Extending Coherence for Optimal Detection of Nonstationary Harmonic Signals

Scott Wisdom*, James Pitton†*, and Les Atlas*
*Department of Electrical Engineering, University of Washington, Seattle, WA, USA
†Applied Physics Laboratory, University of Washington, Seattle, WA, USA

***Abstract*—This paper describes an improved detector for nonstationary harmonic signals. The performance improvement is accomplished by using a novel method for extending the coherence time of such signals. This method applies a transformation to a noisy signal that attempts to fit a simple model to the signal's slowly changing fundamental frequency over the analysis duration. By matching the change in the signal's fundamental frequency, analysis is more coherent with the signal over longer durations, which allows the use of longer windows and thus improves detection performance.**

I. Introduction

Nonstationary harmonic signals, which are composed of narrowband components at integer multiples of a slowly changing fundamental frequency $f(t)$, abound in science and engineering. Examples of nonstationary harmonic signals include voiced speech and signals encountered in passive sonar applications, such as ship engine noise and marine mammal vocalizations. Detection of nonstationary harmonic signals is an important problem. For example, the strong energy contained in voiced speech harmonics is a useful feature for voice activity detection [1]. In passive sonar, detection of nonstationary harmonics is essential in identification of contacts. Spectrograms of two examples of nonstationary harmonic signals are shown in figure 1.

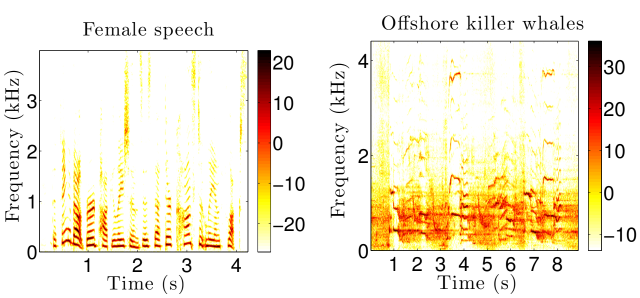

Fig. 1. Spectrograms in dB of two examples of nonstationary harmonic signals. Left: female speech, right: hydrophone recording of offshore killer whales, which consists of multiple nonstationary harmonic signals.

Conventional processing of nonstationary harmonic signals is typically done on a frame-by-frame basis. Such processing is based on the assumption that the signal's frequency content does not change significantly over the analysis frame duration. The result of this assumption limits performance, because short frames reduce the amount of data available to statistical estimators. Since efficient statistical estimators produce better estimates when provided with longer data records (assuming stationary signals), their performance should improve when provided with more stationary data.

First, we examine a formulation of the optimal detectors for nonstationary harmonic signals in terms of the classic detectors originally described by Scharf [2] and a generalized likelihood ratio test (GLRT). This formulation unifies several classic detectors [3]–[5]. Then, we present our main contribution, which is a novel method of extending the duration of analysis frames, which provides improved performance of statistical estimators for nonstationary harmonic signals. This method is inspired by pioneering work on warped wavelets [6] and the fan-chirp transform [7]–[10]. Recently, we used our method to improve detection of modulated random processes [11] and enhancement of noisy and reverberant speech [12]. In this paper, we proceed in a similar fashion, and show how our method can improve detection of nonstationary harmonic signals. To validate our approach, we demonstrate substantial performance improvements on synthetic signals.

II. Background

This section describes the signal model of interest and the theory and implementation of the conventional detector for nonstationary harmonic signals.

A. Signal model

In this paper, we consider real-valued nonstationary harmonic signals of the form

$$x(t) = \sum_{k=1}^{K} A_k \cos\left(k2\pi \int_0^t f(\tau)d\tau + \phi_k\right) \tag{1}$$

$$= \mathrm{Re}\left\{\sum_{k=1}^{K} a_k \exp j\, k2\pi \int_0^t f(\tau)d\tau\right\} \tag{2}$$

where $\mathbf{a} \in \mathbb{C}^K$ are complex-valued amplitudes that give the amplitude and phase of each harmonic (i.e., $a_k = A_k e^{j\phi_k}$), K is the number of harmonics, and $f(t)$ is the instantaneous fundamental frequency of $x(t)$. If T is the duration of the analysis window in seconds and f_s is the sampling rate, then the $N = Tf_s$ discrete-time samples of the noisy signal are modeled as

$$\mathbf{y} = \mathbf{x} + \mathbf{v}, \tag{3}$$

This work is funded by ONR contract N00014-12-G-0078, delivery order 0013, and ARO grant number W911NF1210277.

with $\mathbf{y}, \mathbf{x}, \mathbf{v} \in \mathbb{R}^N$, where $\mathbf{v}$ is additive white Gaussian noise with variance σ^2.

