FAST PARTIAL TRACKING OF AUDIO WITH REAL-TIME CAPABILITY THROUGH LINEAR PROGRAMMING

Julian Neri

SPCL* , CIRMMT†
McGill University, Montréal, Canada
julian.neri@mail.mcgill.ca

Philippe Depalle

SPCL* , CIRMMT†
McGill University, Montréal, Canada
philippe.depalle@mcgill.ca

ABSTRACT

This paper proposes a new partial tracking method, based on linear programming, that can run in real-time, is simple to implement, and performs well in difficult tracking situations by considering spurious peaks, crossing partials, and a non-stationary short-term sinusoidal model. Complex constant parameters of a generalized short-term signal model are explicitly estimated to inform peak matching decisions. Peak matching is formulated as a variation of the linear assignment problem. Combinatorially optimal peak-to-peak assignments are found in polynomial time using the Hungarian algorithm. Results show that the proposed method creates high-quality representations of monophonic and polyphonic sounds.

1. INTRODUCTION

The sinusoidal model proves beneficial for its capacity to represent non-stationary sounds. The sinusoidal model represents a sound signal as a sum of P time-varying sinusoids, called *partials*, with instantaneous log-amplitude $a_p(t)$, phase $\phi_p(t)$, and frequency $f_p(t)$,

$$s(t) = \sum_{p=1}^{P} \exp\big(a_p(t) + i\phi_p(t)\big) \tag{1}$$

$$\phi_p(t) = \phi_p(0) + 2\pi \int_0^t f_p(u)du \tag{2}$$

Decomposing a sound signal into a set of partials, or *partial tracking*, is useful for a variety of applications, including sound synthesis [1], sound source separation [2] [3], audio coding [4], audio effects [5] [6], and automatic music transcription [7] [8].

Partial tracking consists of two operations that are performed either sequentially or jointly. First, instantaneous sinusoidal model parameters are estimated from a short-term analysis of the sound signal. Second, the instantaneous parameters are linked according to their expected temporal progressions, forming partial trajectories. The parameter estimates are interpolated between each short-term analysis frame so that $a_p(t)$ and $\phi_p(t)$ can be evaluated at the sampling rate.

Despite practical applications of partial tracking and its wide use in the field, aspects of the process complicate the potential for a flawless outcome. A complex sound often has hundreds of partials, plus a stochastic component, sculpting its time-varying spectral envelope [9]. Sinusoidal model parameters must be estimated accurately from short-term estimates to ensure appropriate tracking decisions. Polyphonic sounds further complicate the analysis because the frequency trajectories of two partials might cross [10]. Peak matching poses a large combinatorial problem that must be repeated for many, typically thousands, of time frames. Thus, there are not only difficulties associated with the quality of tracking, but also with speed and tractability [11]. Many partial tracking methods have been proposed over the last several decades, as summarized in Section 1.1.

This paper presents a new partial tracking method, based on linear programming, that improves the state of the art of sinusoidal modeling. The proposed method can operate in real-time, is simpler to implement than the McAulay and Quatieri (MCQ) method [12], and creates sinusoidal model representations comparable to the leading hidden Markov model (HMM)-based methods [10] [11]. For parameter estimation, the method considers a generalized non-stationary short-term sinusoidal model. The peak matching procedure is formulated as a variation of the linear assignment problem [13], a fundamental combinatorial optimization problem, allowing for an optimal peak-to-peak assignment solution in polynomial time.

This paper is organized as follows. Section 2 overviews the assignment problem. Section 3 establishes the new method of partial tracking, first by describing short-term analysis additive model parameter estimation, then by deriving the assignment problem costs. Section 4 details the results from experiments that demonstrate the ability of the new partial tracker. Section 5 concludes the paper and proposes future research on the applications of the assignment problem in audio.

1.1. Overview of Previous Work

McAulay and Quatieri (MCQ) [12] developed the first partial tracking algorithm for sinusoidal modeling of speech. The MCQ method connects peaks that have minimum frequency difference between consecutive analysis frames. The MCQ method uses a non-optimal greedy algorithm, does not consider spurious peaks, and assumes a stationary short-term signal model. Modifications of the MCQ method include using a reassigned bandwidth enhanced model [14] and considering an intermediate "sleep" state for every trajectory [15]. A linear prediction coding-based method was proposed in [16] [17] that determines the most probable match using the trajectory's previous samples and can interpolate missing data. A non-causal strategy was proposed in [18] that builds each trajectory starting from a reliable two-point connection then growing it in every direction by appending smaller pieces to it. The adaptive method from [19] uses B-splines to estimate the parameters of the additive model. Adaptive oscillators were used to track partials in [20], and a Kalman filtering approach was described in [21]. The

* Sound Processing and Control Laboratory
† Centre for Interdisciplinary Research in Music Media and Technology

hidden Markov model (HMM) partial tracker [10] optimizes the combination of trajectories within an analysis window, considers spurious peaks, and performs well in several difficult tracking situations. The HMM tracker was improved in [11] for non-stationary and noisy signals by formulating a new peak matching criterion that incorporates explicitly measured frequency slope information.

