A KALMAN-BASED FUNDAMENTAL FREQUENCY ESTIMATION ALGORITHM

Liming Shi,[1*] *Jesper K. Nielsen,*[1] *Jesper R. Jensen,*[1*] *Max A. Little,*[2,3] *Mads G. Christensen*[1*]

[1] Audio Analysis Lab, AD:MT, Aalborg University, {ls, jkn, jrj, mgc}@create.aau.dk
[2] Engineering and Applied Science, Aston University, max.little@aston.ac.uk
[3] Media Lab, Massachusetts Institute of Technology

ABSTRACT

Fundamental frequency estimation is an important task in speech and audio analysis. Harmonic model-based methods typically have superior estimation accuracy. However, such methods usually assume that the fundamental frequency and amplitudes are stationary over a short time frame. In this paper, we propose a Kalman filter-based fundamental frequency estimation algorithm using the harmonic model, where the fundamental frequency and amplitudes can be truly nonstationary by modeling their time variations as first-order Markov chains. The Kalman observation equation is derived from the harmonic model and formulated as a compact nonlinear matrix form, which is further used to derive an extended Kalman filter. Detailed and continuous fundamental frequency and amplitude estimates for speech, the sustained vowel /a/ and solo musical tones with vibrato are demonstrated.

Index Terms— Fundamental frequency estimation, extended Kalman filter, harmonic model

1. INTRODUCTION

The fundamental frequency can be described as the lowest rate for a periodic signal to repeat itself. Fundamental frequency information for voiced speech or audio signals has various applications, such as speech enhancement [1], voice disorder detection [2], automatic speech recognition [3] and music processing [4]. A very large number of fundamental frequency estimation algorithms have been proposed in the past, including those that could be broadly described as *non-parametric* and *parametric* methods. Here we will define non-parametric methods as those which are based on the autocorrelation function obtained within a specified time frame; examples include Yin [5] and RAPT [6]. These methods are computationally simple but they are prone to observation noise and *subharmonic error* (that is, misidentifying multiples of the actual fundamental frequency, a.k.a. octave error). To reduce this subharmonic error problem, a recently devised method – the sawtooth waveform-inspired pitch estimator (SWIPE) [7], and variants – use the cross-correlation function against a sawtooth signal combined with frequency-domain information. By contrast, examples of parametric methods are *harmonic models* which use *nonlinear least squares* (NLS) model parameter estimation [8]. Under appropriate assumptions, such NLS estimators are optimal from a statistical perspective but are very computationally costly to run in practice. To lower this computational cost, recently a fast NLS has been proposed which exploits the matrix structure using a recursive matrix solver [9]. Most parametric harmonic models, as with non-parametric methods, assume signal stationarity at least over short time frames, but in practice this assumption is unrealistic. To account for the non-stationarity of voiced speech signals, a *harmonic chirp* model for voiced speech has been proposed, and the fundamental frequency and chirp rate parameters are obtained iteratively [10]. Another parametric model, the *adaptive quasi-harmonic* model [11] has been proposed to attempt to capture time variation in both frequency and amplitude of voiced speech signals. All of the above methods are on a segment-by-segment basis. Recently, *instantaneous* fundamental frequency estimation algorithms based on the harmonic model which use *nonlinear recursive filters* have been proposed [12]. As the model parameter update results in a nonlinear state equation, classical extended (EKF), unscented and particle Kalman filters have been proposed to perform the parameter estimation in this time-varying model. Continuous variations in fundamental frequency are obtained. However, the size of the state space in [12] is $3K+1$, where K is the harmonic order, leading to high computational effort.

In this paper, we propose to use the harmonic model to fit voiced speech and music signals. A first order Markov chain is used to capture non-stationarity in fundamental frequency and amplitude. By exploiting linear relationships between the phases of different harmonics, the size of the state space is decreased to $K+2$ compared to previous Kalman filtering approach. The resulting nonlinear observation equation is formulated in compact matrix form, and finally an extended Kalman smoother is applied to track instantaneous fundamental frequency and amplitudes.