B. Conventional detector

The goal is to design an optimal detector for $\mathbf{x}$. Most existing approaches assume that that fundamental frequency is constant over short analysis durations of length T; that is, $f(t) = f_0$ for $t_0 \leq t \leq t_0 + T$. Such an assumption is valid for short enough durations. For example, most speech processing algorithms are limited to frame durations of 10 to 30 milliseconds, because the speech signal is assumed to be stationary during this short period. Given these assumptions and the model (2), the measurements (3) can be written as a linear model [2], given by

$$\mathbf{y} = \underline{\mathbf{G}}(\boldsymbol{\psi})\underline{\boldsymbol{\theta}} + \mathbf{v}, \tag{4}$$

where $\underline{\mathbf{G}}(\boldsymbol{\psi}) = [\mathbf{G}(\boldsymbol{\psi}) \;\; \mathbf{G}^*(\boldsymbol{\psi})] \in \mathbb{C}^{N \times 2K}$ is a linear signal subspace, $\boldsymbol{\psi}$ are parameters defining the subspace, and $\underline{\boldsymbol{\theta}} \in \mathbb{C}^{2K}$ are unknown linear signal parameters. Under conventional assumptions, $\boldsymbol{\psi} = \{\omega_0, K\}$ and $\underline{\boldsymbol{\theta}} = [\mathbf{a}^H \mathbf{a}^T]^H$, where $\mathbf{a} \in \mathbb{C}^K$ are the complex amplitudes in (2). The two blocks in $\underline{\mathbf{G}}(\boldsymbol{\psi})$ are built from

$$[\mathbf{G}(\omega_0, K)]_{n,k} = \exp jk\omega_0 n \tag{5}$$

where $\omega_0 = \frac{2\pi f_0}{f_s}$. We will use this notation for frequency of discrete-time signals throughout.

Using the identity

$$A\cos(x + \phi) = A\cos\phi\cos x - A\sin\phi\sin x, \tag{6}$$

the linear model (4) can be written equivalently [1], [13] as

$$\mathbf{y} = \mathbf{H}(\boldsymbol{\psi})\boldsymbol{\theta} + \mathbf{v}, \tag{7}$$

where $\mathbf{H}(\psi) = [\mathbf{H_c}(\psi) \; \mathbf{H_s}(\psi)] \in \mathbb{R}^{N \times 2K}$ and

$$\begin{aligned} [\mathbf{H_c}(\omega_0, K)]_{n,k} &= \cos k\omega_0 n \\ [\mathbf{H_s}(\omega_0, K)]_{n,k} &= \sin k\omega_0 n. \end{aligned} \tag{8}$$

The unknown signal parameters are now $\boldsymbol{\theta} = \left[\mathbf{a_c}^T \; \mathbf{a_s}^T\right]^T \in \mathbb{R}^{2K}$. Thus, the formulation (7) expresses (4) in terms of the real and imaginary parts of the complex amplitudes $a_{c,k} = A_k \cos\phi_k = \operatorname{Re} a_k$ and $a_{s,k} = -A_k \sin\phi_k = -\operatorname{Im} a_k$, and uses a real-valued subspace matrix $\mathbf{H}(\boldsymbol{\psi})$.

