USING OPTIMAL TRANSPORT FOR ESTIMATING INHARMONIC PITCH SIGNALS

Filip Elvander[*]*, Stefan Ingi Adalbjörnsson*[†]*, Johan Karlsson*[‡]*, and Andreas Jakobsson*[*]

[*]Div. of Mathematical Statistics, Lund University, Sweden
[†]Div. of Mathematics, Lund University, Sweden
[‡]Dept. of Mathematics, Royal Institute of Technology, Sweden
emails: {filipelv, sia, aj}@maths.lth.se, johan.karlsson@math.kth.se

ABSTRACT

In this work, we propose a novel multi-pitch estimation technique that is robust with respect to the inharmonicity commonly occurring in many applications. The method does not require any *a priori* knowledge of the number of signal sources, the number of harmonics of each source, nor the structure or scope of any possibly occurring inharmonicity. Formulated as a minimum transport distance problem, the proposed method finds an estimate of the present pitches by mapping any found spectral line to the closest harmonic structure. The resulting optimization is a convex and highly tractable linear programming problem. The preferable performance of the proposed method is illustrated using both simulated and real audio signals.

***Index Terms*—** Multi-pitch estimation, frequency clustering, inharmonicity, optimal transport distance, convex optimization.

1. INTRODUCTION

The problem of estimating the fundamental frequency, or pitch, of a harmonic, or close-to-harmonic, signal occurs in a wide range of applications [1–9]. Often, the problem is complicated by the number of sources being unknown, as is the number of components detailing each source. Furthermore, some sources, such as, e.g., audio signals resulting from stringed instruments, exhibits inharmonicity, implying that higher order components may deviate from the harmonic model, often with increasing deviation for the higher harmonics [10–12]. In such scenarios, a naive approach exploiting the sinusoidal frequency model in the time domain results in a cumbersome high dimensional optimization problem, as the uncertainty due to the inharmonicity will occur in the nonlinear frequency parameter. Previously, this problem has been approached by approximate optimization in the time domain [12, 13], approximating the frequency uncertainty with an uncertainty in the functional form of the sinusoid [10], or via a subspace-based framework robust to such deviations [14]. For certain applications, there also exists source specific pitch estimators that rely on the inharmonicity following a parametric model, see, e.g., [15]. However, such estimators are generally unable to resolve cases when harmonics from different sources overlap, as commonly occurs, for instance, in western music playing in harmony.

In order to handle such situations, while still allowing for an unknown number of sources, we here formulate the multi-pitch problem such that the estimated pitches are obtained as the ones minimizing a particular (convex) Monge-Kantorovich optimal transportation problem. These methods have also earlier been shown useful for problems in signal analysis, e.g., for clustering, tracking, registration, and robust identification [16–19]. Transport problems have a rich history going back to questions concerning how to most efficiently transport soil from one location to another, and has since attracted attention in various fields (see [20] and references therein). An example of this is the facility localization problem, where for a set of customers one seeks to determine locations of facilities that minimizes the sum of the distances from each customer to its closest facility. As we will see, the multi-pitch estimation problem can be reformulated as a facility location problem [20].

In this setting, the harmonic model (facilities) should be selected so that the spectral components (customers) can be transported to the closest harmonic model with minimal total cost. In this case, the mass to be moved constitutes the amplitude of the observed spectral component at a given frequency; as this amplitude may originate from two or more sources which have overlapping harmonics at the given frequency, we should allow the optimization to transport parts of the observed amplitude to different harmonic candidates. We further wish to introduce restrictions on the allowed mass transport problem such that ambiguity with different suboctaves are avoided, promoting spectrally smooth solutions similar to those proposed in [2, 3, 21, 22]. As we show in the following, the desired optimization problem can be formulated as a linear programming (LP) problem, for which powerful solvers are available, even for big data applications [23]. In the numerical section, we illustrate the preferable performance of the proposed method as compared to several previously suggested methods, for both simulated and real audio signals.

This work was supported in part by the Swedish Research Council, Carl Trygger's foundation, and the Royal Physiographic Society in Lund.

