Robust Subspace-based Fundamental Frequency Estimation

Christensen, Mads Græsbøll; Vera-Candeas, Pedro; Somasundaram, Samuel D.; Jakobsson, Andreas

Published in:
Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing

DOI (link to publication from Publisher):
10.1109/ICASSP.2008.4517556

Publication date:
2008

Document Version
Accepted author manuscript, peer reviewed version

ROBUST SUBSPACE-BASED FUNDAMENTAL FREQUENCY ESTIMATION

Mads G. Christensen[1], *Pedro Vera-Candeas*[2], *Samuel D. Somasundaram*[3], *and Andreas Jakobsson*[3]

[1] Dept. of Electronic Systems, Aalborg University, Denmark
[2] Telecom. Engineering Dept., University of Jaén, Spain
[3] Dept. of Electrical Engineering, Karlstad University, Sweden

ABSTRACT

The problem of fundamental frequency estimation is considered in the context of signals where the frequencies of the harmonics are not exact integer multiples of a fundamental frequency. This frequently occurs in audio signals produced by, for example, stiff-stringed musical instruments, and is sometimes referred to as inharmonicity. We derive a novel robust method based on the subspace orthogonality property of MUSIC and show how it may be used for analyzing audio signals. The proposed method is both more general and less complex than a straight-forward implementation of a parametric model of the inharmonicity derived from a physical instrument model. Additionally, it leads to more accurate estimates of the individual frequencies than the method based on the parametric inharmonicity model and a reduced bias of the fundamental frequency compared to the perfectly harmonic model.

Index Terms— Acoustic signal analysis, spectral analysis, frequency estimation

1. INTRODUCTION

The problem of estimating the fundamental frequency, or pitch period, of a set of harmonically related sinusoids is one of the classical problems of signal processing, not least in speech and audio processing where it is important to many applications ranging from analysis and compression to separation and enhancement. In recent years, it has also found new applications in music information retrieval and it remains an active research topic. In the ideal case, the frequencies of the harmonics are integer multiples of the fundamental frequency. For many musical instruments, though, the frequencies are not exact integer multiples of the fundamental. This phenomenon is known as inharmonicity and this is the problem we are concerned with in this paper. For particular instruments, like stiff-stringed instruments such as the piano, the inharmonicity is very pronounced and has to be taken into account when tuning the instrument [1, 2]. Also for speech signals, inharmonicity has been observed to be of importance for modeling and coding purposes (see, e.g., [3]). There are many reasons why this inharmonicity should be taken into account when estimating the fundamental frequency. Firstly, the assumption of the frequencies of the harmonics being exact integer multiples of the fundamental may lead to a bias in the estimated fundamental frequency. It may also lead to significant bias of estimated amplitudes since the integer multiples may not capture the peaks of the spectrum. This, in turn, may lead to audible artifacts when resynthesizing the audio. Similarly, biased estimates can result in a bad fit of the signal model to the data and low likelihoods, causing incorrect model and order selections. Therefore, one would expect a fundamental frequency estimator that takes this phenomenon into account to be more robust than the estimators based on the idealized model. We will now proceed to define the problem at hand mathematically. We are here concerned with a set of sinusoids having frequencies $\{\omega_l\}$ corrupted by an additive white complex circularly symmetric Gaussian noise, $\epsilon(n)$, for $n = 0, \ldots, N-1$,

$$x(n) = \sum_{l=1}^{L} \alpha_l e^{j\omega_l n} + \epsilon(n), \tag{1}$$

where $\{\alpha_l\}$ are the complex amplitudes (here considered nuisance parameters); these may easily be found given the frequencies. In this work, it is assumed that the number of harmonics L is known or found a priori. For the perfectly harmonic case, the frequencies of the harmonics are exact integer multiples of a fundamental frequency ω_0, i.e., $\omega_l = \omega_0 l$ with $l \in \mathbb{N}$, in which case the signal model in (1) is characterized by a single nonlinear parameter. As already discussed, this model is not always a good fit. Depending on the instrument, different parametric models of the inharmonicity of the harmonics can be derived from physical models (see, e.g., [4]). An example of such a model for stiff-stringed instruments is $\omega_l = \omega_0 l\sqrt{1 + Bl^2}$ where $B \ll 1$ is an unknown, positive stiffness parameter. In this case, the model in (1) now contains two unknown nonlinear parameters, namely ω_0 and B, which complicates matters. This model, which we will refer to as the parametric model of the inharmonicity, has been used in Bayesian fundamental frequency estimation for audio analysis in [5] and sinusoidal audio coding in [6].