When subspace parameters $\boldsymbol{\psi}$ are known, and signal parameters $\boldsymbol{\theta}$ and noise variance σ^2 are unknown, the uniformly most powerful (UMP) detector is the constant false-alarm rate (CFAR) matched subspace detector [2, §4.12]. When subspace parameters $\boldsymbol{\psi}$ are also unknown, the generalized likelihood ratio test (GLRT) statistic is given by [1]

$$D(\mathbf{y}) = \frac{\mathbf{y}^T \mathbf{P}_{\mathbf{H}(\hat{\boldsymbol{\psi}})} \mathbf{y}}{\mathbf{y}^T \left(\mathbf{I} - \mathbf{P}_{\mathbf{H}(\hat{\boldsymbol{\psi}})}\right) \mathbf{y}}. \tag{9}$$

The matrix $\mathbf{P}_{\mathbf{H}(\hat{\boldsymbol{\psi}})}$ is a projection onto the most likely linear signal subspace $\mathbf{H}(\hat{\boldsymbol{\psi}})$, where $\hat{\boldsymbol{\psi}}$ are the maximum likelihood estimates of the unknown subspace parameters $\boldsymbol{\psi}$. The projection for a matrix $\mathbf{H}$ is given by $\mathbf{P_H} = \mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$, and the projection onto the complement $\mathbf{H}^\perp$ of $\mathbf{H}$ is given by $\mathbf{P}_\mathbf{H}^\perp = \mathbf{I} - \mathbf{P_H}$. Thus, the detection statistic (9) is a ratio of the energy of $\mathbf{y}$ that lies in the most likely signal subspace $\mathbf{H}(\hat{\boldsymbol{\psi}})$ to the energy of $\mathbf{y}$ that lies in the complement $\mathbf{H}^\perp(\hat{\boldsymbol{\psi}})$ of the most likely signal subspace.

C. Implementation of conventional detector

A practical and efficient implementation of (9) makes use of the fast Fourier transform (FFT). Note that this implementation assumes that ω_0 is an integer multiple of the FFT center frequencies ω. The spectrum $Y(\omega)$ of $\mathbf{y}$ is estimated by applying a window function $\mathbf{h}$ and computing the discrete Fourier transform (DFT) using the FFT:

$$Y(\omega) = \sum_{n=0}^{N-1} h_n y_n \exp -j\omega n, \tag{10}$$

for $\omega = 2\pi\frac{n_{FFT}}{N_{FFT}}$, $n_{FFT} \in \{0, ..., N_{FFT} - 1\}$. When K is known, then the maximum likelihood estimate of ω_0 can be shown (see Appendix) to be

$$\hat{\omega}_0 = \underset{\omega_0}{\operatorname{argmax}} \sum_{k=1}^{K} |Y(k\omega_0)|^2. \tag{11}$$

When K is not known, estimation becomes more complicated, since K determines the model order of the system. Popular methods for estimating K include the Akaike information criterion (AIC), minimum description length (MDL), and maximum *a posteriori* (MAP) probability [14]. All these methods consist of minimizing a cost function plus a penalty term that prevents overfitting. Whipps and Moses [15] found that MDL yields consistent estimates for large sample records, and provide an efficient algorithm to jointly estimate $\hat{\omega}_0$ and $\hat{K}$ by minimizing [15, (32)]

$$\left\{\hat{\omega}_0, \hat{K}\right\} = \underset{\omega_0, K}{\operatorname{argmin}} \, N\log\left(\mathbf{y}^T\mathbf{P}^\perp_{\mathbf{H}(\omega_0,K)}\mathbf{y}\right) + K\log N. \tag{12}$$

Given $\hat{\omega}_0$ and $\hat{K}$, the detection statistic (9) can be approximated as

$$\hat{D}(\mathbf{y}) = \frac{\sum_{\omega_s \in \Omega_s} |Y(\omega_s)|^2}{\sum_{\omega_v \in \Omega_v} |Y(\omega_v)|^2}, \tag{13}$$

where $\Omega_s = \left\{\hat{\omega}_0, 2\hat{\omega}_0, ..., \hat{K}\hat{\omega}_0\right\}$ is a set of frequencies corresponding to a "harmonic sieve" and $\Omega_v = \Omega \setminus \Omega_s$.

Note that given $\hat{\omega}_0$ and $\hat{K}$, the maximum likelihood estimates of the complex amplitudes a_k are simply (see Appendix)

$$\hat{a}_k = \frac{2}{N} Y(k\hat{\omega}_0), \text{ for } k = 1, ..., \hat{K}. \tag{14}$$

However, the maximum likelihood estimates $\hat{a}_k$ are not required for computing the detection statistic (13).

