Bayesian model comparison is to compute the Bayes' factor for competing pairs of models. However, before we can use Bayes' theorem to make inference about the fundamental frequency and the model order, we first have to turn our prior information I into a sampling distribution and prior distributions on the model parameters. The amount of prior information I we assume can be stated in the following way.

Assumption 2.1. *We are given N data points $\{x(n)\}_{n=0}^{N-1}$ from a zero-mean complex-valued signal which has been sampled at a known and uniform sampling frequency. The signal is wide-sense stationary (WSS) and consists of a systematic part which is periodic, corrupted by additive noise, and bandlimited to the known angular frequency interval $[\omega_a, \omega_b]$.*

Using Jaynes' principles of maximum entropy and transformation groups [15], the following probability model can be shown to automatically follow from the prior information I if N and the signal-to-noise ratio (SNR) are large enough. It is given by

$$p(\boldsymbol{\alpha}_l, \sigma^2, \omega, l, g|I) \propto \mathcal{CN}(\boldsymbol{\alpha}_l; \boldsymbol{0}, g\sigma^2(\boldsymbol{Z}_l^H\boldsymbol{Z}_l)^{-1})(\sigma^2(1+g))^{-1}$$
$$\times (LW_l)^{-1}\mathbb{I}_{\Omega_l}(\omega)\mathbb{I}_{\mathbb{R}^+}(g)\mathbb{I}_{\mathbb{R}^+}(\sigma^2)\mathbb{I}_{\mathcal{L}}(l) \tag{10}$$

$$p(\boldsymbol{x}|\boldsymbol{\alpha}_l, \sigma^2, \omega, l, I) = \mathcal{CN}(\boldsymbol{x}; \boldsymbol{Z}_l\boldsymbol{\alpha}_l, \sigma^2\boldsymbol{I}_N) \tag{11}$$

where $(\cdot)^H$ denotes complex transposition and

$$\boldsymbol{x} \triangleq \begin{bmatrix} x(0) & \cdots & x(N-1) \end{bmatrix}^T \tag{12}$$

$$\boldsymbol{\alpha}_l \triangleq \begin{bmatrix} \alpha_1 & \cdots & \alpha_l \end{bmatrix}^T \tag{13}$$

$$\boldsymbol{z}_i \triangleq \begin{bmatrix} \exp(ji\omega 0) & \cdots & \exp(ji\omega(N-1)) \end{bmatrix}^T \tag{14}$$

$$\boldsymbol{Z}_l \triangleq \begin{bmatrix} \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_l \end{bmatrix} \tag{15}$$

Moreover, $W_l = \omega_b/l - \omega_a$, $\mathbb{I}_S(\cdot)$ is the indicator function on the set S, and $\mathcal{L} = \{1, 2, \cdots, L\}$, $\Omega_l = [\omega_a, \omega_b/l]$, and $\mathbb{R}^+$ are the set of candidate model orders, the set of candidate fundamental frequencies, and the set of positive real numbers, respectively. The parameter g is an auxiliary variable which is introduced to make the inference more analytically tractable. The joint prior on $\boldsymbol{\alpha}_l$ and σ^2 from (10) is known as the Zellner's g-prior, and it has been extensively studied for Bayesian linear regression [16].

3. INFERENCE

From the prior distribution in (10) and the sampling distribution in (11), we now use Bayes' theorem to compute the posterior distributions in (3). The joint posterior pdf marginalised over the complex amplitudes and the noise variance is [1]

$$p(\omega, g, l|\boldsymbol{x}) \propto \int_0^\infty \int_{\mathbb{C}^l} p(\boldsymbol{x}|\boldsymbol{\alpha}_l, \sigma^2\omega, g, l)p(\boldsymbol{\alpha}_l, \sigma^2\omega, g, l)d\boldsymbol{\alpha}_l d\sigma^2$$
$$\propto \frac{\Gamma(N)f_l(\omega, g)\mathbb{I}_{\Omega_l}(\omega)\mathbb{I}_{\mathbb{R}^+}(g)\mathbb{I}_{\mathcal{L}}(l)}{(\pi\boldsymbol{x}^H\boldsymbol{x})^N LW_l} \tag{16}$$