2. HARMONIC MODEL ESTIMATION

Consider the following general signal observation model

$$y_n = s_n + v_n, \qquad (1)$$

where y_n is the observation signal and v_n denotes zero mean Gaussian noise with variance r_v, and n is the integer time index. We assume that the voiced speech or audio signal s_n is produced by a time-varying harmonic model, i.e.,

$$s_n = \underset{k=1}{\overset{K}{}} A_{n,k}\cos(\theta_{n,k}), \qquad (2)$$

$$\theta_{n,k} = k\omega_n n + \theta_{0,k}, \quad k = 1, \cdots, K, \qquad (3)$$

where $A_{n,k}$ is the instantaneous amplitude of the k^{th} harmonic at time instant n, $\theta_{n,k}$ is the instantaneous phase, $\omega_n = 2\pi f_n/F_s$ is the instantaneous normalized digital radian frequency, $F_s = 1/T_s$ is the sampling rate, T_s is the sampling period, and $\theta_{0,k}$ is the initial phase, and K is the number of harmonics. Our objective is to estimate the fundamental frequency ω_n and amplitudes $A_{n,k}$, $1 \le k \le K$, simultaneously.

[*] This work was funded by the Danish Council for Independent Research, grant ID: DFF 4184-00056.

978-1-5386-1631-4/17/$31.00 ©2017 IEEE

Assume that the fundamental frequency and amplitudes are time-invariant in a short time frame with a length N, and thus the time index n can be ignored, i.e. $\omega_n = \omega_0$ and $A_{n,k} = A_k$, $1 \leq k \leq K$. Combining (1), (2) and (3), and using Euler's formula, we obtain

$$y_n = \mathop{}_{k=1}^{K} \quad a_k z_n^k + a_k^* z_n^{-k} \quad + v_n, \tag{4}$$

where the superscript $*$ denotes complex conjugation, the complex amplitude a_k is defined as $a_k = \frac{A_k}{2} e^{j\theta_{0,k}}$, and $z_n = e^{j\omega_0 n}$. Collecting N observation signals into a vector and writing (4) in matrix form yields

$$\mathbf{y} = \mathbf{Za} + \mathbf{v}_n, \tag{5}$$

where $\mathbf{y}_n = [y_1, y_2, \cdots, y_N]^T$ and $\mathbf{v}_n$ is defined in the same form, $\mathbf{a} = [a_1, a_1^*, a_2, a_2^*, \cdots, a_K, a_K^*]^T$, and where $\mathbf{Z} = [\mathbf{z}(1), \mathbf{z}(-1), \mathbf{z}(2), \mathbf{z}(-2), \cdots, \mathbf{z}(K), \mathbf{z}(-K)]$ with $\mathbf{z}(k)$ defined as $\mathbf{z}(k) = [z_1^k, z_2^k, \cdots, z_N^k]^T$. With i.i.d. Gaussian noise assumptions on the elements of $\mathbf{v}_n$ and fixed fundamental frequency ω_0, the maximum likelihood (ML) estimate of the complex amplitude vector $\mathbf{a}$ can be found using the normal equations, $\hat{\mathbf{a}} = \left(\mathbf{Z}^H\mathbf{Z}\right)^{-1}\mathbf{Z}^H\mathbf{y}$ [8]. Replacing $\mathbf{a}$ in (5) with the ML estimate $\hat{\mathbf{a}}$, the ML estimator of the fundamental frequency can be formulated as the least squares problem

$$\begin{aligned} \hat{\omega}_0 =& \arg\min_{\omega_0} \quad \mathbf{y} - \mathbf{Z} \quad \mathbf{Z}^H\mathbf{Z} \quad^{-1} \quad \mathbf{Z}^H\mathbf{y} \quad_2^2 \\ =& \arg\max_{\omega_0} \mathbf{y}^T\mathbf{Z} \quad \mathbf{Z}^H\mathbf{Z} \quad^{-1} \mathbf{Z}^H\mathbf{y}, \end{aligned} \tag{6}$$

where $1/\cdot 1/_2^2$ is the squared 2-norm. The above NLS maximization problem is solved by a coarse grid search followed by a gradient ascent refinement process.