III. Proposed Method

The nonstationary harmonic signal detection in the previous section made the assumption that $f(t) \approx f_0$ over a short analysis interval T. However, this assumption limits the amount of data available to the detection statistic (9). To increase the amount of data available during an analysis frame, we add a

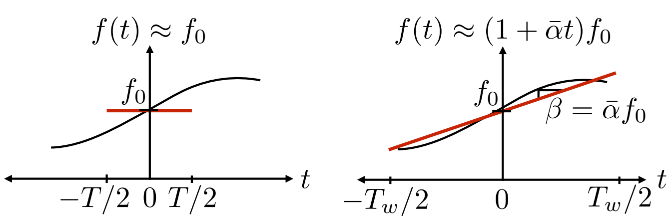

Fig. 2. Left: conventional assumption that $f(t)$ is approximately constant over a short duration T. Right: new assumption of a linear fit to $f(t)$ that allows a longer analysis duration $T_w \geq T$.

parameter α to the set of subspace parameters $\boldsymbol{\psi}$ that allows a linear fit to $f(t)$ over a longer analysis duration $T_w \geq T$. Thus, under this new model, the instantaneous fundamental frequency is modeled as

$$f(t) \approx (1 + \bar{\alpha} t)\, f_0, \tag{15}$$

where $\bar{\alpha} = \beta / f_0$ is a continuous normalized chirp rate and β is the slope in units of Hertz per second of the linear fit. Figure 2 shows a comparison between the conventional assumption of constant frequency f_0 over the shorter duration T and this new assumption of a linear fit to $f(t)$ over a longer duration T_w.

A. Proposed detector

The linear approximation to fundamental frequency (15) yields the instantaneous phase $\phi(t) = \phi_{\bar{\alpha}}(t) f_0$, where

$$\phi_{\bar{\alpha}}(t) = \int_0^t (1 + \bar{\alpha}\tau) d\tau = \left(t + \frac{1}{2}\bar{\alpha} t^2\right). \tag{16}$$

For discrete-time samples $y_n = y(t_n)$ where $t_n = n/fs$, this phase can be written as

$$\phi_\alpha(n) = \phi_{\bar{\alpha}}(t_n) = \frac{1}{f_s}\left(n + \frac{1}{2}\alpha n^2\right), \tag{17}$$

where $\alpha = \bar{\alpha}/f_s$ is a discrete-time normalized chirp rate. The phase of the kth harmonic in (2) is then

$$k 2\pi f_0 \frac{1}{f_s}(n + \frac{1}{2}\alpha n^2) = k\omega_0 \left(n + \frac{1}{2}\alpha n^2\right). \tag{18}$$

Thus, the detector using the linear fit (15) uses the detection statistic (9) with the signal subspace matrix $\mathbf{H}(\psi) = [\mathbf{H_c}(\psi)\ \mathbf{H_s}(\psi)] \in \mathbb{R}^{N \times 2K}$, where

$$\begin{aligned} [\mathbf{H_c}(\omega_0, \alpha, K)]_{n,k} &= \cos k\omega_0 \left(n + \frac{1}{2}\alpha n^2\right) \\ [\mathbf{H_s}(\omega_0, \alpha, K)]_{n,k} &= \sin k\omega_0 \left(n + \frac{1}{2}\alpha n^2\right). \end{aligned} \tag{19}$$

B. Implementation of proposed detector

In (19), the subspace matrix $\mathbf{H}$ for the proposed detector is formulated using basis functions with nonstationary, linearly-varying frequency. An efficient and equivalent implementation of the proposed detector consists of a time-warping (which compensates for the nonlinear phase (16), (17) of the signal), followed by a FFT. Then the detection statistic is computed using a harmonic sieve, as in (13).

