Multi-Pitch Estimation of Inharmonic Signals

Tommy Nilsson

Master's thesis
2013:E15

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

Abstract

A signal with harmonic structure is often characterized by its fundamental frequency, or pitch. This single parameter contain vital information of a range of applications such as musical transcription, tuning of stringed instruments, speech processing and more. Some signal sources, for instance a stiff vibrating string, exhibit waveforms with slightly deviating harmonic structure. This phenomenon is known as inharmonicity and it complicates the matter of estimating the fundamental frequency. If several signals with this inharmonic structure are added together, for instance for a musical chord from a stringed instrument, we arrive at the problem formulation of this work. How does one estimate the fundamental frequency of a multi-pitch signal containing inharmonicities?

In this work we present a multi-pitch estimator able to handle inharmonic signals. This algorithm uses a source separating technique in order to apply a single pitch estimator to the resulting sub-problems. The performance of the estimator is evaluated and compared to other multi-pitch estimators with good results.

CONTENTS

Acknowledgements

I would like to take this opportunity to thank my supervisor Andreas Jakobsson for all the great support he has provided through out the course of this thesis. Also, I would like to thank Stefan Ingi Adalbjörnsson and Naveed Razzaq Butt. Your participation and advices has been very helpful.

Thank you,

Tommy Nilsson

-Lund, Tuesday 2nd April, 2013

Chapter 1

INTRODUCTION

Signal processing is the mathematical field where signals of various kind are manipulated and analyzed for different purposes. A common type of signals within this field are the periodic ones. An example of such a signal can be seen in figure 1.1.a. A periodic signal repeats itself after an period time τ and with a repetitive frequency $f = \frac{1}{\tau}$. Areas where these signals occur are for instance recordings of sound, data from medical applications and information from radar experiments. In each of these examples, the desired information lies within the signal and it is this data that needs to be extracted in a structured and reproducible way.

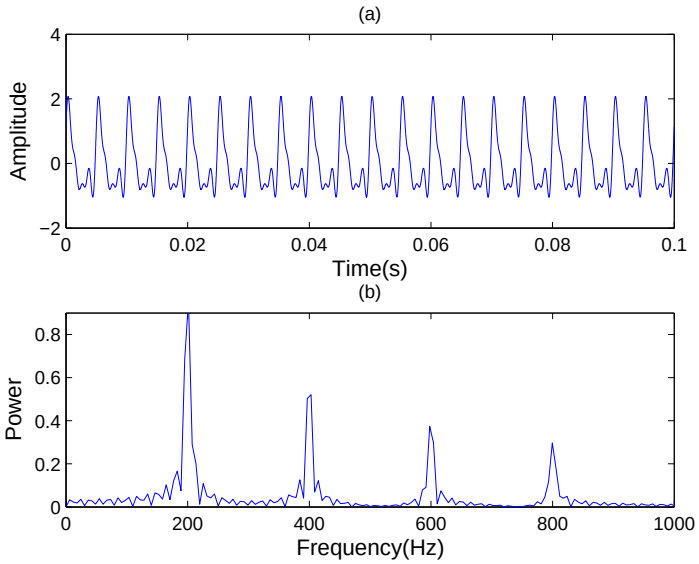

Figure 1.1. A periodic signal consisting of four components with $f_1 = 200$ and $f_2 = 400$, $f_3 = 600$ and $f_4 = 800$ Hz.

This thesis deals with signals with a certain structure which is common for

instance in music and speech. This signal has the property that all periodic components in the waveform can be identified to have a frequency which is a integer multiple of some fundamental frequency, f_0, known as *pitch*. This situation is called the *single-pitch* problem. A signal consisting of two or more pitch sources with this structure is known as *multi-pitch* problems [1] [2]. This occur for instance when two people speak at the same time or when more than one string is sounding in a musical chord.

There are a lot of methods for estimating the fundamental frequency in the single pitch case, and many of these do so very well. However, for the multi-pitch case there are fewer... Some of these methods are presented in [1] and rely on different approaches such as filtering methods, orthogonality of subspaces of covariance matrix and more. In this work, a methodology has been developed which iteratively approximate the multi-pitch problem into separate single-pitch problems. By doing so, a powerful method for single-pitch estimation can be applied for each of the subproblem in order to extract this valuable information.

1.1 Single-pitch

The signals treated in this work will all have a certain structure, as mentioned earlier, known as *harmonic signals*. Firstly there will be a lowest frequency called $fundamental frequency$ or *pitch* and is denoted f_0. Then there will be higher frequency components, f_l, which are called *harmonics*. These harmonics will have a frequency which is a multiple of the pitch

$$f_l = f_0 l \tag{1.1.1}$$

or expressed in angular frequency

$$\omega_l = \omega_0 l \tag{1.1.2}$$

where $\omega_0 = f_0 2\pi$ and $l \in [1, 2, \ldots, L]$. The number of harmonics, L, is termed the *model order*. This parameter is determined by the physical properties of the source. These type of signals are, as mentioned before, common in speech and music since vibrating strings [3] and glottis [5] which produce sound with this specific structure. A typical spectrum of a vibrating string is shown in figure 1.1(b) where the harmonic structure is obvious.

A commonly used model to represent a harmonic signal at a sampling instance is [1]

$$x(n) = \sum_{l=1}^{L} a_l e^{j\omega_l n} + e(n) \tag{1.1.3}$$

where n is the index of a certain time sample in the range $n \in [1\,, 2, \ldots N]$, N is the number of samples, $a_l = A_l e^{i\phi_l}$ is the complex amplitude containing phase ϕ_l and amplitude A_l information of the lth component and $e(n)$ is white Gaussian

noise with variance σ_e^2. The signal is uniformly sampled with a sampling rate f_s. By taking a sub vector with length M of the sampled data and putting them in a vector, the model can be extended to include M consecutive samples

$$\boldsymbol{x}(n) = [x(n) \cdots x(n+M-1)]^T \tag{1.1.4}$$

where $(\cdot)^T$ dentoes the *transpose* operator. These samples can be expressed in a simple form using the matrices $\boldsymbol{z} \in \mathbb{C}^{Mx1}$, $\boldsymbol{Z} \in \mathbb{C}^{MxL}$, $\boldsymbol{a} \in \mathbb{C}^{Lx1}$ and $\boldsymbol{D} \in \mathbb{C}^{LxL}$ defined as

$$\boldsymbol{z}(\omega) = \begin{bmatrix} 1 & e^{j\omega} & \cdots & e^{j\omega(M-1)} \end{bmatrix}^T \tag{1.1.5}$$

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}(\omega) & \cdots & \boldsymbol{z}(\omega L) \end{bmatrix} \tag{1.1.6}$$

$$\boldsymbol{a} = \begin{bmatrix} a_1 & \cdots & a_L \end{bmatrix}^T \tag{1.1.7}$$

$$\mathbf{D} = \begin{pmatrix} e^{j\omega n} & & 0 \\ & \ddots & \\ 0 & & e^{j\omega nL} \end{pmatrix} \tag{1.1.8}$$

The specific structure of $\boldsymbol{Z}$ is known as a $Vandermonde$ structure. Now the vector, $\boldsymbol{x}(n)$, from eq. (1.1.4) containing the sums defined in (1.1.3), can be expressed as

$$\boldsymbol{x}(n) = \boldsymbol{Z}\boldsymbol{D}(n)\boldsymbol{a} + \boldsymbol{e}(n) = \boldsymbol{Z}\boldsymbol{a}(n) + \boldsymbol{e}(n) \tag{1.1.9}$$

where the time variation along the sampling vector has been included in the amplitude term $\boldsymbol{a}(n) = \boldsymbol{D}(n)\boldsymbol{a}$. This model can now be used to represent a single pitch signal with start at sample n through sample $n + M - 1$.

1.2 Multi-pitch

The multi-pitch case, as mentioned earlier, occurs when more then one source is present in a signal at the same time. A typical spectrum of this situation is depicted in figure 1.2. The model presented in section 1.1 can be expanded to include any number of K sources. It could be for instance K persons speaking or K vibrating strings. The corresponding multi-pitch model of (1.1.3) is

$$x(n) = \sum_{k=1}^{K} \sum_{l=1}^{L_k} a_{l,k} e^{j\omega_{l,k} n} + e(n) \tag{1.2.1}$$

where L_k is the model order of source k, k is in the interval $[1, K]$, $a_{l,k}$ is the complex amplitude of the lth component of source k and $\omega_{0,k}$ is the pitch of source k. This model can be expressed with matrices in a similar way to (1.1.9)

$$\boldsymbol{x}(n) = [\boldsymbol{Z}_1 \dots \boldsymbol{Z}_K][\boldsymbol{a}_1(n)^T \dots \boldsymbol{a}_K(n)^T]^T + \boldsymbol{e}(n) \tag{1.2.2}$$

In the general situation some or all model parameter $a_{l,k}$, $\omega_{0,k}$, L_k and K are unknown. The presented models will be used as a generic structure which will be fitted to data via parameter estimation using different algorithms, presented in chapter 2.