Recently, it has been shown that the subspace orthogonality principle known from the MUltiple SIgnal Classification (MUSIC) estimation method (see, e.g., [7]) can be used for finding the fundamental frequency and the model order L jointly for both a single source [8] and multiple sources [9]. In this paper, we consider the problem of taking inharmonicity into account in a method based on the orthogonality property. We extend the method in [8] to the parametric model of the inharmonicity but this leads to a computationally complex method. Instead, we use an alternative model that at first sight will appear more complicated and derive a novel robust estimator. More specifically, we use a model where the frequencies $\{\omega_l\}$ are modeled as $\omega_l = \omega_0 l + \Delta_l$ with $\{\Delta_l\}$ being a set of small unknown perturbations that are to be estimated along with the fundamental frequency ω_0. We refer to this frequency model as the perturbed model. The perturbations $\{\Delta_l\}$ should be small since arbitrarily large perturbations will result in meaningless estimates of ω_0. Such unstructured perturbations have also been previously used in a Bayesian framework [10]. One would expect that the introduction of additional L nonlinear parameters in the model (1) would result in a more complicated estimator, but, as it turns out, the resulting estimator is, considering the complexity of the problem, rather simple. Aside from the computational complexity, there is another important reason why one would prefer the perturbed model over the

The work of M. G. Christensen is supported by the Parametric Audio Processing project, Danish Research Council for Technology and Production Sciences grant no. 274–06–0521.

parametric one, namely that it is more general, meaning that it can be expected to hold for a more general class of signals. For example, the parametric inharmonicity models are not the same for strings with clamped or pinned ends (see [4]). While it is here assumed that the model order L in (1) is known, it can in fact be found jointly with the other nonlinear parameters using the proposed method (c.f. [8]), but this is beyond the scope of this paper.

The rest of this paper is organized as follows. First, in Section 2, we will review the fundamentals of the covariance matrix model and the subspace orthogonality property and show how this can be used for finding the fundamental frequency and stiffness parameters of the parametric inharmonicity model. In Section 3, we then present the new robust estimator associated with the perturbed model. Some experimental results and examples are given in Section 4 before the conclusion in Section 5.

2. SOME PRELIMINARIES

In this section, we present the fundamentals of the MUSIC algorithm (see, e.g., [7]) and introduce some useful vector and matrix definitions. We will do this for the signal model in (1) based on a set of unconstrained frequencies $\{\omega_l\}$ and then show how this can be applied to the parametric inharmonicity model. We start out by defining $\tilde{\mathbf{x}}(n)$ as a signal vector containing M samples of the observed signal, i.e., $\tilde{\mathbf{x}}(n) = [\, x(n) \; \cdots \; x(n+M-1) \,]^T$ with $(\cdot)^T$ denoting the transpose. Then, assuming that the phases of α_l are independent and uniformly distributed on the interval $(-\pi, \pi]$, the covariance matrix $\mathbf{R} \in \mathbb{C}^{M \times M}$ of the signal in (1) can be written as

$$\mathbf{R} = \mathrm{E}\left\{\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^H(n)\right\} = \mathbf{A}\mathbf{V}\mathbf{A}^H + \sigma^2\mathbf{I}, \tag{2}$$

where $\mathrm{E}\{\cdot\}$ and $(\cdot)^H$ denote the statistical expectation and the conjugate transpose, respectively. Note that for this decomposition to hold, the noise need not be Gaussian. Furthermore, $\mathbf{V}$ is a diagonal matrix containing the squared amplitudes, i.e., $\mathbf{V} = \mathrm{diag}([\, |\alpha_1|^2 \; \cdots \; |\alpha_L|^2 \,])$, and $\mathbf{A} \in \mathbb{C}^{M \times L}$ a rank L Vandermonde matrix, i.e.,

$$\mathbf{A} = \left[\; \mathbf{a}(\omega_1) \quad \cdots \quad \mathbf{a}(\omega_L) \;\right], \tag{3}$$

