EUSIPCO 2013 1569742069

MULTI-PITCH ESTIMATION OF INHARMONIC SIGNALS

Tommy Nilsson, Stefan I. Adalbjörnsson, Naveed R. Butt, and Andreas Jakobsson

Dept. of Mathematical Statistics, Lund University, Sweden

ABSTRACT

This work presents a relaxation-based multi-pitch estimation technique for harmonic signals suffering from inharmonicity. Different from most earlier works, the proposed method does not require a priori knowledge of the number of sources present, nor of their respective number of harmonics, or the inharmonicity structure of the expected deviations. Using a recent group-sparse multi-pitch estimation method to form initial coarse pitch estimates, the number of sources and their harmonics are estimated using a BIC-based formulation, whereafter an iterative, relaxation-based, technique is formed to separately estimate the inharmonicity of each source using a recently proposed robust single-pitch estimation technique. The proposed algorithm is evaluated and compared to other existing methods using both simulated and real audio signals, clearly illustrating the improved performance.

1. INTRODUCTION

The estimation of the fundamental frequency, or pitch, of harmonically related sinusoidal signals is a problem finding applications in a wide range of applications, such as, for instance, electrocardiography (ECG), parametric coding of audio and speech, automatic music transcription, musical genre classification, tuning of musical instruments, and separation and enhancement of audio and speech sources, and the topic has attracted a notable attention during the recent decades (see, e.g., [1, 2] and the references therein). Commonly, the pitch estimate is formed assuming only the presence of a single source, i.e., signals containing only a single fundamental frequency and its harmonics, often assuming a priori knowledge also of the number of harmonics of the source, using some form of similarity measures, such as the cross-correlation, cepstrum, or the average squared difference function, or being based on second order statistics (see, for example, [1–4]). Furthermore, such estimators typically also assume that the harmonics are formed as exact integer multiples of the fundamental frequency (see, e.g., [5, 6]). However, this is not always the case, and the deviation of the higher frequencies from exact integer multiples of the fundamental frequency, a phenomenon called inharmonicity, is often observed in real-world signals. For instance, it is well known that inharmonicity arises in piano tones due to the stiffness in the piano strings [7]. Inharmonicity has also been considered in the modeling and coding of speech signals, and several different models of inharmonicity have been developed [8, 9], as, if not properly compensated for, the frequency deviations will lead to poor amplitude and pitch estimates [10]. To alleviate this problem, several robust single-source fundamental frequency estimation algorithms have been proposed in the recent literature, allowing for inharmonicity in the observed signal, both being based on a known frequency deviation function that depend functionally on a single unknown stiffness parameter [4, 11–13], or using some form of robust formulation to allow for an unstructured perturbations [10, 14]. In this work, we consider the combination of these problems, treating the joint estimation of the fundamental frequencies of an unknown number of sources, each with an unknown number of harmonics, suffering from some form of unstructured inharmonicity perturbations.

This work was supported in part by the Swedish Research Council and Carl Trygger's foundation.

2. SIGNAL MODEL

Consider a measured (complex-valued) signal, $x(n)$, consisting of K separate sources, with each source being formed as a sum of harmonically related sinusoids, such that (see also [1])

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

for $n = 1, \ldots, N$, where L_k denotes the model order of the kth source, and $a_{\ell,k} = A_{\ell,k} e^{i\phi_{\ell,k}}$ is the complex-valued amplitude of the ℓth component of this source, with $A_{\ell,k} \geq 0$, $\phi_{\ell,k}$, and $\omega_{\ell,k}$ denoting the (real-valued) amplitude, phase, and angular frequency of the kth source's ℓth component, and $e(n)$ is a complex-valued circularly symmetric white Gaussian noise. Commonly, a regular harmonic structure is assumed, such that $\omega_{\ell,k} = \ell\omega_{0,k}$, for $\ell = 1, \ldots, L_k$, where $\omega_{0,k}$ denotes the fundamental frequency of the kth source, whereas for signals exhibiting inharmonicity, such as signals originating from a stiff stringed instrument, it is com-

mon to use a parametric inharmonicity model, such that [15]