Maximum likelihood estimation of ω_0 and α for a frame of data can be shown (see Appendix) to be a joint maximization over the gathered magnitude-squared spectrum:

$$\{\hat{\omega_0}, \hat{\alpha}\} = \underset{\omega_0, \alpha}{\operatorname{argmax}} \sum_{k=1}^{\hat{K}} |Y(k\omega_0, \alpha)|^2, \tag{20}$$

where $Y(\omega, \alpha)$ is given by

$$Y(\omega, \alpha) = \sum_{n=0}^{N-1} y_n e^{-j\omega(1 + \frac{1}{2}\alpha n)n}. \tag{21}$$

$Y(\omega, \alpha)$ is related to the discrete fan-chirp transform [7], [8], given by

$$Y^{fc}(\omega, \alpha) = \sum_{n=0}^{N-1} y_n \sqrt{|\phi'_\alpha(n)|} e^{-j\omega\phi_\alpha(n)}. \tag{22}$$

Using a variable substitution $\tau = \phi_\alpha(n)$, (22) can be written [7], [8] in an equivalent form

$$Y^{fc}(\omega, \alpha) = \sum_{m=0}^{M-1} \widetilde{y}_{\alpha,m} e^{-j\omega m}, \tag{23}$$

which is the DFT of the time-warped signal $\widetilde{y}_{\alpha,m}$. The M-length time-warped signal $\widetilde{y}_{\alpha,m}$ is given by [8, (66)], [9]

$$\widetilde{y}_{\alpha,m} = \sum_{\ell=0}^{N-1} \frac{y_\ell}{\sqrt{1 + \alpha t_\ell}} h\left(\phi_{\bar{\alpha}}^{-1}(\widetilde{t}_m) - t_\ell\right), \tag{24}$$

where $h(t)$ is an interpolation filter, which can be an ideal sinc function or—to reduce computation—shorter filters, such as a cubic Hermite spline or a linear interpolation kernel. Upsampling y_n before time-warping is advantageous. The times t_n and $\widetilde{t}_m$ are evenly spaced and centered about 0. That is, $t_n = n/f_s$ for $n \in \{-N/2, ..., N/2\}$ and $t_m = m/f_s$ for $m \in \{-M/2, ..., M/2\}$, and the inverse mapping $\phi_{\bar{\alpha}}^{-1}$ in (24) is given by [8, (20)], [10, (5)]

$$\phi_{\bar{\alpha}}^{-1}(t) = -\frac{1}{\bar{\alpha}} + \frac{\sqrt{1 + 2\bar{\alpha} t}}{\bar{\alpha}}. \tag{25}$$

Figure 3 illustrates the effect of time-warping on a short clip of voiced speech.

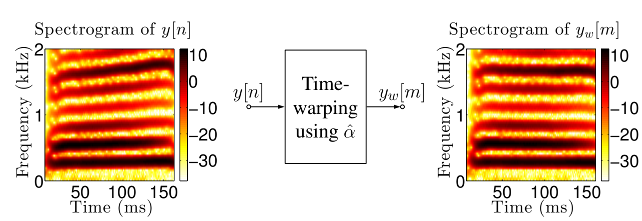

Fig. 3. Illustration of the effect of time-warping. When a short frame of a nonstationary harmonic signal with approximately linear fundamental frequency variation (left) is warped using the maximum likelihood chirp rate $\hat{\alpha}$, the harmonic chirps are converted to harmonic tones with approximately constant frequency (right).

