Sparse Multi-Pitch and Panning Estimation of Stereophonic Signals

Kronvall, Ted; Jakobsson, Andreas; Hansen, Martin Weiss; Jensen, Jesper Rindom; Christensen, Mads Græsbøll

Published in:
11th IMA International Conference on Mathematics in Signal Processing

Publication date:
2016

Document Version
Accepted author manuscript, peer reviewed version

SPARSE MULTI-PITCH AND PANNING ESTIMATION OF STEREOPHONIC SIGNALS

T. Kronvall[*]*, A. Jakobsson*[*]*, M. W. Hansen*[†]*, J. R. Jensen*[†]*, M. G. Christensen*[†]

[*]Dept. of Mathematical Statistics, Lund University, Sweden
[†]Audio Analysis Lab, AD:MT, Aalborg University, Denmark

ABSTRACT

In this paper, we propose a novel multi-pitch estimator for stereophonic mixtures, allowing for pitch estimation on multi-channel audio even if the amplitude and delay panning parameters are unknown. The presented method does not require prior knowledge of the number of sources present in the mixture, nor on the number of harmonics in each source. The estimator is formulated using a sparse signal framework, and an efficient implementation using the ADMM is introduced. Numerical simulations indicate the preferable performance of the proposed method as compared to several commonly used multi-channel single pitch estimators, and a commonly used multi-pitch estimator.

Index Terms— Sparse modeling, multi-pitch estimation, stereophonic signals, amplitude and delay panning, ADMM

1. INTRODUCTION

The problem of pitch (or fundamental frequency) estimation is important in a wide variety of applications, such as separation, enhancement, transcription, classification, and source localization (see, e.g., [1–4]). Given the importance of the field, the area has attracted notable attention, and a range of estimators have been developed (see, e.g., [5]). These estimators may be grouped in two categories, namely parametric and non-parametric methods. Examples of non-parametric methods are those based on autocorrelation [6, 7], the average magnitude difference function (AMDF) [8], and the harmonic product spectrum [9]. A common drawback of these methods is that they cannot distinguish between the fundamental pitch period and multiples of it, and they exhibit poor performance under noisy conditions. An example of a parametric method is the maximum likelihood (ML) pitch estimator [9] (see also [5] for further examples). It should be noted that a lot of recorded material is available in stereo, thereby allowing pitch estimation algorithms to exploit both channels. Several methods have been developed for this case, including the work presented in [10], which is based on a multi-microphone periodicity function (MPF), [11], presenting the multi-microphone maximum a posteriori (MAP) approach, and [12], wherein a multi-channel maximum likelihood (MC ML) pitch estimator is presented. It may also be noted that the idea of separating sources from a multi-channel mixture is used within the source separation [13] and array processing [14] research communities. Neither of these works explicitly make use of the stereophonic mixture created in recording studios when mixing several stereophonic signals. In such a case, each of the signals might have different mixing parameters, such as panning and equalization. The panning can be well described as consisting of amplitude and delay panning, wherein the former details the different gains that are applied to the channels to alter the perception of direction [15]. Delay panning will further alter the perception of direction, where a delay of more than 1 ms places the source mostly in the channel where the signal arrives first [16]. According to [17], delays in the 12-40 ms range, added to one of the channels of a signal, can enhance the spatial quality of the signal and add depth; the so-called Haas effect [18]. Recently, it has been investigated how knowing the amplitude and delay pan parameters influence the pitch estimation of stereophonic mixtures [19]. This study was limited to mixtures of single-pitch signals, and assumed perfect knowledge about the mixing parameters. In this paper, we propose an extension of this work, combining it with the ideas presented in [20]. The resulting estimator estimates the pitches in each of the stereo channels, as well as any inter-channel structure, which, in this work, we restrict to the amplitude and delay pannings, while allowing for an unknown number of sources, with an unknown number of harmonics. This is done by forming a two-step approach, first estimating the fundamental frequencies present in the multi-channel recording using a multi-pitch estimator based on the block sparse estimation framework, without taking the pan parameters into account. Then, in a second step, we propose to use the multi-pitch estimates to estimate amplitude and delay panning parameters for each signal in the mixture by solving a non-linear least squares (NLS) problem. The performance of the proposed method is verified against a number of multi-channel single pitch estimators, as well as a publicly available multi-pitch estimator, for both synthetic and real audio signals.