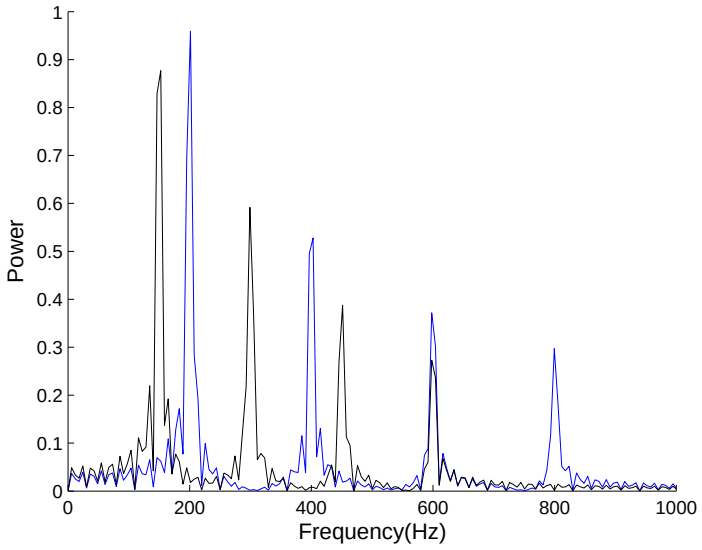

Figure 1.2. Spectrum of a signal containing two sources. The black source contains $f_0 = 150$, $f_1 = 300$, $f_2 = 450$, and $f_3 = 600Hz$. The blue source contains $f_0 = 200$, $f_1 = 400$, $f_2 = 600$, and $f_3 = 800$.

1.3 Inharmonicity

The models described in sections 1.1 and 1.2 rely on perfect harmonic structure, *i.e.*, the harmonics are perfect multiples of the pitch. However, there are exceptions, common for instance in sound where a small deviation from each harmonic is present. This phenomenon is known as *inharmonicity*. One example when these deviations are present is when the signal source is a vibrating string [3], occurring for instance in music containing stringed instruments. In this case, the deviations are due to stiffness of the string. The size of the deviation can be modeled using a physical parameter B which is determined by the dimensions of string k. The shifted frequency can be calculated as

$$\omega_{l,k}(\omega_{0,k}, B_k) = l\omega_0\sqrt{1 + l^2 B_k} \tag{1.3.1}$$

where the parameter B_k is the inharmonicity coefficient of string K and is typically

in the range $[10^{-5}, 10^{-3}]$ [4]. A more general model, including a deviation with unknown structure, is

$$\omega_{l,k}(\omega_{0,k}, \Delta_{l,k}) = \omega_{0,k} l + \Delta_{l,k} \tag{1.3.2}$$

where $\Delta_{l,k}$ denotes the deviation of the lth harmonic of source k and is assumed small; meaning it gives a slight change to the original signal. (1.3.2) can now be substituted into (1.1.3) or (1.2.1) to give a model of a signal with slightly deviating harmonics. To give a feeling of the impact of the inharmonicity caused by a inharmonicity coefficient $B = 0.0005$, the deviation of the harmonics corresponding to the pitches $200, 400, 600, 800$ and 1000 is depicted in figure 1.3. In this picture it is easy to see that this phenomena grows fast for higher model orders.

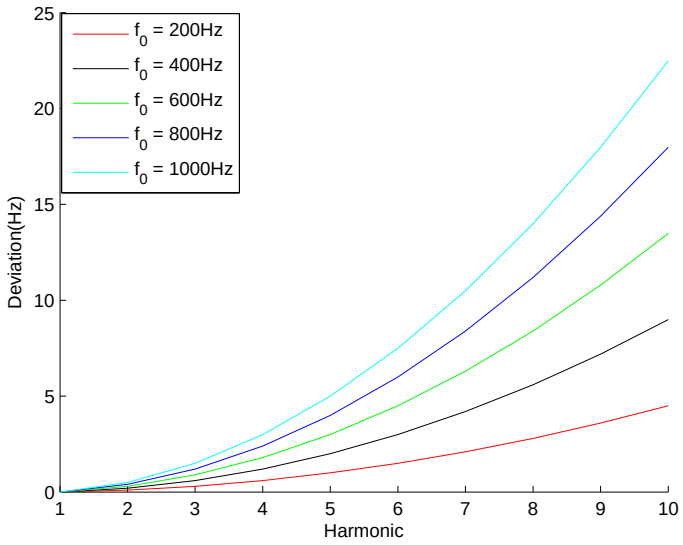

Figure 1.3. The impact of the inharmonicity coefficient $B = 0.0005$ with respect to different harmonics for five pitches, 200, 400, 600, 800 and 1000Hz.

1.4 Aim of the thesis

A drawback of a lot of the existing multi-pitch estimators is that they rely on a perfect harmonic structure of the signal. Therefore, when the inharmonicities are present, problems occur. The produced estimates risks being biased or even not close to the true pitches [6]. The aim of this thesis is to introduce a multi-pitch estimator which is able to estimate the number of sources K, their pitches $\omega_{0,k}$ and the corresponding model order L_k from an inharmonic signal. Different properties

of the estimator will be evaluated such as source detection, model order detection, robustness to noisy environments and more. Comparison with existing multi-pitch estimators will be performed to give a picture of the difficulties that arise when inharmonics are present and how the proposed algorithm overcome them.

Chapter 2

THE PROPOSED ALGORITHM

2.1 Outline of the proposed algorithm

In this chapter, the proposed algorithm is presented. The aim of the method is to estimate the unknown pitches, $\omega_{0,k}$, of a inharmonic multi-pitch signal. In order to do so, the other unknown parameters in (1.2.1), $a_{l,k}$, L_k and K, will also be estimated along the way.

The proposed method for multi-pitch estimation consists of five different sub-algorithms. The theory behind each of these algorithms will be explained in sections 2.2-2.6. However, to set the reader into context before going into details, a outline of the proposed estimator will now be presented. The sub-algorithms used will be treated as *blackbox*-models in this outline, *i.e.*, for a given input, some desired output is provided. The details of how this output is generated is not presented initially, rather only the purpose of each block is outlined. A flowchart of the proposed method can be seen in Figure 2.1. The different steps from input to output of the proposed method will now be described:

1. The first step is based on the recently introduced $PEBS$-algorithm. This algorithm takes the signal to be analyzed as input and output a set of estimates of candidate pitches, $\omega_{0,k}$, and their respective model orders, L_k. Candidate pitches refer to frequencies that might be a possible true pitch. $PEBS$ do not give any attention to the inharmonicity. This phenomenon will be accounted for in step 3. Therefore, the result of this step can be viewed as initial estimates, which are to be refined.

2. Next is to decide which of the frequencies in the candidate set from the previous step that are likely to be true pitches of the signal. This is done with a BIC-based criterion. This step takes the estimated set of candidate frequencies and their model orders as input and outputs a statement about which pitches might be the true fundamental frequencies.

3. The third step in the proposed method is to take the coarse estimates of the

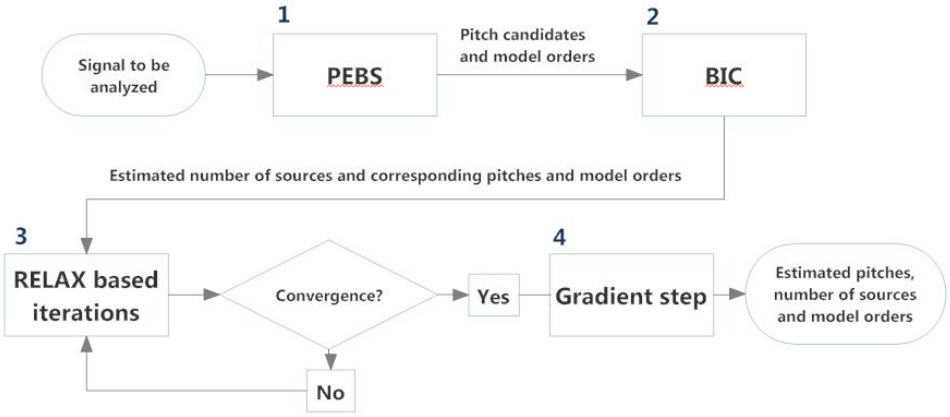

Figure 2.1. Flowchart of the proposed algorithm.

assumed true pitches from previous step and refine them using the $RELAX$ and RCP-algorithms in a iterative combination with each other. RCP is an algorithm which estimates the pitch of a single source signal suffering from inharmonicity. The $RELAX$ step is used to approximate the multi-pitch problem as a set of single-pitch problems. This source separation procedure will yield better results if good estimates of the pitch and harmonics are available. The iterations starts out by separating the different sources, using $RELAX$ with the rough initial estimates from step 2. Thereafter, RCP is applied to the resulting sub-problems in order to extract their pitch information. The first iteration is now done. Next iteration is carried out in the same way, but with the new refined pitch estimations from previous step as initial values which gives a more accurate separation procedure, yielding better conditions for RCP to improve the estimates further. These iterations goes on until some convergence criterion is met.

4. A last step of narrow range *gradient search* is performed to refine the estimates further. The input is the estimated pitches from step 3 and output is refined estimates.

Note that it is not until step 3 that the inharmonicity is given any attention.