The assignment problem [22], presented in Section 2, is a fundamental combinatorial optimization problem that describes a variety of real-world problems. Variations of the assignment problem, especially the multidimensional assignment problem [23], have been used to describe the problem of multi-target tracking [24], jointly estimating the number of targets and their trajectories from sensor measurements. Although the assignment problem has been successfully applied to such problems for over a half century, to the extent of our knowledge, it has not been applied to tracking problems in audio.

2. THE ASSIGNMENT PROBLEM

2.1. Problem statement

The assignment problem is a fundamental combinatorial optimization problem in the field of operations research [13].

The problem involves assigning R members of one set, *agents*, to another, *tasks*. Any agent can perform any task. An agent-task assignment incurs a *cost* that may vary depending on the assignment. The goal is to assign an agent to perform one task, and assign a task to one agent, such that the sum of individual costs is minimized.

The assignment problem is formally expressed as a linear programming problem with the following mathematical model:

$$\text{minimize} \quad \sum_{i=1}^{R}\sum_{j=1}^{R} C_{ij}X_{ij} \tag{3a}$$

$$\text{subject to} \quad \sum_{i=1}^{R} X_{ij} = 1 \qquad j = 1,\dots,R \tag{3b}$$

$$\sum_{j=1}^{R} X_{ij} = 1 \qquad i = 1,\dots,R \tag{3c}$$

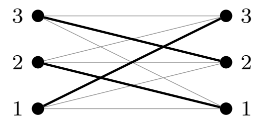
where C_{ij} is the cost of assigning agent i to task j, and X_{ij} is a binary variable that equals 1 if agent i is assigned to task j and 0 otherwise. The first constraint (3b) ensures that every agent is assigned to one task, while the second constraint (3c) ensures that every task is assigned to one agent.

In terms of graph theory, this is equivalent to finding the minimum cost assignment in a weighted bipartite graph [22]. Figure 1 represents the assignment of agents to tasks as a graph and as an annotated cost matrix.

Linear programming problems can be solved by the simplex algorithm [25], however, more efficient algorithms have been developed that take advantage of the assignment problem's specific structure.

2.2. Hungarian algorithm

The Hungarian algorithm is a combinatorial algorithm that can solve the assignment problem in polynomial time [26]. The algorithm takes as an input the cost matrix **C** and outputs the optimal assignments matrix **X**. If the number of agents does not equal the



$$\begin{pmatrix} C_{11} & C_{12} & \boldsymbol{C_{13}} \\ \boldsymbol{C_{21}} & C_{22} & C_{23} \\ C_{31} & \boldsymbol{C_{32}} & C_{33} \end{pmatrix}$$

Figure 1: Assignments represented as bold lines in a bipartite graph (left) and as bold elements in a cost matrix (right).

number of tasks, dummy variables are appended to the cost matrix to make it square. The Hungarian algorithm consists of the following three steps.

1. For each row, subtract the row's minimum value from every value in that row. For each column, subtract the column's minimum value from every value in that column.
2. Cover the zeros in the resulting matrix with the minimum number of vertical and horizontal lines. If R lines are required, an optimal assignment of zeros exists and the algorithm stops. If less than R lines are required, proceed to Step 3.
3. Find the minimum value in the matrix that is not covered by the lines from Step 2. Subtract the value from every uncovered element and add the value to every covered element. Return to Step 2.

This popular algorithm's implementation is freely available online (commonly as a single function) in several software languages [27].

2.3. Variations of the Assignment Problem

Variations of the assignment problem use different objectives, constraints, or dimensions. A survey of such variations is in [13]. For example, a *one-to-many* assignment problem has a looser constraint that allows an agent to perform more than one task. A variation that is particularly applicable to multi-target tracking is the *multidimensional* assignment problem.

2.4. Multidimensional Assignment Problem

Multidimensional assignment problems consist of assigning the members of three or more sets [28]. A type of multidimensional problem that has been applied to multi-target tracking is the axial three-dimensional assignment problem. This type of problem involves assigning members over three sets, where each assignment incurs a cost C_{hij}, such that the total cost is minimized.

Multidimensional assignment problems are NP-hard [23]. The simplest way to solve a multidimensional assignment problem is to enumerate every possible combination of assignments then choose the one with the lowest cost [28], however, this solves the problem in factorial time. Research has led to algorithms that either solve or approximately solve the problem with improved tractability. For example, [29] details a branch and bound algorithm that approximately solves the axial three-dimensional case. Alternatively, [23] shows that an axial three-dimensional problem can be solved in polynomial time if the cost C_{hij} can be split into the sum of two sub-costs, $C_{hij} = C_{hi} + C_{ij}$.

3. PROPOSED METHOD

3.1. Overview

The proposed partial tracking method sequentially performs two processes. First, short-term sinusoidal model parameters are estimated for each peak j in frame k. Second, the short-term parameter estimates are connected over consecutive analysis frames, $k-1$ and k, by solving an assignment problem, forming trajectories.