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

where B_k is the stiffness, or inharmonicity, coefficient of source k. Typically, even for stringed instruments, the stiffness coefficient is unknown and needs to be estimated from the data or set based on physical properties of the source. However, in general, many forms of sources will exhibit a more irregular inharmonicity, which is then more properly modeled as an unknown frequency perturbation, such that

$$\omega_{\ell,k}(\Delta_{\ell,k}) = \ell\omega_{0,k} + \Delta_{\ell,k} \tag{3}$$

where $\Delta_{\ell,k}$ denotes the perturbation of the ℓth harmonic, of the kth source.

3. ESTIMATION ALGORITHM

In order to allow for both an unknown number of signal sources, and the number of their harmonics, as well as the general inharmonicity model in (3), we introduce a relaxation-based estimation scheme that combines the recent multi-pitch PEBS estimation method, presented in [16], which allows for the estimation of the fundamental frequency of perfectly harmonically related sources, with the RCP inharmonicity estimation method, proposed in [14], which allows for the estimation of a single fundamental frequency, allowing for the here considered general inharmonicity model. The resulting estimation scheme is accomplished using a 4-step procedure, where a set of candidate pitches and their model order is initially formed using PEBS. Then, in a second step, a BIC-based order estimation scheme determines which of these candidate pitches that are likely present in the signal, as well as refines the estimate of the number of harmonics for each source, whereafter, in a third step, an iterative greedy estimation procedure forms refined pitch estimates, allowing for a general inharmonicity, using RCP. Finally, in the fourth step, the set of pitch frequencies are refined using a least-squares gradient step. A schematic overview of the proposed algorithm is presented in Algorithm 1. In order to detail the above summarized estimation steps, let

$$\boldsymbol{y} = \left[\begin{array}{ccc} x(1) & \cdots & x(N) \end{array}\right]^T \tag{4}$$

which implies that

$$\boldsymbol{y} = \left[\begin{array}{ccc} \boldsymbol{Z}_1 & \ldots & \boldsymbol{Z}_K \end{array}\right] \left[\begin{array}{c} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_K \end{array}\right] + \boldsymbol{e} \tag{5}$$

where $\boldsymbol{e}$ has been formed similar to $\boldsymbol{y}$, and

$$\boldsymbol{Z}_k = \left[\begin{array}{ccc} \boldsymbol{z}(\omega_{0,k}) & \cdots & \boldsymbol{z}(\omega_{L_k,k}) \end{array}\right] \tag{6}$$

$$\boldsymbol{z}(\omega) = \left[\begin{array}{cccc} 1 & e^{j\omega} & \cdots & e^{j\omega(N-1)} \end{array}\right]^T \tag{7}$$

$$\boldsymbol{a}_k = \left[\begin{array}{ccc} a_{1,k} & \cdots & a_{L_k,k} \end{array}\right]^T \tag{8}$$

Algorithm 1 Outline of proposed method

STEP 1 : Estimate a set of candidate pitches and their model orders, L_k, using (9) and (13).
STEP 2 : Determine the likely candidate frequencies using (14).
STEP 3 :
while not converged **do**
  for $k = 1$ to $\hat{K}$ **do**
    Subtract all but source k from the signal
    Estimate $\omega_{0,k}$ and $\omega_{0,l}$, $l \in [1, \hat{L}_k]$, using RCP.
  end for
end while
STEP 4 : Gradient step

with $(\cdot)^T$ denoting the transpose. Then, in order to form the initial coarse pitch estimate, we apply the PEBS algorithm, which exploits a group sparse modeling approach, such that, for each candidate pitch, the harmonically related frequency components are grouped into a block, whereafter the blocks best modeling the signal are found as [16]

$$\min_{\boldsymbol{a}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{a}\|_2^2 + \lambda\|\boldsymbol{a}\|_1 + \alpha\sum_{k=1}^{P}\sqrt{\Delta_k}\|\boldsymbol{a}_k\|_2 \tag{9}$$