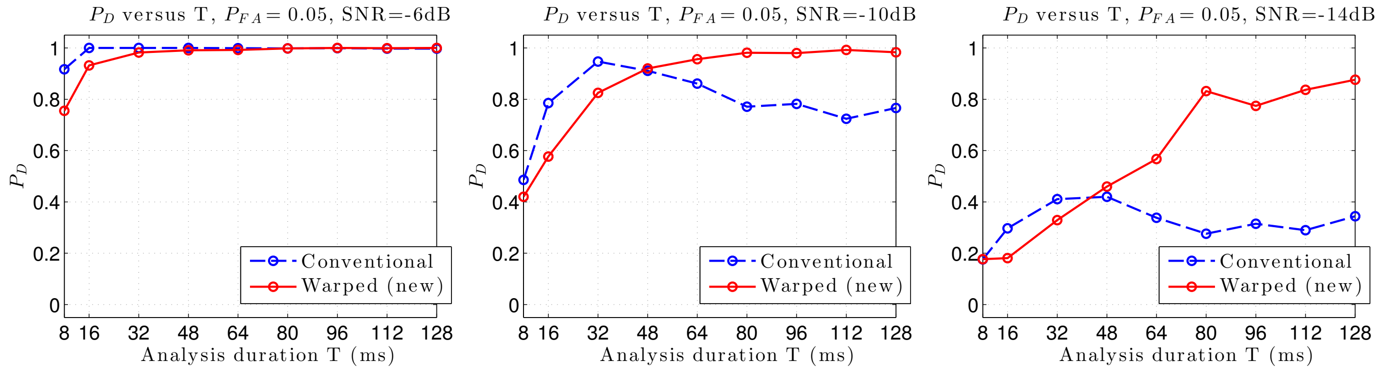

Fig. 4. Monte Carlo experiments with A_k, ϕ_k, ω_0, α, and σ^2 unknown. The conventional detector, which assumes constant fundamental frequency $f(t)$ over the analysis duration, is compared to the new warped detector, which assumes a linear model of $f(t)$ over the analysis duration. Left: -6dB SNR, center: -10dB SNR, right: -14dB SNR.

The figure shows that when time-warping is performed with a matching chirp rate α, harmonics with approximately linearly-varying $f(t)$ are converted into harmonics with approximately constant $f(t)$.

Thus, the warped spectrum $Y(\omega, \alpha)$ from (21) can be efficiently computed as a time-warping (24) followed by a FFT. Using this efficient implementation, a grid search can be performed over a set of frequencies Ω_0 and a set of chirp rates $\mathcal{A}$ to determine the maximum likelihood estimates $\hat{\omega}_0$ and $\hat{\alpha}$ using (20). Then, the detection statistic (9) can be approximated as in (13):

$$\hat{D}_{new}(\mathbf{y}) = \frac{\sum_{\omega_s \in \Omega_s} |Y(\omega_s, \hat{\alpha})|^2}{\sum_{\omega_v \in \Omega_v} |Y(\omega_v, \hat{\alpha})|^2}. \tag{26}$$

IV. Results

Monte Carlo experiments using synthetic data demonstrate the improved performance of the new warped detector (26) as compared to the conventional detector (13). The synthetic signal $x(t)$ to be detected is given by (1). It is sampled at $f_s = 16$ kHz and has $K = 4$ harmonics, uniform random amplitudes A_k between 0 and 1, uniform random phases ϕ_k between 0 and 2π, a center frequency of $f_0 = 300$ Hz, and a normalized chirp rate of $\bar{\alpha} = \frac{\Delta f}{T f_0} = \frac{\beta}{f_0} = 3.125$. That is, the signal's instantaneous fundamental frequency changes linearly at a rate of $3.125 f_0$ Hz per second, which corresponds to a total change of 30 Hz over a 32 ms duration.

In terms of implementation parameters, for both detectors the maximum likelihood estimate ω_0 was found by searching over a grid of 281 frequencies ranging from $f_{0,min} = 80$ Hz to $f_{0,max} = 360$ Hz. For the new warped detector, the maximum likelihood chirp rate $\hat{\alpha}$ was found by searching over 101 chirp rates between $\bar{\alpha}_{min} = -31.25$ and $\bar{\alpha}_{max} = 31.25$. The length M of the warped signal is equal to the original number of samples N. The measurements y_n are upsamped by 8 before warping, and linear interpolation is used to implement (24).

For the experiments, the samples x_n are embedded in additive white Gaussian noise at -6dB, -10dB, and -14dB SNR, where SNR is defined as

$$\text{SNR} \triangleq \frac{\sum_{k=1}^{K} A_k^2}{2\sigma^2} \tag{27}$$

and the analysis window duration T is varied. For each SNR and T, average results were computed using 1000 Monte Carlo trials.

For these 3 different SNRs, the panels of figure 4 show probability of detection P_D versus analysis window duration T, given a false alarm probability $P_{FA} = 0.05$. Notice that for all SNRs, our new warped detector performs worse than the conventional detector for short ($T \leq 48$ ms) analysis frames, while the new warped detector improves performance for longer analysis durations. In very low SNRs (right panel), our new detector can achieve substantial performance improvement over the conventional detector, given a long enough analysis duration. Note, though, that for real-world signals, the warped detector will only improve performance so long as the fundamental frequency of the signal changes approximately linearly over the analysis duration.