2. SIGNAL MODEL

Consider N samples of a (reasonably) stationary signal, $y(t)$, that may be well described as a sum of close-to-harmonic sources, $x(t)$, corrupted by an additive broadband noise, $e(t)$, such that $y(t) = x(t) + e(t)$, where[1]

$$x(t) = \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} a_{k,\ell} e^{i2\pi(f_k \ell + \Delta_{k,\ell})t} \,. \tag{1}$$

Here, K denotes the number of sources, each containing L_k close-to-harmonic signal components. The constant f_k denotes the pitch of the kth source, and the constants $a_{k,\ell}$ and $\Delta_{k,\ell}$ denote the complex amplitude and frequency deviation, respectively, of the ℓth harmonic of the kth source. The deviation will thus be zero for fully harmonic sources, whereas $\Delta_{k,\ell}$ otherwise details the inharmonicity. Depending on the source, one may have models for such inharmonicities, such as the model used for pianos (see, e.g., [11]). In the frequency domain, the assumed signal may thus be represented as

$$X(f) = \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} a_{k,\ell} \delta(f - f_k \ell - \Delta_{k,\ell}) \tag{2}$$

where $\delta(\cdot)$ denotes the Dirac delta function. In this work, we aim at estimating both the number of sources, K, and their pitches, f_k, while allowing for unknown frequency deviations, $\Delta_{k,\ell}$. In order to do so, we consider the transport cost (see, e.g., [20]) associated with assigning each spectral component to a set of candidate pitches, i.e., the transport cost of moving the component onto the assumed harmonic structure related to each candidate pitch.

In order to introduce notation, let $\mathcal{F}$ denote the set of observed spectral components in the signal of interest, whereas $\mathbf{\Omega}$ denotes the set of all considered candidate pitches. Furthermore, let M and P denote the number of elements of the sets $\mathcal{F}$ and $\mathbf{\Omega}$, respectively. Here, the number of candidate pitches are assumed to be much larger than the number of sources, such that $P \gg K$. Finally, each candidate pitch is assumed to have at most $L_{\max} \geq \max_k L_k$ harmonics.

3. OPTIMAL TRANSPORT

In order to find an optimal assignment of the amplitudes corresponding to the observed line spectrum frequencies to the set of pitch candidates, one needs to define a function describing the cost of a certain assignment and then minimize this function over all possible assignments. In order to do this, let the function $c(f, f_p)$ describe the cost of moving one unit of amplitude from the line spectral frequency f to the pitch candidate f_p. For example, the cost of assigning all amplitudes in the line spectrum $Y(f)$, defined as $Y(f) = \sum_{f_m \in \mathcal{F}} a_{f_m} \delta(f - f_m)$, where a_{f_m} denotes the amplitude of the spectral line at frequency f_m, to the candidate pitch f_p is

$$\sum_{f_m \in \mathcal{F}} |a_{f_m}| \, c(f_m, f_p) \,. \tag{3}$$

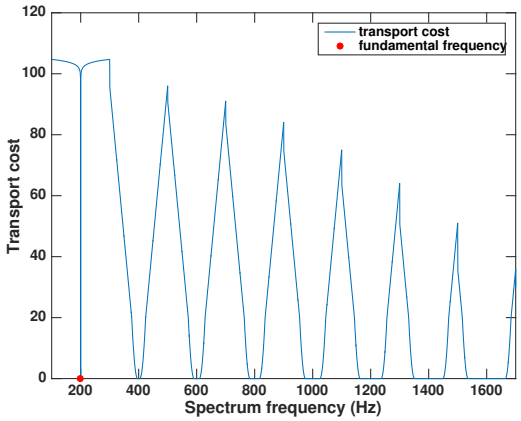

Fig. 1. Transportation cost for candidate pitch with fundamental frequency 200 Hz.

To describe the cost of a general assignment, let $\mathbf{C}$ be the $P \times M$ matrix whose (p, m):th element is equal to $c(f_m, f_p)$. Also, let $\mathbf{W}$ be the $P \times M$ matrix describing the amplitude assignment, i.e., the (p, m):th element of $\mathbf{W}$ describes how much of the magnitude $|a_m|$ that is assigned to candidate pitch f_p. Thus, to ensure that all the estimated spectral content is mapped to some pitch, the sum of the m:th column of $\mathbf{W}$ must be equal to $|a_m|$. With this, the cost of an assignment described by $\mathbf{W}$ may be expressed as $\text{tr}\left(\mathbf{C}^T\mathbf{W}\right)$, where $(\cdot)^T$ denotes the transpose, and $\text{tr}(\cdot)$ denotes the trace of a matrix. Defining the $M \times 1$ vector $\mathbf{a} = \begin{bmatrix} |a_1| & \dots & |a_M| \end{bmatrix}^T$, and letting $\mathbf{1}_P$ be a $P \times 1$ vector of ones, one may formulate the desired optimal transport problem as