This paper considers a trajectory to be a time-sequence of spectral peaks with short-term sinusoidal model parameters, defined in Section 3.2, that satisfy continuity constraints at the midpoint of consecutive analysis frames. Accordingly, *useful* assignments satisfy those continuity constraints while *spurious* assignments do not. Section 3.3 defines a cost for both assignment types. The assignment type with the lowest cost is the most probable. The optimal combination of assignments is found using the Hungarian algorithm.

3.2. Short-Term Additive Parameter Estimation

Parameters are estimated over short-term analysis frame k at time $t_k = kH/f_s$, where H is the hop size and f_s is the sampling frequency. The frame's time index n ranges from $-N/2$ to $N/2$, where $N+1$ is the frame's duration. The center of the frame is at $n = 0$ and aligned with t_k.

The short-term signal model over frame k is a sum of R_k generalized sinusoids

$$s(n) = \sum_{j}^{R_k} \exp\left(\sum_{i=0}^{Q} \alpha_{ij} n^i\right) \tag{4}$$

where α_{ij} are the complex constants of sinusoid j and Q is the order of the polynomial [30]. The instantaneous log-amplitude and phase of sinusoid j are

$$a_j^k(n) = \Re\left(\sum_{i=0}^{Q} \alpha_{ij} n^i\right) \tag{5}$$

$$\phi_j^k(n) = \Im\left(\sum_{i=0}^{Q} \alpha_{ij} n^i\right) \tag{6}$$

Since the sinusoid's normalized angular frequency is the time derivative of the phase,

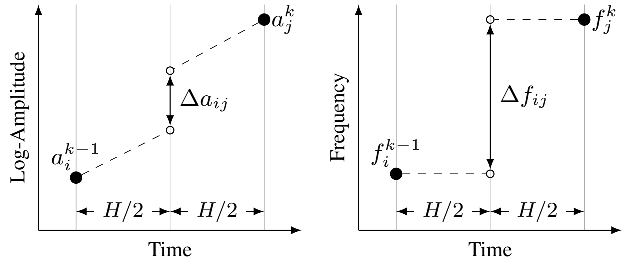
$$f_j^k(n) = \frac{f_s}{2\pi}\Im\left(\sum_{i=0}^{Q} \alpha_{ij} i n^{i-1}\right) \tag{7}$$

There are several options for estimating α_{ij}. A comparison of sinusoidal model parameter estimators is in [31]. Using the distribution derivative method (DDM) [30] allows for the estimation of α_{ij} up to an arbitrary polynomial order Q.

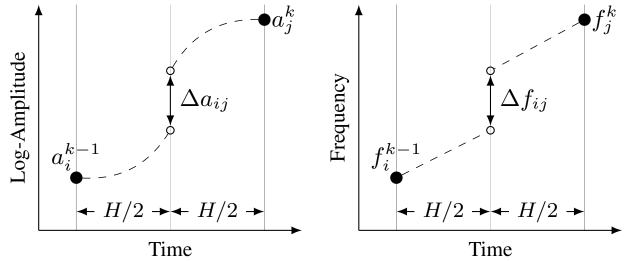
3.3. Costs of Useful and Spurious Assignments

An assignment cost is quantified by a multivariate Gaussian, similarly to the "matching criterion" defined in [10]. The cost of assigning peak i in frame $k-1$ to peak j in frame k is

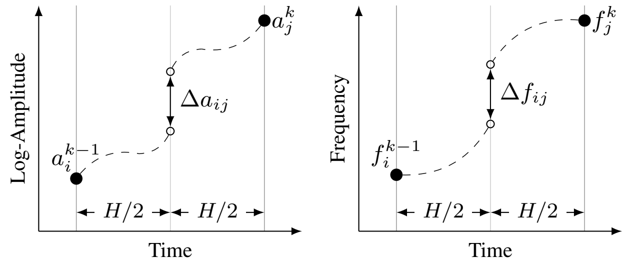
$$A_{ij} = 1 - \exp\left(-\frac{\Delta f_{ij}^2}{2\sigma_f^2} - \frac{\Delta a_{ij}^2}{2\sigma_a^2}\right) \tag{8}$$



(a) $Q = 1$.



(b) $Q = 2$.



(c) $Q = 3$.

Figure 2: Illustration of equation (10) (left) and (11) (right) for different settings of polynomial order Q.

for useful assignments and

$$B_{ij} = 1 - (1-\delta)A_{ij} \tag{9}$$

for spurious assignments, where

$$\Delta a_{ij} = a_i^{k-1}(H/2) - a_j^k(-H/2) \tag{10}$$

$$\Delta f_{ij} = f_i^{k-1}(H/2) - f_j^k(-H/2) \tag{11}$$

Figure 2 illustrates that equations (10) and (11) evaluate the midpoint continuity over peak i and j by extending their short-term sinusoidal model amplitude and frequency trajectories.

