Modeling and Frequency Tracking of Marine Mammal Whistle Calls

by

Jared Severson

B.S., Colorado School of Mines, 2001

Submitted in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

February 2009

Author .....

Joint Program in Oceanography/Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
September 25, 2008

Certified by .....

James C. Preisig
Associate Scientist, Woods Hole Oceanographic Institution
Thesis Supervisor

Accepted by .....

David E. Hardt
Chairman, Mechanical Engineering Departmental Committee on Graduate Students
Massachusetts Institute of Technology

Accepted by .....

James C. Preisig
Chair, Joint Committee for Applied Ocean Science and Engineering
Massachusetts Institute of Technology/
Woods Hole Oceanographic Institution

Modeling and Frequency Tracking of Marine Mammal Whistle Calls

by

Jared Severson

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

Marine mammal whistle calls present an attractive medium for covert underwater communications. High quality models of the whistle calls are needed in order to synthesize natural-sounding whistles with embedded information. Since the whistle calls are composed of frequency modulated harmonic tones, they are best modeled as a weighted superposition of harmonically related sinusoids. Previous research with bottlenose dolphin whistle calls has produced synthetic whistles that sound too "clean" for use in a covert communications system. Due to the sensitivity of the human auditory system, watermarking schemes that slightly modify the fundamental frequency contour have good potential for producing natural-sounding whistles embedded with retrievable watermarks. Structured total least squares is used with linear prediction analysis to track the time-varying fundamental frequency and harmonic amplitude contours throughout a whistle call. Simulation and experimental results demonstrate the capability to accurately model bottlenose dolphin whistle calls and retrieve embedded information from watermarked synthetic whistle calls. Different fundamental frequency watermarking schemes are proposed based on their ability to produce natural sounding synthetic whistles and yield suitable watermark detection and retrieval.

Thesis Supervisor: James C. Preisig
Title: Associate Scientist

Acknowledgments

This thesis started as a bold idea without the guarantee of success, which is typical of all true research. Jim Preisig was uniquely suited to work with me on the project, and from the start, we were both captivated by its potential. Jim's focus and persistence in developing a thorough understanding of the frequency estimation problem were essential to applying the recent advances in structured total least squares to the well-known but rarely-used Prony's method. When it comes to running a complex underwater acoustic communications experiment, Jim's proficiency cannot be beat.

I want to thank Art Baggeroer for the support he has given me and previous Navy Joint Program students throughout the years. His suggestions were crucial for developing my theoretical understanding of the frequency estimation problem.

Milica Stojanovic brought a burst of energy in the final stages of the work, at the expense of her own research. Her enthusiasm helped turn a potential crisis into the promising watermarking scheme involving continuous phase modulation.

Stacy DeRuiter and Laela Sayigh helped me to understand marine mammal whistle calls from a biological perspective, and gave me access to high quality recordings that were necessary for good frequency estimation.

Marsha Gomes has been extremely helpful answering all of my administrative questions and looking after the general well-being of the Joint Program students.

Finally, I thank my Lord and Savior, Jesus Christ: by whom, through whom, and for whom are all things; who has called me by name unto a holy calling according to His own purpose and grace; who, while strengthening me to complete this project, has enabled me to lead others into a deeper trust in Him; and who testifies, "Behold, I come quickly, and my reward is with me, to give every man according as his work shall be."

To the God of unbroken promises

&

To you who will not be withheld

Contents

List of Figures

Chapter 1

Introduction

Due to the challenges of the underwater ocean environment, minimal progress has been made in the field of covert underwater acoustic communications. In some applications, low data rates would be an acceptable tradeoff for a sufficiently low probability of detection. Current robust underwater acoustic communication systems rely on a relatively high Signal-to-Noise Ratio (SNR) that precludes a covert posture. Marine biologics provide a significant source of background noise that any underwater acoustic communications system needs to overcome. However, if the communications scheme was able to mimic marine biologics in their natural environment, a covert posture may be retained while operating at a relatively high SNR.

Marine mammal whistle calls are an attractive medium for masking underwater acoustic communications due to their low frequency range, relatively sustained duration and regular harmonic structure. High-quality synthetic models are needed to effectively mimic marine mammal whistle calls with an embedded information signal. This thesis focuses on developing techniques for processing and embedding information in bottlenose dolphin whistle calls, but the techniques derived are applicable to other harmonically-structured tonal signals, including other marine mammal whistle calls.

1.1 Prior Work

1.1.1 Classification of Bottlenose Dolphin Whistle Calls

Christian [9] compiled a database of bottlenose dolphin whistle calls for his research on using generic signal compression for the identification, characterization and repetition detection of various signals. His approach estimated the fundamental frequency contour of a whistle call, recorded with a nominal 50 kHz sample rate, using 512 point blocks with no overlap, as higher resolution was not considered necessary. He compared the periodogram and Burg's autoregressive (AR) methods of spectral estimation, and concluded that the periodogram provided sufficient resolution of the fundamental frequency when compared to the computationally expensive Burg technique. Five major spectral peaks from each block were retained from which a tracking algorithm resolved the fundamental frequency contour. A 16-dimension coding space was then developed using the fundamental frequency contour to generate a dictionary of unique whistles. Single dolphins were found to reproduce their signature whistles very precisely, and were estimated to be capable of producing over a billion unique whistles.

1.1.2 Prior Models of Bottlenose Dolphin Whistle Calls

Although some methods used in human speech analysis and synthesis have been tested on marine mammals [3, 46], Buck *et al.* [5, 24] have been behind the effort to model bottlenose dolphin whistle calls for synthesis and modification purposes. A parametric model that can synthesize natural-sounding whistles can be used to study how dolphins communicate by modifying the whistle frequency contour and observing the response of dolphins.

Weighted Superposition of Sinusoidal Harmonics

Buck *et al.* [5] initially proposed a whistle model characterized as the weighted superposition of harmonically related sinusoids,

$$s[n] = \sum_{r=1}^{R} a_r[n] \sin(2\pi\phi_r[n]) \quad , \tag{1.1}$$

which embodies their typical description as frequency-modulated tonal calls. The fundamental frequency contour is extracted using a peak-picking algorithm detailed in [6], which was found to work well for recordings of individual animals at high SNR. The signal is broken into short blocks for which it is assumed to be relatively constant in amplitude and frequency. Frequency and energy contours for each harmonic are constructed from analyzing each block. Different modification strategies are proposed that modify different characteristics of the frequency and energy contours. Finally, whistles are synthesized at the original sample rate by interpolating phase and amplitude contours from the compressed frequency and energy contours. This technique differed from other speech processing algorithms [2, 46, 64] primarily in that discrete-time upsampling was performed instead of linear or polynomial interpolation between blocks. Example whistles recorded at 81.92 kHz were synthesized using a block length of 512 samples with 50% overlap. Human testing could distinguish between the original and unmodified synthetic whistles using quarter-speed in-air playbacks. The synthetic whistles were characterized as "clean sounding" and "not enough noise" when compared to the original whistles.

Autoregressive Model

Based on the distinct perceptual differences between original dolphin whistles and their synthetic counterparts produced with the sinusoidal model, Buck's student Huang [24] proposed using an AR synthesis model to generate more natural-sounding synthetic whistles. The whistle was broken into blocks of length 512 samples with

75% overlap, which are smoothly recombined during synthesis using a half-amplitude Hamming window. Each block was then modeled using a high order ($p = 60$) AR model. It was noted that the signal residue power spectrum contained a noticeable component of the original frequency contour. For each block, the resulting system poles were compared to the frequency contour used in the sinusoidal model for selecting signal poles corresponding to each harmonic. The whistles are then synthesized by driving the corresponding all-pole filter for each block with the signal residue for unmodified whistles and a white noise residue for modified whistles. While the AR synthesis whistles sounded more "natural" than the cleaner sinusoidal synthesis whistles, a study has not been performed to assess the overall quality of the AR synthesis whistles. Some problems encountered were the high computational load and the need to choose algorithm parameters such as block length, amount of overlap and AR system individually for each dolphin whistle.

1.1.3 Related Work in Human Speech Processing

Generally, human speech processing has focused on a stochastic model for speech production that seeks to design filters that imitate the physical dynamics of speech [15, 41]. These filters are then driven by combinations of two basic forms of excitation, periodic impulses for voiced speech and white noise for unvoiced speech. Linear prediction analysis is usually used to design all-pole filters that describe short blocks of similar speech patterns. Cepstral analysis was developed to separate the impulse response of the vocal system model from the excitation sequence, but its application is limited based on its computational complexity.

The basic sinusoidal superposition model in Eq. (1.1) used by Buck *et al.* [5] has been researched in human speech processing with excellent results. Serra and Smith [64] note that additive synthesis algorithms were among the first techniques used in computer-based synthesis, with the introduction of the heterodyne filter in the early 1970's, followed by the digital phase vocoder. McAulay and Quatieri [46, 53]

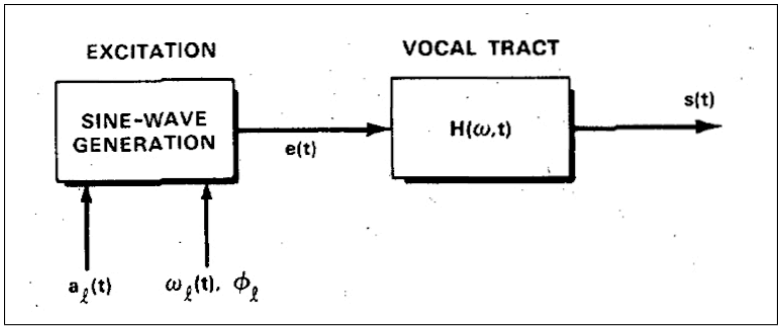

Figure 1-1: Quatieri and McAulay's speech production model [53]

and Smith and Serra [65] developed similar algorithms at about the same time that addressed inharmonic and pitch-changing sounds. Essentially, each algorithm used the same sinusoidal model while developing new methods to track relevant frequency contours and smoothly vary amplitude and phase from block to block. The signal was broken into analysis blocks, with overlap ranging from 50% to 75%, and relevant frequencies selected based on peaks in the discrete Fourier transform. McAulay and Quatieri included a time-varying filter model of the vocal tract at the output of the sinusoidal representation, as seen in Fig. 1-1. For a variety of sounds, including some whale sounds, their algorithm was reported to produce synthetic signals "essentially perceptually indistinguishable" from the original signal. Serra and Smith [64] updated their algorithm to better incorporate noise-like aspects of speech by removing the sinusoidal representation from the original signal and then applying stochastic modeling to the residual, but found that combining the sinusoidal and stochastic components sometimes produced undesirable results. The deterministic plus stochastic model was refined by Levine [36] by further decomposing the stochastic component into a quasi-stationary "noise" part and a rapidly changing "transient" part, resulting in a coding scheme that is both efficient and expressive [38].

1.2 Introduction to Information Hiding

The field of information hiding [11, 27] has largely grown out of the field of cryptography to include the additional aspect of keeping the existence of the information secret. A lot of the techniques that are used in information hiding draw upon the experience gained from cryptography, and in many cases the lines between the two are blurred, since any cryptographic system would be more robust to attack if its very existence was a secret. However, the practical wisdom of cryptography teaches that sensitive information should also be protected by a secret key, to safeguard against the information hiding techniques being discovered [50]. In general, information hiding techniques can be divided into four categories, which either include or exclude the separate principles of steganography and watermarking based on their application [11].

1.2.1 Steganography

Steganography is the art of concealed communication, in which the very existence of a message is secret [11]. Most applications of steganography follow the same general principle [26] described as follows. Alice, who wants to share a secret message m with Bob, randomly chooses a harmless message c, called *cover-object*, which can be transmitted to Bob without raising suspicion. With the potential use of a secret key k, a *stego-object* s is generated by embedding m into c in a careful way so that a third party cannot detect the existence of a secret in the apparently harmless message s. Alice then transmits s to Bob over an insecure channel, hoping that Wendy, a suspicious person with access to s, will not notice the embedded message. Bob can reconstruct m, since he knows the embedding method used by Alice and has access to the key k used in the embedding process. The extraction of m from s should be possible without access to the original cover c. In a "perfect" system, a normal cover should be indistinguishable from a stego-object, either by a human

or computer looking for a statistical pattern. There are basically three types of steganographic protocols that differ based on the choice of k. Pure steganography does not incorporate the prior exchange of secret information, so a key is not used in the embedding process. Secret key and public key steganography bolster security by using a secret or public key in the embedding process, although both use a secret key to reconstruct the secret message [26].

1.2.2 Watermarking

Watermarking, while closely related to steganography, is based on different underlying philosophies, requirements, and applications that result in techniques that clearly distinguish themselves from steganography. Essentially, the purpose of a watermark is to embed self-identifying information within a cover-object that can be used for copyright protection or tracking purposes. While the existence of a watermark does not normally need to be kept secret, the watermark should be permanently attached to the cover-object. Thus, watermarking has the notion of being robust to both malicious and benign attacks to remove the identifying information. In practical commercial applications, the watermark should be perceptually transparent enough to not annoy consumers or reduce the value of the product [32].

1.2.3 Applicable Digital Audio Watermarking Techniques

Watermarking of digital audio signals is more challenging compared to watermarking image or video sequences due to the wide dynamic range of the human auditory system (HAS). The HAS perceives sounds over a range of power greater than $10^9 : 1$ and a range of frequencies greater than $10^3 : 1$. In particular, the HAS has a high sensitivity to additive white Gaussian noise, which can be detected as low as 80 dB below ambient level in a sound file. However, there are some "holes" available in which to place a watermark. While the HAS has a wide dynamic range, it has a small differential range, meaning loud sounds generally tend to mask out quiet

sounds. Additionally, the HAS is insensitive to a constant relative phase shift in a stationary audio signal. Finally, some environmental distortions are so common that they are ignored by the listener in most cases [4, 12].

Due to the sensitivity of the HAS, digital audio watermarking techniques apply directly to steganographic applications, since on a perceptual basis the existence of an embedded message needs to be kept a secret. In a covert communications scenario, the robustness against intentional attacks is not usually required, although signal processing modifications, channel-induced signal distortion and additive ambient noise should not prevent retrieval of the watermark. In these applications, the watermark is expected to achieve a higher data rate and use blind detection schemes for watermark detection and reconstruction [12].

Fig. 1-2 shows a basic model depicting watermarking as a communications process, as described by He and Scordilis [23]. A secret message is modulated into a watermark waveform using a secret key. The watermark is embedded imperceptibly into a host signal to form the stego-signal. Transmission through a channel adds noise and distortion to the stego-signal. The watermark detector reconstructs the watermark from the received signal using the secret key, and in some cases, the host signal. Blind detection, in which the host signal is not available, is more flexible in operation, but lowers the achievable data rate by making detection more complex.