3. KALMAN FILTER-BASED FUNDAMENTAL FREQUENCY ESTIMATION ALGORITHM

We now proceed to consider the time-varying fundamental frequency and amplitude scenario. We first formulate the state and observation equations based on the time-varying harmonic model (2) and (3), and the observation model (1), respectively. Then, the extended Kalman smoother framework is applied to solve the nonlinear observation equation problem.

3.1. State and observation equations

Assuming that the continuous phase can be written as $\Theta_{t,k} = k\Omega_t t + \Theta_{0,k}$, at a typical sampling rate F_s, we obtain that the instantaneous frequency of the k^{th} harmonic is

$$\begin{aligned} k\omega_n = k\Omega_t T_s|_{t=nT_s} =& \frac{T_s \partial\Theta_{t,k}}{\partial t}|_{t=nT_s} \\ \approx& T_s \frac{\Theta_{nT_s,k} - \Theta_{nT_s-T_s,k}}{T_s} \\ =& \theta_{n,k} - \theta_{n-1,k}, \end{aligned} \tag{7}$$

where Ω_t is the continuous radian frequency, $\omega_n = \Omega_{nT_s}T_s$ and $\theta_{n,k} = \Theta_{nT_s,k}$.

We collect the frequency, amplitudes and phase $\theta_{n-1,1} - \theta_{0,1}$ as a $(K+2) \times 1$ state vector

$$\mathbf{x}_n = [\omega_n, A_{n,1}, \cdots, A_{n,K}, \theta_{n-1,1} - \theta_{0,1}]^T. \tag{8}$$

From (7) and (8), we can further derive that the phases of different harmonics for $n \geq 1$ are related by

$$\begin{aligned} \theta_{n,k} =& \theta_{n-1,k} + k\omega_n \\ =& \theta_{0,k} + k \mathop{}_{i=1}^{n} \quad \omega_i \\ =& \theta_{0,k} + k(\theta_{n-1,1} - \theta_{0,1} + \omega_n) \\ =& kx_{n,1} + kx_{n,K+2} + \theta_{0,k}, \quad k = 1, \cdots, K, \end{aligned} \tag{9}$$

where $x_{n,i}$ denotes the i^{th} component of the vector $\mathbf{x}_n$. Substituting (8) and (9) into (2), the harmonic model can be re-formulated as

$$s_n = \mathop{}_{k=1}^{K} \quad x_{n,k+1}\cos(kx_{n,1} + kx_{n,K+2} + \theta_{0,k}). \tag{10}$$

We assume the frequency and amplitudes are changing in time according to a first order Markov chain random walk model

$$x_{n,k} = x_{n-1,k} + m_{n,k}, \quad k = 1, \cdots, K+1, \tag{11}$$

where $m_{n,k}$ are K zero mean, i.i.d. Gaussian processes. Moreover, based on the phase update (7) and definition (8), we have

$$\begin{aligned} x_{n,K+2} =& \theta_{n-1,1} - \theta_{0,1} \\ =& \theta_{n-2,1} + \omega_{n-1} - \theta_{0,1} \\ =& x_{n-1,K+2} + x_{n-1,1}. \end{aligned} \tag{12}$$