where we have defined

$$f_l(\omega, g) \triangleq (1+g)^{N-l-1}[1 + g(1 - R_l^2(\omega))]^{-N} \tag{17}$$

$$R_l^2(\omega) \triangleq \boldsymbol{x}^H\boldsymbol{P}_l\boldsymbol{x}(\boldsymbol{x}^H\boldsymbol{x})^{-1} \tag{18}$$

$$\boldsymbol{P}_l \triangleq \boldsymbol{Z}_l(\boldsymbol{Z}_l^H\boldsymbol{Z}_l)^{-1}\boldsymbol{Z}_l^H . \tag{19}$$

The matrix $\boldsymbol{P}_l$ is the projection matrix onto the space spanned by the columns of $\boldsymbol{Z}_l$, and $R_l^2(\omega)$ resembles the coefficient of determination from linear regression analysis where it is used to measure the prediction performance. We can also marginalise (16) over g and obtain the marginal posterior pdf on ω under model order l as

$$p(\omega|\boldsymbol{x}, l) = \int_0^\infty \frac{p(\omega, g, l|\boldsymbol{x})}{p(l)}dg \propto \int_0^\infty f_l(\omega, g)\mathbb{I}_{\Omega_l}(\omega)dg$$
$$\propto {}_2F_1(N, 1; l+1; R_l^2(\omega))\mathbb{I}_{\Omega_l}(\omega) \tag{20}$$

---

[1]Note that the expressions from now on do not explicitly depend on the prior information I to keep the notation uncluttered.

where $_2F_1$ is the Gaussian hypergeometric function [17, Ch. 15]. Moreover, an unnormalised marginal likelihood is given by the following integral representation

$$m_l(\boldsymbol{x}) = \frac{\Gamma(N)}{(\pi \boldsymbol{x}^H \boldsymbol{x})^N W_l} \int_{\Omega_l} \int_0^\infty f_l(\omega, g) dg d\omega \qquad (21)$$

$$= \frac{\Gamma(N)}{(\pi \boldsymbol{x}^H \boldsymbol{x})^N l W_l} \int_{\Omega_l} {}_2F_1(N, 1; l+1; R_l^2(\omega)) d\omega . \quad (22)$$

Unfortunately, the statistical quantities of the posterior pdf on the fundamental frequency such as the mode or the variance are not available in closed-form due to the non-linear way that ω parametrises the pdf in (20). Moreover, the posterior model order probabilities are not available in closed-form since the integral in (22) cannot be computed analytically. In Sec. 4, we discuss one approximate way of finding these quantities and posterior probabilities.

4. APPROXIMATION

Our approximation is based on a normal approximation of the pdf on the fundamental frequency in (20), and a full Laplace approximation of the double integral in (21). Since the function $f_l(\omega, g)$ is not symmetric and in order to avoid edge effects near $g = 0$, we use the re-parametrisation $\tau = \ln g$ with the Jacobian $dg/d\tau = \exp(\tau)$ [16]. This re-parametrisation suggest that the posterior distribution over g is approximately a log-normal distribution. With this, the Laplace approximation of the integral in (21) is

$$\int_{\Omega_l} \int_0^\infty f(\omega, g) dg d\omega = \int_{\Omega_l} \int_{-\infty}^\infty \exp(\tau) f(\omega, \exp(\tau)) d\tau d\omega$$

$$= 2\pi \exp(\hat{\tau}) f(\hat{\omega}, \exp(\hat{\tau})) \sqrt{s_l(\hat{\omega}| \exp(\hat{\tau})) \gamma_l(\hat{\tau}|\hat{\omega})} \qquad (23)$$