where $\mathbf{a}(\omega) = \left[\; 1 \quad e^{j\omega} \quad \cdots \quad e^{j\omega(M-1)} \;\right]^T$. Also, σ^2 denotes the variance of the additive noise, $\epsilon(n)$, and $\mathbf{I}$ is the $M \times M$ identity matrix. We also note that $\mathbf{A}\mathbf{V}\mathbf{A}^H$ has rank L. For notational simplicity, we have omitted the dependency of $\mathbf{A}$ on the unknowns. Let $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ be the eigenvalue decomposition (EVD) of the covariance matrix. Then, $\mathbf{U}$ contains the M orthonormal eigenvectors of $\mathbf{R}$, i.e., $\mathbf{U} = [\, \mathbf{u}_1 \; \cdots \; \mathbf{u}_M \,]$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the corresponding eigenvalues, λ_k, with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$. Let $\mathbf{G}$ be formed from the eigenvectors corresponding to the $M-L$ least significant eigenvalues, i.e., $\mathbf{G} = [\, \mathbf{u}_{L+1} \; \cdots \; \mathbf{u}_M \,]$. The noise subspace spanned by $\mathbf{G}$ will then be orthogonal to the Vandermonde matrix $\mathbf{A}$, i.e., $\mathbf{A}^H\mathbf{G} = \mathbf{0}$ which is what we refer to as the subspace orthogonality property. In practice, only an estimate of the covariance matrix is available from which we can obtain an estimate of the noise subspace $\mathbf{G}$. From this matrix, an estimate of the these parameters can be obtained by minimizing the Frobenius norm as

$$J(\{\omega_l\}) = \mathrm{Tr}\left\{\mathbf{A}^H\mathbf{G}\mathbf{G}^H\mathbf{A}\right\}, \tag{4}$$

with $\mathrm{Tr}\{\cdot\}$ denoting the trace. Based on this cost function, the fundamental frequency ω_0 and the stiffness parameter B of the parametric inharmonicity model can be estimated in a straight-forward manner. Specifically, the $\mathbf{A}$ matrix is constructed from the two parameters as

$$\mathbf{A} = \left[\; \mathbf{a}(\omega_0\sqrt{1+B}) \quad \cdots \quad \mathbf{a}(\omega_0 L\sqrt{1+BL^2}) \;\right]. \tag{5}$$

Based on (4), we then suggest to obtain an estimate of the fundamental frequency and stiffness parameter as

$$(\hat{\omega}_0, \hat{B}) = \arg\min_{\omega_0, B} J(\omega_0, B), \tag{6}$$

which has to be evaluated for a large range of combinations of the two parameters. We note that the estimator (6) has not previously appeared in the literature.

3. ROBUST SUBSPACE METHOD

The question is now how the fundamental frequency and the individual frequencies can be found for the perturbed model where $\omega_l = \omega_0 l + \Delta_l$, i.e., the Vandermonde matrix containing the complex sinusoids is now characterized by ω_0 and $\{\Delta_l\}$ as

$$\mathbf{A} = \left[\; \mathbf{a}(\omega_0 + \Delta_1) \quad \cdots \quad \mathbf{a}(\omega_0 L + \Delta_L) \;\right]. \tag{7}$$

However, direct minimization of the cost function in (4) will not only be very computationally demanding since we now have $L+1$ nonlinear parameters, but it will also not lead to any meaningful estimates of the parameters since we have no control over the distribution of the perturbations $\{\Delta_l\}$. Instead, we propose to estimate the parameters by redefining the cost function in (4) as

$$J(\omega_0, \{\Delta_l\}) = \mathrm{Tr}\left\{\mathbf{A}^H\mathbf{G}\mathbf{G}^H\mathbf{A}\right\} + P(\{\Delta_l\}), \tag{8}$$

where $P(\{\Delta_l\})$ is the penalty function which is a non-decreasing function of a metric with $P(\{0\}) = 0$. Also, it is desirable that the penalty function is additive over the harmonics. Therefore, a natural choice is $P(\{\Delta_l\}) = \sum_{l=1}^{L} \nu_l |\Delta_l|^p$ with $p \geq 1$ which penalizes large perturbations Δ_l. Also, $\{\nu_l\}$ is a set of positive regularization constants, the meaning of which will be discussed later. In arriving at a computationally efficient estimator, we first note that the Frobenius norm is additive over the columns of $\mathbf{A}$, i.e.,

$$J(\omega_0, \{\Delta_l\}) = \mathrm{Tr}\left\{\mathbf{A}^H\mathbf{G}\mathbf{G}^H\mathbf{A}\right\} + \sum_{l=1}^{L} \nu_l |\Delta_l|^p \tag{9}$$

$$= \sum_{l=1}^{L} \mathbf{a}^H(\omega_0 l + \Delta_l)\mathbf{G}\mathbf{G}^H\mathbf{a}(\omega_0 l + \Delta_l) + \sum_{l=1}^{L} \nu_l |\Delta_l|^p. \tag{10}$$