where $\|\cdot\|_p$ denotes the p-norm, λ and α are tuning parameters which decide how the penalties are weighted, Δ_k is the number of harmonics in block k, P the number of considered frequencies, $\boldsymbol{W}$ a matrix containing blocks of the Vandermonde matrices, $\boldsymbol{Z}_k$, which represent each possible source, and $\boldsymbol{a}$ is a vector consisting of the amplitudes of all the considered blocks, i.e.,

$$\boldsymbol{Z}_k = \left[\begin{array}{ccc} \boldsymbol{z}(\omega_k) & \cdots & \boldsymbol{z}(\omega_k L_k) \end{array}\right] \tag{10}$$

$$\boldsymbol{W} = \left[\begin{array}{ccc} \boldsymbol{Z}_1 & \cdots & \boldsymbol{Z}_P \end{array}\right] \tag{11}$$

$$\boldsymbol{a} = \left[\begin{array}{ccc} \boldsymbol{a}_1^T & \cdots & \boldsymbol{a}_P^T \end{array}\right]^T \tag{12}$$

Thus, the minimization is formed such that the distance between the measured signal, $\mathbf{y}$, and all considered blocks of candidate pitch frequencies, and their harmonics, while being penalized by such that the optimization attempts to minimize the number of blocks having a non-zero constitution, as well as penalizing non-full blocks (in order to limit the halfing/doubling problem). As noted in [16], the restriction of the allowed frequency range implies that the number of harmonics for each source, L_k, are restricted as a function of the fundamental frequency, such that $L_k < \lfloor 2\pi/\omega_k \rfloor$, where $\lfloor\cdot\rfloor$ denotes the round-down to nearest integer operation. Setting the maximum allowed number of harmonics to $L_{\max}$, the normalization Δ_k, introduced to avoid the otherwise natural tendency to always favor lower candidate frequencies over the blocks corresponding to their double frequencies, is selected as $\Delta_k = \min(L_{\max}, L_k)$. The resulting amplitude vector, $\boldsymbol{a}$, thus forms an estimate of the

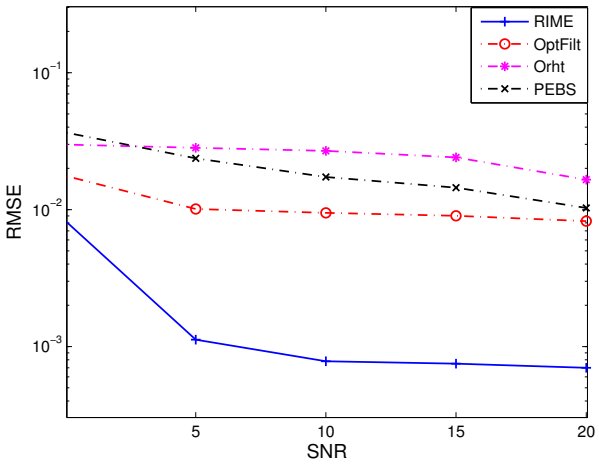

Fig. 1. The RMSE of the discussed estimators as a function of the SNR.

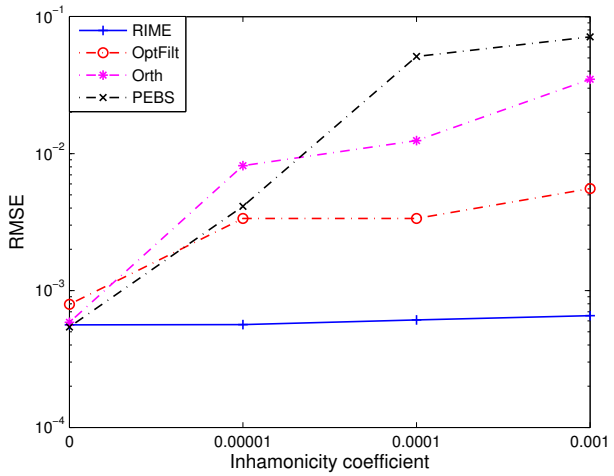

Fig. 2. The RMSE of the discussed estimators as a function of the relative level of inharmonicity.