$$\begin{aligned} \underset{\mathbf{W},\mathbf{x}}{\text{minimize}} \quad & \text{tr}\left(\mathbf{C}^T\mathbf{W}\right) \\ \text{subject to} \quad & \mathbf{W}^T\mathbf{1}_P = \mathbf{a}, \quad \mathbf{x}^T\mathbf{1}_P = K \\ & \mathbf{W} \leq \mathbf{x}\mathbf{a}^T, \quad \mathbf{W} \geq \mathbf{0} \\ & x_i \in \{0, 1\} \;,\; i = 1, \dots, P \end{aligned} \tag{4}$$

where the inequalities for matrices and vectors should be interpreted element-wise. The binary vector $\mathbf{x}$ here controls whether a pitch candidate f_p is present in the solution or not, i.e., if $x_p = 1$, then f_p is present and if $x_p = 0$, then it is not. However, as x_i are binary variables, this problem is not convex. Furthermore, this formulation assumes precise knowledge of the number of sources, K, which in general is unknown. In order to remedy this, we consider the convex relaxation (cf. [16])

$$\begin{aligned} \underset{\mathbf{W},\mathbf{x}}{\text{minimize}} \quad & \text{tr}\left(\mathbf{C}^T\mathbf{W}\right) + \lambda \mathbf{1}_P^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{W}^T\mathbf{1}_P = \mathbf{a}, \quad \mathbf{W} \leq \mathbf{x}\mathbf{a}^T \\ & \mathbf{x} \geq 0, \quad \mathbf{W} \geq \mathbf{0} \end{aligned} \tag{5}$$

[1] For computational and notational simplicity, we here use the time-discrete analytical version of the measured data.

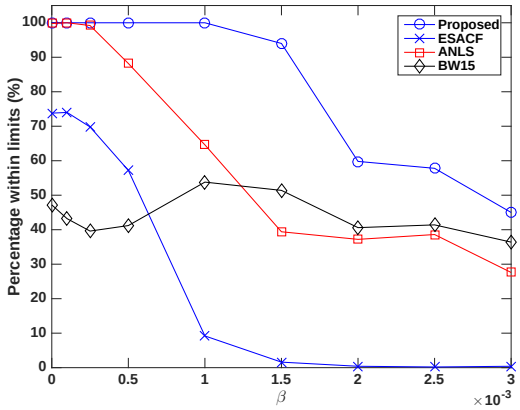

Fig. 2. Percentage of pitch estimates found within $\pm 3\%$ of the ground truth in the simulated data case.

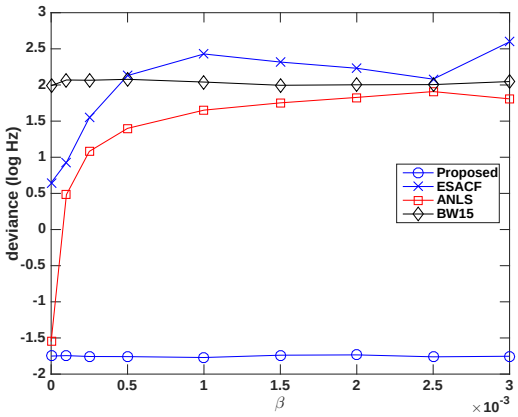

Fig. 3. Expected maximal absolute deviation of pitch estimates from the ground truth in the simulated data case.

with $\lambda > 0$. The second term of the objective function in (5) allows for an implicit choice of the sparsity of $\mathbf{x}$ via the regularization parameter λ. However, using the relaxation in (5), the cost function is unable to distinguish between sub-octaves, i.e., the row of C corresponding to some f_0 that may be greater or equal to the row corresponding to $f_0/2$. Fortunately, this may be included in the modeling by considering the structure of the amplitude assignment. Specifically, for each candidate pitch f_p, define an $L_{\max} \times M$ matrix $\mathbf{L}^{(p)}$ that describes the mapping between the line spectral frequencies and the harmonics corresponding to f_p. That is, the (ℓ, m):th element of $\mathbf{L}^{(p)}$ is equal to one if $f_p \ell$ is the harmonic of pitch f_p that is closest in frequency to the line spectral frequency f_m, and zero otherwise. As each spectral line is mapped to precisely one harmonic, each column of $\mathbf{L}^{(p)}$ has exactly one element equal to one, whereas all the rest are zero. This linear mapping thus allows for the inclusion of constraints on the relative amplitudes of each pitch. For example, it may be used to promote spectral smoothness in each pitch. In this work, we restrict our attention to only requiring active pitches to have non-zero amplitude in the first harmonic. As this constraint is then convex it can easily be included in (5), yielding