In the underwater channel, the primary sources of distortion are multipath arrivals and Doppler spreading [29, 51]. In order to combat these effects and maintain the fidelity of the stego-signal, the best watermarking scheme appears to be based on slight modifications of the fundamental frequency contour that result in natural-sounding stego-signals. Liu [38] has focused on a parametric approach to digital audio watermarking that is heavily based on the sinusoidal synthesis model and the work of Smith, Serra and Levine [36, 64]. Fig. 1-3 shows the watermarking scheme based on parametric analysis and synthesis proposed by Liu [38]. To embed a binary watermark W, the host signal is first decomposed into $\mathbf{s} = \mathbf{s}_{|\theta\rangle} + \mathbf{r}$, where $\mathbf{s}_{|\theta\rangle}$ is

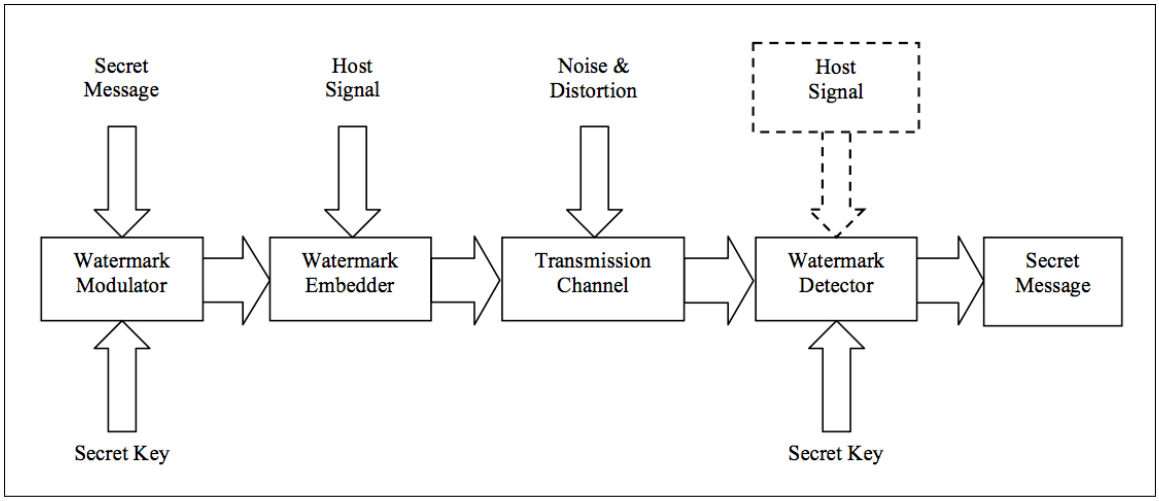

Figure 1-2: Communication model for watermarking [23]

perfectly parameterized and $\mathbf{r}$ is a residual orthogonal to $\mathbf{s}_{|\theta\rangle}$. Then, the parameter set $|\theta\rangle$ is modified to $|\theta^*\rangle$ to carry the watermark W. The new signal $\mathbf{s}_{|\theta^*\rangle}$, constructed from the watermarked parameter set, is combined with $\mathbf{r}$ to form the stego-signal $\mathbf{x}$ which is transmitted through a channel with unknown noise and distortion. Upon the reception of a corrupted copy $\mathbf{y}$, parameters are estimated so as to decode W. The attempt at watermarking is successful if the estimated parameters $|\hat{\theta}\rangle$ are close enough to $|\theta^*\rangle$ such that the decoded binary message $\hat{W}$ is identical to W. There is an inherent tradeoff when determining how θ is modified to θ^*: the modification should be small enough to not introduce perceptible distortion, but it should also be as big as possible to maximize robustness against attacks. In the case of a digital audio signal, the parametric component $\mathbf{s}_{|\theta\rangle}$ matches the sinusoidal model perfectly and receives the watermark, while the stochastic component $\mathbf{r}$ is removed during watermarking but then added back in for transmission to minimize perceptible alteration from the host signal $\mathbf{s}$.

Chen and Wornell [8] designed a class of digital watermarking techniques called quantization index modulation (QIM) that were shown to reach or nearly reach embedding rate capacity for important classes of models. However, this simplest form

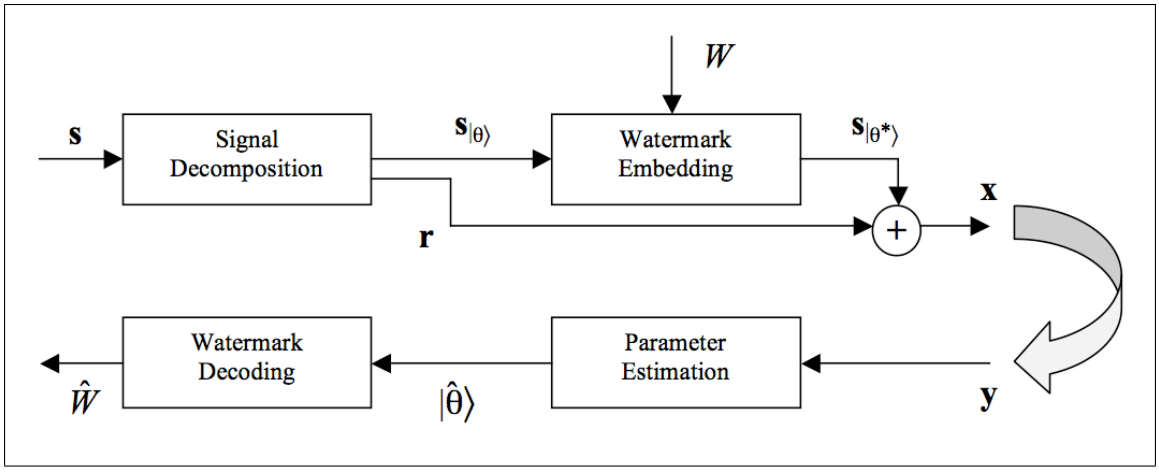

Figure 1-3: Parametric watermarking scheme [38]

of QIM was not robust to amplitude scaling, which is a common operation in music processing. Liu is currently working on the development of a F-QIM watermarking scheme that applies QIM techniques to the frequency parameters in the sinusoidal synthesis model [38].

Krishnan *et al.* have proposed a watermarking scheme based on joint time frequency analysis of the audio signal [30]. Most of the other watermarking techniques analyze audio in either the time or frequency domains separately, which does address the nonstationarity of audio signals. Krishnan *et al.* calculate the instantaneous mean frequency (IMF) of the audio signal using the Wigner-Ville distribution (WVD). The WVD is a time frequency distribution that gives a clear picture of the instantaneous frequency and group delay of a signal, but suffers from confusing artifacts when the signals are multicomponent [10]. The IMF for short blocks of the signal is determined, and then a spread spectrum watermarking scheme is implemented; to recover the watermark the IMF for the original signal is needed. Krishnan *et al.* also propose a chirp based spread spectrum watermarking scheme that reduces the complexity of watermark detection relative to the IMF scheme. The detector extracts the watermarking bits and uses the WVD and a chirp detection algorithm to decode the watermark [30].

1.3 Objectives

This thesis proposes a new approach for determining the parameters of the sinusoidal superposition model of Eq. (1.1) to represent recorded marine mammal whistle calls. To achieve high quality results, the recordings are assumed to consist solely of tonal whistle calls at high SNR produced by a single animal, without contamination by high frequency clicks. A new method for tracking the nonlinear fluctuations in a whistle call's fundamental frequency contour is developed based on the structured total least squares method. Amplitude contours for each harmonic are then determined using the estimated fundamental frequency contour and Prony's method. Different methods of watermarking the fundamental frequency contour are examined in terms of human imperceptibility and complexity of watermark reconstruction in the underwater environment. Experimental data is presented demonstrating the ability to track a whistle's fundamental frequency contour in an underwater communications scenario. In summary, the ability to communicate at low data rates using a natural-sounding synthetic marine mammal whistle call is demonstrated.

1.4 Organization

The remainder of this thesis consists of five chapters. Chapter 2 develops the progression of linear prediction techniques to model exponentially damped sinusoidal data. Chapter 3 describes a new approach to estimate the frequency and amplitude contours of chirp signals. Simulation results demonstrate the performance of the new approach, and other frequency estimation methods are compared to the structured total least squares method. Chapter 4 applies the results of Chapter 3 to building synthetic bottlenose dolphin whistle calls and examines different approaches to watermarking synthetic whistles. Chapter 5 presents data from a shallow water ocean experiment testing watermarked chirps and synthetic whistle calls. Finally, Chapter 6 closes with conclusions and indicates future directions for research.

Chapter 2

Sinusoidal Modeling Using Linear Prediction

The term *linear prediction* as a method for time series analysis dates back to Wiener in 1949 [41, 74]. Since then, it has been widely applied in many fields for the modeling, parameterization, prediction, and control of dynamic systems and signals [42], and has been used in speech analysis and synthesis since 1966 [41]. Generally, the work focuses on discrete stochastic models of autoregressive (AR) systems whose value at any point in time is a linear combination of a finite number of past samples plus additive noise. Signals are parameterized in the linear prediction or autoregressive coefficients, and can then be synthesized by driving a corresponding all-pole filter with white noise [15, 21, 42]. Spectral estimation is performed by fitting an AR model to the data's autocorrelation sequence and transforming into the frequency domain. Although it is not a spectral estimation technique, Prony's method has a close relationship to the least squares linear prediction algorithms used for AR parameter estimation. In contrast to AR methods that seek to fit a random model to the second order statistics, the modern version of Prony's method seeks to fit a deterministic exponentially damped sinusoidal model to the data [43]. Based on the sustained tonal characteristic of a marine mammal whistle call, applying a deterministic sinusoidal

model is an intuitive starting point for estimating whistle call frequency contours.

2.1 Prony's Method

Gaspard Riche, Baron de Prony's paper [14, 43] proposed in 1795 a method for exactly fitting damped exponentials to available data points for his research on the expansion of various gases. The modern form of Prony's method generalizes to identifying the amplitudes A_k, damping factors α_k, sinusoidal frequencies f_k, and initial phases θ_k of a linear combination of complex exponentials,

$$x[n] = \sum_{k=1}^{p} A_k \exp[(\alpha_k + j2\pi f_k)(n-1)T + j\theta_k] \tag{2.1}$$

for $1 \leq n \leq p$, where T is the sample interval. In the case of real data, the complex exponentials must occur in complex conjugate pairs of equal amplitude, reducing Eq. (2.1) to

$$x[n] = \sum_{k=1}^{p} 2A_k \exp[(\alpha_k(n-1)T] \cos[2\pi f_k(n-1)T + \theta_k] \quad . \tag{2.2}$$

Eq. (2.1) can be written in the form

$$x[n] = \sum_{k=1}^{p} h_k z_k^{n-1} \quad , \tag{2.3}$$

where the complex constants h_k and z_k are defined as

$$h_k = A_k \exp(j\theta_k) \quad , \tag{2.4}$$

$$z_k = \exp[(\alpha_k + j2\pi f_k)T] \quad . \tag{2.5}$$

Expressing Eq. (2.3) in matrix form as a set of simultaneous equations for $1 \leq n \leq p$ results in

$$\begin{bmatrix} z_1^0 & z_2^0 & \dots & z_p^0 \\ z_1^1 & z_2^1 & \dots & z_p^1 \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{p-1} & z_2^{p-1} & \dots & z_p^{p-1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix} = \begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[p] \end{bmatrix} \quad . \tag{2.6}$$

Prony discovered a method to separately solve for the exponential z_k elements, from which Eq. (2.6) can then be solved for the vector of unknown constants h_k. Appendix A shows that Eq. (2.3) is the solution to a homogeneous constant-coefficient difference equation

$$\sum_{m=0}^{p} w[m]x[n-m] = 0 \quad , \tag{2.7}$$

where $w[m]$ are the coefficients of the polynomial $\phi(z)$ with roots z_k,

$$\phi(z) = \prod_{k=1}^{p}(z - z_k) = z^p + \sum_{m=1}^{p} w[m]z^{p-m} \quad . \tag{2.8}$$

The p equations for which Eq. (2.7) is valid, $p+1 \leq n \leq 2p$, can be expressed in matrix form as

$$\begin{bmatrix} x[p] & x[p-1] & \dots & x[1] \\ x[p+1] & x[p] & \dots & x[2] \\ \vdots & \vdots & \ddots & \vdots \\ x[2p-1] & x[2p-2] & \dots & x[p] \end{bmatrix} \begin{bmatrix} w[1] \\ w[2] \\ \vdots \\ w[p] \end{bmatrix} = - \begin{bmatrix} x[p+1] \\ x[p+2] \\ \vdots \\ x[2p] \end{bmatrix} . \tag{2.9}$$

Prony's method to fit p exponentials to $2p$ data points can be summarized in three steps. First, Eq. (2.9) is solved to determine the coefficients of the polynomial $\phi(z)$ in Eq. (2.8). Second, the roots z_k of $\phi(z)$ are calculated. Third, Eq. (2.6) is solved to determine the parameters h_k.