candidate pitches, with the kth index in $\boldsymbol{a}$ corresponding to the pitch candidate ω_k. As we are here, different from in [16], considering sources suffering from inharmonicity, it is important to note that the considered grid of possible pitch candidates should be selected as coarse set, as for on a fine grid, the inharmonicity structure will result in that the amplitude vector, $\boldsymbol{a}$, will suffer from line-splitting, such that closely spaced candidate pitches will be deemed to well model the source, as both pitch frequencies will, due to the inharmonicity, fit into (9). Using a coarse initial candidate grid avoids this problem for all cases when the pitches may be considered to be well separated. In the second step, using the so-obtained block amplitude estimate, we refine the estimate of the number of harmonics for each of the candidate pitches. These are formed using the initial block size of $L_k = \lfloor 2\pi/\omega_k \rfloor$, but may be reduced to a more realistic model order by omitting weak harmonics from the blocks. Here, we do so by reducing the active block size for the kth candidate pitch, $\hat{L}_k$, such that

$$\hat{L}_k = \sum_{\ell=1}^{\min(L_{\max},\, L_k)} u\Big[a_{\ell,k} - 0.01\max(\boldsymbol{a}_k)\Big] \qquad (13)$$

where $a_{\ell,k}$ is the ℓth amplitude in the vector $\boldsymbol{a}_k$, and $u[x]$ is the indicator function taking the values one if $x > 0$, and otherwise zero, i.e., only those amplitudes larger than a percentage of the largest amplitude in the block are considered significant. Clearly, this is a rather ad hoc measure, but one that, in our experience, works well to give a reasonable rough model order estimate for each block. Using the thus resulting rough estimate of the number of significant harmonics in each block, the number of likely candidate pitches, $\hat{K}$, is selected from $\boldsymbol{a}_k$ using the BIC-style selection rule formed as the minimum of [16] (see also [17])

$$BIC(\ell) = 2N\ln(\sigma^2_{\boldsymbol{y},\ell}) + (5H_\ell + 1)\ln(N) \qquad (14)$$

where $H_\ell = \sum_{k=1}^{\ell} \hat{L}_k$ and $\sigma^2_{\boldsymbol{y},k}$ is the variance of the signal after the k first harmonic components have been removed, over the range of considered sources $k \in [1, K_{max}]$, where K_{max} denotes the assumed largest number of sources. The candidate pitch frequencies are then selected as the largest (absolute) valued candidate pitches among $\boldsymbol{a}_k$.Then, having determined the likely number of sources, $\hat{K}$, and the number of their harmonics, $\hat{L}_k$, we proceed to the third, iterative, step, wherein each candidate pitch estimate is refined to allow for the general inharmonicity model in (3). This is done by approximating the multi-pitch estimation problem as a set of single-pitch estimation problem, where each refined estimate may be formed using the recent robust Capon pitch (RCP) estimator [14]. Reminiscent to RELAX-algorithm presented in [18], we therefore proceed to form an estimate of the qth pitch as

$$\boldsymbol{y}_q = \boldsymbol{y} - \sum_{k=1, k\neq q}^{\hat{K}} \boldsymbol{Z}_k \hat{\boldsymbol{a}}_k \qquad (15)$$

where $\hat{\boldsymbol{a}}_{\hat{k}}$ represents the least squares estimate of the complex amplitudes of source k, with $\boldsymbol{Z}_k$ denoting the corresponding source. The resulting signal, $\boldsymbol{y}_q$, may then be treated as a single source signal with its (possibly) $\hat{L}_q$ perturbed harmonics. This allows for the use of the RCP algorithm, which is based on a multi-dimensional covariance fitting criterion that attempts to maximally explain the observed signal power, while allowing for uncertainties in the frequency vectors. Defining $\hat{\boldsymbol{Z}}_q \in \mathbb{C}^{M\times \hat{L}_q}$ as the Fourier