$$\begin{aligned} \underset{\mathbf{W},\mathbf{x}}{\text{minimize}} \quad & \operatorname{tr}\left(\mathbf{C}^T\mathbf{W}\right) + \lambda \mathbf{1}_P^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{W}^T \mathbf{1}_P = \mathbf{a}, \quad \mathbf{W} \leq \mathbf{x}\mathbf{a}^T \\ & \mathbf{x} \geq 0, \quad \mathbf{W} \geq \mathbf{0} \\ & \left((Q+1)\,\mathbf{e}_1 - \mathbf{1}_M\right)^T \mathbf{L}^{(p)} \left[\mathbf{W}\right]_{p\cdot}^T \leq 0 \end{aligned} \tag{6}$$

for $p = 1, \ldots, P$. Here, $Q > 1$ assures that a scaled version of the amplitude assigned to the first harmonic dominates the amplitude assigned to the rest of the harmonics, thus enforcing solutions where active pitches have non-zero amplitude assigned to their first harmonics. In our simulations, we use $Q = 3L_{\max}$. Here, $\mathbf{e}_1$ denotes the $M \times 1$-vector with its first element equal to one, and the rest zero, with $[\mathbf{W}]_{p\cdot}$ denoting row p of $\mathbf{W}$. It is worth noting at the resulting problem is an LP, which may thus be solved using standard convex solvers.

4. CHOICE OF TRANSPORT COST FUNCTION

To model the amplitude distribution of a pitch, the transport cost function $c(\cdot,\cdot)$ should assign the cost of associating amplitude at a frequency f_m to a candidate pitch f_p depending on the distance between f_m and the closest harmonic of f_p, e.g.,

$$c(f_m, f_p) = \min_{\ell \in \mathbb{N}} c(f_m, f_p \ell) = \min_{\ell \in \mathbb{N}} |f_m - f_p \ell|^2 . \tag{7}$$

However, this function would too harshly penalize inharmonicity, as the higher harmonics of inharmonic pitches could typically deviate significantly from integer multiples of the pitch. We therefore propose to only have harsh penalties for the pitch, while allowing subsequent harmonics to deviate somewhat more. Specifically, for the first harmonic, let

$$c_1(f_m, f_p) = \rho s_+ \left(|f_p - f_m|, \frac{\Delta_f}{2} \right)^\nu \tag{8}$$

where $s_+(\cdot)$ is the soft threshold function defined as

$$s_+\left(x, \frac{\Delta_f}{2}\right) = \left| \max\left(x - \frac{\Delta_f}{2}, 0\right) \right| \tag{9}$$

and Δ_f is the spacing of the candidate pitch grid. Thus, we allow for a deadzone corresponding to the grid resolution, while penalizing larger deviances according to a scaled, highly non-convex, pseudo-norm. To allow for increasing deviations with higher harmonics, we instead use

$$c_\ell(f_m, f_p) = \min\left(\epsilon_\ell(f_m, f_p), \xi \epsilon_\ell(f_m, f_p)^2\right) \tag{10}$$

where $\epsilon_\ell(f_m, f_p) = s_+(|f_p - f_m|, \psi f_p \ell^2)$. Thus, the width of the deadzone is dependent on the harmonic order ℓ as well as being scaled by a small number, ψ. In our simulations, $\rho = 100$, $\xi = 0.01$, $\nu = 0.05$, and $\psi = 0.005$. An illustration of the transport cost function is shown in Figure 1, where the cost of assigning frequencies on the interval $(100, 1700)$ Hz is shown for a pitch of 200 Hz. Here, the width of the deadzone scales quadratically with the harmonic order.

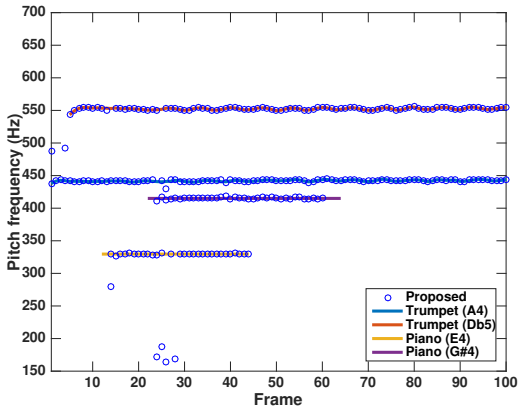

Fig. 4. Estimated fundamental frequencies for a signal containing two trumpet notes as well as two piano notes.