2.2 Step 1 - Estimate candidate pitches using PEBS

The $PEBS$ (Pitch Estimation using Block Sparsity) algorithm was recently introduced in [7]. The idea of the algorithm is to, given a set of test frequencies, create a sparse solution to some cost function that indicates which sources that are present in the signal, *i.e.*, the algorithm should be able to select a few pitches from a set

of test frequencies which is capturing dominant properties in the waveform. The solution is obtained via minimization over the amplitudes, corresponding to every pitch candidate in the test set and corresponding harmonics, using the following optimization problem

$$\min_{\boldsymbol{a}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{a}\|_2^2 + \lambda\|\boldsymbol{a}\|_1 + \alpha\sum_{k=1}^{P}\sqrt{\Delta}_k\|\boldsymbol{a}_k\|_2 \tag{2.2.1}$$

where $\boldsymbol{y}$ is the observed data, $\|\cdot\|_p$ denotes the p-norm, λ and α are tuning parameters which decides how the penalties are weighted, P is the number of test frequencies, $\boldsymbol{W}$ is a matrix containing blocks of $Vandermonde$ matrices, $\boldsymbol{Z}_k$, which represent each pitch and its harmonics in the test set, $\boldsymbol{Z}_k$ is the block corresponding to test frequency ω_k. To meet the Nyquist criterion for the harmonics, the block size of blocks corresponding to higher pitch candidates will decrease. This decaying block size will give a disadvantage to higher pitch frequencies since fewer harmonics can be fitted in the cost function in (2.2.1). Therefore, the penalty $\sqrt{\Delta}_k$ is introduced to even out this drawback, where Δ_k is the number of harmonics in block k. Lastly, $\boldsymbol{a}$ is a vector with sub vectors containing amplitudes of corresponding $\boldsymbol{Z}_k$-block. Thus,

$$\boldsymbol{Z}_k = \begin{bmatrix} \boldsymbol{z}(\omega_k) & \cdots & \boldsymbol{z}(\omega_k L) \end{bmatrix} \tag{2.2.2}$$

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{Z}_1 & \cdots & \boldsymbol{Z}_P \end{bmatrix} \tag{2.2.3}$$

$$\boldsymbol{a} = \begin{bmatrix} \boldsymbol{a}_1^T & \cdots & \boldsymbol{a}_P^T \end{bmatrix}^T \tag{2.2.4}$$

The result after minimization of the cost function (2.2.1) can be visualized to the user as a pseudo spectrum formed by taking the 2-norm of every amplitude block, $\boldsymbol{a}_k$ and plot it against the corresponding test frequency value. An example of such a spectrum can be seen in Figure 2.2. A true pitch is likely to be close to the pitch candidate, ω_k, corresponding to block $\boldsymbol{a}_k$ when its 2-norm yields a large value.

A very nice feature of the $PEBS$-algorithm is that it estimates the model order, L_k, of each source very well. The estimation is performed by looking at the estimates, $a_{l,k}$, of each amplitude block, $\boldsymbol{a}_k$, and approve all elements above some threshold as a harmonic. The model order of each block is then estimated as the number of elements that satisfies this condition, *i.e.*,

$$\hat{L}_k = \sum_{l=1}^{L_{max}} u\big[a_{l,k} - 0.01\max(\boldsymbol{a}_k)\big] \tag{2.2.5}$$

where L_{max} is the number of columns in $\boldsymbol{Z}_k$ and $u[x]$ is the indicator function taking the value 1 if $x > 0$and zero otherwise. The threshold is set to a hundredth of the maximum value within each block.

The minimization in 2.2.1 can be solved using $ADMM$(Alternating Directions Method of Multipliers). For more details on how the optimization problem is solved,

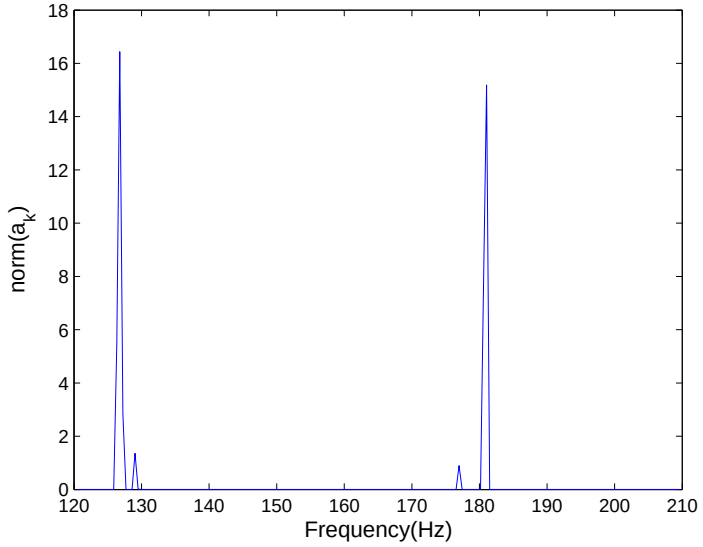

Figure 2.2. A typical resulting pseudo spectrum after pitch estimation with $PEBS$.

the reader is referred to [7] and [9]. In this work the tuning parameters where set to $c = 1/3$, $\chi = 0.2$, $\alpha = c\chi$ and $\lambda = (1 - c)\chi$ in accordance with [7].

Lastly, it is worth noting that the maximal number of harmonics is restricted by

$$L_k < \left\lfloor \frac{1}{f_0} \right\rfloor \text{ or } \left\lfloor \frac{2\pi}{\omega_0} \right\rfloor \tag{2.2.6}$$

since all harmonics must fit into the normalized frequencies interval [-0.5 0.5] in order to avoid aliasing effects.

2.3 Step 2 - Select the number of sources using BIC

The algorithm presented in section 2.2 gives a set of pitch candidates and their model orders, L_k. To be able to determine the actual number of sources in the signal, *i.e.*, which of the candidates should be considered true pitches, this work uses a BIC-like criterion [14,17], proposed in [8]. BIC (Bayesian Information Criterion) is used for model order estimation in a range of statistical applications, such as control theory [18] and time series analysis [14]. The desired model order in this case is the number of sources, K. This should not be confused with the model order of each source, which is L_k, which was estimated in section 2.2. The criterion, in its

general form, is used to estimate the unknown model order for some signal with known structure. The criterion is [14]

$$BIC(n) = -2\ln(p_n(\boldsymbol{y}, \hat{\boldsymbol{\theta}}^n)) + \ln(|I_{\theta_n}|) \tag{2.3.1}$$

where n is the assumed model order, $\boldsymbol{y}$ is the data vector with length N, $p_n(\boldsymbol{y}, \hat{\boldsymbol{\theta}}^n)$ is the *probability density function* of $\boldsymbol{y}$ with parameter vector $\hat{\boldsymbol{\theta}}^n$ of size n and I_{θ_n} is the $Fisher\,Information$ (FI). Now this criterion can be applied to a range parameter vector sizes $n \in [1, \bar{n}]$, where $\bar{n}$ is the maximum assumed number of unknown parameters. The n that minimize (2.3.1) is chosen as the model order. The general criterion in (2.3.1) can be applied to the model of a periodic signal (for details of how this is done, the reader is referred to [17]). Doing this, the BIC-function takes the following form

$$BIC(n_c) = 2N\ln(\sigma_{\boldsymbol{y}}^2) + (5n_c + 1)\ln(N) \tag{2.3.2}$$

where n_c is the number of periodic components in the signal, $\sigma_{\boldsymbol{y}}^2$ is the variance of the signal after the first n_c periodic components has been removed. Here, the main interest lies in estimating the number of sources. To do so, we proceed as follows: Instead of subtracting a single component, a pitch and all corresponding harmonics is removed at the same time, *i.e.*, eliminating a whole source. This would suggest that in order to estimate the number of sources, one may form the following BIC-function

$$BIC(K) = 2N\ln(\sigma_{\boldsymbol{y}}^2) + (5H_K + 1)\ln(N) \tag{2.3.3}$$

where $H_K = \sum_{k=1}^{K} L_k$. Using this methodology, the estimated number of sources, $\hat{K}$ is given as the minimum of (2.3.3). Figure 2.3 gives an example of such a BIC-curve in the case of $K = 2$ sources, clearly showing that the curve is minimized for the correct number of sources. The performance of the suggested method will be further investigated in chapter 3.