The desired parameters are then determined by the relationships

$$\alpha_k = \ln|z_k|/T \tag{2.10}$$

$$f_k = \tan^{-1}[\mathrm{Im}\{z_k\}/\,\mathrm{Re}\{z_k\}]/2\pi T \tag{2.11}$$

$$A_k = |h_k| \tag{2.12}$$

$$\theta_k = \tan^{-1}[\mathrm{Im}\{h_k\}/\,\mathrm{Re}\{h_k\}] \quad . \tag{2.13}$$

2.2 Least Squares Prony Method

In practical situations, the presence of some noise in the data sequence prevents obtaining an exact exponential fit to the data, so the number of data points N usually exceeds the $2p$ data points used in the original Prony method. In this overdetermined case, the data is approximated as an exponential sequence,

$$\hat{x}[n] = \sum_{k=1}^{p} h_k z_k^{n-1} \quad , \tag{2.14}$$

for $1 \le n \le N$, with observation error $\epsilon[n] = x[n] - \hat{x}[n]$. Applying standard linear least squares (LS) procedures [19] to the original Prony method results in the three-step LS Prony method. First, forming the linear prediction relation

$$\mathbf{A}\mathbf{w} \approx \mathbf{b} \quad , \tag{2.15}$$

$$\mathbf{A} = \begin{bmatrix} x[p] & x[p-1] & \dots & x[1] \\ x[p+1] & x[p] & \dots & x[2] \\ \vdots & \vdots & \ddots & \vdots \\ x[N-1] & x[N-2] & \dots & x[N-p] \end{bmatrix}, \ \mathbf{w} = \begin{bmatrix} w[1] \\ w[2] \\ \vdots \\ w[p] \end{bmatrix}, \text{ and } \mathbf{b} = - \begin{bmatrix} x[p+1] \\ x[p+2] \\ \vdots \\ x[N] \end{bmatrix},$$

the LS solution is given by

$$\mathbf{w_{LS}} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{b} \quad . \tag{2.16}$$

Second, the roots z_k of $\phi(z)$ in Eq. (2.8) are calculated. Third, the LS solution for the parameters h_k is given by

$$\mathbf{h_{LS}} = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H\mathbf{x} \quad , \tag{2.17}$$

where

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_p \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \dots & z_p^{N-1} \end{bmatrix}, \ \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}, \text{ and } \mathbf{x} = \begin{bmatrix} x[1] \\ x[2] \\ \vdots \\ x[N] \end{bmatrix}.$$

Unfortunately, the LS Prony method doesn't perform well in the presence of significant additive noise because it assumes the data matrix $\mathbf{A}$ is error free and models the observation error in $\mathbf{b}$ as white noise. Different methods that have been used to improve the performance of the Prony method include employing high prediction orders and reduced rank approximations of the data matrix via singular value decomposition (SVD) [31, 43, 68, 69]. The higher prediction order improves the estimation of signal parameters by adding extra exponentials to model the additive noise. The poles z_k related to the true signal exponentials cluster closer to their correct values, while the extraneous poles fluctuate widely to account for the noise. The noise contribution to the data matrix $\mathbf{A}$ can be reduced by using its reduced rank approximation

$$\mathbf{A}_K = \mathbf{U}_K\mathbf{\Sigma}_K\mathbf{V}_K^H \tag{2.18}$$

composed of the largest K singular values and singular vectors of $\mathbf{A}$, where K is the number of signal exponentials, and

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \tag{2.19}$$

where

$$\mathbf{U} = [\mathbf{u}_1,\ \ldots,\ \mathbf{u}_{N-p}], \quad \mathbf{u}_i\ \in\ \mathbb{R}^{N-p},$$
$$\mathbf{V} = [\mathbf{v}_1,\ \ldots,\ \mathbf{v}_p], \quad \mathbf{v}_i\ \in\ \mathbb{R}^{p},$$
$$\mathbf{\Sigma} = \mathrm{diag}(\sigma_1,\ \ldots,\ \sigma_p), \quad \sigma_1 \geq \cdots \geq \sigma_{\min(N-p,\,p)}\ ,$$

is the SVD of $\mathbf{A}$ with $\mathbf{U}^H\mathbf{U} = \mathbf{I}_{N-p}$ and $\mathbf{V}^H\mathbf{V} = \mathbf{I}_p$. The principle eigenvector (PE) method developed by Tufts and Kumaresan [68, 69] uses both a high prediction order and the reduced rank approximation of Eq. (2.18) to improve Prony's method in the presence of noise. More recent work has applied a modified LS Prony method to the frequency estimation problem [25, 39, 66].

2.3 Total Least Squares Approach

In the classical LS problem of Eq. (2.15), there is an underlying assumption [18] that all of the errors are confined to the vector $\mathbf{b}$, i.e., that the data matrix $\mathbf{A}$ has no errors. Since both $\mathbf{A}$ and $\mathbf{b}$ contain values from the data sequence $x[n]$ for $1 \leq n \leq N$, errors in $\mathbf{b}$ will also appear in $\mathbf{A}$. The total least squares (TLS) method [18, 73] compensates for error in both $\mathbf{A}$ and $\mathbf{b}$, and should be expected to give a better solution than Eq. (2.15).

2.3.1 Solution to the Total Least Squares Problem

A good way to motivate the TLS method is to state the ordinary LS problem as follows:

$$\underset{\mathbf{\Delta b}\ \in\ \mathbb{R}^{N-p}}{\text{minimize}}\ \|\mathbf{\Delta b}\|_2 \qquad (2.20)$$
$$\text{subject to } \mathbf{b} + \mathbf{\Delta b} \in \mathrm{Range}(\mathbf{A})$$

where $\|\cdot\|_2$ denotes the l_2 norm given by

$$\|\mathbf{\Delta b}\|_2 = \sqrt{\sum_i {\Delta b_i}^2} \quad . \tag{2.21}$$

The LS problem amounts to perturbing the observation $\mathbf{b}$ by a minimum amount $\mathbf{\Delta b}$ so $\mathbf{b} + \mathbf{\Delta b}$ can be predicted by the columns of $\mathbf{A}$. The TLS problem accounts for perturbation in both $\mathbf{b}$ and $\mathbf{A}$, i.e.,

$$(\mathbf{A} + \mathbf{\Delta A})\, \mathbf{w} = \mathbf{b} + \mathbf{\Delta b} \quad , \tag{2.22}$$

or expressing Eq. (2.22) in a different form,

$$\left(\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} + \begin{bmatrix} \mathbf{\Delta A} & \mathbf{\Delta b} \end{bmatrix} \right) \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix} = 0$$

or

$$(\mathbf{C} + \mathbf{\Delta C})\, \mathbf{z} = 0 \tag{2.23}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}, \quad \mathbf{\Delta C} = \begin{bmatrix} \mathbf{\Delta A} & \mathbf{\Delta b} \end{bmatrix}, \quad \text{and} \quad \mathbf{z} = \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix}.$$

The TLS problem seeks to

$$\underset{\mathbf{\Delta C} \,\in\, \mathbb{R}^{(N-p)\times(p+1)}}{\text{minimize}} \|\mathbf{\Delta C}\|_F \tag{2.24}$$
$$\text{subject to } (\mathbf{b} + \mathbf{\Delta b}) \in \text{Range}(\mathbf{A} + \mathbf{\Delta A})$$

where $\|\cdot\|_F$ denotes the Frobenius norm given by

$$\|\mathbf{\Delta C}\|_F = \sqrt{\sum_{i,j} |\Delta c_{ij}|^2} \quad . \tag{2.25}$$

Eq. (2.23) shows that the TLS problem involves finding a perturbation matrix $\mathbf{\Delta C} \in \mathbb{R}^{(N-p)\times(p+1)}$ having minimum norm such that $\mathbf{C} + \mathbf{\Delta C}$ is rank deficient. The SVD can be used for this purpose. Let

$$\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \tag{2.26}$$

where

$$\begin{aligned}
&\mathbf{U} = [\mathbf{u}_1,\ \ldots,\ \mathbf{u}_{N-p}], \quad \mathbf{u}_i \ \in\ \mathbb{R}^{N-p},\\
&\mathbf{V} = [\mathbf{v}_1,\ \ldots,\ \mathbf{v}_{p+1}], \quad \mathbf{v}_i \ \in\ \mathbb{R}^{p+1},\\
&\mathbf{\Sigma} = \mathrm{diag}(\sigma_1,\ \ldots,\ \sigma_{p+1}), \quad \sigma_1 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_{p+1} \geq 0,
\end{aligned}$$

be the SVD of $\mathbf{C}$ with $\mathbf{U}^H\mathbf{U} = \mathbf{I}_{N-p}$ and $\mathbf{V}^H\mathbf{V} = \mathbf{I}_{p+1}$. It is assumed here that the problem is overdetermined, i.e., $N > 2p$. Two cases arise in the TLS solution. In the first case, when $\sigma_p > \sigma_{p+1}$, a unique solution exists. The solution can be thought of as finding a matrix $(\mathbf{C} + \mathbf{\Delta C})$ of rank p that satisfies Eq. (2.24). A reduced rank approximation to

$$\mathbf{C} = \sum_{i=1}^{p+1} \sigma_i \mathbf{u}_i \mathbf{v}_i^H \tag{2.27}$$

is obtained by removing one or more σ_i terms from Eq. (2.27). The smallest perturbation $\mathbf{\Delta C}$ that reduces the rank of $\mathbf{C}$ by one is

$$\mathbf{\Delta C} = -\sigma_{p+1}\mathbf{u}_{p+1}\mathbf{v}_{p+1}^H \quad . \tag{2.28}$$

Inserting Eq. (2.28) into Eq. (2.23) yields $\mathbf{z} = \alpha \mathbf{v}_{p+1}$, since $\mathbf{v}_{p+1}$ is now in the nullspace of

$$(\mathbf{C} + \mathbf{\Delta C}) = \sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{v}_i^H \quad . \tag{2.29}$$

Thus, provided $(\mathbf{v}_{p+1})_{p+1} \neq 0$, the TLS solution is given by

$$\mathbf{w_{TLS}} = \frac{-1}{(\mathbf{v}_{p+1})_{p+1}} \begin{bmatrix} (\mathbf{v}_{p+1})_1 \\ \vdots \\ (\mathbf{v}_{p+1})_p \end{bmatrix} \quad . \tag{2.30}$$

The TLS solution does not exist if $(\mathbf{v}_{p+1})_{p+1} = 0$, but this doesn't commonly arise in engineering applications. In the second case, when $\sigma_p = \sigma_{p+1}$, a solution may still exist, but it is not unique. However, a unique solution is chosen in the sense of minimum norm [18, 73].

An alternative expression for the TLS solution $\mathbf{w_{TLS}}$ in Eq. (2.30) can be derived as follows.

$$\begin{aligned} \mathbf{C}^H \mathbf{C} \mathbf{v}_{p+1} &= (\mathbf{V \Sigma U}^H)(\mathbf{U \Sigma V}^H)\mathbf{v}_{p+1} \\ &= (\mathbf{V \Sigma}^2 \mathbf{V}^H)\mathbf{v}_{p+1} \\ &= \sigma_{p+1}^2 \mathbf{v}_{p+1} \quad . \end{aligned} \tag{2.31}$$

Inserting $\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ and $\mathbf{v}_{p+1} = \begin{bmatrix} \mathbf{v}'_{p+1} \\ (\mathbf{v}_{p+1})_{p+1} \end{bmatrix}$ into Eq. (2.31) gives the expression:

$$\begin{bmatrix} \mathbf{A}^H \mathbf{A} & \mathbf{A}^H \mathbf{b} \\ \mathbf{b}^H \mathbf{A} & \mathbf{b}^H \mathbf{b} \end{bmatrix} \begin{bmatrix} \mathbf{v}'_{p+1} \\ (\mathbf{v}_{p+1})_{p+1} \end{bmatrix} = \sigma_{p+1}^2 \begin{bmatrix} \mathbf{v}'_{p+1} \\ (\mathbf{v}_{p+1})_{p+1} \end{bmatrix} . \tag{2.32}$$

Expanding Eq. (2.32) gives the set of equations,

$$(\mathbf{A}^H\mathbf{A} - \sigma_{p+1}^2\mathbf{I}_p)\mathbf{v}_{p+1}^{'} + (\mathbf{v}_{p+1})_{p+1}\mathbf{A}^H\mathbf{b} = 0 \tag{2.33}$$

$$\mathbf{b}^H\mathbf{A}\mathbf{v}_{p+1}^{'} + (\mathbf{b}^H\mathbf{b} - \sigma_{p+1}^2)(\mathbf{v}_{p+1})_{p+1} = 0 \quad . \tag{2.34}$$

But if $(\mathbf{v}_{p+1})_{p+1} \neq 0$, $\mathbf{w_{TLS}} = \frac{-\mathbf{v}_{p+1}^{'}}{(\mathbf{v}_{p+1})_{p+1}}$ so Eq. (2.33) reduces to

$$(\mathbf{A}^H\mathbf{A} - \sigma_{p+1}^2\mathbf{I}_p)\mathbf{w_{TLS}} = \mathbf{A}^H\mathbf{b} \quad . \tag{2.35}$$

If $(\mathbf{A}^H\mathbf{A} - \sigma_{p+1}^2\mathbf{I}_p)$ is invertible, the alternative expression for the TLS solution is

$$\mathbf{w_{TLS}} = (\mathbf{A}^H\mathbf{A} - \sigma_{p+1}^2\mathbf{I}_p)^{-1}\mathbf{A}^H\mathbf{b} \quad . \tag{2.36}$$

2.3.2 Prony's Method and Total Least Squares

The TLS solution $\mathbf{w_{TLS}}$ is the maximum likelihood (ML) estimate for Eq. (2.15) only if the errors in $\mathbf{C} = \begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}$ are independently and identically normally distributed with common covariance matrix proportional to the identity matrix with zero mean [35, 73]. Due to the Toeplitz structure of the matrix $\mathbf{A}$, the errors are not independently distributed, so the TLS solution is not optimum. However, the TLS solution does tend to reduce the effects of noise in the linear prediction formulation, and provides improvements over the LS solution. Rahman and Yu [56] applied the TLS method to the linear prediction frequency estimation problem and demonstrated better performance than the LS-based principal eigenvector (PE) method [69] for the same prediction order. The TLS method yielded the greatest improvement relative to the PE method at minimal prediction orders, although both solutions improve with higher prediction orders. As the prediction order is increased, additional correlated errors are added to the matrix $\mathbf{C}$, reducing the benefit of the TLS method. At maximal prediction order, with $p = \frac{N}{2}$ for even N, both the TLS and PE solutions

converge to the same performance.

The matrix $\mathbf{Z}$ in Eq. (2.17) used in the third step of the Prony Method for determining the parmeters h_k has a Vandermonde structure [19]. Assuming that relatively good estimates are available for the system poles z_k, the major source of error will be in the observation vector $\mathbf{x}$. Thus, the LS solution of Eq. (2.17) appropriately accounts for errors in the model.

2.4 Structured Total Least Squares Approach

Structured Total Least Squares (STLS) is a natural extension to the TLS approach when the same observations occur in multiple rows of the matrix $\mathbf{C}$ in Eq. (2.23). In order to find an ML estimate of $\mathbf{w}$, $\begin{bmatrix}\mathbf{\Delta A} & \mathbf{\Delta b}\end{bmatrix}$ needs to have the same structure as $\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}$ [1]. This leads to the following formulation of the STLS problem [35]:

$$\begin{aligned}&\underset{\mathbf{\Delta A},\mathbf{\Delta b},\mathbf{w}}{\text{minimize}} \left\| \begin{bmatrix}\mathbf{\Delta A} & \mathbf{\Delta b}\end{bmatrix} \right\|_X \\ &\text{such that} (\mathbf{A}+\mathbf{\Delta A})\mathbf{w} = (\mathbf{b}+\mathbf{\Delta b}), \\ &\text{and } \begin{bmatrix}\mathbf{\Delta A} & \mathbf{\Delta b}\end{bmatrix} \text{ has the same structure as } \begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix},\end{aligned} \tag{2.37}$$

where $\|\cdot\|_X$ denotes the l_2 norm defined on the unique entries of $\begin{bmatrix}\mathbf{\Delta A} & \mathbf{\Delta b}\end{bmatrix}$. Many different formulations have been proposed for the STLS problem involving linearly structured matrices: the Constrained Total Least Squares (CTLS) approach [1], the Structured Total Least Norm (STLN) approach [60, 72], and the Riemannian Singular Value Decomposition (RiSVD) approach [13]. Each approach uses iterative numerical algorithms to find the solution, but all of them suffer from inherent multiple local minima that occur in the STLS problem [34]. When the noise level is relatively low, the STLS problem is not difficult to solve, and simple starting values will suffice. However, when STLS is used for its rank reducing properties and there is not a solution nearby in an l_2 norm sense, the starting values need to be chosen with care.