5. NUMERICAL RESULTS

We proceed to examine the performance of the proposed method using both simulated and measured audio signals. In both settings, the line spectrum is estimated using the MUSIC estimator [24], with $M \gg \sum_k L_k$. The amplitudes a_m are then estimated using least-squares. Initially, we examine a simulated signal consisting of two pitches, with pitches f_1 and f_2, with varying degrees of inharmonicity. The harmonics of the pitches are modelled using the piano model (see, e.g., [11]), i.e., $f_{k,\ell} = f_k \ell \sqrt{1 + \beta \ell^2}$, for $\ell = 1, \ldots, L_k$ and $k = 1, 2$, where the parameter $\beta \ll 1$ controls the level of inharmonicity. The frequencies f_1 and f_2 are drawn uniformly on the intervals $(300, 390)$ Hz and $(400, 540)$ Hz, respectively. The harmonic orders L_k are drawn uniformly on $[8, 12]$, whereas the magnitude of each harmonic is drawn uniformly on $(0.75, 1.25)$, with phases drawn uniformly on $[0, 2\pi)$. We thereafter add an additive white Gaussian noise to the signal, resulting in a signal-to-noise-ratio of 30 dB. The signal is then sampled for 30 ms at 40 kHz. This is done for 500 Monte Carlo simulations and for varying values of β. Performance is then measured as the percentage of the simulations in which both pitch estimates are found within $\pm 3\%$ of their respective ground truths and where no erroneous extra pitch estimates are produced. For the proposed method, we set $L_{\max} = 20$ and $\lambda = 15$. As comparison, we include three other types of pitch estimators; the approximate non-linear least squares estimator (ANLS) (see, e.g., [5]); the autocorrelation-based enhanced summary autocorrelation (ESACF) estimator [25]; and the method presented in [9], which is based on probabilistic latent component analysis. The latter method, hereafter referred to as BW15, is specifically designed for multi-pitch estimation for music signals, with pitch estimates restricted to the chromatic Western scale, i.e., to the keys of the piano. This frequency resolution corresponds precisely to the chosen accuracy limit of $\pm 3\%$ of the ground truth pitches. The method is based on extensive training on a database of various forms of signals[2]. As ANLS requires knowledge of both the number of sources and the number of harmonics for each source, it is here provided with oracle model order knowledge. For all methods, the algorithm settings recommended by their respective authors have been used. As shown in Figure 2, the proposed method outperforms the other methods for all considered levels of inharmonicity. It may be noted that the performance of the BW15 method is not strictly decreasing with the inharmonicity parameter β; rather, the best performance is achieved for the value $\beta = 10^{-3}$, arguably due to this being the best match to the method's training library. We also evaluate the accuracy of the pitch estimates, measured as the maximum absolute deviation of each estimate from its corresponding ground truth, conditioned on that the estimates are found within $\pm 3\%$ of their respective ground truths. The results are shown in Figure 3, with deviation shown in log-scale. Again, the proposed method outperforms all comparison methods.

	Proposed	ESACF	BW15
Accuracy	0.928	0.691	0.366
Precision	0.974	0.984	0.391
Recall	0.952	0.699	0.849

Table 1. Performance measures for the proposed method as well as the ESACF and BW15 methods.

In Figure 4, we study a real audio signal consisting of two harmonic trumpet signals and two piano signals with some inharmonicity. Specifically, the signal is composed of two trumpet signals, with pitches 440 and 554.37 Hz, corresponding to the notes A4 and D♭5, and of two piano notes, with pitches 329.65 and 415.3 Hz, corresponding to the notes E4 and G♯4. Ground truth estimates for the trumpet pitches have been obtained by applying the YIN estimator [26] to the single channel recordings. Ground truths for the pianos are known as the signals are simulated using software synthesizers. As can be seen in Figure 4, the proposed method is able to correctly group the frequencies into the correct pitches, with only small errors during the onset phase, where the frequency content is highly transient and non-sparse. The recording was sampled at 44.1 kHz and was subdivided into non-overlapping estimation frames of length 30 ms. The settings for the proposed method was $L_{\max} = 10$ and $\lambda = 15$. Table 1 compares the proposed method to the ESACF and BW15 methods, while excluding ANLS as exact model order information of the number of harmonics of each source is unavailable. The table presents the performance measures *Accuracy*, *Precision*, and *Recall* [27]. As can be seen, the performance of the proposed method is clearly better than that of the comparison methods; likely, this results from ESACF having problems with estimating the pitches of the inharmonic pianos, whereas BW15 suffers from not being able to accurately estimate the trumpets, perhaps caused by bad match to its training data set.