When evaluating the BIC-function for different model orders, $\sigma_{\boldsymbol{y}}^2$ needs to be calculated from the reduced original signal, $\boldsymbol{y}$. How to obtain the reduced signal will now be explained. Firstly, the amplitudes of the components to be subtracted must be estimated. The estimation is done using Least Squares(LS) [11]. The idea of LS is to find the estimates of the complex amplitude $\boldsymbol{a}$ that minimizes the error norm between the signal and a model, *i.e.*,

$$\|\boldsymbol{e}\|_2 = \|\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{a}\|_2 \tag{2.3.4}$$

for a certain vector of frequencies $\boldsymbol{\omega}$, entering in the structure matrix $\boldsymbol{Z}(\boldsymbol{\omega})$. The estimates of $\boldsymbol{a}$ is found as [11]

$$\hat{\boldsymbol{a}} = (\boldsymbol{Z}^H\boldsymbol{Z})^{-1}\boldsymbol{Z}^H\boldsymbol{y} \tag{2.3.5}$$

where $\hat{\boldsymbol{a}}$ is the estimated complex amplitudes and $(\cdot)^H$ is the complex transpose. The reduced signal can now be calculated in the following way

$$\boldsymbol{y}_{reduced} = \boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{a}} \tag{2.3.6}$$

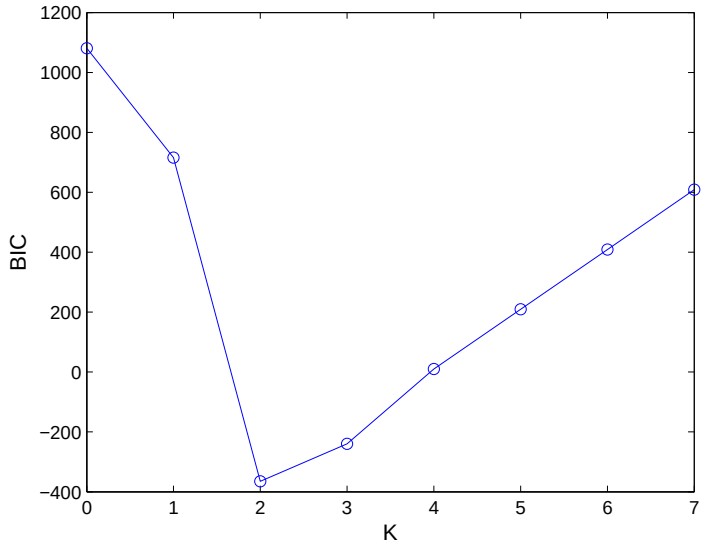

Figure 2.3. A BIC-curve showing the evaluated cost function for different number of sources. In this case, there are two pitches present in the signal, indicated by the location of the minimum of the curve.

2.4 Preliminaries of step 3 - Accounting for inharmonicity using RCP

The RCP (Robust Covariance fitting for Pitch estimation) algorithm is a convex optimization technique used to find the pitch and harmonics of a single source signal suffering from inharmonicity of the form (1.3.2). This algorithm is introduced in this section, and will be used in step 3, presented section 2.5. Since this estimator only applies to single pitch problems, all source indexing will be dropped in this section, *i.e.*, k and K will be omitted.

The formulation of the convex optimization problem, leading to the frequency estimates, will now be presented. For more details on how the optimization problem is solved, the reader referred to [8]. The idea of the algorithm is to find the pitch and harmonics which maximally explains the observed signal power defined as [8]

$$\log(\det(\boldsymbol{R})) = \log(\det(\mathrm{E}(\boldsymbol{y}(n)\boldsymbol{y}(n)^H))) = \log(\det(\boldsymbol{Z}_\Delta \boldsymbol{P} \boldsymbol{Z}_\Delta^H + \sigma_e^2 \boldsymbol{I})) \tag{2.4.1}$$

where $\boldsymbol{R}$ is the covariance matrix of the signal $\boldsymbol{y}$ estimated as

$$\hat{\boldsymbol{R}} = \frac{1}{N-M+1} \sum_{n=0}^{N-M} \boldsymbol{y}(n)\boldsymbol{y}^H(n) \tag{2.4.2}$$

Further, $\mathrm{E}(\cdot)$ is the expectation value operator, $\boldsymbol{P} = \mathrm{diag}(|A_1|^2 \dots |A_L|^2)$, σ_e^2 is the noise variance, $\boldsymbol{Z}_\Delta \in \mathbb{C}^{M \times L}$ is the *Vandermonde*-structured matrix from (1.1.6) with a small deviation Δ_l added to each frequency argument ω_l, such that

$$\boldsymbol{Z}_\Delta = [\boldsymbol{z}(\omega_1 + \Delta_1) \cdots \boldsymbol{z}(\omega_L + \Delta_L)] \tag{2.4.3}$$

The maximization problem (2.4.1) will be be subjected to three constraints in order to ensure a valid solution:

First constraint : The purpose of the first constraint is to set a bound on how large the deviations from the initial points can be, *i.e.*, how much the pitch and each harmonic is allowed to be altered during the optimization. This is regulated by some ϵ_l, restricting the size of the deviation via the 2-norm of the difference between the initial and the optimized point. This constraint can be written in a mathematical form as

$$\|(\boldsymbol{Z}_\Delta - \boldsymbol{Z}_{init})\boldsymbol{e}_l\|_2 \leq \epsilon_l \tag{2.4.4}$$

where $\boldsymbol{Z}_{init} \in \mathbb{C}^{M \times L}$ is again a *Vandermonde*-structured matrix, now constituting the initial point of the optimization

$$\boldsymbol{Z}_{init} = [\boldsymbol{z}(\omega_1) \cdots \boldsymbol{z}(\omega_L)] \tag{2.4.5}$$

and $\boldsymbol{e}_l$ is the lth column vector of the $L \times L$ identity matrix. ϵ_l should be chosen to reflect the assumed level of inharmonicity.

Second constraint : Next constraint ensures that the optimized parameters preserve the positive semi-definiteness property, which is necessary for a valid covariance matrix [14]

$$\boldsymbol{Z}_\Delta \boldsymbol{P} \boldsymbol{Z}_\Delta^H + \sigma_e^2 \boldsymbol{I} \preceq \hat{\boldsymbol{R}} \tag{2.4.6}$$

with $\boldsymbol{A} \preceq \boldsymbol{B}$ denoting that $\boldsymbol{B} - \boldsymbol{A}$ is positive semidefinite.

Third constraint : The last constraint ensures the positiveness of each estimated amplitude and the diagonality the the matrix $\boldsymbol{P}$ which is necessary for the decomposition of the covariance matrix to be valid, *i.e.*,

$$\boldsymbol{P} = \boldsymbol{P} \odot \boldsymbol{I} \succeq \boldsymbol{0} \tag{2.4.7}$$

where I is the $L \times L$ identity matrix and $\odot$ is the $Schur - Hadamard$ element wise operator.

Finally, the convex optimization problem can be formulated as

$$\begin{aligned} \max_{\boldsymbol{Z}_\Delta, \boldsymbol{P}, \sigma_e^2} \quad & \log(\det(\boldsymbol{Z}_\Delta \boldsymbol{P} \boldsymbol{Z}_\Delta^H + \sigma_e^2 \boldsymbol{I})) \\ \text{subject to} \quad & \boldsymbol{Z}_\Delta \boldsymbol{P} \boldsymbol{Z}_\Delta^H + \sigma_e^2 \boldsymbol{I} \preceq \hat{\boldsymbol{R}} \\ & \|(\boldsymbol{Z}_\Delta - \boldsymbol{Z})\boldsymbol{e}_l\| \leq \epsilon_l \\ & \boldsymbol{P} = \boldsymbol{P} \odot \boldsymbol{I} \succeq \boldsymbol{0} \end{aligned} \tag{2.4.8}$$

As mentioned earlier, for details on how this optimization problem is solved, the reader is referred to [8].

2.5 Step 3 - RELAX-based iterations with RCP

We now proceed to present the proposed $RELAX$-based estimation scheme with purpose of transforming the multi-pitch problem into separate single-pitch problems. These sub-problems can thereafter be treated with the single-pitch estimation algorithm RCP, presented in section 2.4. The method is inspired by the $RELAX$-algorithm used in [10].

To help the reader grasp the concept of the following section, an algorithmic presentation of step 3 can be seen in Algorithm 1. After step 2, presented in section 2.3, coarse estimates of the true pitches and their model orders are given. The idea of the this step is to subtract all but one of these sources from the signal to obtain a approximately single-pitch problem (the procedure of subtracting a source from the signal is described in section 2.3). The frequency content of this reduced signal is estimated with the single-pitch estimator RCP. The same procedure is applied to all sources. These estimates are assumed to be better than the rough initial estimates. Now the first iteration of the algorithm is finished. The following iterations is performed in the same way, but with the new refined frequency estimates in the source removal process. Since the refined estimates ought to be closer to the true pitches, the source removal procedure should exhibit a better approximation of the multi to single-pitch problem, giving better conditions for the RCP-algorithm to produce better estimates.

In order to know if the algorithm has converged, a stopping or convergence criterion is needed. The one used is the following, proposed in [10],

$$\frac{|\ \|y_{i-1}\| - \|y_i\|\ |}{\|y_{i-1}\|} < \epsilon \tag{2.5.1}$$

where y_i and y_{i-1} are the original signals with pitches and corresponding harmonics, found after iteration i respective $i-1$, removed. The idea of this criterion is to indicate the degree of change between every iteration. If the change is smaller than some ϵ, then the algorithm is assumed to have found some stationary points hopefully close to the true pitches. In this work, ϵ was chosen as 0.01.

Algorithm 1 Step 3; refine rough estimates.

```
Initialize with rough estimates of pitches formed using PEBS
while not converged do
  for k = 1 to K̂ do
    Subtract all but source k from the signal
    Estimate ω_{0,k} and ω_{l,k}, l, ∈ [1, L̂_k] using RCP
  end for
end while
```

2.6 Step 4 - Gradient search

As a final step of the proposed algorithm, a *gradient search* [12], is performed to enhance the estimates of the pitches further. This step is applied to each of the approximated single-pitch problems after the convergence criterion of $RELAX$, explained in section 2.5, has been meet, *i.e.*, the best possible approximation of the multi to single-pitch transformation. Therefore, all source indexing is omitted in this section.