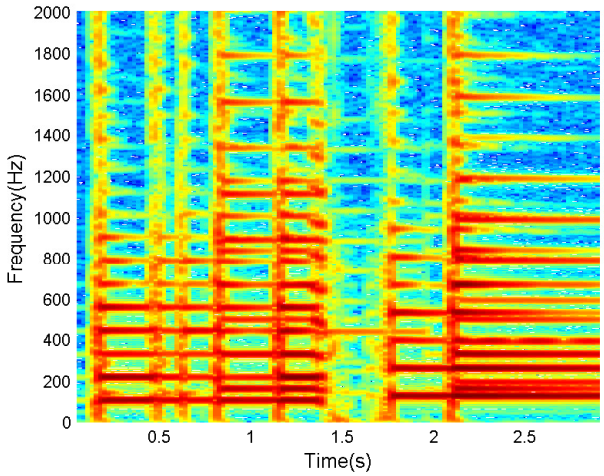

Fig. 3. Spectrogram of recorded guitar sound.

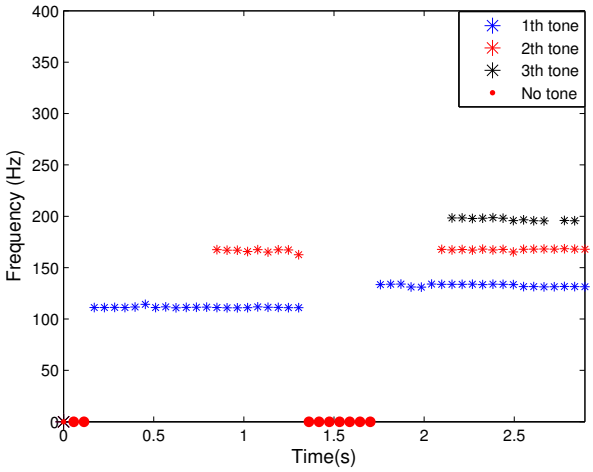

Fig. 4. The RIME estimate of the guitar recording.

matrix of the form in (6), evaluated using the PEBS estimates of the frequency components of source q, and $\boldsymbol{Z}_{q,\Delta}$ as its perturbed counterpart, and forming a set of $M \times 1$ overlapping sub-vectors

$$\boldsymbol{y}_q^{(n)} = \left[\begin{array}{ccc} \boldsymbol{y}_q(n) & \cdots & \boldsymbol{y}_q(n+M-1) \end{array} \right]^T \tag{16}$$

for $n = 1, \ldots, N-M+1$, where $\boldsymbol{y}_q(\ell)$ denotes the ℓth element of $\boldsymbol{y}_q$, the covariance fitting formulation of RCP may be expressed as the optimization problem

$$\begin{array}{ll} \underset{\boldsymbol{Z}_{q,\Delta}, \boldsymbol{P}_q, \sigma^2_{e,q}}{\max} & \log\Big(\det(\boldsymbol{Z}_{q,\Delta}\boldsymbol{P}_q\boldsymbol{Z}^*_{q,\Delta} + \sigma^2_{e,q}\boldsymbol{I}_L)\Big) \\ \text{subject to} & \boldsymbol{Z}_{q,\Delta}\boldsymbol{P}_q\boldsymbol{Z}^*_{q,\Delta} + \sigma^2_{e,q}\boldsymbol{I}_L \preceq \hat{\boldsymbol{R}}_{\mathbf{y},q} \\ & \|(\boldsymbol{Z}_{q,\Delta} - \hat{\boldsymbol{Z}}_q)\boldsymbol{e}_l\| \leq \epsilon_l \\ & \boldsymbol{P}_q = \boldsymbol{P}_q \odot \boldsymbol{I}_L \succeq \boldsymbol{0} \end{array} \tag{17}$$