This work was supported in part by the Swedish Research Council, Carl Trygger's foundation, the Royal Physiographic Society in Lund, the Villum Foundation, and the Danish Council for Independent Research, grant ID: DFF 1337-00084.

Algorithm 1 The proposed PEBS-Pan algorithm

initialize $k := 0$, $\mathbf{u}(0) = \mathbf{u}_0, \mathbf{z}(0) = \mathbf{z}_0$, and $\mathbf{d}(0) = \mathbf{d}_0$
repeat {Multi-pitch estimation via ADMM}
  $\mathbf{z}(k) = \left(\mu \boldsymbol{I} + \mathbf{V}^H \mathbf{V}\right)^{-1} \left(\mu(\mathbf{u}(k) - \mathbf{d}(k)) + \mathbf{V}^H \mathbf{y}\right)$
  $\mathbf{u}(k+1) = \mathcal{T}\left(T\left(\mathbf{z}(k) - \mathbf{d}(k), \lambda_2/\mu\right), \lambda_3/\mu\right)$
  $\mathbf{d}(k+1) = \mathbf{d}(k) - (\mathbf{z}(k+1) - \mathbf{u}(k+1))$
  $k \leftarrow k + 1$
until convergence
for $p \in \mathcal{I}$ **do**
  Estimate $\boldsymbol{\Psi}_p$ using (15)
end for

2. SIGNAL MODEL

Consider a measured M-channel snapshot of an audio signal, $\mathbf{y}(t)$, formed as

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) & \dots & y_M(t) \end{bmatrix}^T \tag{1}$$

for $t = t_1, \dots, t_N$, where $(\cdot)^T$ denotes the transpose. In order to detail the proposed estimator, we begin by separating the signal into its voiced and its unvoiced parts, respectively, i.e.,

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{e}(t) \tag{2}$$

where $\mathbf{x}(t)$ denotes the voiced or tonal M-channel part of the audio, and $\mathbf{e}(t)$ denotes the unvoiced part, which includes, e.g., broadband noise and transients. The voiced part of the audio signal is assumed to be well described by a sum of K harmonically related complex-valued[1] signals, such that [5]

$$\mathbf{x}(t) = \sum_{k=1}^{K} \sum_{\mathcal{L}_k} \mathbf{b}_{k,\ell} e^{i2\pi f_k \ell t} \tag{3}$$

where f_k is the fundamental frequency of the k:th pitch, and ℓ is a frequency component in the integer set $\mathcal{L}_k$ of harmonics present in the k:th pitch, respectively. Furthermore, $\mathbf{b}_{k,\ell} \in \mathbb{C}^{M\times 1}$ denotes the unconstrained set of complex amplitudes representing the magnitude and phase of each frequency component, in each channel. The amplitudes $\mathbf{b}_{k,\ell}$ will contain information about the spatial connection of the channels, such as the location of the source and the M microphones. As shown in [4, 20], such structure may be exploited in forming joint estimators of the pitches and the source locations. In cases where instead the sources have been spatially enhanced virtually, by manually adding amplitude and delay panning for each source in the mixing process, the amplitudes may be expressed as [19]

$$\mathbf{b}_{k,\ell} = a_{k,\ell} \mathbf{h}_{k,\ell} \tag{4}$$

$$\mathbf{h}_{k,\ell} = \begin{bmatrix} h_{k,\ell,1} & \dots & h_{k,\ell,M} \end{bmatrix}^T \tag{5}$$

$$h_{k,\ell,m} = a_{k,\ell}\, g_{k,m} e^{-i2\pi f_k \ell \tau_{k,1}} \tag{6}$$

[1] For notational and computational convenience, we use the analytic representation of the audio.