The general idea of gradient search is to evaluate some cost function for a set of parameter values in the vicinity of a initial point. The parameter value in this set that yields the maximum or minimum (whatever is desired) of the cost function is taken as the new parameter value. In this thesis, the estimated pitches $\omega_{0,k}$, from previous step, will be taken as initial points and the 2-norm of the difference between the observed signal $\boldsymbol{y}$ and the estimated model $\boldsymbol{Za}$, denoted $\boldsymbol{e}$, will constitute the cost function

$$\|\boldsymbol{e}\|_2^2 = \|\boldsymbol{y} - \boldsymbol{Za}\|_2^2 = (\boldsymbol{y} - \boldsymbol{Za})^H(\boldsymbol{y} - \boldsymbol{Za}) \tag{2.6.1}$$

If the model resembles the true content of the signal, then (2.6.1) should be close to zero. The aim is to find a pitch within a set $\bar{\boldsymbol{\omega}}_0$, defined below, that exhibit the lowest value of the cost function.

There are different approaches to choose the set of values in the vicinity of the initial point, at which the cost function is evaluated. In this work the set is selected as the values along the direction of the negative gradient of (2.6.1), $\boldsymbol{d}$. This choice is known as *steepest decent* [12]. The set is now constituted by

$$\bar{\boldsymbol{\omega}}_0 = \omega_0 + d\boldsymbol{\theta} \tag{2.6.2}$$

where $\boldsymbol{\theta}$ is a tight equidistant grid (in this work, $\boldsymbol{\theta}$ consists of 200 values with spacing 10^{-8}) and the direction $\boldsymbol{d}$

$$\boldsymbol{d} = -\nabla_{\omega_0}\|\boldsymbol{e}\|_2 = -\nabla_{\omega_0}\|\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{a}\|_2 = -\nabla_{\omega_0}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{a})^H(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{a}) \tag{2.6.3}$$

To write the expression for $\boldsymbol{d}$ in a compact form, the following vectors are used

$$\mathbf{Y} = \begin{bmatrix} 0 & i & \cdots & i(M-1) \end{bmatrix} \tag{2.6.4}$$

$$\boldsymbol{z}(\omega) = [1\ e^{j\omega} \cdots\ e^{j\omega(M-1)}]^T \tag{2.6.5}$$

where (2.6.5) is the same vector defined in (1.1.5). Now, (2.6.3) may be expanded into

$$\nabla_{\omega_0}\|\boldsymbol{e}\|_2 = \nabla_{\omega_0}(\boldsymbol{y}^H\boldsymbol{y} - \boldsymbol{y}^H\boldsymbol{Z}\boldsymbol{a} - \boldsymbol{a}^H\boldsymbol{Z}^H\boldsymbol{y} + \boldsymbol{a}^H\boldsymbol{Z}^H\boldsymbol{Z}\boldsymbol{a}) \tag{2.6.6}$$

yielding the gradients

$$\nabla_{\omega_0}(\boldsymbol{y}^H\boldsymbol{y}) = 0 \tag{2.6.7}$$

$$\nabla_{\omega_0}(\boldsymbol{y}^H\boldsymbol{Z}\boldsymbol{a}) = a_1\boldsymbol{Y}\boldsymbol{y} \odot \boldsymbol{z}(\omega_0) \tag{2.6.8}$$

$$\nabla_{\omega_0}(\boldsymbol{x}^H\boldsymbol{Z}^H\boldsymbol{y}) = -a_1\boldsymbol{Y}\boldsymbol{y} \odot \boldsymbol{z}(-\omega_0) \tag{2.6.9}$$

$$\nabla_{\omega_0}(\boldsymbol{x}^H\boldsymbol{Z}^H\boldsymbol{A}\boldsymbol{x}) = -a_1^H\sum_{l=2}^{L} a_l\boldsymbol{Y}\boldsymbol{z}(\omega_l - \omega_0) + a_1\sum_{l=2}^{L} a_l^H\boldsymbol{Y}\boldsymbol{z}(\omega_0 - \omega_l) \tag{2.6.10}$$

The gradient, $\boldsymbol{d}$, is calculated as the sum of the expressions (2.6.7) - (2.6.10). Now, the value θ_{min} in $\boldsymbol{\theta}$ which exhibits the minimum, in combination with the direction $\boldsymbol{d}$ and the initial point ω_0, gives the refined estimate $\bar{\omega}_0$ via

$$\bar{\omega}_0 = \omega_0 + \boldsymbol{d}\theta_{min} \tag{2.6.11}$$

Chapter 3

TESTING OF THE DIFFERENT SUB ALGORITHMS

In this chapter, the performance of vital parts of the proposed algorithm will be investigated. Some advantages and disadvantages of the different parts will be pointed out and discussed.

3.1 PEBS

3.1.1 Estimate model order

The rate at which the algorithm is able to find the right model order, L_k, is investigated via simulations. The experiment is performed for different number of sources to show how the rate differs with respect to K. A total of 250 simulations are made for each number of sources from $K = 1$ to 5. The number of harmonics corresponding to each pitch is a random integer $U \in [3, 9]$. The result is calculated as the percentage of the different deviations from the true model order, *e.g.*, if the deviation $L_k - \hat{L}_k = 1$ occurs 50 times for $K = 2$, the percentage is $50/(250 * 2) = 0.1 * 100\%$ since there are a total of $250 * 2$ model orders to be estimated. The result can be seen in Figure 3.1. From this figure, one can see that when the number of sources increase, the accuracy of the algorithm decrease. This makes sense since more sources will give a more messy signal in contrast to the single pitch case (K=1) when the model order is well estimated.

The ability to estimate the model order perfectly every time would of course be a very nice feature. This is unfortunately not achieved with the proposed algorithm which can be seen from the simulations in Figure 3.1. However, in most of the signals observed during this work, the amplitudes of the higher harmonics are just a small fraction of the pitches amplitudes. If the model order is slightly off, this will not have any major impact on the final result since the missed out components have minor impact on the signal. From Figure 3.1 it can be seen that the density

of the estimated model orders lies in ±1 from the true value.

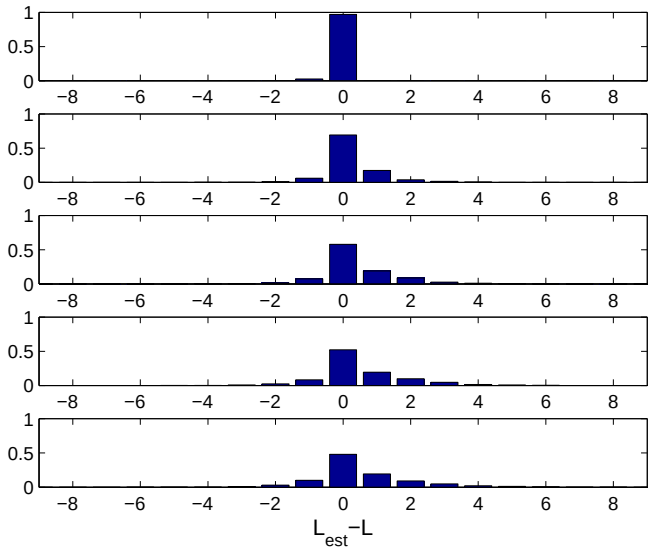

Figure 3.1. Percentage of difference between true and estimated model order, as the number of sources grow from one, at the top, to five, at the bottom.

3.1.2 Peak-splittings for large inharmonicities

When signals deviate from the perfect harmonic structure, presented in section 1.1, $PEBS$ will exhibit some problems. A simulation of a signal containing two pitches, both with inharmonicity coefficient $B = 0.0005$ and $L = 8$, is performed to illustrate the consequences. The result can be seen in the lower part of figure 3.2, in which it can be seen that where there should be one peak, indicating a pitch, is now multiple peaks. This might now be interpreted as multiple pitches, which is incorrect. The splitting is due to the fact that when the location of the harmonics drift, more frequencies in the test set will fit better into the cost function (2.2.1). To avoid this problem when estimating pitch candidates, the number of frequencies in the test set is decreased, making the set more sparse. In this work, 40 data points in the normalized frequency range [0.02 0.045] were used. The result of doing this can be seen in the upper part of figure 3.2. Decreasing the resolution is no problem since this step of the algorithm is supposed only to give a coarse estimation of the pitches.

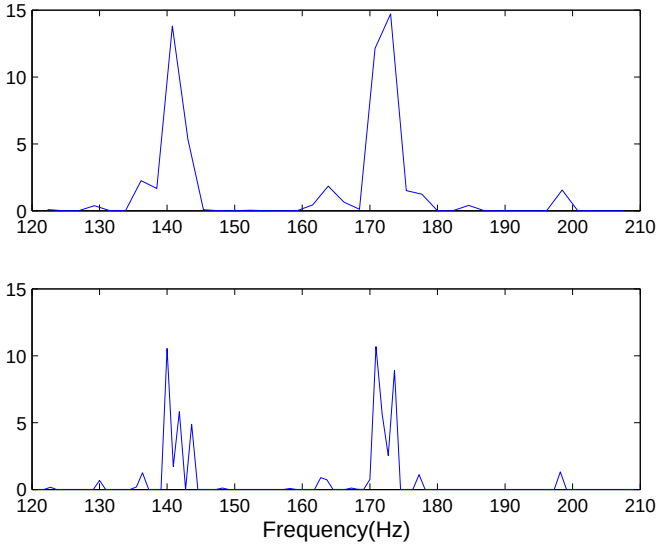

Figure 3.2. A peak, indicating a pitch, splits into two peaks when large inharmonics are present in the signal.