Based on (11) and (12), we can write the state equation in matrix form

$$\mathbf{x}_n = \mathbf{F}\mathbf{x}_{n-1} + \mathbf{\Gamma}\mathbf{m}_n, \tag{13}$$

where $\mathbf{F}$ is a $(K+2) \times (K+2)$ lower triangular Toeplitz matrix with first column $[1, 0, \cdots, 0, 1]^T$, $\mathbf{\Gamma}$ is a $(K+2) \times (K+1)$ Toeplitz matrix with first column $[1, 0, \cdots, 0]^T$ and the first row as $[1, 0, \cdots, 0]$, and the state noise vector is defined as $\mathbf{m}_n = [m_{n,1}, m_{n,2}, \cdots, m_{n,K+1}]^T$ with a covariance matrix $\mathbf{Q}_m$. Combining (1) and (10), we can write the observation equation in matrix form

$$y_n = (\mathbf{G}\mathbf{x}_n)^T \cos(\mathbf{B}\mathbf{x}_n + \boldsymbol{\theta}_0) + v_n, \tag{14}$$

where $\mathbf{G}$ is a $K \times (K+2)$ Toeplitz matrix with first column as a zero vector and first row as $[0, 1, 0, \cdots, 0]$, $\mathbf{B}$ is a $K \times (K+2)$ zero matrix except that the first and last columns are $[1, 2, \cdots, K]^T$, and $\boldsymbol{\theta}_0 = [\theta_{0,1}, \theta_{0,2}, \cdots, \theta_{0,K}]^T$.

3.2. Linearization via Taylor approximation

We linearise the nonlinear observation equation (14) using the first-order Taylor expansion around estimate $\mathbf{x}_n = \hat{\mathbf{x}}_{n|n-1}$

$$y_n \approx h(\hat{\mathbf{x}}_{n|n-1}) + \mathbf{H}_n(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1}) + v_n, \tag{15}$$

$$h(\hat{\mathbf{x}}_{n|n-1}) = (\mathbf{G}\hat{\mathbf{x}}_{n|n-1})^T \cos(\mathbf{B}\hat{\mathbf{x}}_{n|n-1} + \boldsymbol{\theta}_0), \tag{16}$$

Algorithm 1 Extended Kalman smoohter for fundamental frequency estimation

```
Initiate harmonic order $K$, state vector $\mathbf{x}_1$ and initial phase $\boldsymbol{\theta}_0$ with the NLS and Amp-LS algorithms
Choose initial state covariance $\mathbf{P}_{1|1}$, state noise covariance $\mathbf{Q_m}$ and background noise variance $r_v$
Filtering step (forward, online):
for $n = 2, 3, \cdots, N$ do
    $\mathbf{x}_{n|n-1} = \mathbf{F}\mathbf{x}_{n-1|n-1}$
    $\mathbf{P}_{n|n-1} = \mathbf{F}\mathbf{P}_{n-1|n-1}\mathbf{F}^T + \boldsymbol{\Gamma}\mathbf{Q_m}\boldsymbol{\Gamma}^T$
    Calculate $\mathbf{H}_n$ based on (17)
    $\mathbf{K}_n = \mathbf{P}_{n|n-1}\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_{n|n-1}\mathbf{H}_n^T + r_v)^{-1}$
    Obtain $h(\hat{\mathbf{x}}_{n|n-1})$ based on (16)
    $\mathbf{x}_{n|n} = \mathbf{x}_{n|n-1} + \mathbf{K}_n(y_n - h(\hat{\mathbf{x}}_{n|n-1}))$
    $\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n\mathbf{H}_n\mathbf{P}_{n|n-1}$
end for
Smoothing step (backward, offline):
for $n = N, N-1, \cdots, 2$ do
    $\mathbf{S}_{n-1} = \mathbf{P}_{n-1|n-1}\mathbf{F}^T\mathbf{P}_{n|n-1}^{-1}$
    $\mathbf{x}_{n-1|N} = \mathbf{x}_{n-1|n-1} + \mathbf{S}_{n-1}(\mathbf{x}_{n|N} - \mathbf{x}_{n|n-1})$
    $\mathbf{P}_{n-1|N} = \mathbf{P}_{n-1|n-1} + \mathbf{S}_{n-1}(\mathbf{P}_{n|N} - \mathbf{P}_{n|n-1})\mathbf{S}_{n-1}^T$
end for
```