Standard deviations σ_f and σ_a are defined by the formulas

$$\sigma_f^2 = \zeta_f^2 / \big(2\ln(\delta-2) - 2\ln(\delta-1)\big) \tag{12}$$

$$\sigma_a^2 = \zeta_a^2 / \big(2\ln(\delta-2) - 2\ln(\delta-1)\big) \tag{13}$$

The parameter δ changes the relative preference towards spurious or useful assignments: smaller values promote spurious ones and larger values promote useful ones. Parameters ζ_f and ζ_a are values of Δf and Δa, respectively, that mark the point of transition between a useful or spurious assignment. Figure 3 shows how the cost functions change with respect to these parameters.

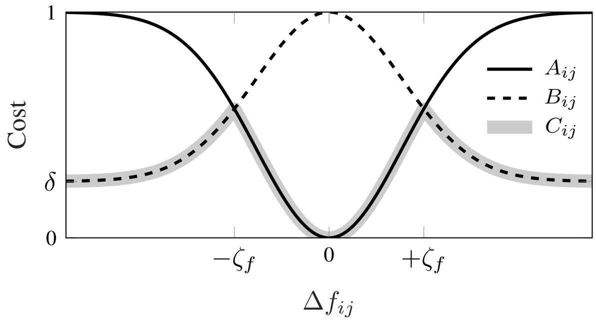


Figure 3: Illustration of how the parameters ζ_f and δ change the useful cost A_{ij}, spurious cost B_{ij}, and cost C_{ij} defined in equation (14).

3.4. Cost Matrix

This is a *multi-criteria* assignment problem [13] because it consists of minimizing an objective function that has two decision criteria. There is not only a cost A_{ij} of connecting two peaks as a useful trajectory, but also a cost B_{ij} of not connecting them. We can recognize this decision model's multiple criteria simply by constructing a single cost matrix whose values are

$$C_{ij} = \min\{A_{ij}, B_{ij}\} \tag{14}$$

3.5. Solving the Assignment Problem

The optimal assignments matrix **X** is retrieved by inputting the cost matrix **C** into the Hungarian algorithm. Following equation (14), an assignment $X_{ij} = 1$ is useful if A_{ij} is less than B_{ij}.

3.6. Partial Labeling

A trajectory is an unbroken (continuous) path from a peak in some frame to a peak in a future frame. Therefore, a useful assignment that continues an existing trajectory from the previous observation is labeled with that trajectory's index. On the other hand, if a useful assignment does not continue a path but rather starts one, it is labeled with a new index. Figure 4 illustrates the labeling of useful assignments over time.

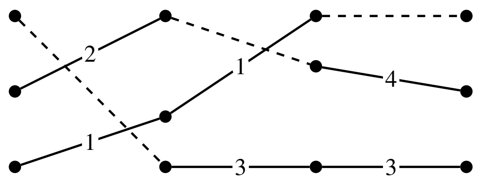


Figure 4: Illustration of labeling useful assignments (solid lines) over a sequence of frames. Dashed lines show spurious assignments.

3.7. Computation Cost & Implementation

True real-time operation is possible not only because the peak matching method has a low computational cost, but also because peak assignment and labeling only depends on the current frame k and previous frame $k-1$. Solving the assignment problem is a polynomial time operation, $O(R^3)$, where R is the largest of the two number of peaks R_{k-1} and R_k. The proposed method can run in real-time in many practical situations, depending on R, the hop size H, and the speed of the parameter estimator.

Implementing this partial tracker is simpler than other ones, including the MCQ method. Peak matching only consists of defining a cost matrix with equation (14) and running the Hungarian algorithm.

3.8. Recasting Previous Partial Tracking Methods

The McAulay and Quatieri (MCQ) method matches peaks over consecutive frames based on a minimum frequency difference criterion. In terms of an assignment problem, the cost is simply

$$C_{ij} = |f_i - f_j| \tag{15}$$

Rather than use the MCQ method's non-optimal greedy algorithm, optimal assignments can be found using the Hungarian algorithm. This approach avoids all the heuristics associated with the MCQ method. The MCQ method recast as an assignment problem is a simple case of the proposed method with $Q = 1$ that does not consider amplitude information or spurious assignments.

The hidden Markov model (HMM)-based method considers the peak connections over two frames as a hidden state. State transition probabilities are the product of matching criteria. Each matching criterion θ_{hij} quantifies how well peaks h, i, and j, over frames $k-2$, $k-1$, and k (two states), satisfy parameter slope continuity constraints,

$$\theta_{hij} = \exp\left(-\frac{(\Delta f_{hi} - \Delta f_{ij})^2}{\sigma_f^2} - \frac{(\Delta a_{hi} - \Delta a_{ij})^2}{\sigma_a^2}\right) \tag{16}$$

where $\Delta f_{ij} = f_i - f_j$ and $\Delta a_{ij} = a_i - a_j$.