where

$$\boldsymbol{P}_q = \text{diag}\Big(\begin{array}{ccc} |A_{1,q}|^2 & \cdots & |A_{\hat{L}_q,q}|^2 \end{array} \Big) \tag{18}$$

$$\hat{\boldsymbol{R}}_{\mathbf{y},q} = \frac{1}{N-M+1} \sum_{n=0}^{N-M} \boldsymbol{y}_q^{(n)} \boldsymbol{y}_q^{(n)} \tag{19}$$

and $\sigma^2_{e,q}$ denotes the noise variance corresponding to source q, $\boldsymbol{e}_l$ is the lth column of an $L \times L$ identity matrix, $\boldsymbol{I}_L$, $\odot$ the Schur-Hadamard (element-wise) product, and $\mathbf{A} \preceq \mathbf{B}$ implies that $\mathbf{A} - \mathbf{B}$ is positive semi-definite. Thus, the first constraint requires that the matrix $\hat{\boldsymbol{R}}_{\mathbf{y},q} - \boldsymbol{Z}_{q,\Delta}\boldsymbol{P}_q\boldsymbol{Z}^*_{q,\Delta} - \sigma^2_{e,q}\boldsymbol{I}_L$ is positive semidefinite, while the second constraint ensures that each column of the perturbed matrix, $\boldsymbol{Z}_{q,\Delta}$, lies within a small hyper-sphere of radius ϵ_l around the assumed frequency vectors in $\hat{\boldsymbol{Z}}_q$. The third and final constraint requires $\boldsymbol{P}_q$ to be positive definite and diagonal. Using the approach developed in [14], this maximization yields robust estimates of all the frequency components of source q. In the interest of brevity, we refer the reader to [14] for the full details of the RCP estimator. Next, the above process is repeated for all the $\hat{K}$ sources, leading to the first set of refined pitch estimates. Once the frequency components of all the sources have been estimated, one may choose to do further iterations of the refinement process, using, in each iteration, the previous iteration's refined pitch estimates to form improved nominal (assumed) frequency vectors in RCP. This may be repeated until a desired convergence criterion is met. Finally, a gradient search is performed for each of the separated sub problems from step three to further refine the pitch estimates. These are formed by evaluating the error norm

$$\|\boldsymbol{e}\|_2^2 = \|\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{a}\|_2^2 = (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{a})^*(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{a}) \tag{20}$$

for a set of frequencies, $\bar{\boldsymbol{\omega}}_{0,k}$, in the vicinity of the estimated pitches, $\hat{\omega}_{0,k}$. The element in this set that minimize (20) is chosen as the refined estimate of $\omega_{0,k}$. The narrow ranged set, $\bar{\boldsymbol{\omega}}_{0,k}$, is selected as

$$\bar{\boldsymbol{\omega}}_{0,k} = \hat{\omega}_{0,k} + d_k\boldsymbol{\theta} \tag{21}$$

where $\boldsymbol{\theta}$ is a 1-dimensional closely spaced grid and d_k the direction of the negative gradient of (20) with respect to $\hat{\omega}_{0,k}$, i.e.,

$$d_k = -\nabla_{\hat{\omega}_{0,k}} \|\boldsymbol{e}\|_2 \tag{22}$$

$$= -\nabla_{\hat{\omega}_{0,k}} \|\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{a}\|_2 \tag{23}$$

which may be expressed as (24), given at the top of the next page, where $\nabla_{\hat{\omega}_{0,k}}$ is the gradient operator with respect to $\hat{\omega}_{0,k}$. We term the resulting estimator the *R*obust *I*nharmonicity-based *M*ulti-pitch *E*stimator (RIME).

$$d_k = a_1 \boldsymbol{Y}\boldsymbol{y} \odot \boldsymbol{z}(\hat{\omega}_{0,k}) - a_1 \boldsymbol{Y}\boldsymbol{y} \odot \boldsymbol{z}(-\hat{\omega}_{0,k}) - a_1^* \sum_{l=2}^{L} a_l \boldsymbol{Y}\boldsymbol{z}(\hat{\omega}_{l,k} - \hat{\omega}_{0,k}) + a_1 \sum_{l=2}^{L} a_l^* \boldsymbol{Y}\boldsymbol{z}(\hat{\omega}_{0,k} - \hat{\omega}_{l,k}) \quad (24)$$