3.1.3 Halvlings

When PEBS is used to estimate pitch candidates, there are some test frequencies that are more likely to get wrongly favored than others by the cost function in (2.2.1). One particular false pitch candidate that occur more often then others are called *halvlings* which have a frequency half of the true pitch. The reason is that this candidate frequency share every other harmonic, starting from the second of the true pitch, with the true fundamental frequency. Therefore, the halvlings will have a good fit into the minimization criterion (2.2.1). In Figure 3.3, a typical situation can be seen where a halvling is located at $80Hz$ and the true pitch at $160Hz$. At a first glance, one would wrongly mistake the former peak for a pitch. This error can be avoided by studying the estimated complex amplitudes in the $\boldsymbol{a}_k$ vector described in chapter 2, section 2.2. If the first element of this vector is considerably smaller then the second element, then the peak is likely to be a halvling. In this work, a peak was assumed to be a halvling if the first element was less than a hundredth of the second one.

3.2 BIC

To evaluate the proposed BIC-based model order estimator's ability to find the right number of pitch sources, 250 simulations are performed for each number of

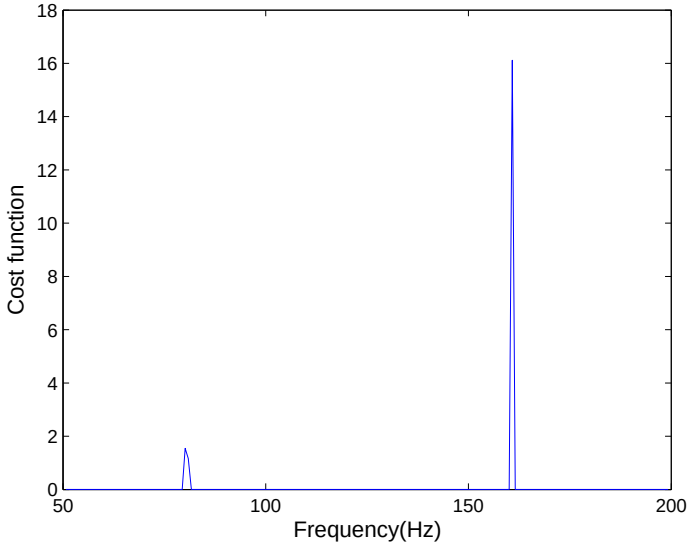

Figure 3.3. A case where the problem of halvlings occur. The true pitch is located at $160Hz$ and the halvling at $80Hz$.

sources, K, in the interval $[1, 5]$. In every simulation, the deviation of $\hat{K}$ from the true value, K, is calculated in a similar fashion to how the estimation of L_k was evaluated in section 3.1.1. The outcome of the experiment is displayed as the percentage of each deviation occurring for every K. The test signal suffers from inharmonicity with an inharmonicity coefficient, B, which is $U \in [0, 0.0005]$. The number of harmonics of each pitch is $U \in [3, 9]$. The result can be seen in figure 3.4. The conclusion from this test is that it is easier to find the right number of sources if they are fewer.

3.3 RELAX and RCP

3.3.1 Source separation

The purpose of the $RELAX$-algorithm is, as mentioned before, to separate a multi-pitch case into separate single-pitch cases. In Figure 3.5 the original spectrum of a recording of a guitar playing a C chord, containing 3 sources, and their separations are depicted. The three sources, constituting the C chord, is $C = 130.8Hz$, $E = 164.8Hz$ and $G = 196.0Hz$. It can be seen that a busy multi-pitch problem has been divided into three sub problems where the dominant harmonic structure of each single pitch is obvious. After the source separation procedure, there are still some residuals of other sources left, as can be seen in the pictures. However, the

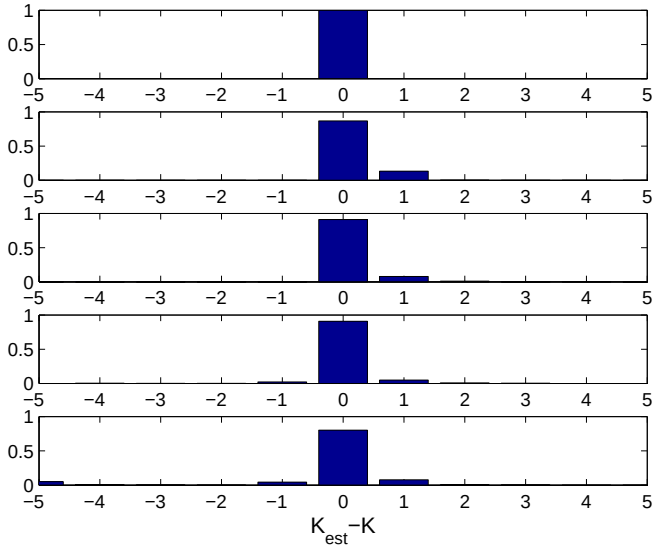

Figure 3.4. Evaluation of BIC criterion as the number of sources grows from single-source, at the top, to five sources, at the bottom. On the x-axis, the deviation from the true number of sources and on the y-axis percentage of each deviation.

aim is to remove dominant characteristics of other sources and that each pitch is separated in a satisfying way.

3.3.2 Convergence to the true pitches

Throughout the $RELAX$-based iterations, the estimated pitches will ideally converge towards the true fundamental frequencies. This convergence typically starts out with a large step from the initial value, given by $PEBS$, towards the true pitch. The following iterations are typically smaller, adjusting the pitch until the convergence criterion (2.5.1) is met. In Figure 3.6 two typical convergence curves are displayed for a signal containing two pitches, showing how the estimates converges towards the true fundamental frequencies. In the upper figure, the first iteration yield a small improvement in the estimate of $\omega_{0,1}$. However, in the second iteration, the sources separation procedure has been improved since a better estimate of $\omega_{0,2}$ is provided from the first step, now allowing $\omega_{0,1}$ to be estimated more accurate. There is a possibility that $\omega_{0,1}$ could be refined further by performing more iterations. However, since the convergence criterion is met after iteration 4, the procedure is terminated. To obtain more accurate estimates, the user parameter controlling the level of change between iterations, ϵ, in (2.5.1) can be lowered, yielding better estimates but a longer execution time.

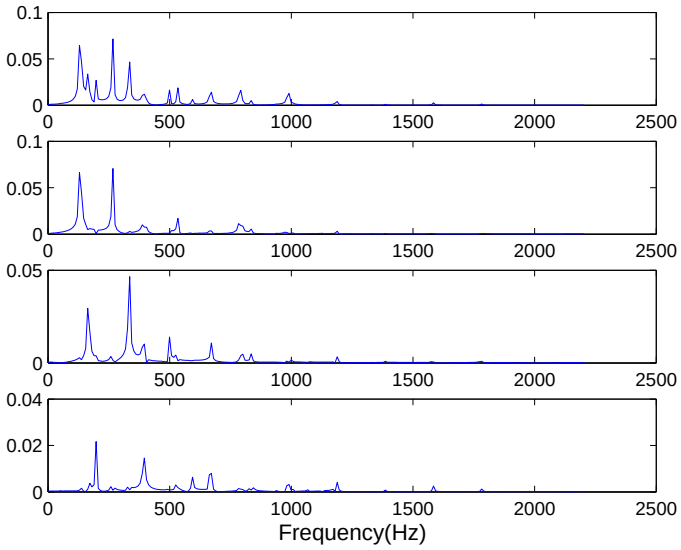

Figure 3.5. *RELAX*-based iterations separating the multi pitch case into three single pitch cases.

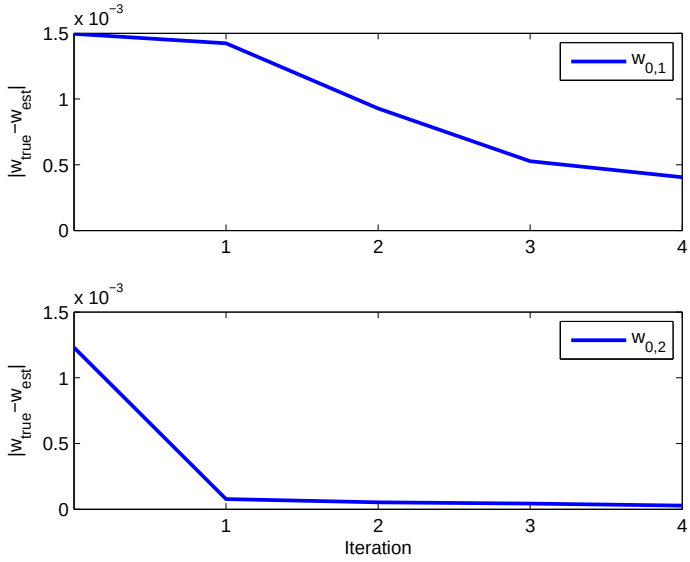

Figure 3.6. The iterative convergence of two pitches during step 3 of the algorithm.

Chapter 4

EVALUATION OF RIME

The algorithm developed in this work is given the name $RIME$ (Robust Inharmonicity-based Multi-pitch Estimator). In this chapter, the proposed algorithm will be tested on real and simulated signals to evaluate its performance.