2.4.1 STLS Solution for Hankel/Toeplitz Matrices

The linear prediction relation of Eq. (2.15) can be written with a Hankel structure by reordering the columns of matrix $\mathbf{A}$ and reversing $\mathbf{w}$:

$$\mathbf{A}\mathbf{w} \approx \mathbf{b} \quad , \tag{2.38}$$

$$\mathbf{A} = \begin{bmatrix} x[1] & x[2] & \dots & x[p] \\ x[2] & x[3] & \dots & x[p+1] \\ \vdots & \vdots & \ddots & \vdots \\ x[N-p] & x[N-p+1] & \dots & x[N-1] \end{bmatrix},$$

$$\mathbf{w} = \begin{bmatrix} w[p] \\ w[p-1] \\ \vdots \\ w[1] \end{bmatrix}, \text{ and } \mathbf{b} = -\begin{bmatrix} x[p+1] \\ x[p+2] \\ \vdots \\ x[N] \end{bmatrix},$$

so that $\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ has a Hankel structure. The solution $\mathbf{w}$ is then reversed for determining the poles z_k in Step 2 of the Prony method. The Hankel STLS problem can be solved using the Hankel STLN formulation:

$$\begin{aligned} &\underset{\mathbf{\Delta x},\, \mathbf{w}}{\text{minimize}} \sum_{n=1}^{N} (\Delta x[n])^2 \\ &\text{such that } (\mathbf{A} + \mathbf{\Delta A})\mathbf{w} = (\mathbf{b} + \mathbf{\Delta b}), \\ &\text{and } \begin{bmatrix} \mathbf{\Delta A} & \mathbf{\Delta b} \end{bmatrix} \text{ has a Hankel structure,} \end{aligned} \tag{2.39}$$

where $\Delta x[n]$ for $1 \leq n \leq N$ are the unique entries of the Hankel matrix $\begin{bmatrix} \mathbf{\Delta A} & \mathbf{\Delta b} \end{bmatrix}$. The STLN approach solves the STLS problem as a nonlinear optimization problem with nonlinear constraints [60, 72]. Lemmerling and van Huffel [35] propose the following STLN algorithm for solving Eq. (2.39):

STLN Algorithm

Input: extended Hankel data matrix $[\mathbf{A}\ \mathbf{b}] \in \mathbb{R}^{m\times(n+1)}$ $(m > n)$ of full rank $n+1$ and identity weighting matrix $\mathbf{I}_{m+n}$

Output: the parameter vector $\mathbf{w} \in \mathbb{R}^{n\times 1}$ and vector $\boldsymbol{\Delta x} \in \mathbb{R}^{(m+n)\times 1}$ composed of the unique entries of the matrix $[\boldsymbol{\Delta A}\ \boldsymbol{\Delta b}]$

Step 1: Initialize $\boldsymbol{\Delta x}$, $\mathbf{w}$, and Lagrange multiplier vector $\boldsymbol{\gamma} \in \mathbb{R}^{m\times 1}$

Step 2: While *stop criterion* not satisfied

Step 2.1: Solve the following system of equations:

$$\begin{bmatrix} \mathbf{S} & \mathbf{J}^T \\ \mathbf{J} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta}\tilde{\mathbf{x}} \\ \boldsymbol{\Delta}\tilde{\mathbf{w}} \\ \boldsymbol{\Delta}\tilde{\boldsymbol{\gamma}} \end{bmatrix} = - \begin{bmatrix} \mathbf{g} + \mathbf{J}^T\boldsymbol{\gamma} \\ \mathbf{r}(\boldsymbol{\Delta x}, \mathbf{w}) \end{bmatrix}$$

Step 2.2: $\boldsymbol{\Delta x} \leftarrow \boldsymbol{\Delta x} + \boldsymbol{\Delta}\tilde{\mathbf{x}}$

$\mathbf{w} \leftarrow \mathbf{w} + \boldsymbol{\Delta}\tilde{\mathbf{w}}$

$\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma} + \boldsymbol{\Delta}\tilde{\boldsymbol{\gamma}}$

End

where $\mathbf{S} = \begin{bmatrix} \mathbf{I}_{m+n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(m+2n)\times(m+2n)}$, $\mathbf{J} = \begin{bmatrix} \mathbf{W} & \mathbf{A} + \boldsymbol{\Delta A} \end{bmatrix} \in \mathbb{R}^{m\times(m+2n)}$ is the Jacobian of the constraints $\mathbf{r}(\boldsymbol{\Delta x}, \mathbf{w})$ in Eq. (2.39),

$$\mathbf{r}(\boldsymbol{\Delta x}, \mathbf{w}) = (\mathbf{A} + \boldsymbol{\Delta A})\mathbf{w} - (\mathbf{b} + \boldsymbol{\Delta b}) \quad ,$$

$\mathbf{g} = \begin{bmatrix} \mathbf{I}_{m+n} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta x} \\ \boldsymbol{\Delta w} \end{bmatrix} \in \mathbb{R}^{(m+2n)\times 1}$ is the gradient of the objective function in Eq. (2.39), and $\mathbf{W} \in \mathbb{R}^{m\times(m+n)}$ is defined by

$$\mathbf{W}\boldsymbol{\Delta x} = [\boldsymbol{\Delta A}\ \boldsymbol{\Delta b}] \begin{bmatrix} \mathbf{w} \\ -1 \end{bmatrix} \quad ,$$

which for the Hankel-structured matrix $[\mathbf{\Delta A}\ \mathbf{\Delta b}]$ has the form

$$\mathbf{W} = \begin{bmatrix} w[p] & \dots & w[1] & -1 & 0 & \dots & \dots & 0 \\ 0 & w[p] & \dots & w[1] & -1 & 0 & & \vdots \\ \vdots & & \ddots & & \ddots & \ddots & & \vdots \\ 0 & & & \ddots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & w[p] & \dots & w[1] & -1 \end{bmatrix} .$$

The *stop criterion*, chosen based on the application, tests for convergence of the STLN algorithm. With each iteration, the algorithm updates parameter estimates for $\mathbf{\Delta x}$ and $\mathbf{w}$ in an attempt to drive the constraint $\mathbf{r}(\mathbf{\Delta x}, \mathbf{w})$ to zero. If the iterative solution approaches close to one of many local minima, the algorithm will not converge to the actual STLS solution. The system of equations in Step 2.1 is solved using the LDL^T factorization of the matrix $\begin{bmatrix} \mathbf{S} & \mathbf{J}^T \\ \mathbf{J} & \mathbf{0} \end{bmatrix}$.

A natural choice for the initialization parameters in the STLN algorithm would be to set $\mathbf{\Delta x}_{initial} = \mathbf{0}$, $\boldsymbol{\gamma} = \mathbf{0}$, and $\mathbf{w}_{initial} = \mathbf{w_{LS}}$ or $\mathbf{w_{TLS}}$. It turns out that $\mathbf{w}_{initial} = \mathbf{w_{LS}}$ is the better choice, but both take a large number of iterations for the STLN algorithm to converge to a solution that is often a local minima. Lemmerling *et al.* [34] propose a better initialization procedure based on the Hankel Total Least Squares (HTLS) subspace algorithm developed for Nuclear Magnetic Resonance (NMR) data fitting [71]. The HTLS algorithm is suboptimal in the sense that while it gives a good estimate of the solution, it is not the closest rank-deficient Hankel matrix to $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$. The STLN algorithm is then initialized close to the global solution for the STLS problem using the values $\mathbf{\Delta x}_{initial} = \mathbf{\Delta x_{HTLS}}$, $\boldsymbol{\gamma} = \mathbf{0}$, and $\mathbf{w}_{initial} = \mathbf{w_{HTLS}}$.

HTLS Algorithm Description

The HTLS algorithm [70] is based on the fact that Eq. (2.14) can be modeled by an autonomous linear state-space model of order p,

$$\begin{aligned} \mathbf{y}[n+1] &= \mathbf{B}\mathbf{y}[n] \\ x[n] &= \mathbf{h}^T\mathbf{y}[n] + \epsilon[n], \end{aligned} \tag{2.40}$$

where $\mathbf{y}[n]$ is a complex state vector, $\mathbf{h}^T$ is a complex row vector, and $x[n]$ are noisy observations with observation error $\epsilon[n] = x[n] - \hat{x}[n]$. Equating Eq. (2.14) and Eq. (2.40) for $1 \leq n \leq N$ yields

$$\hat{x}[n] = \mathbf{h}^T\mathbf{B}^{n-1}\mathbf{y}[1] = \sum_{k=1}^{p} h_k z_k^{n-1}, \tag{2.41}$$

where $\hat{x}[n]$ has zero observation error, and defines

$$\mathbf{B} = \begin{bmatrix} z_1 & 0 & \dots & 0 \\ 0 & z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & z_p \end{bmatrix}, \quad \mathbf{y}[1] = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}.$$

Essentially, the modern Prony method is described in a state-space model which is used to estimate the parameters z_k and h_k. A Hankel matrix $\mathbf{H} \in \mathbb{R}^{(L \times M)}$, as square as possible for best parameter accuracy [71], such that $L = M(+1) \approx N/2$, is formed using the N data points,

$$\mathbf{H} = \begin{bmatrix} x[1] & x[2] & \dots & x[M] \\ x[2] & x[3] & \dots & x[M+1] \\ \vdots & \vdots & \ddots & \vdots \\ x[L] & x[L+1] & \dots & x[N] \end{bmatrix}.$$

If the observation error $\epsilon[n]$ is zero, $\mathbf{H}$ decomposes into an observability matrix $\mathcal{O}$ and a controllability matrix $\mathcal{C}$ given by:

$$\mathbf{H} = \mathcal{O}\mathcal{C} = \begin{bmatrix} \mathbf{h}^T \\ \mathbf{h}^T\mathbf{B} \\ \vdots \\ \mathbf{h}^T\mathbf{B}^{L-1} \end{bmatrix} \begin{bmatrix} \mathbf{y}[1] & \mathbf{B}\mathbf{y}[1] & \dots & \mathbf{B}^{M-1}\mathbf{y}[1] \end{bmatrix}. \tag{2.42}$$

In practice, the observation error in Eq. (2.41) is non-zero. $\mathbf{H}_p$, the SVD reduced-rank approximation of $\mathbf{H}$, is computed as

$$\mathbf{H}_p = \mathbf{U}_p\mathbf{\Sigma}_p\mathbf{V}_p^H \quad , \tag{2.43}$$

where

$$\begin{aligned} &\mathbf{H}_{L\times M} = \mathbf{U}_{L\times L}\mathbf{\Sigma}_{L\times M}\mathbf{V}^H_{M\times M} \quad , \\ &\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \dots, \sigma_{\min(L,M)}), \quad \sigma_1 \geq \cdots \geq \sigma_{\min(L,M)}, \end{aligned}$$

and $\mathbf{U}_p$, $\mathbf{\Sigma}_p$, and $\mathbf{V}_p$ are the first p columns of $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}$. $\mathbf{H}_p$ is used to estimate $\mathcal{O}$ and $\mathcal{C}$ up to a similarity transformation matrix $\mathbf{S}$,

$$\mathbf{H}_p = \hat{\mathbf{U}}_p\hat{\mathbf{V}}_p^H \approx (\mathcal{O}\mathbf{S}^{-1})(\mathbf{S}\mathcal{C}) \quad , \tag{2.44}$$

where $\hat{\mathbf{U}}_p = \mathbf{U}_p$ and $\hat{\mathbf{V}}_p = \mathbf{V}_p\mathbf{\Sigma}_p$ if *unbalanced* splitting is used, and $\hat{\mathbf{U}}_p = \mathbf{U}_p\mathbf{\Sigma}_p^{\frac{1}{2}}$ and $\hat{\mathbf{V}}_p = \mathbf{V}_p\mathbf{\Sigma}_p^{\frac{1}{2}}$ if *balanced* splitting is used. Substituting $\mathbf{B} = \mathbf{S}^{-1}\mathbf{Q}\mathbf{S}$ into Eq. (2.44), where $\mathbf{Q} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$ has the same eigenvalues as $\mathbf{B}$, yields:

$$\mathbf{H}_p = \hat{\mathbf{U}}_p \hat{\mathbf{V}}_p^H \approx \begin{bmatrix} \mathbf{h}^T\mathbf{S}^{-1} \\ \mathbf{h}^T\mathbf{S}^{-1}\mathbf{Q} \\ \vdots \\ \mathbf{h}^T\mathbf{S}^{-1}\mathbf{Q}^{L-1} \end{bmatrix} \begin{bmatrix} \mathbf{Sy}[1] & \mathbf{QSy}[1] & \dots & \mathbf{Q}^{M-1}\mathbf{Sy}[1] \end{bmatrix} \quad . \tag{2.45}$$

The TLS solution $\mathbf{Q_{TLS}}$ is computed for the incompatible set

$$\underline{\hat{\mathbf{U}}}_p \mathbf{Q} \approx \overline{\hat{\mathbf{U}}}_p \quad , \tag{2.46}$$

where $\underline{\hat{\mathbf{U}}}_p$ and $\overline{\hat{\mathbf{U}}}_p$ are derived from $\hat{\mathbf{U}}_p$ by omitting its first and last row,

$$\underline{\hat{\mathbf{U}}}_p = \begin{bmatrix} \mathbf{h}^T\mathbf{S}^{-1}\mathbf{Q} \\ \mathbf{h}^T\mathbf{S}^{-1}\mathbf{Q}^2 \\ \vdots \\ \mathbf{h}^T\mathbf{S}^{-1}\mathbf{Q}^{L-1} \end{bmatrix} \quad \text{and} \quad \overline{\hat{\mathbf{U}}}_p = \begin{bmatrix} \mathbf{h}^T\mathbf{S}^{-1} \\ \mathbf{h}^T\mathbf{S}^{-1}\mathbf{Q} \\ \vdots \\ \mathbf{h}^T\mathbf{S}^{-1}\mathbf{Q}^{L-2} \end{bmatrix} \quad .$$

Provided $\tilde{\mathbf{V}}_{22}$ is non-singular, the TLS solution is given by

$$\mathbf{Q_{TLS}} = -\tilde{\mathbf{V}}_{12}(\tilde{\mathbf{V}}_{22})^{-1} \quad , \tag{2.47}$$

in which $\tilde{\mathbf{V}}_{12}$ and $\tilde{\mathbf{V}}_{22}$ are obtained from the SVD of the augmented matrix

$$\begin{bmatrix} \underline{\hat{\mathbf{U}}}_p & \overline{\hat{\mathbf{U}}}_p \end{bmatrix} = \tilde{\mathbf{U}}_{(L-1)\times(L-1)} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{V}}^H_{2p\times 2p} \quad , \tag{2.48}$$

where

$$\tilde{\mathbf{V}} = \begin{bmatrix} \tilde{\mathbf{V}}_{11} & \tilde{\mathbf{V}}_{12} \\ \tilde{\mathbf{V}}_{21} & \tilde{\mathbf{V}}_{22} \end{bmatrix} \begin{matrix} p \\ p \end{matrix} \quad . \\ \begin{matrix} p & p \end{matrix}$$

If $\tilde{\mathbf{V}}_{22}$ is close-to singular in Eq. (2.47), it is replaced by its pseudo-inverse $\tilde{\mathbf{V}}_{22}^{\dagger}$. The system pole estimates $\hat{z}_k$ are equal to the eigenvalues of $\mathbf{Q_{TLS}}$. It is not necessary to find the similarity transformation matrix $\mathbf{S}$. Finally the parameter estimates $\hat{h}_k$ are obtained by inserting the pole estimates $\hat{z}_k$ into Eq. (2.17),

$$\mathbf{h_{LS}} = (\mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H\mathbf{x} \quad . \tag{2.49}$$

STLS Initialization using HTLS

Once the estimates $\hat{z}_k$ and $\hat{h}_k$ are obtained using the HTLS Algorithm with unbalanced splitting in Eq. (2.44), the resulting adjusted data values are calculated as

$$(x[n] + \Delta x_{HTLS}[n]) = \sum_{k=1}^{p} \hat{h}_k \hat{z}_k^{n-1} \quad , \tag{2.50}$$

from which the initial values for $\mathbf{\Delta A_{HTLS}}$, $\mathbf{\Delta b_{HTLS}}$, and $\mathbf{w_{HTLS}}$ in Eq. (2.39) are found.