where $\mathbf{H}_n$ is a $1 \times (K+2)$ Jacobian matrix

$$\begin{aligned}\mathbf{H}_n =& \frac{\partial(\mathbf{G}\mathbf{x}_n)^T \cos(\mathbf{B}\mathbf{x}_n + \boldsymbol{\theta}_0)}{\partial \mathbf{x}_n^T} \Big|_{\mathbf{x}_n=\hat{\mathbf{x}}_{n|n-1}} \\ =& \left[\cdots, \frac{\partial(\mathbf{G}\mathbf{x}_n)^T \cos(\mathbf{B}\mathbf{x}_n + \boldsymbol{\theta}_0)}{\partial x_{n,k}}, \cdots\right]_{\mathbf{x}_n=\hat{\mathbf{x}}_{n|n-1}} \\ =&[\cdots, \mathbf{i}_k^T\mathbf{G}^T \cos(\mathbf{B}\hat{\mathbf{x}}_{n|n-1} - \boldsymbol{\theta}_0) \\ & + (\mathbf{G}\hat{\mathbf{x}}_{n|n-1})^T(\sin(\mathbf{B}\hat{\mathbf{x}}_{n|n-1} + \boldsymbol{\theta}_0)\ 8\ \mathbf{B}_{\cdot,k}), \cdots] \\ =&\cos((\mathbf{B}\hat{\mathbf{x}}_{n|n-1} + \boldsymbol{\theta}_0)^T)\mathbf{G} \\ & - (\mathbf{G}\hat{\mathbf{x}}_{n|n-1})^T \mathrm{diag}(\sin(\mathbf{B}\hat{\mathbf{x}}_{n|n-1} + \boldsymbol{\theta}_0))\mathbf{B}, \qquad (17)\end{aligned}$$

where $\mathbf{i}_k$ is a zero vector except that the k^{th} element is 1, 8 denotes the element-wise product, $\mathbf{B}_{\cdot,k}$ denotes the k^{th} column of the matrix $\mathbf{B}$, $\mathrm{diag}(\mathbf{z})$ denotes converting a column vector $\mathbf{z}$ to a diagonal matrix with the $(i,i)^{\text{th}}$ diagonal element set as the i^{th} element of $\mathbf{z}$.

3.3. Kalman-based fundamental frequency estimation

We use the extended Kalman filter (EKF) smoother to estimate the mean and covariance of the state vector $\mathbf{x}_n$. The filtering and smoothing steps of the EKF are shown in Algorithm 1. For real-time applications, the forward filtering step should be used without the backward smoothing step. Using only the forward filtering step leads to larger uncertainty over the parameter estimates [13]. This algorithm can be initialized with the NLS estimate and the complex amplitude estimator using least-squares (Amp-LS) [4]. For Kalman filter parameter tuning we refer the reader to [14] and [15].

4. RESULTS

In this section, we test the performance of the proposed Kalman-based fundamental frequency tracking algorithm for real speech and music signal.