where $(\hat{\omega}, \hat{\tau})$, $s_l(\hat{\omega}| \exp(\hat{\tau}))$, and $\gamma_l(\hat{\tau}|\hat{\omega})$ are the mode and the variances of the bivariate normal approximation of $g(\omega, \tau) \triangleq \exp(\tau) f(\omega, \exp(\tau))$, respectively. Note that the cross-covariance is zero. As shown in [16], $\hat{\tau}$ and $\gamma_l(\hat{\tau}|\hat{\omega})$ are for $\hat{g} \triangleq \exp(\hat{\tau})$ given by

$$\hat{\tau} = \ln \left[\frac{\sqrt{\beta_\tau^2 - 4\alpha_\tau} + \beta_\tau}{-2\alpha_\tau} \right] \qquad (24)$$

$$\gamma_l(\hat{\tau}|\hat{\omega}) = \left[\frac{N(1 - R_l^2(\hat{\omega}))\hat{g}}{[1 + \hat{g}(1 - R_l^2(\hat{\omega}))]^2} - \frac{(N - l - 1)\hat{g}}{(1 + \hat{g})^2} \right]^{-1} \qquad (25)$$

with

$$\alpha_\tau \triangleq -(1 - R_l^2(\hat{\omega}))l \qquad (26)$$

$$\beta_\tau \triangleq (N - 1)R_l^2(\hat{\omega}) - l + 1 . \qquad (27)$$

The mode $\hat{\omega}$ is given by

$$\hat{\omega} = \arg\max_{\omega \in \Omega_l} g(\omega, \tau) = \arg\max_{\omega \in \Omega_l} \boldsymbol{x}^H \boldsymbol{P}_l \boldsymbol{x} \qquad (28)$$

and is the same as the ML-estimate [11] and independent of τ. Unfortunately, it is hard to find $\hat{\omega}$ since the cost-function in (28) has a complicated multi-modal shape and is very sharply peaked around $\hat{\omega}$. Typically, $\hat{\omega}$ is found by first evaluating the cost-function on a fine grid and then performing a local optimisation around the maximum value of the cost-function on this grid. However, this procedure might be too costly from a computational point of view since
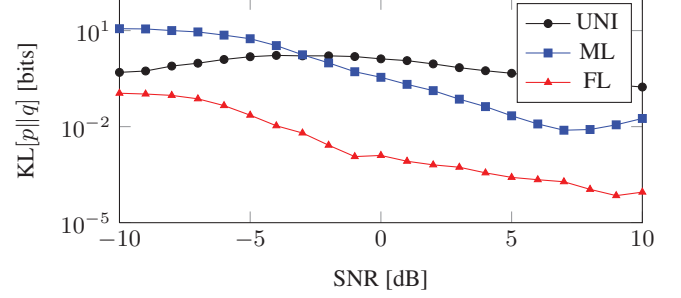

Fig. 1. The KL divergence between the true posterior distribution on the model order and various approximate distributions on the model order.

we have to evaluate the projection matrix $\boldsymbol{P}_l$ for every candidate frequency. We can speed up the computation significantly by making the approximation

$$\hat{\omega} \approx \arg\max_{\omega \in \Omega_l} \boldsymbol{x}^H \boldsymbol{Z}_l \boldsymbol{Z}_l^H \boldsymbol{x} = \arg\max_{\omega \in \Omega_l} \|\boldsymbol{Z}_l^H \boldsymbol{x}\|_2^2 \qquad (29)$$

which can be computed efficiently using a single FFT [11]. To get the exact estimate in (28), we may use the approximate estimate as the starting point of a local optimisation using the exact cost-function in (28). To find the variance $s_l(\hat{\omega}|\hat{g})$ of the fundamental frequency, we need to find the second order derivatives of $\ln g(\omega, \tau)$ w.r.t. ω and evaluate it at the mode $(\hat{\omega}, \hat{\tau})$. It can be shown that it can be approximated by