HTLS Algorithm

Input: extended Hankel data matrix $[\mathbf{A}\ \mathbf{b}] \in \mathbb{R}^{m\times(n+1)}$ $(m > n)$ of full rank $n+1$

Output: extended Hankel noise data matrix $[\mathbf{\Delta A_{HTLS}}\ \mathbf{\Delta b_{HTLS}}]$ and parameter vector $\mathbf{w_{HTLS}}$, such that $[\mathbf{A} + \mathbf{\Delta A_{HTLS}}\ \mathbf{b} + \mathbf{\Delta b_{HTLS}}]$ is a rank-deficient Hankel matrix.

Step 1: $\mathbf{y} \leftarrow [\mathbf{A}(:,1)^T\ \mathbf{A}(m,2:n)\ \mathbf{b}(m)]^T$

Step 2: $M \leftarrow ceil((m+n)/2)$

Step 3: $\mathbf{H} \leftarrow hankel(\mathbf{y}(1 : m+n-M+1),\ \mathbf{y}(m+n-M+1 : m+n))$

Step 4: Calculate the left singular vectors $\mathbf{U}(:, i),\ i = 1, \ldots, n$ of $\mathbf{H}$,
corresponding to the n largest singular values

Step 5: Calculate the TLS solution of the system
$\mathbf{U}(2 : M, 1 : n)\mathbf{Q} \approx \mathbf{U}(1 : M-1, 1 : n)$.
The eigenvalues of $\mathbf{Q}$ are the estimated signal poles $\hat{z}_l,\ l = 1, \ldots, n$

Step 6: Solve the complex amplitudes $\hat{h}_l,\ l = 1, \ldots, n$, from the system of equations:
$y(k) \approx \sum_{l=1}^{n} \hat{h}_l \hat{z}_l^k,\ k = 1, \ldots, m+n$

Step 7: $\hat{y}(k) \leftarrow \sum_{l=1}^{n} \hat{h}_l \hat{z}_l^k,\ k = 1, \ldots, m+n$

Step 8: $[\mathbf{\Delta A_{HTLS}}\ \mathbf{\Delta b_{HTLS}}] \leftarrow hankel(\hat{\mathbf{y}}(1 : m),\ \hat{\mathbf{y}}(m : m+n)) - [\mathbf{A}\ \mathbf{b}]$

Step 9: Solve the square system
$(\mathbf{A}(1 : n,\ 1 : n)\ \mathbf{\Delta A}(1 : n,\ 1 : n))\mathbf{w_{HTLS}} = \mathbf{b}(1 : n) + \mathbf{\Delta b}(1 : n)$

The STLN algorithm is then initialized using $\mathbf{\Delta x_{HTLS}}$ and $\mathbf{w_{HTLS}}$. The improved initialization procedure enhances both the solution quality and calculation time by starting the iterative search routine close to the global minimum for the Hankel STLS problem [34]. Lemmerling *et al.* [33] demonstrated the improved accuracy of the STLN algorithm using HTLS parameter initialization in a speech compression application. Even with the improved HTLS initialization procedure, the computational load of the STLN algorithm is large compared to standard speech coding algorithms. Various methods have been used to produce faster STLS algorithms [44], but current algorithms are are still too slow for real-time application.

Chapter 3

Simulation Results

As described in Chapter 1, the different techniques for modeling acoustic signals based on the sinusoidal superposition model of Eq. (1.1) differ primarily in the method by which the interpolation of amplitude and phase contours is performed between analysis blocks. Frequency estimation is generally performed by taking the Fourier transform (DFT) of a windowed block of data of length N samples, with $N = 512$ being common in practice, although some algorithms adaptively vary N. Different windowing functions are used to provide better spectral peak estimation performance. The data sample advance between analysis blocks, known as the hop size H, is usually chosen to have some overlap between blocks to produce smoother results across time at the expense of higher computational loading [64]. Choosing $H = 1$ is generally not used since parameters are assumed to be slowly-varying and accumulation of excess data is not desirable [17]. In most applications, data storage is an important design criteria, and while optimal synthesis quality is desired, some amount of signal compression is acceptable.

In the case of modeling marine mammal whistle calls, computational loading and signal compression is not a design criteria in generating high-quality synthesis models. Since the frequency contours of a marine mammal whistle call vary with time, a method of closely tracking the frequency contour is desired to improve the synthesis

quality. This is achieved by using a hop size of $H = 1$ and reducing the effective window size by applying the parametric approach of linear prediction to estimate the instantaneous frequency. Based on the harmonic structure of marine mammal whistle calls, estimation of the fundamental frequency contour should provide adequate estimates of higher harmonics, as assumed in [5]. In a communications scenario, good frequency tracking performance is desired even at relatively low SNR to ensure capability of reconstructing the embedded watermark.

The rest of this chapter proceeds as follows. The algorithm for tracking the fundamental frequency contour and amplitude contours using weighted STLS and Prony's method is described. Simulation results are presented for tracking frequency contours of chirp signals with constant amplitude, and are compared to other frequency estimation methods. Finally, simulation results are presented for tracking both the frequency and amplitude of a chirp signal with variable amplitude harmonics.

3.1 Algorithm Description

The algorithm applies a sliding block window of size M samples to a harmonically structured whistle recording $s[n]$, where $p = 2R$ is the model order, R is the number of harmonics in $s[n]$, and $M - p$ is the number of linear prediction equations used to estimate the AR parameters of $s[n]$. Thus, $s[n]$ is modeled as

$$s[n] = \sum_{r=1}^{R} a_r[n] \cos\left(2\pi\phi_r[n] + \theta_r\right) + v[n], \quad \text{for} \quad 1 \leq n \leq N, \tag{3.1}$$

or explicitly writing each exponential component,

$$s[n] = \sum_{r=1}^{R} \frac{a_r[n]}{2} \exp\left(j\theta_r\right)\left[\exp\left(j2\pi\phi_r[n]\right) + \exp\left(-j2\pi\phi_r[n]\right)\right] + v[n], \tag{3.2}$$

where $f_r[n]$ is the instantaneous frequency of the rth harmonic at time n such that

$$\phi_r[n] = \sum_{i=1}^{n} f_r[i]/f_s \quad , \tag{3.3}$$

$a_r[n]$ is the amplitude of the rth harmonic at time n, θ_r is the initial phase of the rth harmonic, f_s is the sample rate, and $v[n]$ is additive ambient noise. The lth analysis block, using a hop size of $H = 1$, is expressed as

$$\begin{aligned} & x_l[m] = W[m]s[m + l - 1], \\ & \text{for} \quad 1 \leq l \leq L = N - M + 1, \\ & \text{and} \quad 1 \leq m \leq M, \end{aligned} \tag{3.4}$$

where $W[m]$, discussed on page 51, is a window of length M applied to the data. Setting up the first step of Prony's method using the Hankel structure in Eq. (2.38) gives

$$\mathbf{A}_l \mathbf{w}_l \approx \mathbf{b}_l \quad , \tag{3.5}$$

$$\mathbf{A}_l = \begin{bmatrix} x_l[1] & x_l[2] & \dots & x_l[p] \\ x_l[2] & x_l[3] & \dots & x_l[p+1] \\ \vdots & \vdots & \ddots & \vdots \\ x_l[M-p] & x_l[M-p+1] & \dots & x_l[M-1] \end{bmatrix},$$

$$\mathbf{w}_l = \begin{bmatrix} w_l[p] \\ w_l[p-1] \\ \vdots \\ w_l[1] \end{bmatrix}, \text{ and } \mathbf{b}_l = - \begin{bmatrix} x_l[p+1] \\ x_l[p+2] \\ \vdots \\ x_l[M] \end{bmatrix}.$$

Eq. (3.5) is solved using the STLS method if $v[n] \neq 0$, but in simulations where $v[n] = 0$, the LS method is sufficient. The system pole estimates $\hat{z}_{k,l}$ are then found

as the roots of the polynomial

$$\hat{\phi}_l(z) = z^p + \sum_{k=1}^{p} \hat{w}_l[k] z^{p-k} \quad , \tag{3.6}$$

keeping in mind that $\mathbf{w}_l$ is written in reverse order when $\mathbf{A}_l$ has a Hankel structure. In the presence of noise, the poles $\hat{z}_{k,l}$ fluctuate back and forth across the unit circle as the analysis block $\mathbf{x}_l$ moves through the data, giving a better frequency estimate than if the poles were constrained to be on the unit circle. However, the underlying model in Eq. (3.1) assumes that the original dolphin whistle has an undamped sinusoidal structure, so only the frequency component

$$\tilde{f}_{k,l} = \frac{f_s}{2\pi} \tan^{-1}\left(\frac{\mathrm{Im}\{\hat{z}_{k,l}\}}{\mathrm{Re}\{\hat{z}_{k,l}\}}\right) \tag{3.7}$$

is retained while scaling the pole estimates to the unit circle, i.e.,

$$\tilde{z}_{k,l} = \frac{\hat{z}_{k,l}}{|\hat{z}_{k,l}|} \quad . \tag{3.8}$$

In the STLN formulation [34], the HTLS algorithm is used to initialize the iterative search for the closest rank-deficient Hankel matrix $\begin{bmatrix}\mathbf{A}_l & \mathbf{b}_l\end{bmatrix}$. However, simulation results show that both the STLN [35] and extended structured least squares (ES-TLS) [75] algorithms do not improve upon the frequency estimate $\tilde{f}_{k,l}$ provided by the HTLS algorithm. Thus, the poles $\tilde{z}_{k,l}$ are found as the normalized eigenvalues of the matrix $\mathbf{Q}_{\mathbf{TLS},l}$ (Eq. (2.47)). The pole estimates $\tilde{z}_{k,l}$ are then used in Step 3 of the Prony method to calculate the parameters $\tilde{h}_{k,l}$ using Eq. (2.17),

$$\tilde{\mathbf{h}}_l = (\tilde{\mathbf{Z}}_l^H \tilde{\mathbf{Z}}_l)^{-1} \tilde{\mathbf{Z}}_l^H \mathbf{x}_l \quad , \tag{3.9}$$

where

$$\tilde{\mathbf{Z}}_l = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \tilde{z}_{1,l} & \tilde{z}_{2,l} & \dots & \tilde{z}_{p,l} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{z}_{1,l}^{M-1} & \tilde{z}_{2,l}^{M-1} & \dots & \tilde{z}_{p,l}^{M-1} \end{bmatrix}, \ \tilde{\mathbf{h}}_l = \begin{bmatrix} \tilde{h}_{1,l} \\ \tilde{h}_{2,l} \\ \vdots \\ \tilde{h}_{p,l} \end{bmatrix}, \text{ and } \mathbf{x}_l = \begin{bmatrix} x_l[1] \\ x_l[2] \\ \vdots \\ x_l[M] \end{bmatrix}.$$

The least squares estimate of the amplitude of the kth harmonic exponential is

$$\tilde{A}_{k,l} = |\tilde{h}_{k,l}| \quad , \tag{3.10}$$

meaning that for each analysis block, the amplitudes are chosen to minimize the residual mean square error (MSE) between the sinusoidal model and the observed data.

An important aspect of this approach is selecting the window $W[m]$ and measuring the corresponding estimation delay between the leading edge of the analysis window and the effective estimation point of the algorithm. Since there is not currently a recursive implementation of the STLS method, the type of window is restricted to a constant-length analysis of the data, known as a sliding window approach. In general, the window that is chosen is an exponential sliding window,

$$W[m] = \begin{cases} \lambda^{M-m} & 1 \leq m \leq M, \text{ where } 0 < \lambda \leq 1 \\ 0 & \text{elsewhere.} \end{cases} \tag{3.11}$$

If $\lambda = 1$, W is a rectangular window. For $0 < \lambda < 1$, the weights decay at an exponential rate, gradually decreasing the effect of old data on current parameter estimates, which is why λ is called the *forgetting factor* [22]. The resulting rectangular and exponential sliding window approaches using STLS are analogous to the sliding window least squares (SWLS) and exponentially weighted least squares (EWLS) approaches compared by Niedźwiecki [47]. For estimators with the same effective window length,

EWLS has better parameter tracking characteristics due to the window's higher degree of concentration at the leading edge of the window, while the rectangular SWLS has better parameter matching properties due to the linearity of its phase characteristic. Essentially, reducing the forgetting factor λ allows the algorithm to track fast parameter changes better, but lowers the estimation accuracy attainable when parameters are slowly-varying. In terms of AR modeling, the exponential window applies an artificial damping factor to the data in order to improve tracking performance, causing the corresponding system poles to shift to $\hat{z}_k \approx z_k/\lambda$. The linear prediction relation in Eq. (3.5) can also be applied in the backward direction with respect to time. For a sinusoid with poles on the unit circle, choosing $\lambda_f > 1$ in the forward direction scales the system poles within the unit circle and is the same as choosing $\lambda_b = 1/\lambda_f$ in the backward direction.

The effective sample estimation point $\hat{t}_e$ of the analysis window is the weighted time average of the window $W[m]$ for which a linear prediction equation is valid, i.e. $p+1 \leq m \leq M$,

$$\begin{aligned} \hat{t}_e &= \frac{\sum_{m=p+1}^{M} mW[m]}{\sum_{m=p+1}^{M} W[m]} \\ &= \frac{\sum_{m=p+1}^{M} m\lambda^{M-m}}{\sum_{m=p+1}^{M} \lambda^{M-m}} \quad . \end{aligned} \tag{3.12}$$

The corresponding sample estimation delay τ_e is

$$\tau_e = M - \hat{t}_e \quad , \tag{3.13}$$

and the effective window length is $l_{eff} \approx 2\tau_e$. Taking advantage of knowing the point in time $\hat{t}_{e,l}$ for which an estimate $\tilde{f}_{k,l}$ is valid, where

$$\hat{t}_{e,l} = M + l - 1 - \tau_e = \hat{t}_e + l - 1, \tag{3.14}$$

a more localized estimate of the amplitude contours in Eq. (3.9) can be made by contracting $\mathbf{x}_l$ about $\hat{t}_{e,l}$ and reducing the number of rows in $\tilde{\mathbf{Z}}_l$. The weighted average frequency for the lth analysis block,

$$\overline{f}_{k,l} = \frac{\sum_{m=p+1}^{M} f_k[m+l-1]\lambda^{M-m}}{\sum_{m=p+1}^{M} \lambda^{M-m}} \quad , \tag{3.15}$$

where $f_k[n]$ for $1 \leq n \leq N$ is the underlying frequency contour for the kth exponential, provides a measure of the smoothing effect of the sliding window. However, $\tilde{f}_{k,l}$ will usually track closer to $f_r[n]$ than $\hat{f}_{k,l}$ when the frequency contour changes faster than linearly.