V. Conclusion

In this paper, we have described a novel method of improving detection of nonstationary harmonic signals. Our method works by fitting a simple model to fundamental frequency variation over the analysis frame. By estimating the maximum likelihood parameters of this simple model, we are able to perform a time-warping of the data, which is an efficient procedure that renders fundamental frequency approximately constant and thus extends the coherence time of the detector, which improves detection performance.

Many extensions to this work are possible. Future work will examine extending beyond linear fits of fundamental frequency to either higher-order polynomials or other nonlinear models. Also, here we have only considered frame-by-frame detection. Enforcing temporal constraints between frames for the model parameters promises additional performance improvements. Finally, allowing the amplitudes of the harmonics to be time-varying in the model could further improve performance.

Appendix

In this appendix, we show that maximizing the gathered magnitude-squared spectrum in (11) and (20) yields the maximum likelihood estimates of the fundamental frequency ω_0 and the chirp rate α. The derivation is similar to Scharf's proof [2, §6.12] for a single tone with constant frequency in additive white noise.

The signal model is given by (3), which represents discrete samples of the continuous signal (1), where $f(t)$ is given by (15). The unknown parameters are $\boldsymbol{\theta} = \{A_k, \phi_k, \omega_0, \alpha, \sigma^2\}$. Since v_n is zero-mean, independent, and identically distributed white Gaussian noise, the distribution of the measurements $\mathbf{y}$ is

$$f_{\boldsymbol{\theta}}(\mathbf{y}) = (2\pi\sigma^2)^{-N/2} \exp\left(\frac{1}{-2\sigma^2} \sum_{n=0}^{N-1} (y_n - x_n)^2 \right) \quad (28)$$

which corresponds to a log likelihood of

$$L(\boldsymbol{\theta}, \mathbf{y}) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (y_n - x_n)^2. \quad (29)$$

Now set the derivative of (29) with respect to σ^2 to zero:

$$\begin{aligned} \frac{\delta L(\boldsymbol{\theta}, \mathbf{y})}{\delta \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{N-1} (y_n - x_n)^2 = 0 \\ \Rightarrow \hat{\sigma}^2 &= \frac{1}{N} \sum_{n=0}^{N-1} (y_n - x_n)^2, \end{aligned} \quad (30)$$

which shows that the maximum likelihood estimate of variance is the averaged squared residuals between the measurements $\mathbf{y}$ and the signal model $\mathbf{x}$. If (30) is plugged in to (29), it is easy to see that maximizing likelihood requires minimizing the sample variance.

Thus, we desire to minimize $\hat{\sigma}^2$ with respect to A_k, ϕ_k, ω_0, and α. Expanding out (30) and using the approximation $\sum_{n=0}^{N-1} x_n^2 \approx \frac{N}{2} \sum_{k=1}^{K} A_k^2$ yields

$$\begin{aligned} \hat{\sigma}^2 \approx& \frac{1}{N} \sum_{n=0}^{N-1} y_n^2 + \frac{1}{2} \sum_{k=1}^{K} A_k^2 \\ &- \frac{2}{N} \mathrm{Re} \left\{ \sum_{n=0}^{N-1} y_n \sum_{k=1}^{K} A_k e^{j\left(k\omega_0(n+\frac{1}{2}\alpha n^2)+\phi_k\right)} \right\}. \end{aligned} \quad (31)$$

The derivative of (31) with respect to ϕ_k and set to 0 is

$$\frac{\delta \hat{\sigma}^2}{\delta \phi_k} = -\frac{2}{N} \mathrm{Re} \left\{ j e^{-j\phi_k} \sum_{n=0}^{N-1} y_t e^{-jk\omega_0(1+\frac{1}{2}\alpha n^2)} \right\} = 0. \quad (32)$$