where $a_{k,\ell}$ denotes the complex amplitude of harmonic ℓ in pitch k, for the m:th channel, $g_{k,m} \in \mathbb{R}$ the amplitude panning, and $\tau_{k,m} \in \mathbb{R}$ the delay panning, respectively. For the stereophonic case, the amplitude panning $g_{k,m}$ may be well modeled as [21]

$$g_{k,m} = \begin{cases} \cos\theta_k, & \text{for } m = 1 \\ \sin\theta_k, & \text{for } m = 2 \end{cases} \tag{7}$$

for some angle $\theta_k \in [0, 90]^\circ$.

3. SPARSE PITCH AND PANNING ESTIMATION

To allow for an unknown number of sources, K, as well as for the integer sets for each source, $\mathcal{L}_k$, $\forall k$, to be unknown, we build on the sparse frequency estimation framework presented in [22], and the subsequent extensions to group sparsity presented [20, 23, 24]. In order to allow for reliable pitch estimates without using any explicit knowledge about the model orders or the panning parameters, we approximate the signal model in (3) using an extended dictionary over P candidate sources, with $P \gg K$, with each containing up to $L_{\text{max}} > \max_k \mathcal{L}_k$ harmonics, i.e.,

$$\mathbf{x}(t) \approx \sum_{p=1}^{P} \sum_{\ell=1}^{L_{\text{max}}} \mathbf{b}_{p,\ell} e^{i2\pi f_p \ell t} \tag{8}$$

Here, it is assumed that P has been selected so large that the resulting grid of candidate pitches is so finely spaced that some of the candidate pitches may be assumed to be close with the actual pitch frequencies. Thus, one may assume that only K of the candidate amplitudes, $\mathbf{b}_{k,\ell}$, will be non-zero, representing the true amplitudes, whereas all the $P - K$ remaining amplitudes are zero. Denote the set of candidate amplitudes as

$$\boldsymbol{\Phi} = \{\boldsymbol{\Phi}_p\}_{p=1\dots P}, \quad \boldsymbol{\Phi}_p = \{\mathbf{b}_{p,\ell}\}_{\ell=1\dots L_{\text{max}}} \tag{9}$$

As a result, we strive to estimate the linear amplitudes in $\boldsymbol{\Phi}$ such that the sum of squared residuals is minimized, i.e.,

$$\underset{\boldsymbol{\Phi}}{\text{minimize}} \sum_{t=t_1}^{t_N} \left\| \mathbf{y}(t) - \mathbf{x}_{\boldsymbol{\Phi}}(t) \right\|_2^2 \tag{10}$$

where $\|\cdot\|_2$ denotes the Euclidean norm and with $\mathbf{x}_{\boldsymbol{\Phi}}(t)$ as in (8). Clearly, as stated, (10) will not yield the desired solution. By instead introducing appropriate penalties, one may form a minimization that induce a suitably sparse solution. Here, we propose that (10) is reformulated as the (convex) minimization

$$\underset{\boldsymbol{\Phi}}{\text{minimize}} \sum_{t=t_1}^{t_N} \left\| \mathbf{y}(t) - \mathbf{x}_{\boldsymbol{\Phi}}(t) \right\|_2^2 + N\lambda_2 \sum_{p=1}^{P} \sum_{\ell=1}^{L_{\text{max}}} \left\| \mathbf{b}_{p,\ell} \right\|_2 + N\lambda_3 \sum_{p=1}^{P} \sqrt{\sum_{\ell=1}^{L_{\text{max}}} \|\mathbf{b}_{p,\ell}\|_2^2} \tag{11}$$