3.2 Frequency Tracking of Chirp Signals

This section presents simulation results demonstrating the ability to track the frequency of harmonic chirp signals in the presence of white noise, and comparison is made with other frequency estimation methods. The simulated chirp whistles are constructed according to Eq. (3.1) and Eq. (3.3) with $a_r[n] = 1$ for all n, $\theta_r = 0$, $N = 500$ samples, $f_s = 100$ kHz, $v[n]$ is additive white gaussian noise with variance σ_v^2 such that SNR = 5 dB unless specified otherwise, and $f_r[n]$ is specified for each chirp. Unless otherwise specified, the algorithm parameters are chosen as $\lambda = 1$, $M = 101$, and $p = 2R$, with the chirp having R harmonics. In the following figures, f_{HTLS} represents the positive frequency estimate $\hat{f}_k$ of f_r obtained using the HTLS algorithm and f_{AVG} is the weighted average frequency for each analysis block, $\overline{f}_k$.

3.2.1 Single Harmonic Linear Chirp

Fig. 3-1 demonstrates the frequency estimation and tracking performance of the HTLS algorithm for a linear chirp with $f_1[n] = 8000 + 2(n-1)$ (Hz) for $1 \leq n \leq N$. The resulting frequency estimate is essentially unbiased, which can be seen graphically

after adjusting for the estimation delay, where $\tau_e = 49$ samples in this example.

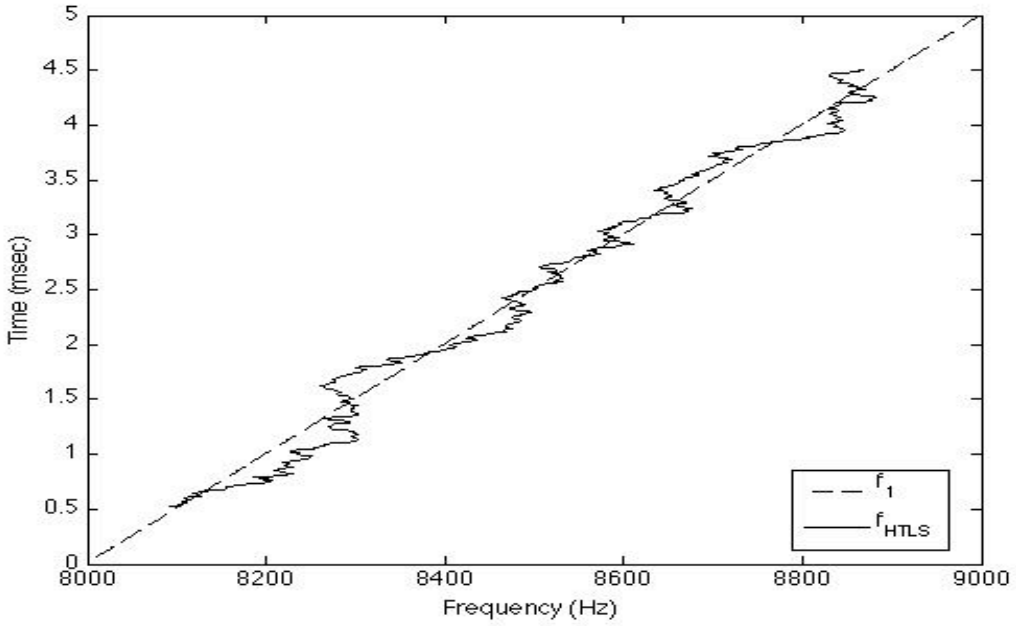

Figure 3-1: HTLS frequency tracking performance for a linear chirp (SNR = 5 dB)

3.2.2 Double Harmonic Linear Chirp

Fig. 3-2 demonstrates the frequency estimation and tracking performance of the HTLS algorithm for a linear chirp with two harmonics ($R = 2$), $f_1[n] = 8000 + 2(n-1)$ (Hz) and $f_2[n] = 16000 + 4(n-1)$ (Hz) for $1 \leq n \leq N$.

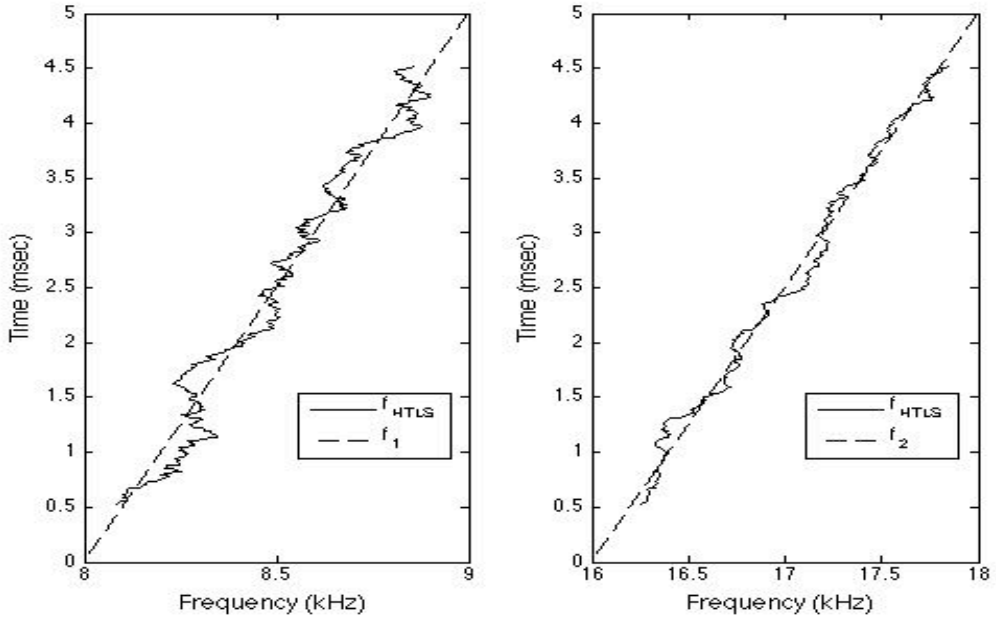

Figure 3-2: HTLS frequency tracking performance for a linear chirp with two harmonics (SNR = 5 dB)

3.2.3 Single Harmonic Linear Chirp with Abrupt Frequency Shifts

Fig. 3-3 demonstrates the frequency estimation and tracking performance of the HTLS algorithm for a linear chirp with an abrupt frequency shift of 250 Hz,

$$f_1[n] = \begin{cases} 8000 + 2.5(n-1) & \text{for } 1 \leq n \leq 250, \\ 7750 + 2.5(n-1) & \text{for } 251 \leq n \leq 500. \end{cases} \tag{3.16}$$

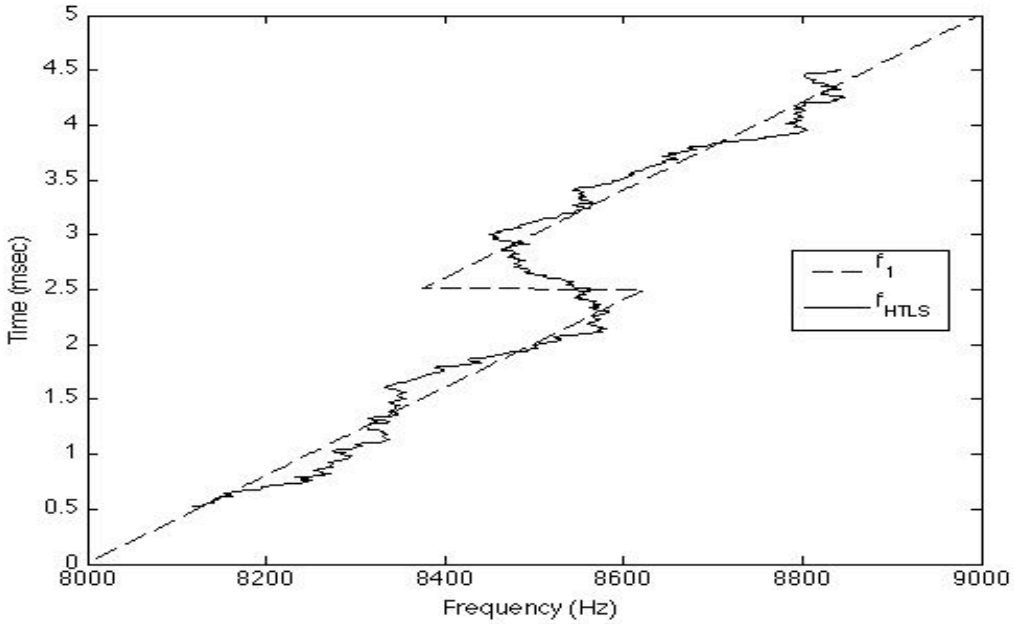

Figure 3-3: HTLS frequency tracking performance for a linear chirp with abrupt frequency shifts (SNR = 5 dB)

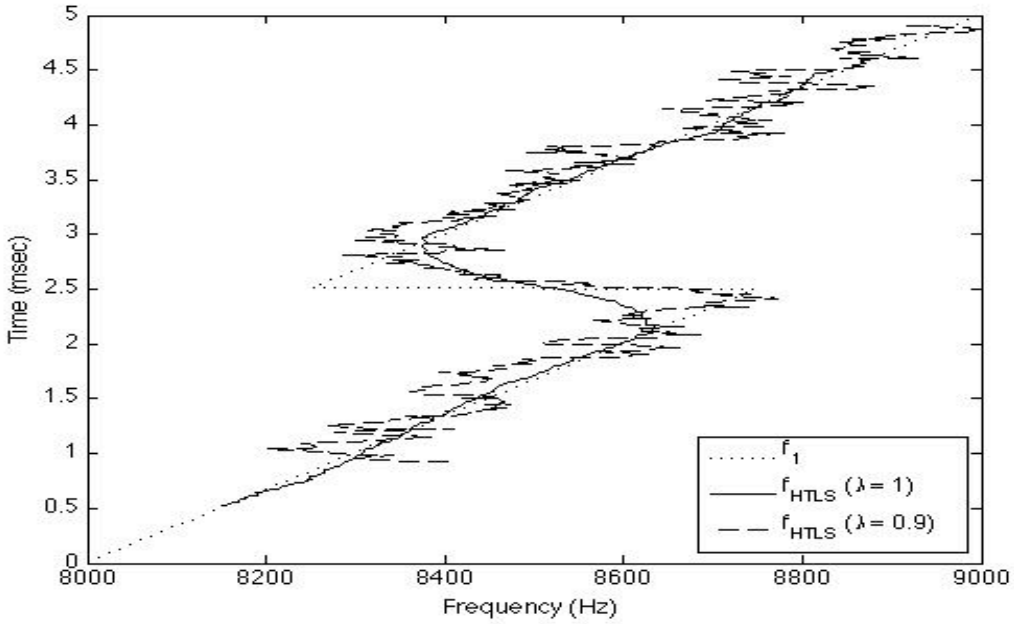

Figure 3-4: HTLS frequency tracking performance vs. λ (SNR = 15 dB)

Fig. 3-4 shows how the tracking performance of the HTLS algorithm is improved by lowering the forgetting factor λ at the expense of estimation accuracy. To clearly demonstrate the tradeoff between tracking performance and estimation error, an SNR of 15 dB and a frequency shift of 500 Hz are simulated, where

$$f_1[n] = \begin{cases} 8000 + 5(n-1) & \text{for } 1 \leq n \leq 250, \\ 7500 + 5(n-1) & \text{for } 251 \leq n \leq 500. \end{cases} \tag{3.17}$$

In the case where $\lambda = 0.9$, the transition between the linear chirp segments is much sharper than for $\lambda = 1$ due to the shorter effective window length. The corresponding estimation point $\hat{t}_e$ is closer to the leading edge of the analysis window, which shifts the frequency estimation region toward the end of the signal. The increased estimation error variance would preclude using $\lambda \neq 1$ for most frequency estimation problems, unless it was necessary to detect abrupt frequency shifts.

3.2.4 Single Harmonic Linear + Sinusoidal Chirp

Fig. 3-5 demonstrates the frequency estimation and tracking performance of the HTLS algorithm for a chirp with a combined linear and sinusoidal frequency contour, $f_1[n] = 8000 + 2(n-1) + 500\sin(\frac{\pi(n-1)}{100})$ (Hz) for $1 \leq n \leq 500$. The frequency estimation error becomes biased at peaks in the underlying frequency contour, $f_1[n]$, due to the smoothing effects of the analysis window. However, the frequency estimator tracks closer to $f_1[n]$ than the weighted average frequency for each analysis window. Thus, while peaks in the actual frequency contour are not fully resolved due to the estimation bias, the existence of peaks in the frequency contour can be detected by the HTLS algorithm with a sliding window. If needed, the actual peaks could be resolved with better accuracy by removing the smoothing effects of the analysis window by deconvolution. In regions where the frequency contour is close to linear, the HTLS frequency estimate is practically unbiased.