Furthermore, by substituting ω_l by $\omega_0 l + \Delta_l$ and Δ_l by $\omega_l - \omega_0 l$ in (10), we get the following simplified cost function

$$J(\omega_0, \{\omega_l\}) = \sum_{l=1}^{L} \mathbf{a}^H(\omega_l)\mathbf{G}\mathbf{G}^H\mathbf{a}(\omega_l) + \nu_l |\omega_l - \omega_0 l|^p. \tag{11}$$

It can be seen that the first term no longer depends on the fundamental frequency or the perturbations but only on the frequency of the l'th harmonic ω_l. Furthermore, we can recognize the first term in (11) as the reciprocal of the MUSIC pseudo-spectrum, which has now to be calculated only once for each segment. It can also be seen that the cost function is additive over independent terms and, therefore, the minimization of the cost function can be performed independently for each harmonic. The fundamental frequency estimator

can thus be rewritten as

$$\hat{\omega}_0 = \arg\min_{\omega_0} \min_{\{\omega_l\}} \left\{ \sum_{l=1}^{L} \mathbf{a}^H(\omega_l)\mathbf{G}\mathbf{G}^H\mathbf{a}(\omega_l) + \nu_l|\omega_l - \omega_0 l|^p \right\}$$
$$= \arg\min_{\omega_0} \sum_{l=1}^{L} \min_{\omega_l} \left\{ \mathbf{a}^H(\omega_l)\mathbf{G}\mathbf{G}^H\mathbf{a}(\omega_l) + \nu_l|\omega_l - \omega_0 l|^p \right\},$$

where the frequencies $\{\omega_l\}$ and thus perturbations are also found implicitly. From this, it is now also clear why we required the penalty function to be additive over the harmonics. We term this estimator robust since it expected to be more robust towards model mismatch than the ideal model and the parametric inharmonicity model. For a given $\hat{\omega}_0$, the frequencies can simply be found for $l = 1, \ldots, L$ as

$$\hat{\omega}_l = \arg\min_{\omega_l} \left\{ \mathbf{a}^H(\omega_l)\mathbf{G}\mathbf{G}^H\mathbf{a}(\omega_l) + \nu_l|\omega_l - \hat{\omega}_0 l|^p \right\}, \quad (12)$$

with only the penalty term changing over l. In the context of statistical estimation, the augmentation of a log-likelihood function by such a penalty term will result in a maximum a posteriori estimate with the penalty term being a log-prior on the perturbations with an implicit uniform prior on the fundamental frequency. For a Gaussian prior, for example, we would have $p = 2$ and $\nu_l = 1/(2\sigma_l^2)$ with σ_l^2 being the variance of the l'th harmonic. The meaning of the regularization constants can also be clarified from the following: For large ν_l, the ensuing perturbation will be small and the estimator is expected to reduce to the perfectly harmonic case, whereas for ν_l close to zero, the estimator will reduce to finding unconstrained frequencies from which no meaningful fundamental frequency estimate can be found. The regularization constants ν_l can also be interpreted as Lagrange multipliers. This means that the estimator can be thought of as a constrained estimator with a set of implicit constraints. In this sense, the method is conceptually related to the robust Capon beamformer of [11] which is based on explicit constraints. It may be worth modifying the penalty such that less emphasis is put on perturbations for higher harmonics or even use an asymmetrical penalty since the parametric inharmonicity model suggest that the harmonics will be higher than the integer multiple of the fundamental, but for now we will defer from further discussion of this.

4. EXPERIMENTAL RESULTS

We will now apply the estimators to analysis of audio signals. Since we assume that the order L is known, we will focus on a single note, namely a piano note $C_5 \sim 523.25$ Hz whose spectrogram and time-domain signal are shown in Figure 1. In the experiments, the following conditions are used. The signal has been down-sampled to 8820 Hz to reduce the computational complexity. The estimates are obtained in the following way. First, segments consisting of 30 ms have been used from which the discrete-time analytic signal and a 66×66 covariance matrix are calculated. The EVD of this matrix is computed and partitioned into signal and noise subspace eigenvectors with $L = 7$. Then, the cost functions of the respective estimators are calculated for a wide range of ω_l (and B for the parametric inharmonicity model) from which coarse estimates are obtained. Finally, these are used to initialize gradient-based refinement methods. For the proposed method, we use $p = 1$ (corresponding to a Laplacian prior) since this is expected to result in most perturbations being close to zero while allowing for a few large ones. The regularization constants with $\nu_l \propto 1/l$ have been used such that larger perturbations are allowed for higher harmonics. The exact perturbation constants were determined empirically for a large set of signals. We note that it is generally safer to choose ν_l too large than too small but other than that, we have observed the choice of ν_l not to be critical.