The HMM-based method can be recast as a multidimensional assignment problem. The peak connections that admit the largest product of matching criteria (state transition probability) are the same ones that admit the smallest sum of assignment costs.

More specifically, the recast HMM method involves a three-dimensional assignment problem because the cost depends on peak parameters (members) over three frames (sets). Recall from Section 2.4 that such a problem is NP-hard. Making the HMM method tractable involves constraining the number of potential states.

Alternatively, if the matching criterion can be expressed as a product of two sub-criteria, $\theta_{hij} = \theta_{hi}\theta_{ij}$, then a polynomial time solution is possible through an assignment problem with cost $C_{hij} = C_{hi} + C_{ij}$. In [11] frequency slope ψ is explicitly estimated and the matching criterion is

$$\theta_{hij} = \exp\left(-\frac{\Delta f_{hi}^2}{\sigma_f^2} - \frac{\Delta f_{ij}^2}{\sigma_f^2} - \frac{(\Delta a_{hi} - \Delta a_{ij})^2}{\sigma_a^2}\right) \tag{17}$$

where $\Delta f_{ij} = (f_i + \psi_i H/2f_s) - (f_j - \psi_j H/2f_s)$. While the frequency slope calculation depends on only two frames, the calculation of amplitude slope depends on parameters over three frames, so the problem is still NP-hard.

The cost function developed in Section 3.3 is expressed in terms of parameters across only two frames, k and $k-1$, by explicitly estimating both the amplitude and the frequency slope, enabling a linear assignment problem formulation and polynomial time solution.

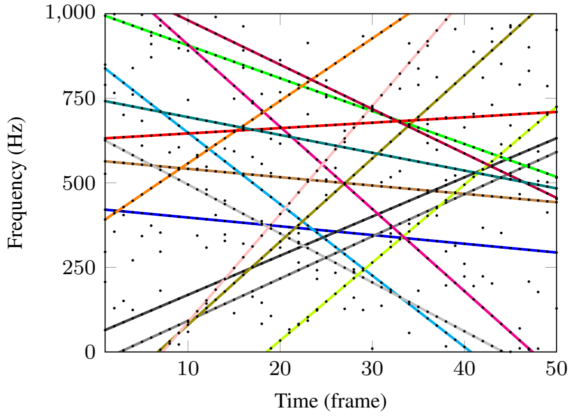


Figure 5: Detected partials (lines) from simulated data (dots) that resemble overlapping chirp sinusoids plus noise.

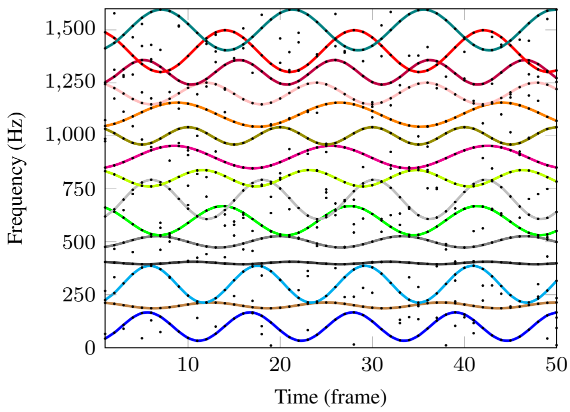
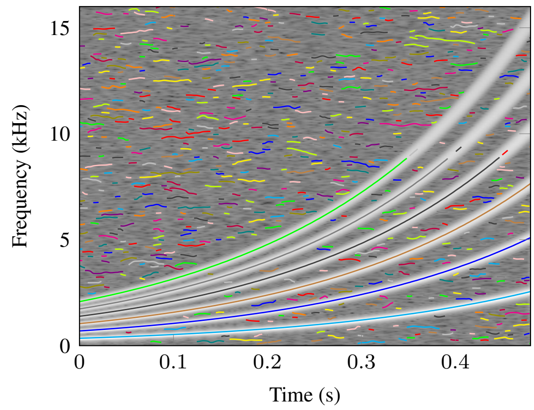
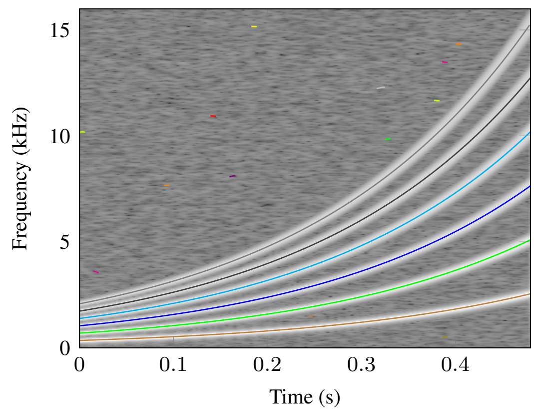


Figure 6: Detected partials (lines) from simulated data (dots) that resemble overlapping frequency modulated sinusoids plus noise.



(a) $Q = 1$. 12 dB R-SNR.



(b) $Q = 2$. 52 dB R-SNR.