4. NUMERICAL RESULTS

In this section, we examine the performance of the proposed RIME estimator, using $N = 500$ samples of a simulated sinusoidal audio signal, consisting of $K = 2$ sources, each with $L_k \in U(3, 10)$ harmonics, suffering from an inharmonicity following (2), with $B_k \in U(0, 0.0005)$. Figures 1 and 2 show the root mean squared error (RMSE), defined as $(RMSE)^2 = \frac{1}{J}\sum_{\ell=1}^{J}\sum_{k=1}^{K}(\omega_{0,k}^{\ell} - \widehat{\omega}_{0,k}^{\ell})$, where $\omega_{0,k}^{\ell}$ and $\widehat{\omega}_{0,k}^{\ell}$ denote the kth true and estimated pitch, for simulation ℓ, respectively, using $J = 250$ Monte-Carlo simulations, as a function of the signal to noise ratio (SNR), defined as $\sigma_y^2\sigma_e^{-2}$, where σ_y^2 denotes the power of the noise free part of $x(n)$, as well as a function of the inharmonicity coefficient B_k (which in this case is the same for both sources). The RIME estimates of the two most dominant sources are compared to the PEBS estimate, the optimal filtering (Opt-Filt), and orthogonality-based (Orth) estimates [1, 2], with the three latter being allowed perfect knowledge of both K and their respective L_k. Finally, we examine the performance of the proposed algorithm for a measured guitar signal consisting of a varying number of sound sources, which in this case corresponds to the number of used strings. Figures 3 and 4 show the spectrogram of the signal as well as the resulting RIME estimates, which can be seen to well follow the true pitch signals.

5. REFERENCES

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

[2] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Signal Processing*, vol. 88, no. 4, pp. 972–983, April 2008.

[3] Z. Zhou, H. C. So, and F. K. W. Chan, "Optimally Weighted Music Algorithm for Frequency Estimation of Real Harmonic Sinusoids ," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, March 25-30 2012.

[4] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "A Robust and Computationally Efficient Subspace-Based Fundamental Frequency Estimator," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 487 –497, March 2010.

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

[6] W. Hess, "Pitch and Voicing Determination," *Advances in Speech Signal Processing*, pp. 3–48, 1992.

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

[8] T. D. Rossing, *The Science of Sound*, Addison-Wesley Publishing Co., 2 edition, 1990.

[9] E. B. George and M. J. T. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 389 –406, Sep 1997.

[10] M. G. Christensen, P. Vera-Candeas, S. D. Somasundaram, and A. Jakobsson, "Robust Subspace-based Fundamental Frequency Estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, March 30-April 4, 2008.

[11] I. Barbancho, L. J. Tardon, S. Sammartino, and A. M. Barbancho, "Inharmonicity-Based Method for the Automatic Generation of Guitar Tablature," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1857 –1868, Aug. 2012.

[12] E. Benetos and S. Dixon, "Joint Multi-Pitch Detection Using Harmonic Envelope Estimation for Polyphonic Music Transcription," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 6, pp. 1111 –1123, Oct. 2011.

[13] V. Emiya, R. Badeau, and B. David, "Multipitch Estimation of Piano Sounds Using a New Probabilistic Spectral Smoothness Principle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 18, no. 6, pp. 1643–1654, August 2010.

[14] N. R. Butt, S. I. Adalbjörnsson, 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.

[15] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, New York, NY, 1988.

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

[17] P. Stoica and Y. Selén, "Model-order selection — A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, July 2004.

[18] J. Li and P. Stoica, "Efficient Mixed-Spectrum Estimation with Applications to Target Feature Extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, February 1996.