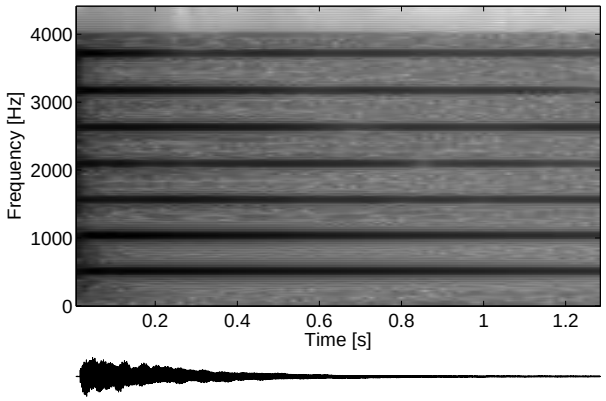

Fig. 1. Spectrogram and time-domain signal of a piano note.

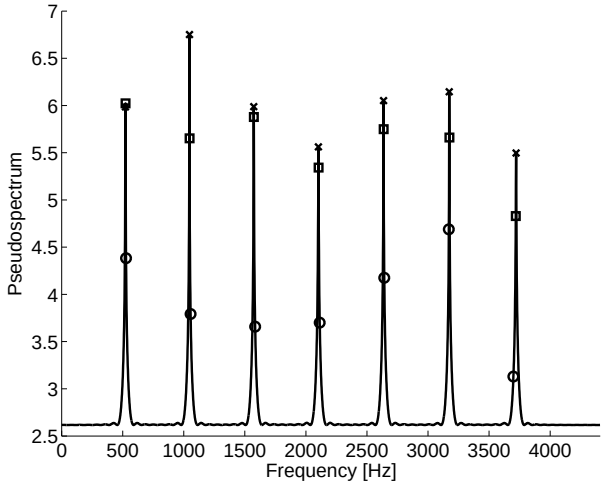

Fig. 2. MUSIC (log) pseudo-spectrum of a segment of the piano note along with the estimated frequencies of the harmonics for the perfectly harmonic model (circles), the parametric inharmonicity model (squares) and using the perturbed model (crosses).

In the first experiment, we will illustrate the ability of the models and estimators to capture the frequencies of the individual harmonics of the piano note. For this signal, the parametric inharmonicity model is expected to perform well. In Figure 2, the logarithm of the MUSIC pseudo-spectrum is shown for a representative stationary 30 ms segment of the signal in Figure 1 along with the frequency estimates obtained using the perfectly harmonic model, the parametric inharmonicity model and the perturbed model. It can be seen that the frequencies obtained using the perfectly harmonic model are biased with the ω_0 estimate being 528.51 Hz. It can also be seen that the parametric inharmonicity model captures the peaks fairly accurately resulting in an ω_0 estimate of 523.59 Hz. The perturbed model, however, appears to capture all the peaks for an estimated fundamental frequency of 526.21 Hz. It can be seen that the proposed method is able to reduce the bias of the perfectly harmonic model while also giving accurate estimates of the individual harmonics. Next, the estimated fundamental frequencies are estimated in steps of 10 ms. The results are shown in Figure 3 for the three methods, namely for the

perfectly harmonic model, the parametric inharmonicity model, and the perturbed model. The general conclusions can be seen to be the same as for Figure 2. The parametric inharmonicity model appears to be the most accurate for this particular signal with the estimates being close to the 523.25 Hz of the note. It can also be seen that the perfectly harmonic model here results in a bias throughout the duration of the signal, and that the perturbed model is able to reduce this bias. The remaining bias is most likely due to the use of a symmetrical penalty function that penalizes negative and positive perturbations equally. We note that the fluctuations of the estimates at about 0.85 s are due to modulations of the 4'th harmonic. Finally, we have investigated how well the models are able to capture the frequencies of the individual harmonics in the following way: Given the frequency estimates, the model is fitted to the data using least-squares and the signal-to-noise ratio (SNR) is calculated. In Figure 4, the results are shown. It can be seen that the perfectly harmonic model does not fit the signal well, and that the proposed method in fact estimates the frequencies more accurately than the parametric model, whereby a better fit is obtained.