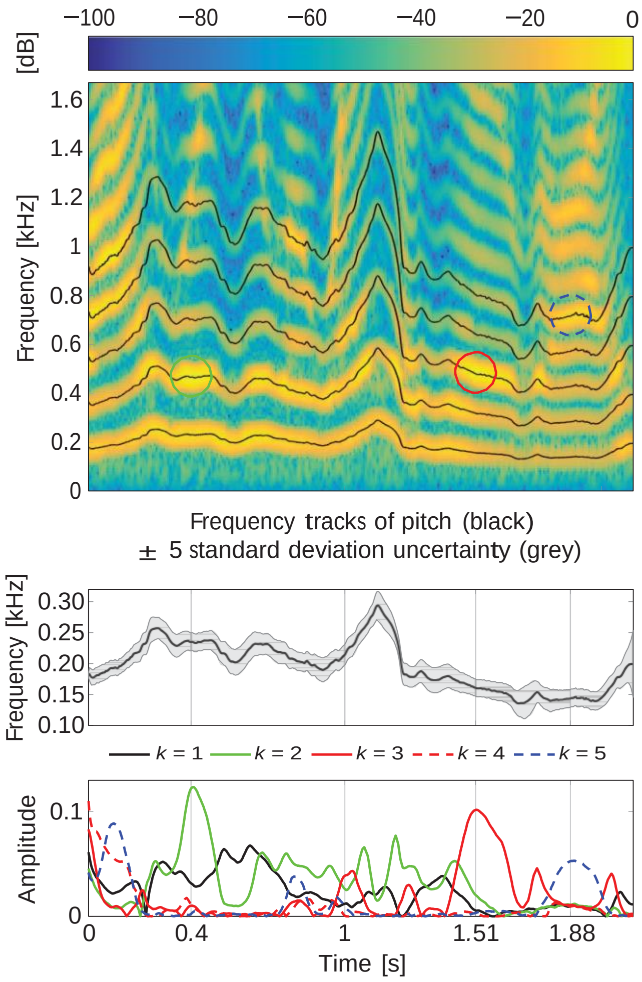

Figure 1: Fundamental frequency estimates of the speech signal "Why were you away a year, Roy?", number of harmonics $K = 5$.

4.1. Speech signal analysis

First, the proposed approach is tested on a speech signal of the spoken sentence "Why were you away a year, Roy?" uttered by a female speaker and sampled at 8000 Hz. The spectrogram of the clean speech signals, fundamental frequency and amplitude estimates are shown in Figure 1, where $K = 5$, $r_v = 10^4$, the SNR for Gaussian white noise is set to 10 dB, $\mathbf{Q_m}$ and $\mathbf{P}_{1|1}$ are set to the identity matrices. As can be seen, the proposed algorithm generates continuous pitch estimates. Large amplitude estimates for harmonics $k = 2$, $k = 3$ and $k = 5$ are obtained in the high energy time-frequency area around 0.4 s, 1.51 s and 1.88 s. However, note that a clear delay in frequency estimate can be seen around 0.3 s (see the 4th and 5th harmonic tracks) due to the fixed harmonic order and r_v we used here. One approach to mitigating this delay is to re-initiate the algorithm with estimated harmonic order K and r_v based on a segmentation approach [16].

Second, the performance of the proposed approach with different harmonic orders is compared with the SWIPE and Fast-NLS algorithms on a sustained /a/ signal from a female with Parkinson's disease [17]. The estimated ground truth fundamental frequencies

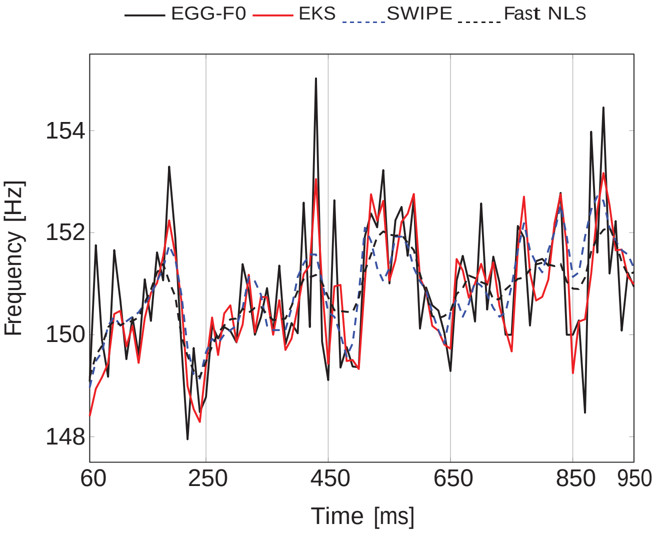

Figure 2: Fundamental frequency estimates of a sustained /a/ signal from a female patient with Parkinson's disease, number of harmonics $K = 4$, noise variance $r_v = 10^5$.