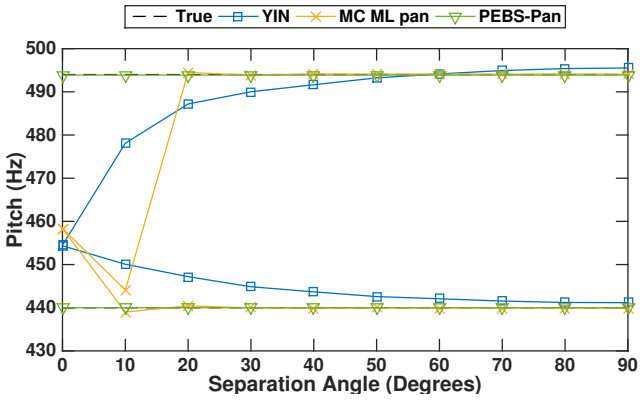

Fig. 1. Average fundamental frequency estimates for synthetic data containing two single-pitch sources, where amplitude and delay panning have been applied to each pitch individually.

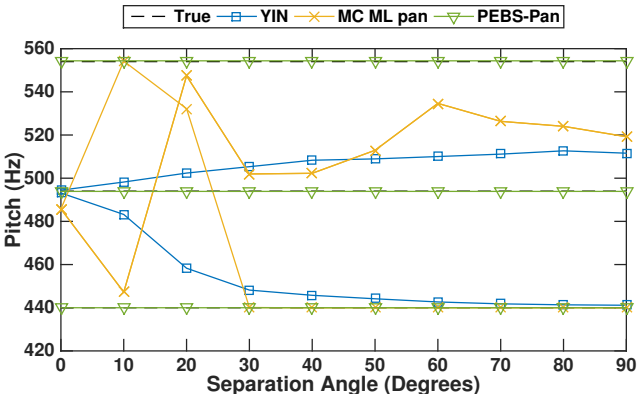

Fig. 2. Average fundamental frequency estimates for synthetic data containing one single-pitch source and one multi-pitch source with two pitches. Amplitude and delay panning have been applied to each source separately.

where the first penalty has been included to reduce the overfitting of the number of sinusoidal components present in each active pitch. Similarly, the second penalty will promote solutions with few pitches, by adding a block sparse constraint to the entire set of harmonics in a pitch candidate. These user-specified penalties implicitly set the model orders by imposing a lower limit on the norm of the amplitudes, $\mathbf{b}_{p,\ell}$. These parameters may, for instance, be chosen using cross-validation or via some simple heuristics from the periodogram. For instance, one may calculate the periodogram for each of the channels, and then calculate the normalized Euclidean norm across all channels at each frequency grid point. As a result, if setting λ_2 to 0.1, the amplitude for all frequencies having a block magnitude less than 0.1 will be set to zero. A similar analysis may then be done for λ_3. The user parameters thus correspond to a minimum power constraint, which implicitly sets the model orders in the signal. We proceed to examine how the estimated $\mathbf{\Phi}$ may be used to estimate the amplitude and delay panning parameters. To that end, consider the integer set of non-zero amplitude blocks

$$\mathcal{I} = \left\{ p : \sum_{\ell=1}^{L_{\max}} \left\| \widehat{\mathbf{b}}_{p,\ell} \right\|_2^2 > 0 \right\} \tag{12}$$

and $\mathbf{\Phi}_{\mathcal{I}}$ as the set of pitches present in the signal, where $\widehat{\mathbf{b}}_{p,\ell}$ denotes the resulting estimates of $\mathbf{b}_{p,\ell}$. Let

$$\mathbf{\Psi}_{\mathcal{I}} = \{\mathbf{\Psi}_p\}_{p\in\mathcal{I}}, \quad \mathbf{\Psi}_p = \{\theta_p, \tau_p\} \tag{13}$$

denote the corresponding sought sets of amplitude and delay panning parameters (for a stereophonic signal, i.e., for $M = 2$). As it is assumed that there is only one panning and delay parameter per pitch, these may be estimated as the solution to an NLS problem based on (4), i.e,