5. CONCLUSION

In this paper, the problem of robust fundamental frequency estimation has been considered for the case where the harmonics are not exact integer multiples of a fundamental frequency, a phenomenon commonly known as inharmonicity that frequently can be observed in signals produced by musical instruments. A new subspace-based method has been proposed for finding the fundamental frequency and a set of small perturbations. We have compared this method to methods based on a commonly used parametric model of the inharmonicity, which is based on a physical musical instrument model, and the perfectly harmonic model where the frequencies are exact integer multiples of the fundamental. The proposed method is both computationally simple and more general than the estimator based on parametric inharmonicity model and has been found to lead to a reduced bias compared to the perfectly harmonic model and more accurate estimates of the individual frequencies than both the other methods for a piano signal.

6. REFERENCES

[1] H. Fletcher, "Normal vibration frequencies of a stiff piano string," in *J. Acoust. Soc. Amer.*, 1964, vol. 36(1).

[2] J. Lattard, "Influence of inharmonicity on the tuning of a piano - measurements and mathematical simulation," *J. Acoust. Soc. Am.*, vol. 94, pp. 46–53, 1993.

[3] 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 and Audio Processing*, vol. 5(5), pp. 389–406, Sept. 1997.

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

[5] S. Godsill and M. Davy, "Bayesian computational models for inharmonicity in musical instruments," in *Proc. IEEE Workshop on Appl. of Signal Process. to Aud. and Acoust.*, 2005, pp. 283–286.

[6] F. Myburg, *Design of a Scalable Parametric Audio Coder*, Ph.D. thesis, Technical University of Eindhoven, 2004.

[7] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, 2005.

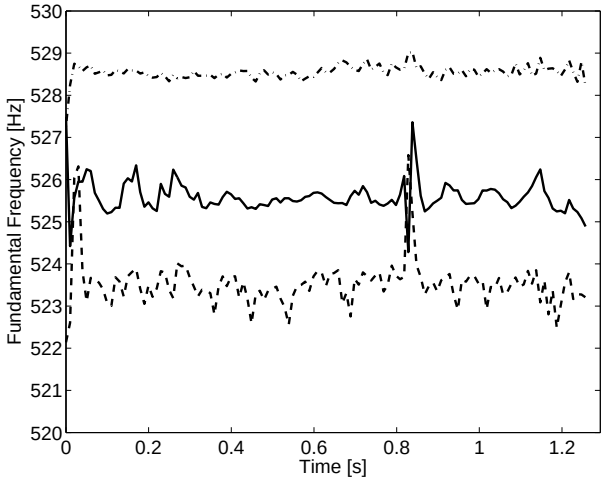

Fig. 3. Fundamental frequencies estimates obtained using the perfectly harmonic model (dash-dotted), the parametric inharmonicity model (dashed), and the perturbed model (solid).

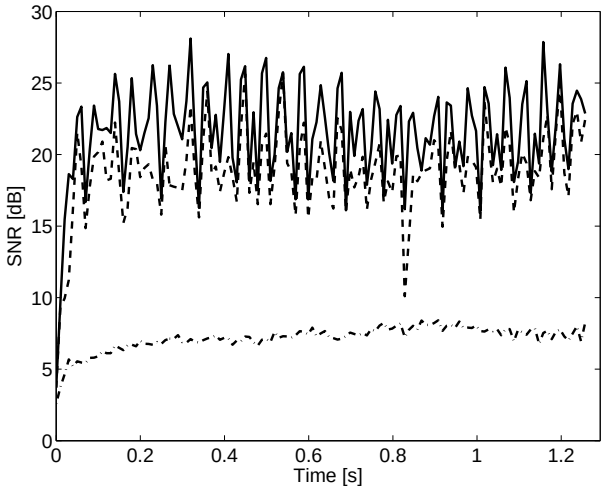

Fig. 4. Reconstruction SNR obtained using the perfectly harmonic model (dash-dotted), the parametric inharmonicity model (dashed), and the perturbed model (solid).

[8] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Joint high-resolution fundamental frequency and order estimation," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15(5), pp. 1635–1644, July 2007.

[9] M. G. Christensen, A. Jakobsson and S. H. Jensen, "Multi-pitch estimation using harmonic music," in *Rec. Asilomar Conf. Signals, Systems, and Computers*, 2006, pp. 521–525.

[10] S. Godsill and M. Davy, "Bayesian harmonic models for musical pitch estimation and analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2002, vol. 2, pp. 1769–1772.

[11] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Processing Lett.*, vol. 10(6), pp. 172–175, June 2003.