Table 1: Performance of the fundamental frequency estimation algorithms for a real, sustained /a/ voice signals of 1 second duration from a female with Parkinson's disease. Performance results here are averaged over 90 frequency values from 60 to 950 ms in steps of 10 ms. MAE: mean absolute error, MRE: mean relative error, RMSE: root mean squared error.

Algorithms	MAE (Hz)	MRE (%)	RMSE (Hz)
EKF, $K = 4$	0.97	0.64	1.26
SWIPE	0.80	0.53	1.05
Fast NLS	0.80	0.53	1.05
EKS, $K = 1$	0.72	0.47	0.95
EKS, $K = 2$	0.71	0.47	0.92
EKS, $K = 3$	0.66	0.44	0.87
EKS, $K = 4$	0.63	0.42	0.86

in 10 ms time frames are extracted from the electroglottography (EGG) and thus referred as EGG-F0. People with Parkinson's tend to exhibit increased vocal breathiness, tremor and roughness, and this presents a challenge for fundamental frequency estimation algorithms. The frequency estimates and the corresponding error measures of mean absolute error (MAE), mean relative error (MRE) and root mean squared errors (RMSE, see definitions in [17]) are obtained, where smaller values of error measures are better (see Figure 2 and Table 1). For the proposed EKS and traditional EKF, the noise variance r_v is set to 10^5 and the fundamental frequency estimates are averaged over every 10 ms segment. As can be seen from Figure 2, the proposed EKS with $K = 4$ achieves the closest approximation to the EGG-F0. Also, from Figure 2 and Table 1 the performance of SWIPE and Fast NLS is similar and tends to obtain a smooth estimate of the fundamental frequency. Furthermore, when $K = 4$, the performance of the proposed EKS is better than for other choices of K.

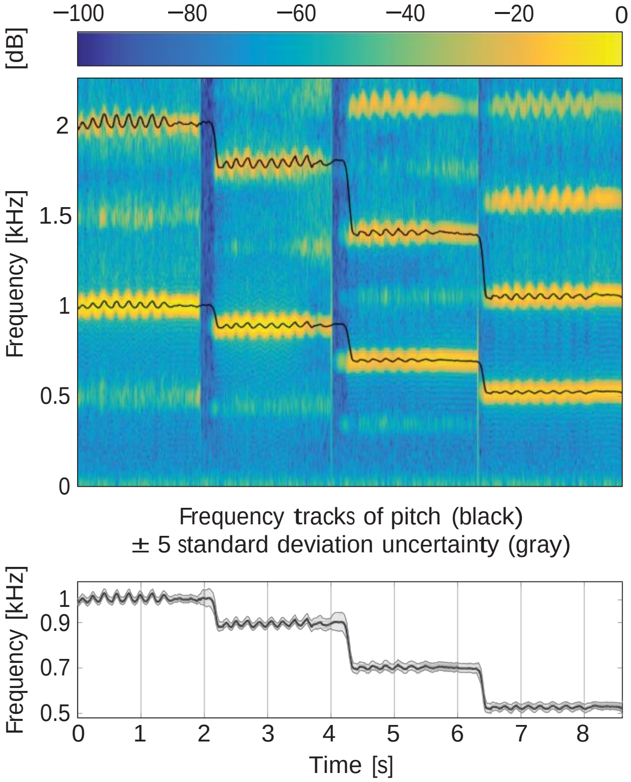

Figure 3: Fundamental frequency estimate of vibrato notes of a flute decreasing from B5 to C5 (eight notes of C-Major scale), number of harmonics $K = 2$.