Figure 7: Detected partials from a synthesized audio signal consisting of harmonically-related logarithmic chirp sinusoids plus noise for different settings of polynomial order Q.

4. RESULTS

4.1. Simulated Data

In these examples, peak parameters are simulated (set "by hand"). Circumventing the short-term analysis highlights the ability of the proposed peak matching method. Useful peaks are simulated by sampling parameter values from analytic expressions of partial trajectories, while spurious peaks are simulated by setting parameter values randomly. Each peak's amplitude-related values are set to zero, which further complicates tracking. Figures 5 and 6 show that the tracker perfectly classifies useful and spurious peaks and resolves overlapping partials that have similar frequency slopes.

4.2. Audio Signals

In the following examples, peak parameters are estimated from a short-term analysis of an audio signal, $s(n)$, using the distribution derivative method (DDM) [30]. The first group of examples involve audio signals that are synthesized from partial trajectories with constant amplitude and corrupted with -40 dB of white noise. The second group of examples involve real speech and musical audio signals. The signal is reconstructed as $\hat{s}(n)$ using the synthesis method described in [12]. The reconstruction signal-to-noise ratio (R-SNR) is used to help quantify the results, given by

$$\text{R-SNR} = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1} s(n)^2}{\sum_{n=0}^{N-1}\left(s(n)-\hat{s}(n)\right)^2}\right) \tag{18}$$

Figure 7 shows the results of tracking a synthetic harmonic signal whose fundamental frequency quickly increases on a logarithmic scale. If frequency slope is not estimated, as shown in Figure 7a, then ζ_f must be large enough to ensure useful peaks are connected over large frequency differences. This results in many false detections of useful assignments from spurious data. Figure 7b shows how the results improve dramatically when frequency slope is estimated.

Figure 8 shows the results of tracking a harmonic signal with strong vibrato (± 5 semitones). A further challenge is posed at 0.5 seconds where the fundamental frequency smoothly steps up by 5 semitones, resulting in close partials with steep slopes.

Figure 9 shows the results of tracking synthetic polyphonic audio that resembles a violin sound with unnaturally strong and fast vibrato superimposed with a trombone sound performing a fast upward glissando.

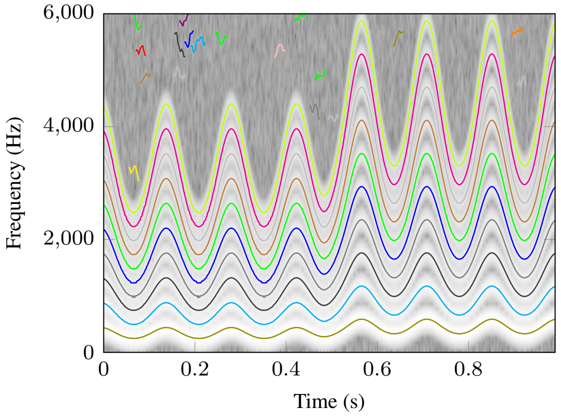


Figure 8: Detected partials from synthesized harmonic audio with vibrato plus noise. The fundamental frequency smoothly steps up by 5 semitones, from 330 Hz to 440 Hz. 26 dB R-SNR.

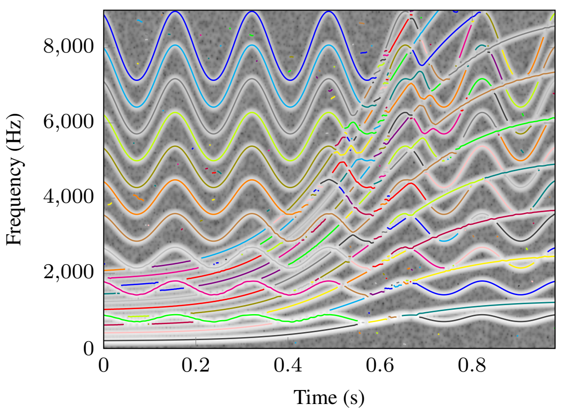
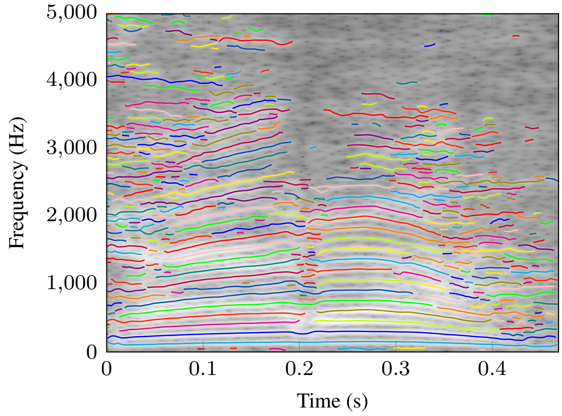


Figure 9: Detected partials from synthesized polyphonic audio plus noise. 19 dB R-SNR.