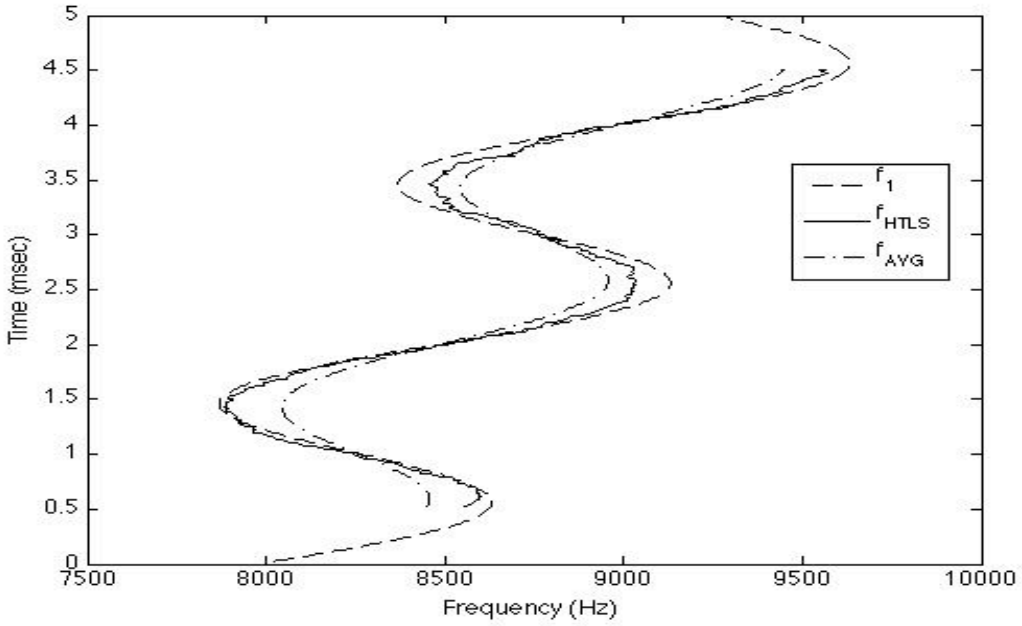

Figure 3-5: HTLS frequency tracking performance for sinusoidal chirp (SNR = 5 dB)

3.2.5 Comparison with Alternative Frequency Estimators

In [43], Marple discusses the important difference between spectral estimation, which attempts to match the spectrum of a signal over a continuous range of frequencies, and frequency estimation, which is only concerned with the behavior of the spectrum local to a specific frequency. Kay [28] reviews the sinusoidal parameter estimation problem, showing how the ML estimate of the frequency of a single complex sinusoid in complex additive white Gaussian noise is found by choosing the frequency at which the periodogram is maximized. The Cramer-Rao lower bound (CRLB) for the unbiased frequency estimator of a single complex exponential of the form

$$s[n] = A_1 \exp[j(2\pi f_1 n + \theta_1)] + v[n], \quad \text{for } 1 \leq n \leq N, \tag{3.18}$$

with unknown parameters A_1, f_1, and θ_1, and complex white Gaussian noise $v[n]$ with variance σ_v^2, was shown by Rife and Boorstyn [57] to be

$$\mathrm{var}(\hat{f}_1) \geq \frac{6\sigma_v^2}{A_1^2\,(2\pi)^2 N(N^2-1)} \quad . \tag{3.19}$$

For a single real sinusoid,

$$\begin{aligned} s[n] &= A_1 \cos(2\pi f_1 n + \theta_1) + v[n] \\ &= \frac{A_1}{2}\Big(\exp[j(2\pi f_1 n + \theta_1)] + \exp[-j(2\pi f_1 n + \theta_1)] \Big) + v[n], \end{aligned} \tag{3.20}$$

for $1 \leq n \leq N$, the frequency CRLB [58] is

$$\mathrm{var}(\hat{f}_1) \geq \frac{6\sigma_v^2}{A_1^2\,\pi^2 N(N^2-1)} \quad . \tag{3.21}$$

When estimating the unknown parameters of a single complex exponential linear chirp sequence, the CRLB of Eq. (3.19) applies to the center frequency of the analysis window [16]. Extending to real linear sinusoidal chirp signals, the CRLB of Eq. (3.21) also applies to the center frequency of the analysis window [58].

Quinn and Hannan [55] present different classes of frequency estimators that can be compared with the HTLS algorithm for linear chirp signals. Fig. 3-6 shows the performance of some of these frequency estimators compared to the CRLB for the linear chirp in Eq. (3.1), with $R = 1$, $a_1[n] = A_1 = 1$ and $f_1[n] = 8000 + (n-1)$ (Hz) for $1 \leq n \leq N$, $N = 1100$, $f_s = 100$ kHz, and $\theta_1 = 0$. The HTLS frequency estimate was computed using a rectangular window ($\lambda = 1$) of length $M = 101$ and a model order of $p = 2$. SNR is defined as

$$SNR = \frac{A_1^2}{2\sigma_v^2} \quad . \tag{3.22}$$

The MSE for each frequency estimator is computed as

$$MSE = \frac{1}{J\,L} \sum_{j=1}^{J} \sum_{l=1}^{L} \left(\frac{\hat{f}_{l,j} - \overline{f}_l}{f_s} \right)^2 \quad , \tag{3.23}$$

where $J = 5$ is the number of independent trials performed for each chirp, $L = 1000$ is the number of frequency estimates computed for each trial, and $\overline{f}_l$ is the center frequency of the lth analysis window for a rectangular window. Each of the frequency estimators applies the same sliding rectangular window to the data to obtain a frequency estimate $\hat{f}_{l,j}$ for each analysis window and trial. The corresponding CRLB is

$$\mathrm{var}(\hat{f}_l) \geq \frac{3}{SNR\,\pi^2 M(M^2-1)} \quad . \tag{3.24}$$

The FTI frequency estimator, using the FTI 3 algorithm of [55], performs an interpolation about the maximiser of the periodogram using three Fourier coefficients.

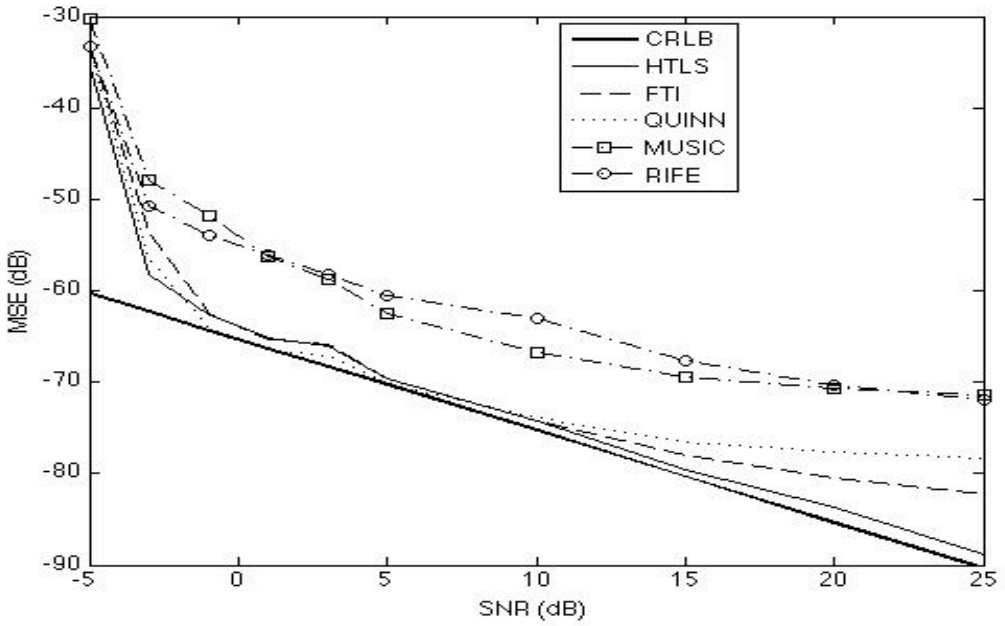

Figure 3-6: Linear chirp frequency estimator performance vs. CRLB

Macleod [40] has developed alternative techniques based on the same approach. The RIFE frequency estimator is an older approach by Rife and Vincent [59] based on quadratic interpolation of the moduli of Fourier coefficients to reduce data storage requirements. The QUINN frequency estimator is an AR-based iterative algorithm developed by Quinn and Fernandes [54]. The multiple signal characterization (MUSIC) frequency estimator developed by Schmidt [63] is based on eigenanalysis of the noise subspace.

Each of the frequency estimators in Fig. 3-6 was developed for quasi-stationary signals for which the frequency could be considered constant in each analysis window. Even though the estimators are used in an unconventional manner when analyzing linear chirps, they provide a baseline to gauge the performance of the HTLS algorithm. As the SNR increases above 10 dB, the HTLS algorithm increasingly outperforms the other frequency estimators and nearly achieves the CRLB for an unbiased estimator. Between 0 and 5 dB, the QUINN frequency estimator outperforms the HTLS and FTI estimators due to an inherent bias that worsens performance at higher SNR. The faster FTI frequency estimator achieves nearly the same performance and can be considered as an alternative to HTLS at lower SNR.

A lot of research has been done on joint ML frequency and chirp rate estimation of linear chirp signals with short data lengths. Djurić and Kay [16] proposed similar estimators based on their ease of on-line or off-line implementation that achieve the CRLB at SNR above 8 dB, with SNR defined as $(\frac{A_1^2}{\sigma_v^2})$ for a single complex sinusoid. Liang and Arun [37] use a method very similar to the HTLS algorithm with balanced splitting to initialize a search for the ML parameter estimates of multiple superimposed chirp signals, with simulation results attaining the CRLB at SNRs above 10 dB. Saha and Kay [61] propose using importance sampling to maximize a compressed likelihood based on frequency and chirp rate to implement joint ML parameter estimation of superimposed chirp signals, demonstrating simulation results that achieve the CRLB at SNRs above 3 dB. At low enough SNR, all of the frequency estimators

depart sharply from the CRLB, as seen in Fig. 3-6 below an SNR of 3 dB. Ultimately, Fig. 3-6 demonstrates that the HTLS algorithm can be used to nearly optimally track the frequencies of chirped signals.

3.3 Amplitude Estimation of Chirp Signals

This section presents simulation results demonstrating the ability of the Prony method to estimate the amplitudes of a double harmonic linear chirp signal based on frequency estimates obtained using the HTLS algorithm. As with Section 3.2, the simulated chirp whistle is constructed according to Eq. (3.1) and Eq. (3.3) with $N = 500$ samples, $f_s = 100$ kHz, $\theta_r = 0$, and $v[n]$ is white Gaussian noise with variance σ_v^2. The frequency and amplitude contours are defined as

$$f_r[n] = \begin{cases} 8000 + 2(n-1), & r = 1 \\ 16000 + 4(n-1), & r = 2 \end{cases} \tag{3.25}$$

and

$$a_r[n] = \begin{cases} \frac{1}{2}(1 + \text{tukey}[n]), & r = 1 \\ \frac{1}{4}(1 + \text{tukey}[n]), & r = 2 \end{cases} \tag{3.26}$$

for $1 \leq n \leq$, where tukey$[n]$ is the N point cosine-tapered Tukey window [20] with parameter $\alpha = 0.5$ shown in Fig. 3-7. The HTLS algorithm parameters are chosen as $\lambda = 1$, $M = 101$, and $p = 4$. The harmonic chirp amplitude estimates are found from a reduced version of Eq. (3.9) by using $W = 20$ data points centered at the estimation point $\hat{t}_{e,l}$ for the lth analysis window,

$$\tilde{\mathbf{h}}_l = (\tilde{\mathbf{Z}}_l^H \tilde{\mathbf{Z}}_l)^{-1} \tilde{\mathbf{Z}}_l^H \mathbf{x}_l \quad , \tag{3.27}$$

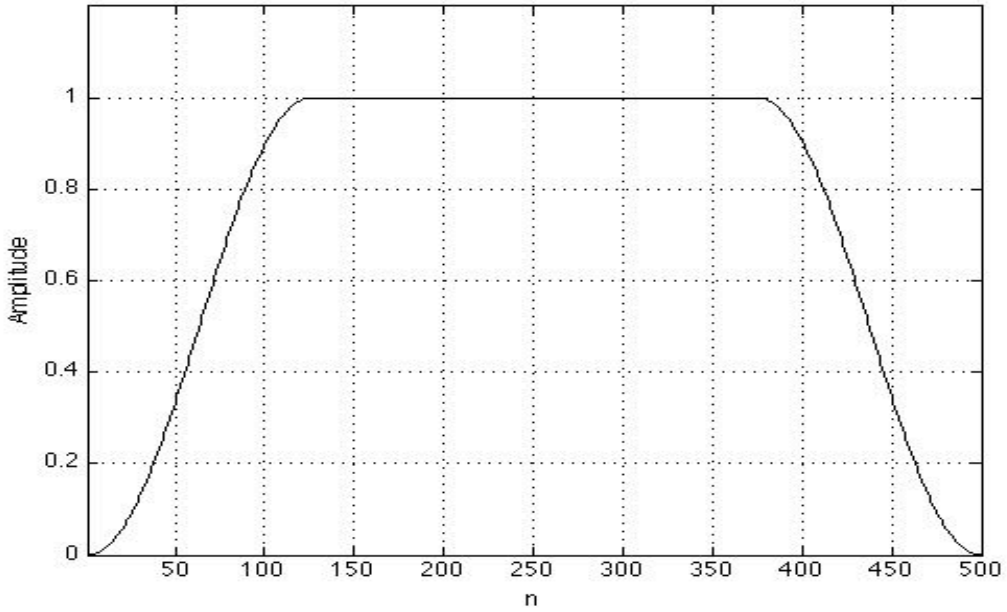

Figure 3-7: Tukey window with $\alpha = 0.5$

where

$$\tilde{\mathbf{Z}}_l = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \tilde{z}_{1,l} & \tilde{z}_{2,l} & \dots & \tilde{z}_{p,l} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{z}_{1,l}^{W-1} & \tilde{z}_{2,l}^{W-1} & \dots & \tilde{z}_{p,l}^{W-1} \end{bmatrix}, \ \tilde{\mathbf{h}}_l = \begin{bmatrix} \tilde{h}_{1,l} \\ \tilde{h}_{2,l} \\ \vdots \\ \tilde{h}_{p,l} \end{bmatrix}, \text{ and } \mathbf{x}_l = \begin{bmatrix} x_l[\lfloor \hat{t}_{e,l} - \frac{W}{2} + 1 \rfloor] \\ x_l[\lfloor \hat{t}_{e,l} - \frac{W}{2} + 2 \rfloor] \\ \vdots \\ x_l[\lfloor \hat{t}_{e,l} - \frac{W}{2} + W \rfloor] \end{bmatrix}.$$

This is done to limit the effect of time-varying frequency and amplitude parameters within the analysis window while providing sufficient averaging to reduce the error variance.