$$\underset{\mathbf{\Psi}_p}{\text{minimize}} \sum_{\ell=1}^{L_{\max}} \left\| \widehat{\mathbf{b}}_{p,\ell} - a_{p,\ell}\mathbf{h}_{p,\ell} \right\|_2^2 \tag{14}$$

for each $p \in \mathcal{I}$, and where $\mathbf{h}_{p,\ell}$ is a function of $\mathbf{\Psi}_p$. Minimizing (14) over $a_{p,\ell}$ yields $\hat{a}_{p,\ell} = \mathbf{h}_{p,\ell}{}^{\dagger}\, \widehat{\mathbf{b}}_{p,\ell}$, where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudo inverse. Inserted into (14), we obtain the NLS optimization problem

$$\underset{\mathbf{\Psi}_p}{\text{maximize}} \sum_{\ell=1}^{L_{\max}} \left\| \left(\mathbf{h}_{p,\ell}\mathbf{h}_{p,\ell}{}^{\dagger} \right) \widehat{\mathbf{b}}_{k,\ell} \right\|_2^2 \tag{15}$$

which solved $\forall p \in \mathcal{I}$ yields an estimate of $\mathbf{\Psi}_{\mathcal{I}}$. Although not required to form the pitch estimates[2], this estimate allows for a reduced computational complexity for long segments of audio where the same panning is used, with $\mathbf{\Psi}$ instead being used to initialize a computationally cheaper algorithm, such as the one proposed in [19].

4. FAST IMPLEMENTATION USING ADMM

By collecting the snapshots over a frame of N samples, the measurement matrix $\mathbf{Y} \in \mathbb{C}^{N\times M}$ may be formed as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}(t_0) & \dots & \mathbf{y}(t_{N-1}) \end{bmatrix}^T \tag{16}$$

such that (11) may be reformulated as

$$\underset{\mathbf{B}}{\text{minimize}} \ \|\mathbf{Y} - \mathbf{W}\mathbf{B}\|_2^2 + N\lambda_2 \sum_{p=1}^{P}\sum_{\ell=1}^{L_{\max}} \|\mathbf{b}_{k,l}\|_2 + N\lambda_3 \sum_{p=1}^{P} \|\mathbf{B}_p\|_{\mathcal{F}}^2 \tag{17}$$

[2]For this reason, due to space limitations, we do not show the resulting performance of these estimate, but note only that they are accurate.

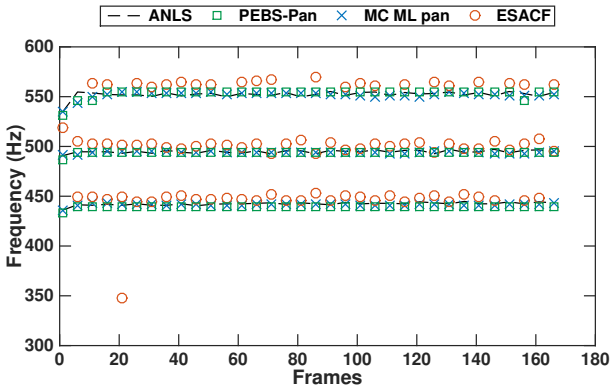

Fig. 3. Pitch tracks for a mixture of three trumpets, each playing a single pitch, where amplitude and delay panning have been individually applied.

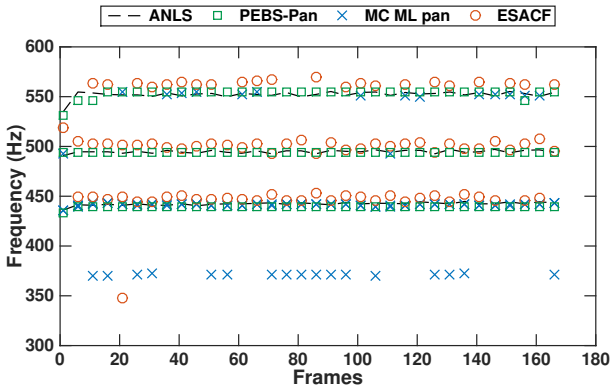

Fig. 4. Pitch tracks for three trumpets grouped as one single-pitch source and one multi-pitch source with two pitches.