$$s_l(\hat{\omega}|\hat{g}) \approx \frac{6(1 + \hat{g})\hat{\sigma}_l^2}{\hat{g} N(N^2 - 1) \sum_{i=1}^l |\hat{\alpha}_i|^2 i^2} \qquad (30)$$

where $\hat{\boldsymbol{\alpha}}_l = (\boldsymbol{Z}^H \boldsymbol{Z})^{-1} \boldsymbol{Z}^H \boldsymbol{x}$ and $\hat{\sigma}_l^2 = N^{-1} \boldsymbol{x}^H (\boldsymbol{I}_N - \hat{g} \boldsymbol{P}_l/(1 + \hat{g})) \boldsymbol{x}$ which for $\hat{g} \to \infty$ are the ML estimate of the i'th complex amplitude and the noise variance, respectively. The exact expression for $s_l(\hat{\omega}|\hat{g})$ is much more complicated, and the expression in (30) is a good approximation for a large N and a high SNR. Moreover, it has the interesting interpretation that it is the asymptotic Cramer-Rao lower bound for $\hat{g} \to \infty$ with the true values of the noise variance and the complex amplitudes replaced by their ML estimates.

A normal approximation of the pdf on the fundamental frequency in (20) now readily follows from (23) as

$$p(\omega|\boldsymbol{x}, l) \approx \int_{-\infty}^\infty \mathcal{N}([\omega, \tau]^T, [\hat{\omega}, \hat{\tau}]^T, \text{diag}[s_l(\hat{\omega}|\hat{g}), \gamma_l(\hat{\tau}|\hat{\omega})]) d\tau$$

$$= \mathcal{N}(\omega, \hat{\omega}, s_l(\hat{\omega}|\hat{g})) \qquad (31)$$

where $\text{diag}[\cdot]$ transforms a vector into a diagonal matrix.

5. SIMULATIONS

We first evaluated the accuracy of the Laplace approximation to the posterior distribution on the model order. To do this, we used the Kullback-Leibler (KL) divergence given by

$$\text{KL}(p\|q) = \sum_{l=1}^L p(l|\boldsymbol{x}) \ln \left[\frac{p(l|\boldsymbol{x})}{q(l|\boldsymbol{x})} \right] , \qquad (32)$$

to measure the distance between the true pdf $p(l|\boldsymbol{x})$ and the approximate pdf $q(l|\boldsymbol{x})$. The true posterior pdf $p(l|\boldsymbol{x})$ was computed from
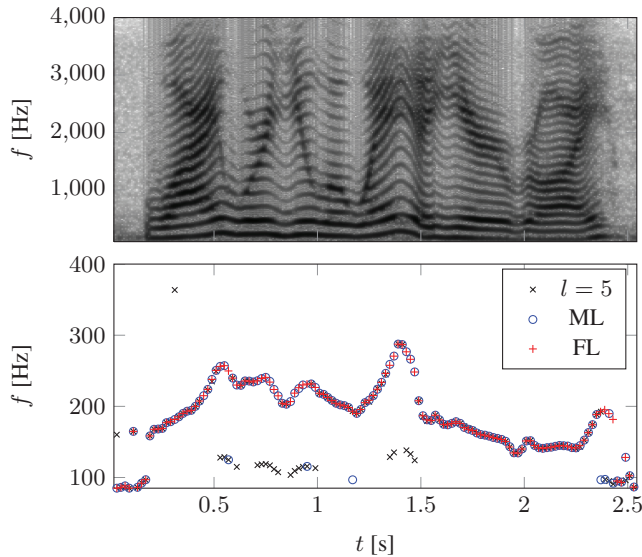
Fig. 2. The spectrogram of a female speech signal and the estimated fundamental frequency.

(9), (8), and (22) with the latter being computed using numerical integration on a fine grid. For $q(l|\boldsymbol{x})$, we instead of numerical integration evaluated the double integral in (21) using the Laplace approximation in (23)^2. For comparison, we also computed $q(l|\boldsymbol{x})$ using the ML-based method suggested in [11, Sec. 2.6] and the uniform distribution. Fig. 1 shows the average KL divergence for 500 Monte Carlo runs at various SNRs in steps of 1 dB from -10 dB to 10 dB. The figure clearly indicates that the proposed method 'FL' was more accurate than the ML-based method. For a low SNR, however, the 'FL' approximation was not very accurate.