4.1 Stationary signals

In the ideal case, the signals analyzed in this work should be stationary. Stationarity means that the data which is to be analyzed should have the same properties throughout the whole data sequence. For a strict definition of stationarity, the reader is referred to [13]. These kind of signals can easily be achieved when one is working with simulated data. However, when working with real signals, things change continuously, and there is no guarantee that the frequency or any other parameter is constant over time. Actually, it is more likely that things do change, *e.g.*, a musician changes notes or a person doesn't speak with a perfect monotonic voice, pronouncing only the same vowel continuously. The way this problem is approached is to divide the data into smaller sets which have a size that hopefully exhibit a stationary behavior. This size depends heavily on the nature of the source. For instance, is it common to play more then 20 notes per second on a guitar or piano? If not, maybe $50ms$ data segments will work well. A new problem that arises in this area is how to decide the time instance when a new note is struck, *i.e.*, when does the $50ms$ data window that does exhibit stationarity start? A few different methods for this problem are proposed in [15].

4.2 Simulated signals

In this section, simulations will be performed to investigate the performance of $RIME$. Different properties are examined via Monte Carlo simulations in order to statistically verify the estimator. This is done by performing a number of J simulations with each having a random phase, amplitude and noise components for every $j \in [1, J]$. Thereafter, the *Root Mean Square Error* is calculated to give a measure of the deviation of the estimated frequency from the true pitch

$$RMSE = \sqrt{\frac{1}{J}\sum_{j=1}^{J}(\omega_{0,1}^{j} - \widehat{\omega}_{0,1}^{j}) + \cdots + (\omega_{0,K}^{j} - \widehat{\omega}_{0,K}^{j})} \tag{4.2.1}$$

where $\omega_{0,k}^{j}$ and $\widehat{\omega}_{0,k}^{j}$ are the true respective the estimated frequency of source k in run j. For all experiments, in the following sections, $J = 250$, $K = 2$, $L_k \in U[3, 9]$, the amplitudes and phases are $U \in [0, 1]$ and $U \in [0, 2\pi]$, respective, with one exception in that $a_{k,1} = \max(\text{rand}(10))$, which gives the largest values out of 10 random numbers in the interval $[0, 1]$ to ensure no ambiguity in pitch location. The properties of the noise sequences is different in the various experiments, therefore details of this matter is explained in the different sections.

In section 4.2.1, a comparison with existing multi-pitch estimators is performed with respect to the level of inharmonicity present. In sections 4.2.2, the proposed estimators ability to estimate the fundamental frequencies when different levels of noise is present will be examined.

4.2.1 Performance with respect to different levels of inharmonicity

In this section, comparison of $RIME$ with other multi-pitch estimators will be performed. Some strengths and weaknesses of $RIME$ will be pointed out. The competing estimators are the ordinary $PEBS$(with dense pitch candidate set and not the sparse set, as explained in section 3.1.2) and two algorithms, *Optimal Filtering* and $MUSIC$, taken from the software package provided with [1]. These algorithms are briefly described as:

Optimal filtering : This method creates a notch filter with stop band at a pitch candidate and its harmonics. This filter is applied to the signal and the power of the resulting waveform is calculated. The same procedure is applied for a set of candidate pitches. If the a pitch candidate in the test set is close to a true pitch, then the signal power will decrease significantly since dominant components has been filtered out. The resulting power vector is plotted against the corresponding frequencies. Where such a figure has a valley, a pitch is assumed to be located.

MUSIC : (MUltiple SIgnal Classification) This method rely on the orthogonality between the noise subspace of a signals covariance matrix R and the $Vandermonde$ matrix $Z(\omega_0)$, where ω_0 is the true pitch and $Z(\omega_0)$ is of the from (1.1.6). The noise subspace of R is denoted U. For a true pitch, ω_0, the orthogonality gives $\|Z^H(\omega_0)U\|_F^2 = 0$. This expression is evaluated for a set of test frequencies $\boldsymbol{\omega}$. The frequencies in $\boldsymbol{\omega}$ that exhibits most orthogonality, *i.e.*, gives values close to zero of the $frobenius$-norm, is chosen as the pitch estimates.

Both of these methods demand prior knowledge of the number of sources, K, and their model orders, L_k. In the following simulations, these numbers are given to *Optimal Filtering*, $MUSIC$ and the dense $PEBS$, while $RIME$ only gets prior knowledge of the number of sources to ensure no ambiguity in the calculation of the $RMSE$. Note that in the real case scenario, it is an additional problem to estimate these parameters, which the proposed estimator deals with.

The experiment is carried out with $J = 250$ simulations for different values of the inharmonicity coefficient in (1.3.1). These values are $B = [0\ 10^{-5}\ 10^{-4}\ 10^{-3}]$. Data length is 500 samples. For every j, two pitches are randomly selected in the range $[0.02\,0.045]$Hz/sample of normalized frequencies. $PEBS$, $Optimal filtering$ and $MUSIC$ are given a set of 200 data points in this range. This yields a resolution of 1.25×10^{-4}Hz/sample. The noise is set to a level which corresponds to $SNR = 25$. The result can be seen in figure 4.1. One can see that the $MUSIC$, $Optimal filtering$ and the ordinary $PEBS$ methods fail as soon as the harmonics starts to deviate due to the increase in the inharmonicity coefficients. However, $RIME$ is able to maintain a low $RMSE$ even for larger values of B.

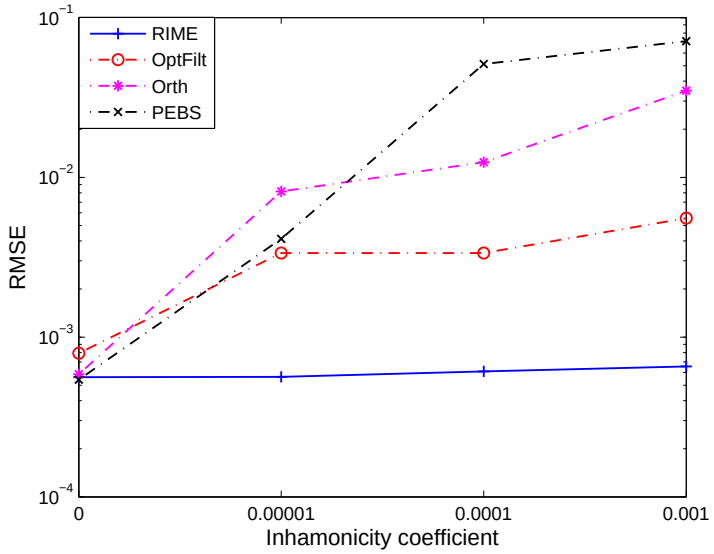

Figure 4.1. Simulation with different inharmonicity coefficients. The blue line is the proposed $RIME$, red is $Optimal filtering$, pink is $MUSIC$ and black is the dense version of $PEBS$.

4.2.2 Performance with respect to different levels of noise

In this section, $RIME$ will be exposed to different levels of noise to see how well it is performing under less optimal circumstances. The performance of the competing esitmators, presented in section 4.2.1, will also be evaluated to see how they perform compared to $RIME$ with respect to noise. The noise level is regulated via the SNR (signal to noise ratio) [13]. This number is defined as

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{4.2.2}$$

where the power P of the signal is measured as

$$P = \sum_{t=1}^{T} y_t y_t^* \tag{4.2.3}$$

with the sample sequence $\boldsymbol{y} = [y_1 \ \dots \ y_T]$ with samples having sampling index t in the range $[1, T]$. When the power of the signal is known, a noise sequence with the right power can be added to meet the SNR criterion. The different SNR values used are [1 5 10 15 20]. The signal contain inharmonicity, with a inharmonicity coefficient $B_k \in U[0\, 0.0005]$. The resulting $RMSE$ can be seen in Figure 4.2.

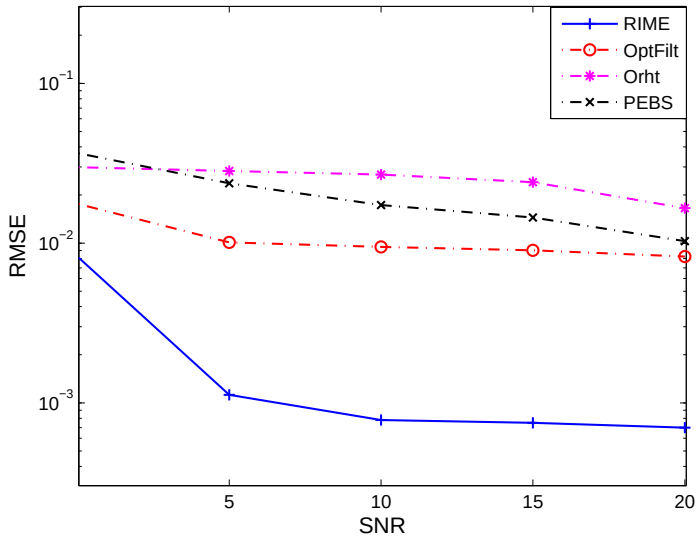

Figure 4.2. $RMSE$ for the SNR values [1 5 10 15 20]. The blue line is the proposed $RIME$, red is *Optimal filtering*, pink is $MUSIC$ and black is the dense version of $PEBS$.

As can be seen in Figure 4.2, $RIME$ is somewhat sensitive to high levels of noise. However, it is still able to perform better then the competing estimators.