4.2. Music signal analysis

The sound of a musical instrument (flute) decreasing in frequency from note B5 to C5 from the University of Iowa Musical Instrument Samples [18] database is tested. The spectrogram of the signals and frequency estimates are shown in Figure 3, with 10 dB SNR for Gaussian white noise. The other parameters have the same settings as in Figure 1. As can be seen, the proposed EKS can obtain a reasonably good estimate of fundamental frequency. Although, the frequency tracks continues almost unchanged when there is no/weak fundamental frequency during transition periods from one note to another (a limitation of the 2-norm based linearized Kalman filtering method), larger frequency uncertainties are obtained there.

5. CONCLUSIONS

In this paper, we have proposed a fundamental frequency estimation algorithm based on a parametric harmonic model. Non-stationary temporal evolution of frequency and amplitude are modeled as first-order Markov chains. Compact nonlinear matrix forms of state and observation equations based are formulated, and an extended Kalman smoother for the problem is derived. The size of the state space is lowered by exploiting the linear relationships between the phases of different harmonics. Continuous fundamental frequency and amplitudes estimates for sustained vowels are compared to ground truth estimates from the EGG, showing that this new algorithm outperforms existing algorithms in terms of accuracy.

6. REFERENCES

[1] M. Krawczyk-Becker and T. Gerkmann, "Fundamental frequency informed speech enhancement in a flexible statistical framework," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 24, no. 5, pp. 940–951, 2016.

[2] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, 2012.

[3] P. Ghahremani, B. BabaAli, D. Povey, K. Riedhammer, J. Trmal, and S. Khudanpur, "A pitch extraction algorithm tuned for automatic speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2014, pp. 2494–2498.

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

[5] 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.

[6] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," *Speech coding and synthesis*, vol. 495, p. 518, 1995.

[7] A. Camacho and J. G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 124, no. 3, pp. 1638–1652, 2008.

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

[9] J. K. Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Fast fundamental frequency estimation: Making a statistically efficient estimator computationally efficient," *Signal Process.*, vol. 135, pp. 188–197, 2017.

[10] M. G. Christensen and J. R. Jensen, "Pitch estimation for non-stationary speech," in *Rec. Asilomar Conf. Signals, Systems, and Computers*. IEEE, 2014, pp. 1400–1404.

[11] Y. Pantazis, O. Rosec, and Y. Stylianou, "Adaptive AM–FM signal decomposition with application to speech analysis," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 19, no. 2, pp. 290–300, 2011.

[12] H. Hajimolahoseini, R. Amirfattahi, H. Soltanian-Zadeh, and S. Gazor, "Instantaneous fundamental frequency estimation of non-stationary periodic signals using non-linear recursive filters," *IET Signal Process.*, vol. 9, no. 2, pp. 143–153, 2015.

[13] D. D. Mehta, D. Rudoy, and P. J. Wolfe, "Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking a," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1732–1746, 2012.

[14] R. Mehra, "On the identification of variances and adaptive kalman filtering," *IEEE Trans. Autom. Control*, vol. 15, no. 2, pp. 175–184, 1970.

[15] S. Formentin and S. Bittanti, "An insight into noise covariance estimation for kalman filter design," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 2358–2363, 2014.

[16] S. M. Nørholm, J. R. Jensen, and M. G. Christensen, "Instantaneous fundamental frequency estimation with optimal segmentation for nonstationary voiced speech," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 24, no. 12, pp. 2354–2367, 2016.

[17] A. Tsanas, M. Zañartu, M. A. Little, C. Fox, L. O. Ramig, and G. D. Clifford, "Robust fundamental frequency estimation in sustained vowels: detailed algorithmic comparisons and information fusion with adaptive kalman filtering," *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2885–2901, 2014.

[18] L. Fritts. [Online]. Available: http://theremin.music.uiowa.edu/index.html