where

$$\mathbf{W} = \left[\begin{array}{ccc} \mathbf{W}_1 & \dots & \mathbf{W}_P \end{array} \right] \tag{18}$$

$$\mathbf{W}_p = \left[\begin{array}{ccc} \mathbf{z}^1 & \dots & \mathbf{z}^{L_{\max}} \end{array} \right] \tag{19}$$

$$\mathbf{z}_p = \left[\begin{array}{ccc} e^{i2\pi f_p t_1} & \dots & e^{i2\pi f_p t_N} \end{array} \right]^T \tag{20}$$

$$\mathbf{B} = \left[\begin{array}{ccc} \mathbf{B}_1^T & \dots & \mathbf{B}_P^T \end{array} \right]^T \tag{21}$$

$$\mathbf{B}_p = \left[\begin{array}{ccc} \mathbf{b}_{p,1} & \dots & \mathbf{b}_{p,L_{\max}} \end{array} \right]^T \tag{22}$$

As the user parameters are non-negative and the functions are convex, (17) is a convex optimization problem, allowing the solution to be found using standard convex minimization technique, such as, CVX [25], SeDuMi [26], or [27]. However, such estimators often scale poorly with increasing data lengths and/or larger P and $L_{\max}$, resulting in an unnecessary high computational complexity. To alleviate this, we here introduce an alternating direction method of multipliers (ADMM) version of (17). Instead of optimizing over the entire problem, the ADMM splits (17) into two significantly simpler subproblems, which may be iteratively solved. Due to lack of space, we omit the derivations of the ADMM, which may be found in, e.g., [28], and instead just outline the implementation of our method, termed Pitch Estimation using Block Sparsity for Panned signals (PEBS-Pan) in Algorithm 1. Here, we have set

$$\mathbf{y} = \operatorname{vec} \mathbf{Y}, \ \mathbf{z} = \operatorname{vec} \mathbf{B}, \ \mathbf{V} = \boldsymbol{I} \otimes \mathbf{W} \tag{23}$$

where $\boldsymbol{I}$ denotes the identity matrix of appropriate size and $\otimes$ denotes the Kronecker product, respectively, and where $\mathcal{T}(\cdot)$ and $T(\cdot)$ denote shrinkage operators similar to the ones used in [20]. An estimate of the present pitches may then be obtained by locating fundamental frequencies of the largest $\|\mathbf{B}_p\|_{\mathcal{F}}, p \in \mathcal{I}$.

5. NUMERICAL RESULTS

We proceed to examine the performance of the proposed method for both synthetic and recorded audio, in comparison with MC ML pan [19], YIN [7], and ESACF [29]. For the simulated data, we consider a stereophonic mixture, with $N = 200$ samples, in Gaussian noise with SNR $= 20$ dB. We first consider a signal containing $K = 2$ sources, each with unit amplitude and $L_k = 8$ harmonics, with pitches given as the notes A4 ($f_1 = 440$ Hz) and B4 ($f_2 = 493.883$ Hz), when the A440 pitch standard is used. These pitches are individually panned and delayed. Figure 1 illustrates the average pitch estimates for 50 Monte-Carlo simulations versus difference in amplitude panning. Clearly, all the estimators resolves the two pitches for higher separation angles, although YIN and MC ML yields poor estimates for low separation angles. In the second simulation, the signal contains $K = 3$ pitches, each with unit amplitude and $L_k = 8$ harmonics, with the abobe pitches and that of C♯5 ($f_3 = 554.365$ Hz). Here, f_2 and f_3 are given the same amplitude and delay panning. Figure 2 illustrates the average pitch estimates, where both the YIN and MC ML pan estimators fail to resolve the three pitches. Examining real audio signals, we consider a recording of three trumpets playing the above notes, each one with different levels of amplitude and delay panning. Figure 3 illustrates the pitch tracks found by the estimators, together with the ANLS estimate, obtained for each channel separately. Here, the MC ML Pan and PEBS-Pan estimators are all able to yield reasonable estimates. In Figure 4, the higher notes have instead been grouped together as a multi-pitch source. In this case, the MC ML pan method fails to distinguish between the pitches, whereas PEBS-Pan yields accurate estimates. In this case, the ESACF method can still resolve the three pitches, albeit yielding poorer estimates. All signals have been sampled at 8 kHz. For PEBS-Pan, $\lambda_2 = 10^{-2}$ and $\lambda_3 = 10^{-2} L_{\max}$ for the synthetic data; $\lambda_2 = 10^{-1}$ and $\lambda_3 = 5 * 10^{-1} L_{\max}$ for the noisier real data.