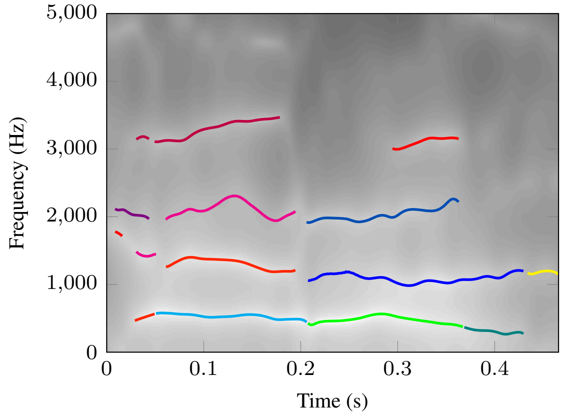
Formant tracking is another time-frequency tracking process that is especially suitable for vocal sounds. The proposed method can be used without modification for formant tracking applications. Figure 10a shows the partials detected from a real male voice sound while Figure 10b shows the results of tracking the formants of the same sound. Linear predictive coding (LPC) with 24 coefficients was used instead of DDM to estimate each formant's short-term amplitude and frequency, corresponding to an order $Q = 1$ polynomial.

Finally, the results of tracking a tango excerpt by Piazolla are shown in Figure 11. This multi-instrumental composition admitted dense short-term spectra with several frames having greater than 150 peaks. For this 14-second long signal over 12,000 partials were detected in a total computational time of 13 seconds on a 2.8 GHz quad-core processor: parameter estimation took 9 seconds and tracking took 4 seconds. The reconstructed sound is perceptually close to the original with a 15 dB R-SNR.

All test signals and reconstructed sounds are available for listening at http://www.music.mcgill.ca/~julian/dafx18.



(a) Detected partials. 22 dB R-SNR.



(b) Detected formants.

Figure 10: Male voice signal tracking results (/kara/).

5. CONCLUSIONS AND FUTURE WORK

This paper developed a new partial tracking method that matches sinusoidal model parameters over consecutive analysis frames by solving a linear assignment problem with the Hungarian algorithm. Results show that the proposed method easily handles exceptionally difficult partial tracking scenarios, involving strongly modulated partials embedded in noise and crossing partials that are common in polyphonic recordings. Moreover, the proposed tracker can operate in real-time and is simple to implement. Other popular methods were recast under the assignment problem framework, revealing them as specific cases of the proposed method. Future work may examine the results of tracking without slope information by solving a multidimensional assignment problem. More generally, other audio applications may be advantageously described as assignment problems.

6. REFERENCES

[1] M. Caetano, G. Kafentzis, A. Mouchtaris, and Y. Stylianou, "Full-band quasi-harmonic analysis and synthesis of musi-

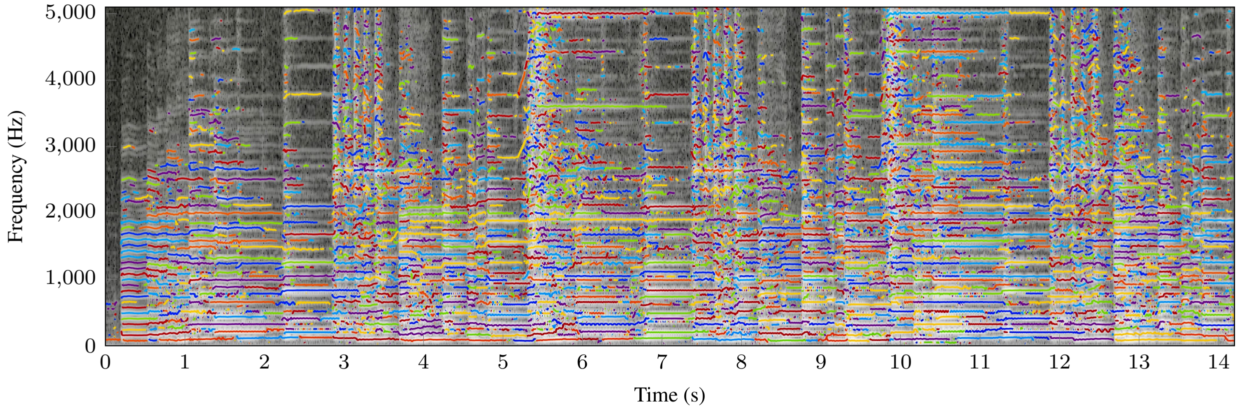


Figure 11: 12,000 detected partials from a polyphonic song sampled at 44.1 kHz. 15 dB R-SNR.

cal instrument sounds with adaptive sinusoids," *Applied Sciences*, vol. 6, no. 5, May 2016.

[2] T. Virtanen and A. Klapuri, "Separation of harmonic sound sources using sinusoidal modeling," in *IEEE Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Istanbul, Turkey, Jun. 2000, vol. 2, pp. 765–768.

[3] E. Creager, N. D. Stein, R. Badeau, and P. Depalle, "Non-negative tensor factorization with frequency modulation cues for blind audio source separation," in *Proc. of the 17th ISMIR Conf.*, New York City, USA, Aug. 2016.