4.3 Real signals

The purpose of all algorithms are eventually to solve some real problem. Therefore, the proposed algorithm is subjected to sound recordings of vibrating strings, in this case, a guitar. It is the pitch of a certain source that decides what musical note is sounding. By estimating this pitch, one could make a statement of what the musician is doing. A common file format, used in this work, for storing recorded sound on a computer is $.wav$ with a sampling frequency $f_s = 44100Hz$. The experiment will be carried out in the following way. Firstly, the signal will be down-sampled with a factor 5. This means taking out every fifth sample and decrease f_s to $44100/5 = 8820Hz$. By doing so, the signal is less dense and the estimated normalized frequencies are not as close to the lower region at 0, where frequency estimation is known to be more troublesome. When this is done, the signal is divided into short sub-vectors containing 500 samples assumed to be approximately stationary, which is necessary as mentioned in section 4.1. Each of these sub vectors are then analyzed with $RIME$.

In section 4.2, the signals were generated with known parameters and the sequences was perfectly stationary. Therefore, it was easy to compare the estimations of the proposed algorithm with the true parameter values. However, when recorded sound is used, it is hard to obtain the true parameter values with a high enough precision to make this comparison. Therefore, in this section the attention will mainly be on the first parts of the algorithm which are $PEBS$ and BIC for estimation of the number of sources, K. The reason for this is that pitch correction made by $RELAX$ and RCP might drown in the uncertainty of the true pitch values.

All recordings are created and owned by the author. The test signal is a mixture of different number of sources. There are zero, one, two, or three sources at different time instance. In order to illustrate the performance of the proposed algorithm the spectrogram and the estimated pitches using $RIME$ are plotted in Figure 4.3 and Figure 4.4. The recording starts out with a single note, then a 2-note chord. Thereafter, it is a break followed by a single note and the recording is finished with a 3-note chord. As can be seen in figure 4.4, the proposed algorithm is able to find the musical structure just described in a satisfying way.

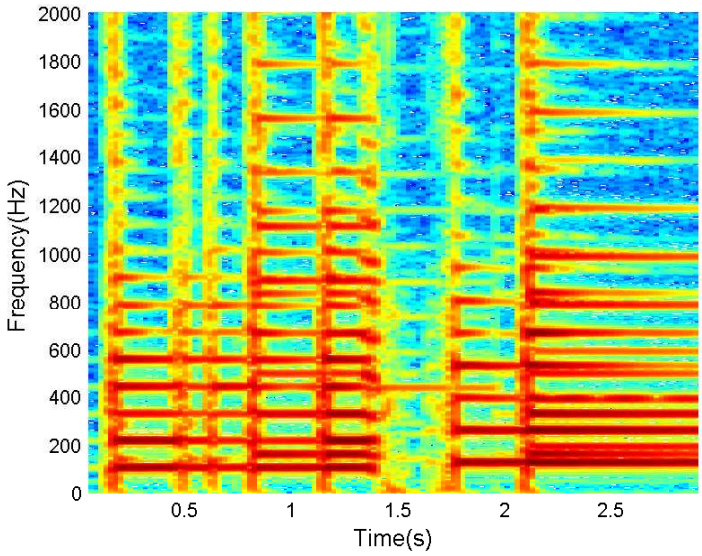

Figure 4.3. Spectrogram of the recorded guitar sound.

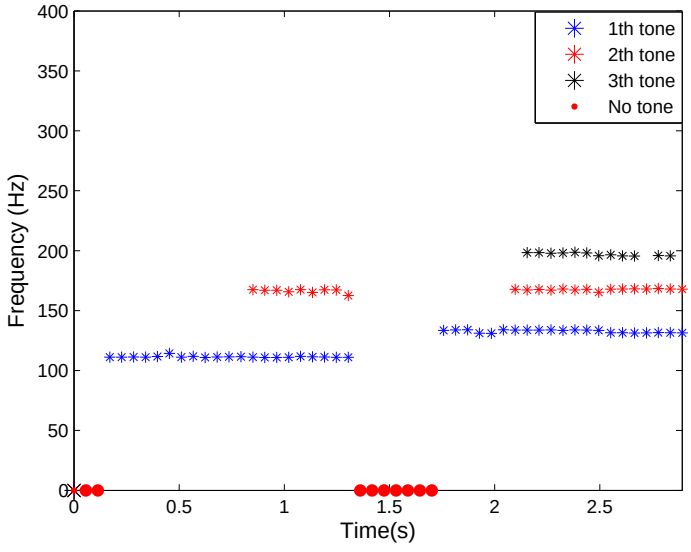

Figure 4.4. Pitches estimated using $RIME$. The over all musical structure of the recording is found as well as the fundamental frequencies.

Chapter 5

DISCUSSION

5.1 Conclusions

In this thesis, a method for high resolution frequency estimation of multi-pitch signals suffering from inharmonicities has been suggested. The estimator has been evaluated via Monte Carlo simulations with satisfying results. The method is able to estimate the number of sources, K, the pitches, $\omega_{0,k}$, and their corresponding model orders, L_k.

Throughout this work, signals with inharmonicites have been given a lot of attention. However, the estimator is of course able to handle waveforms with perfect harmonic structure as well, since this is a special case of the inharmonic structure, *i.e.*, inharmonicity coefficient $B_k = 0$ in (1.3.1) of $\Delta_{l,k} = 0$ in (1.3.2).

An advantage of the algorithm is that it can handle a signal without any prior knowledge of the number of sources and model orders. These parameters are very rarely known beforehand, which makes the method very general and applicable.

5.2 Possible improvements and future work

The proposed algorithm gives satisfying results. However, there are always room for improvements. Some of these will now be addressed.

No attention during the work has been given to computational complexity of the proposed algorithm. This aspect is of great concern when the application is of real-time character. The most computationally heavy part of the proposed algorithm is the RCP since is has to be executed in every iteration of step 3 until convergence of the $RELAX$-based iterations. Also the number of executions increase with the number of present sources in the signal since the pitch has to be estimated for every source in every iteration.

Some of the sub algorithms uses user defined parameters to set weights for optimization problems, convergence criterion and more. The choice of these parameters has not been given much attention. A larger study of these choices might exhibit even better results of the proposed estimator.

Lastly, it is worth noting that the core of the proposed algorithm is the procedure of transforming the multi-pitch problem into separate single-pitch problems.

This methodology is not unique to *RCP*, *i.e.*, future development in single-pitch estimation of signals suffering from inharmonicity can be applied to the proposed estimator, yielding even better results.

BIBLIOGRAPHY

[1] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgon & Claypool, 2009.

[2] H. Kameoka, *Statistical Approach to Multipitch Analysis*, Ph.D. thesis, University of Tokyo, 2007.

[3] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag 1991.

[4] H. Fletcher, *Normal vibrations frequencies of stiff piano string*, Journal of the Acoustical Society of America, vol. 36,no.1, 1962.

[5] E. B. George and M. J. T. Smith, *Speech analysis/synthesis and modifcation using an analysis-by-synthesis/overlap-add sinusoidal model*, IEEE Trans. Speech and Audio Processing, vol. 5(5), pp. 389406, Sept. 1997.

[6] M. G. Chritensen, P. Vera-Candeas, S. D. Somasundaram, and A. Jakobsson, *Robust subspace-based Fundamental frequency Estimation in IEEE Transactions in Audio, Speech and Signal Processing, vol. 20, no. 6, pp. 1857-1868, Aug. 2012.*

[7] S. I. Adalbjörnsson, A. Jakobsson and M. G. Christensen, *"Estimating Multiple Pitches using Block Sparsity"*, in 38th IEEE International Conference on Acoustics, Speech, and Signal Processing, 2013.

[8] N. R. Butt, S. I. Adalbjrnsson, S. D. Somasundaram and A. Jakobsson, *"Robust fundamental frequency estimation in the presence of inharmonicities"*, in 38th IEEE International Conference on Acoustics, Speech, and Signal Processing, 2013.

[9] S. Boyd, N. Parikh, E. Chu , B. Peleato and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Foundation and Trends, Machine learning 1, Jan. 2011.

[10] J. Li and P. Stocia, *"Efficient Mixed-Spectrum Estimation with Applications to Target Feature Extraction"*, IEEE Transactions on signal processing, Vol 44, no. 2, Feb. 1996.

[11] P. Stocia and R. Moses, *Spectral analysis of signals*, Pearson Prentice Hal, 2005.

[12] L. C. Biers, *Mathematical Methods of Optimization*, Holmbergs, 2010, edition 1.

[13] G. Lindgren, H. Rootzén and M. Sandgren, *Stationary stochastic processes*, 1 edition, Sep 2009.

[14] A. Jakobsson, *Time Series Analysis and Signal Modeling*, Student literature to appear.

[15] M. Müller, 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, Oct 2011.

[16] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, Seventh printing, 2009.

[17] P. Stocia and Y. Selén, *A review of information criterion rules*, IEEE Signal processing magazine, 2004, Vol. 21, p. 36-47.

[18] R. Johansson, *System Modeling and Identification*, Englewood Cliffs, NJ: PrenticeHall, 1993.

Master's Theses in Mathematical Sciences 2013:E15
ISSN 1404-6342

LUTFMS-3209-2013

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/