Notice the appearance of $Y(k\omega_0, \alpha)$ from (21) in (32). Thus, to satisfy (32), the maximum likelihood estimate is $\hat{\phi}_k = \angle Y(k\omega_0, \alpha)$. Plugging this estimate into (31) and ignoring the data-dependent term $\frac{1}{N} \sum_{n=0}^{N-1} y_n^2$ yields the expression

$$\hat{\gamma}^2 = -\frac{2}{N} \sum_{k=1}^{K} A_k |Y(k\omega_0, \alpha)| + \frac{1}{2} \sum_{k=1}^{K} A_k^2. \quad (33)$$

Taking the derivative of (33) with respect to A_k and setting it to 0 yields the maximum likelihood estimate of the amplitudes $\hat{A}_k = \frac{2}{N} |Y(k\omega_0, \alpha)|$. Plugging $\hat{A}_k$ into (33) yields

$$\hat{\gamma}^2 = -\frac{2}{N^2} \sum_{k=1}^{K} |Y(k\omega_0, \alpha)|^2. \quad (34)$$

Since minimizing $\hat{\gamma}^2$ maximizes the likelihood, the maximum likelihood estimates of the remaining unknown parameters ω_0 and α are given by the joint maximization

$$\{\hat{\omega}_0, \hat{\alpha}\} = \underset{\omega_0, \alpha}{\mathrm{argmax}} \sum_{k=1}^{K} |Y(k\omega_0, \alpha)|^2 \quad (35)$$

Note that (11) is equivalent to (35) when $\alpha = 0$, and that (20) is equivalent to (35). Thus, (11) and (20) are maximum likelihood estimators for ω_0 and $\{\omega_0, \alpha\}$, respectively.

References

[1] E. Fisher, J. Tabrikian, and S. Dubnov, "Generalized likelihood ratio test for voiced-unvoiced decision in noisy speech using the harmonic model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 502–510, Mar. 2006.

[2] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Reading, MA, 1990.

[3] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete-time observations," *Bell System Technical Journal*, vol. 55, no. 9, pp. 1389–1410, Nov. 1976.

[4] M. Hinich, "Detecting a hidden periodic signal when its period is unknown," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 30, no. 5, pp. 747–750, Oct. 1982.

[5] M. Zeytinoglu and K. M. Wong, "Detection of harmonic sets," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2618–2630, Nov. 1995.

[6] R. Baraniuk and D. Jones, "Warped wavelet bases: unitary equivalence and signal processing," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Minneapolis, MN, Apr. 1993, vol. 3, pp. 320–323 vol.3.

[7] M. Képesi and L. Weruaga, "Adaptive chirp-based time–frequency analysis of speech signals," *Speech Communication*, vol. 48, no. 5, pp. 474–492, May 2006.

[8] L. Weruaga and M. Képesi, "The fan-chirp transform for non-stationary harmonic signals," *Signal Processing*, vol. 87, no. 6, pp. 1504–1522, June 2007.

[9] R. Dunn, T. Quatieri, and N. Malyska, "Sinewave parameter estimation using the fast fan-chirp transform," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, Oct. 2009, pp. 349–352.

[10] P. Cancela, E. López, and M. Rocamora, "Fan chirp transform for music representation," in *Proc. International Conference On Digital Audio Effects (DAFx)*, Graz, Austria, Sept. 2010, pp. 1–8.

[11] S. Wisdom, L. Atlas, and J. Pitton, "Extending coherence time for analysis of modulated random processes," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 340–344.

[12] S. Wisdom, T. Powers, L. Atlas, and J. Pitton, "Enhancement of reverberant and noisy speech by extending its coherence," in *Proc. REVERB Challenge Workshop*, Florence, Italy, May 2014.

[13] J. Tabrikian, S. Dubnov, and Y. Dickalov, "Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model," *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 1, pp. 76–87, Jan. 2004.

[14] P. Djurić, "A model selection rule for sinusoids in white gaussian noise," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1744–1751, July 1996.

[15] G. T. Whipps and R. L. Moses, "A combined order selection and parameter estimation algorithm for coupled harmonics," in *Military Sensing Symposium (MSS) Specialty Group on Battlefield Acoustic and Seismic Sensing, Magnetic and Electric Field Sensors*, Laurel, MD, Oct. 2003.