[4] O. Derrien, R. Badeau, and Gaël Richard, "Parametric audio coding with exponentially damped sinusoids," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 7, pp. 1489–1501, Jul. 2012.

[5] M. Raspaud, S. Marchand, and L. Girin, "A generalized polynomial and sinusoidal model for partial tracking and time stretching," in *Proc. of the 8th Int. Conf. on Digital Audio Effects (DAFx-05)*, Madrid, Spain, Sep. 2005.

[6] S. Kazazis, P. Depalle, and S. McAdams, "Sound morphing by audio descriptors and parameter interpolation," in *Proc. of the 19th Int. Conf. on Digital Audio Effects (DAFx-16)*, Brno, Czech Republic, Sep. 2016.

[7] M. Christensen and A. Jakobsson, "Multi-pitch estimation," *Synthesis Lectures on Speech and Audio Processing*, vol. 5, no. 1, Morgan and Claypool, pp. 1–160, 2009.

[8] J. Burred, A. Röbel, and T. Sikora, "Polyphonic musical instrument recognition based on a dynamic model of the spectral envelope," in *IEEE Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 173–176.

[9] N. Fletcher and T. Rossing, *The Physics of Musical Instruments*, Springer New York, 2nd edition, 1998.

[10] P. Depalle, G. Garcia, and X. Rodet, "Tracking of partials for additive sound synthesis using hidden Markov models," in *IEEE Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr. 1993, vol. 1, pp. 225–228.

[11] C. Kereliuk and P. Depalle, "Improved hidden Markov model partial tracking through time-frequency analysis," in *Proc. of the 11th Int. Conf. on Digital Audio Effects (DAFx-08)*, Espoo, Finland, Sep. 2008.

[12] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 34, no. 4, pp. 744–754, 1986.

[13] D. Pentico, "Assignment problems: A golden anniversary survey," *European Journal of Operational Research*, vol. 176, no. 2, pp. 774–793, Jan. 2007.

[14] K. Fitz and L. Haken, "Bandwidth enhanced sinusoidal modeling in Lemur," in *Proc. Int. Computer Music Conf. (ICMC)*, Banff, Canada, 1995, pp. 154–157.

[15] X. Serra and J. O. Smith, "Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.

[16] M. Lagrange, S. Marchand, and J. Rault, "Tracking partials for the sinusoidal modeling of polyphonic sounds," in *IEEE Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Philadelphia, USA, 2005, vol. 3, pp. 229–232.

[17] M. Lagrange, S. Marchand, and J. Rault, "Enhancing the tracking of partials for the sinusoidal modeling of polyphonic sounds," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 15, no. 5, pp. 1625–1634, Jul. 2007.

[18] M. Bartkowiak and T. Żernicki, "A non-time-progressive partial tracking algorithm for sinusoidal modeling," in *AES 131st International Conference*, New York, USA, Oct. 2011.

[19] A. Röbel, "Adaptive additive modeling with continuous parameter trajectories," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1440–1453, Jul. 2006.

[20] M. Marolt, "A connectionist approach to automatic transcription of polyphonic piano music," *IEEE Transactions on Multimedia*, vol. 6, no. 3, pp. 439–449, Jun. 2004.

[21] H. Satar-Boroujeni and B. Shafai, "A robust algorithm for partial tracking of music signals," in *Proc. of the 8th Int. Conf. on Digital Audio Effects (DAFx-05)*, Madrid, Spain, Sep. 2005, pp. 202–207.

[22] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems, Revised Reprint*, Society for Industrial and Applied Mathematics, 2012.

[23] K. Gilbert and R. Hofstra, "Multidimensional assignment problems," *Decision Sciences*, vol. 19, no. 2, pp. 306–321, Jun. 1988.

[24] A. Poore and S. Gadaleta, "Some assignment problems arising from multiple target tracking," *Mathematical and Computer Modelling*, vol. 43, pp. 1074–1091, May 2006.

[25] D. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, Springer, 4th edition, 2016.

[26] H. Kuhn, "The Hungarian method for the assignment problem," in *Naval Research Logistic Quarterly*, Mar. 1955, vol. 2, pp. 83–97.

[27] Y. Cao, "Hungarian Algorithm for Linear Assignment Problems (V2.3)," MATLAB Central File Exchange, 2011.

[28] W.P. Pierskalla, "The multidimensional assignment problem," *Operations Research*, vol. 16, no. 2, pp. 422–431, Apr. 1968.

[29] E. Balas and M. Saltzman, "An algorithm for the three-index assignment problem," *Operations Research*, vol. 39, no. 1, pp. 150–161, Feb. 1991.

[30] M. Betser, "Sinusoidal polynomial parameter estimation using the distribution derivative," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4633–4645, Dec. 2009.

[31] B. Hamilton and P. Depalle, "Comparisons of parameter estimation methods for an exponential polynomial sound signal model," in *AES 45th International Conference*, Helsinki, Finland, Mar. 2012.