[2]The implementation used was provided online by the authors of [9].

6. REFERENCES

[1] R. B. Randall, *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*, John Wiley & Sons, Chichester, UK, 2011.

[2] 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.

[3] F. Elvander, T. Kronvall, S. I. Adalbjörnsson, and A. Jakobsson, "An Adaptive Penalty Multi-Pitch Estimator with Self-Regularization," *Elsevier Signal Processing*, vol. 127, pp. 56–70, October 2016.

[4] M. S. Reza, M. Ciobotaru, and V. G. Agelidis, "Robust Technique for Accurate Estimation of Single-Phase Grid Voltage Fundamental Frequency and Amplitude," *IET Generation, Transmission Distribution*, vol. 9, no. 2, pp. 183–192, 2015.

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

[6] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–102, April 2009.

[7] P. Somervuo, A. Harma, and S. Fagerlund, "Parametric Representations of Bird Sounds for Automatic Species Recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 2252–2263, November 2006.

[8] 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.

[9] E. Benetos and T. Weyde, "An Efficient Temporally-Constrained Probabilistic Model for Multiple-Instrument Music Transcription," in *Proceedings of the 16th International Society for Music Information Retrieval Conference*, Malaga, Spain, October 2015.

[10] N. R. Butt, S. I. Adalbjörnsson, S. D. Somasundaram, and A. Jakobsson, "Robust Fundamental Frequency Estimation in the Presence of Inharmonicities," in *38th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, May 26–31, 2013.

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

[12] S. M. Nørholm, J. R. Jensen, and M. G. Christensen, "On the Influence of Inharmonicities in Model-Based Speech Enhancement," in *European Signal Processing Conference*, Marrakesh, Sept. 10-13 2013.

[13] T. Nilsson, S. I. Adalbjörnsson, N. R. Butt, and A. Jakobsson, "Multi-Pitch Estimation of Inharmonic Signals," in *European Signal Processing Conference*, Marrakech, Sept. 9-13, 2013.

[14] M. G. Christensen, P. Vera-Candeas, S. D. Somasundaram, and A. Jakobsson, "Robust Subspace-based Fundamental Frequency Estimation," in *33rd IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Las Vegas, March 30-April 4, 2008.

[15] C. Kim, W. Chang, S-H. Oh, and S-Y. Lee, "Joint Estimation of Multiple Notes and Inharmoncity Coefficient Based on f0-Triplet for Automatic Piano Transcription," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1536–1540, December 2014.

[16] F. Carli, L. Ning, and T. Georgiou, "Convex Clustering via Optimal Mass Transport," *arXiv preprint arXiv:1307.5459*, 2013.

[17] X. Jiang, Z-Q. Luo, and T. T. Georgiou, "Geometric methods for spectral analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1064–1074, 2012.

[18] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *International Journal of Computer Vision*, vol. 60, no. 3, pp. 225–240, 2004.

[19] J. Karlsson and L. Ning, "On robustness of ℓ_1-regularization methods for spectral estimation," in *IEEE 53nd Annual Conference on Decision and Control*, Dec 2014.

[20] C. Villani, *Topics in Optimal Transportation*, vol. 58, Graduate studies in Mathematics, AMS, 2003.

[21] A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 804–816, 2003.

[22] V. Emiya, R. Badeau, and B. David, "Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 6, pp. 1643–1654, Aug. 2010.

[23] R. E. Bixby, J. W. Gregory, R. E. Marsten, and D. F. Shanno, "Very Large-Scale Linear Programming: A Case Study in Combining Interior Point and Simplex Methods," *Operations Research*, vol. 40, no. 5, 885-897 1992.

[24] R. O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*, Ph.D. thesis, Stanford University, Stanford, C.A., 1981.

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

[26] 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.

[27] M. Bay, A. F. Ehmann, and J. S. Downie, "Evaluation of Multiple-F0 Estimation and Tracking Systems," in *International Society for Music Information Retrieval Conference*, Kobe, Japan, October 2009.