We also evaluated the proposed method on a speech signal which originates from a female voice uttering *"Why were you away a year, Roy?"*. Fig. 2 shows the spectrogram of the signal and the estimated fundamental frequency for a fixed model order of $l = 5$, the proposed method with a maximum model order of 20, and the ML-based method. The speech signal has a sampling frequency of 8 kHz, and we partitioned it in blocks of 20 ms corresponding to $N = 80$ samples. The ML-based method had a few outliers, whereas the proposed method did not have any outliers. However, when the model order was fixed to $l = 5$ for all blocks, the proposed method had a lot of outliers. This illustrates why model order selection is important even if we are only interested in an estimate of the fundamental frequency.

6. CONCLUSION

In this paper, we have proposed an approximate joint fundamental frequency and model order estimator which can be used for a wide range of applications. From a minimum of prior information and by use of a few approximations, the estimator was developed in a Bayesian framework. Using a Laplace approximation, approximate expressions were developed for the posterior distributions on the fundamental frequency and the model order to lower the computational

---

^{2}A Matlab implementation of the algorithm is available at http://kom.aau.dk/~jkn/publications/publications.php

load of the algorithm. The simulations demonstrated that the proposed method gave a better approximation of the posterior distribution on the model than a state-of-the-art ML-based method.

7. REFERENCES

[1] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer, 2 edition, Jun. 1998.

[2] R. J. Sluijter, *The Development of Speech Coding and the First Standard Coder for Public Mobile Telephony*, Ph.D. thesis, Techniche Universitet Eindhoven, 2005.

[3] V. K. Murthy, L. J. Haywood, J. Richardson, R. Kalaba, S. Salzberg, G. Harvey, and D. Vereeke, "Analysis of power spectral densities of electrocardiograms," *Mathematical Biosciences*, vol. 12, no. 1–2, pp. 41–51, Oct. 1971.

[4] L. R. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 25, no. 1, pp. 24–33, Feb. 1977.

[5] A. M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol. 41, no. 2, pp. 293–309, Feb. 1967.

[6] B. G. Quinn and P. J. Thomson, "Estimating the frequency of a periodic function," *Biometrika*, vol. 78, no. 1, pp. 65–74, Mar. 1991.

[7] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1635–1644, Jul. 2007.

[8] M. G. Christensen, J. L. Højvang, A. Jakobsson, and S. H. Jensen, "Joint fundamental frequency and order estimation using optimal filtering," *EURASIP J. on Advances in Signal Processing*, vol. 13, Jun. 2011.

[9] M. Davy, S. J. Godsill, and J. Idier, "Bayesian analysis of polyphonic western tonal music," *J. Acoust. Soc. Am.*, vol. 119, no. 4, pp. 2498–2517, Apr. 2006.

[10] A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 679–694, 2006.

[11] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, 2009.

[12] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Sinusoidal order estimation using angles between subspaces," *EURASIP J. on Advances in Signal Processing*, 2009.

[13] S. L. Marple, Jr., "Computing the discrete-time "analytic" signal via FFT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2600–2603, Sep. 1999.

[14] J. O. Berger and L. R. Pericchi, "Objective Bayesian methods for model selection: Introduction and comparison," *Institute of Mathematical Statistics Lecture Notes – Monograph Series*, vol. 38, pp. 135–207, 2001.

[15] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst. Sci. Cybern.*, vol. 4, no. 3, pp. 227–241, 1968.

[16] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger, "Mixtures of g priors for Bayesian variable selection," *J. Amer. Statistical Assoc.*, vol. 103, pp. 410–423, Mar. 2008.

[17] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publ., 1970.