6. REFERENCES

[1] M. I. Mandel, R. J. Weiss, and D. P. W. Ellis, "Model-based expectation-maximization source separation and localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 382–394, Feb. 2010.

[2] A. Klapuri and M. Davy, Eds., *Signal Processing Methods for Music Transcription*, Springer, New York, 2006.

[3] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, Jul 2002.

[4] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear least squares methods for joint DOA and pitch estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 923–933, 2013.

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

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

[7] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930, 2002.

[8] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 5, pp. 353–362, Oct 1974.

[9] M. Noll, "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum and a maximum likelihood estimate," in *Proc. Symp. Comput. Process. Commun.* 1969, vol. XIX, pp. pp. 779–797, Polytechnic Press: Brooklyn, New York.

[10] F. Flego and M. Omologo, "Robust f0 estimation based on a multi-microphone periodicity function for distant-talking speech," in *Proc. European Signal Processing Conf.*, 2006.

[11] T. Gerkmann, R. Martin, and D. Dalga, "Multi-microphone maximum a posteriori fundamental frequency estimation in the cepstral domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 4505–4508.

[12] M. G. Christensen, "Multi-channel maximum likelihood pitch estimation," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 409–412, 2012.

[13] B. Gold, N. Morgan, and D. Ellis, *Speech and Audio Signal Processing - Processing and Perception of Speech and Music, Second Edition.*, Wiley, 2011.

[14] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer Topics in Signal Processing. Springer, 2008.

[15] V. Pulkki, *Spatial sound generation and perception by amplitude panning techniques*, Helsinki University of Technology, 2001.

[16] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, 1997.

[17] B. Katz, *Mastering Audio - The Art and the Science*, Focal Press, 2007.

[18] H. Haas, "The influence of a single echo on the audibility of speech," *J. Audio Eng. Soc.*, vol. 20, no. 2, pp. 146–159, 1972.

[19] M. W. Hansen, J. R. Jensen, and M. G. Christensen, "Pitch estimation of stereophonic mixtures of delay and amplitude panned signals," in *Proc. European Signal Processing Conf.*, 2015.

[20] S. I. Adalbjörnsson, T. Kronvall, S. Burgess, K. Åström, and A. Jakobsson, "Sparse Localization of Harmonic Audio Sources," *IEEE Transactions on Audio, Speech, and Language Processing*, 2015, (To appear).

[21] J. C. Bennett, K. Barker, and F. O. Edeko, "A new approach to the assessment of stereophonic sound system performance," *J. Audio Eng. Soc.*, vol. 33, no. 5, pp. 314–321, 1985.

[22] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra," in *17th World Congress IFAC*, Seoul, jul 2008, pp. 10225–10229.

[23] M. Yuan and Y. Lin, "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[24] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Multi-Pitch Estimation Exploiting Block Sparsity," *Elsevier Signal Processing*, vol. 109, pp. 236–247, April 2015.

[25] Inc. CVX Research, "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta," `http://cvxr.com/cvx`, Sept. 2012.

[26] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, August 1999.

[27] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming Ser. B*, vol. 95, pp. 189–217, 2003.

[28] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[29] T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 708–716, Nov 2000.