Fig. 3-8 compares the estimated amplitude contours a_{LS} to the actual contours in Eq. (3.26) for an SNR of 50 dB. There are two noticeable factors which increase the amplitude estimation error at relatively high SNRs. First, even in regions of constant harmonic amplitudes, the second harmonic amplitude estimate shows greater deviation from the known contour. Rife and Boorstyn [58] show that for multiple

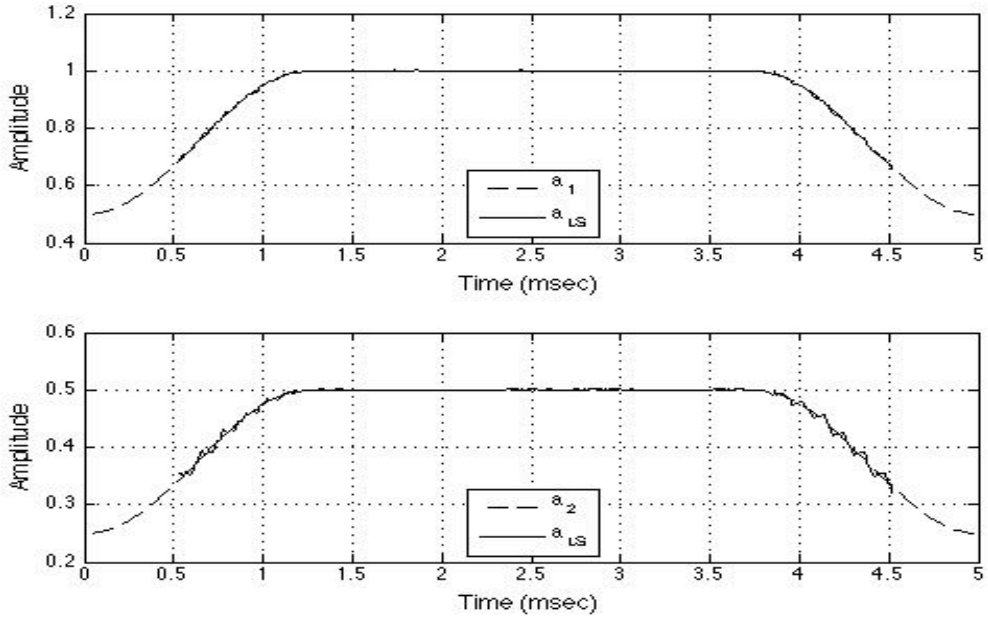

Figure 3-8: Amplitude estimation performance for double harmonic linear chirp (SNR = 50 dB)

tones, the CRLB of a particular tone's frequency estimate depends on its own amplitude but is independent of the other tone amplitudes. The weaker second harmonic results in a less accurate amplitude estimate due to a less accurate frequency estimate in Eq. (3.27). Second, in regions where a tone's amplitude is time-varying, the amplitude estimate is less accurate because Eq. (3.27) assumes the parameters $\tilde{h}_{k,l}$ are constant within the analysis window. The largest estimation error in Fig. 3-8 occurs in regions where both the chirp amplitude is changing and the corresponding frequency estimate is less accurate. A third source of error is due to the assumption that the frequencies are also constant in Eq. (3.27), while the underlying frequency contours are also time-varying. Fig. 3-9 shows the residual MSE in the amplitude estimation problem, computed from Eq. (3.27) as

$$\text{residual MSE} = \frac{\|\mathbf{x}_l - \tilde{\mathbf{Z}}_l \tilde{\mathbf{h}}_l\|_2^2}{W} \quad . \tag{3.28}$$

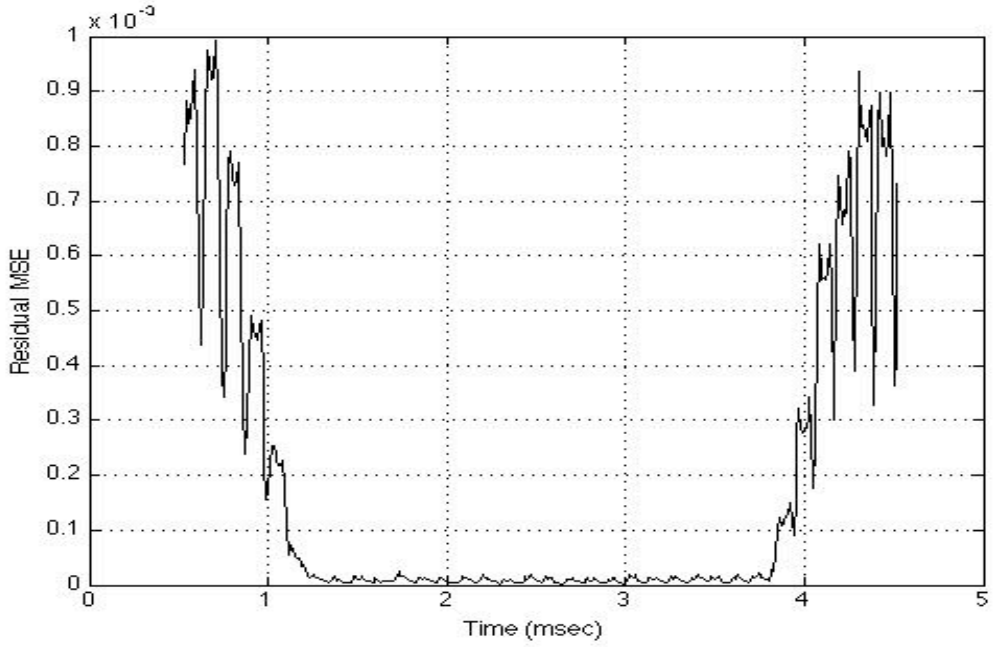

Figure 3-9: Residual MSE for double harmonic linear chirp (SNR = 50 dB)

The residual MSE is characterized as being somewhat periodic and sensitive to rapid changes in the amplitude and frequency contours, with strong dependence on the weaker chirp amplitudes due the corresponding decrease in frequency estimation accuracy.

Fig. 3-10 compares the estimated amplitude contours a_{LS} to the actual contours in Eq. (3.26) for SNR = 25 dB. The increased additive white noise degrades the frequency and amplitude estimation problems, resulting in larger deviations from the underlying amplitude contour for sustained periods of time. Fig. 3-11 shows the corresponding residual MSE. In comparison with Fig. 3-9, the increased additive white noise boosts the residual MSE while reducing the relative performance gain when the amplitudes are held constant.

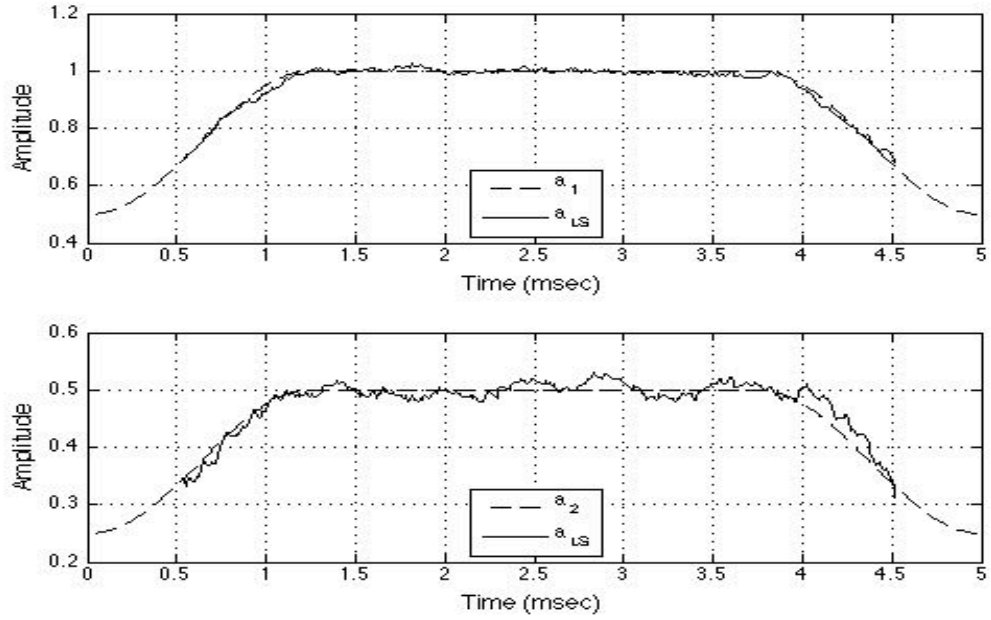

Figure 3-10: Amplitude estimation performance for double harmonic linear chirp (SNR = 25 dB)

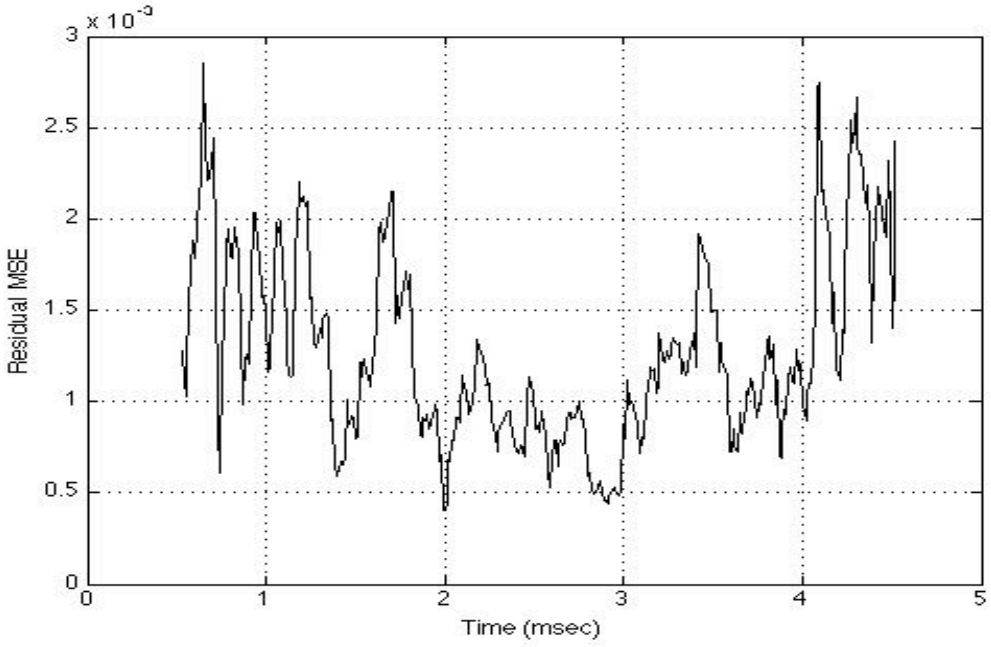

Figure 3-11: Residual MSE for double harmonic linear chirp (SNR = 25 dB)

Chapter 4

Synthetic Marine Mammal Whistle Calls

This chapter applies the experience gained from the parameter estimation of harmonic linear chirps in Chapter 3 to the parameter estimation, modification and synthesis of bottlenose dolphin whistle calls. Section 4.1 focuses on parameter estimation and synthesis of bottlenose dolphin whistle calls. Section 4.2 proposes different strategies for watermarking whistle calls based upon detection capability and exploiting natural variability in the whistle call frequency contours.

4.1 Modeling Recorded Bottlenose Dolphin Whistle Calls

Fig. 4-1 shows a bottlenose dolphin whistle call composed of three separate whistles taken from the Sarasota Bottlenose Dolphin Whistle Catalog maintained at Woods Hole Oceanographic Institution [62]. The whistle call was recorded using a custom built suction cup hydrophone attached to the forehead of the dolphin. The original analog recording at $f_s = 40$ kHz was later digitized using a sample rate of $f_s = 96$ kHz. Fig. 4-2 is a spectrogram of the bottlenose dolphin whistle call in Fig. 4-1

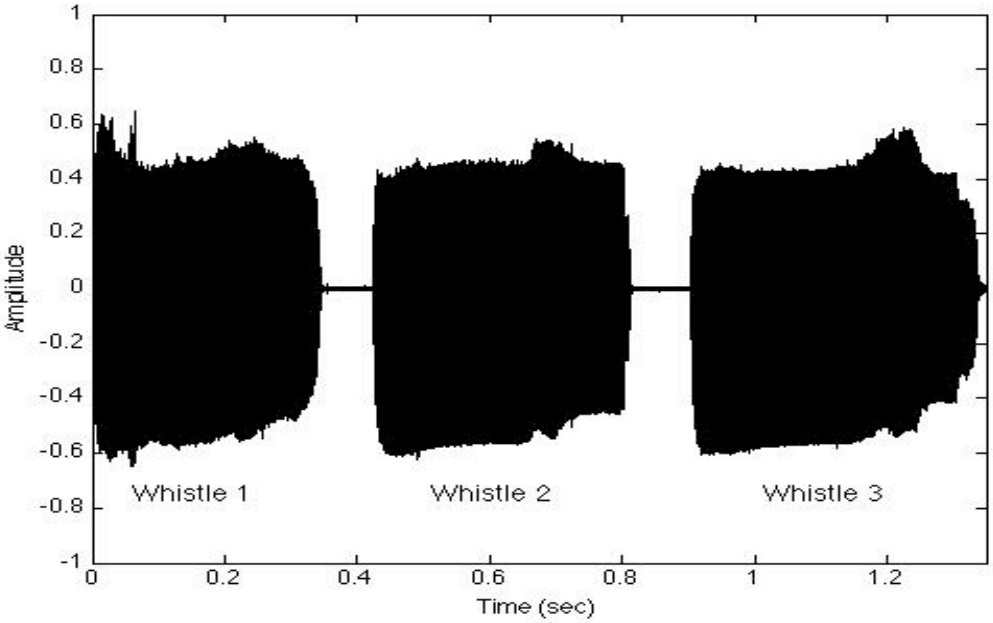

Figure 4-1: Bottlenose dolphin whistle call

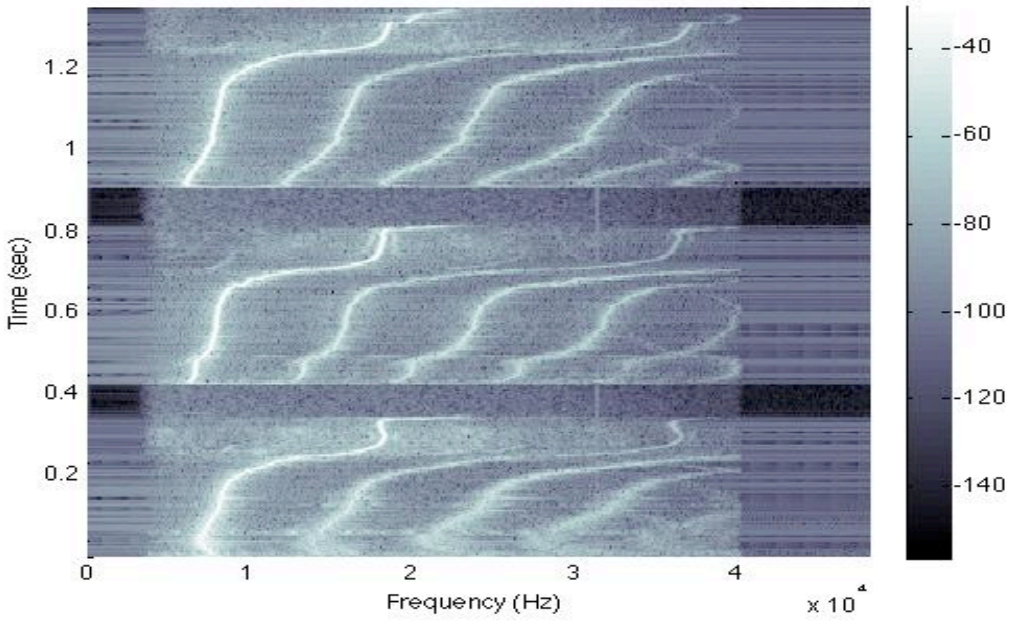

Figure 4-2: Spectrogram of bottlenose dolphin whistle call in Fig. 4-1 (dB)

computed using the short-time Fourier transform with a 750 point Hamming window and 250 samples of overlap [49]. Each whistle contains up to six harmonics with frequency generally increasing throughout the whistle.

The performance of the frequency estimation problem is dependent upon three parameters: the exponential forgetting factor λ, the number of data samples M used in each analysis block, and the model order p. To limit the smoothing effect of the analysis window while achieving optimal frequency matching characteristics, the values $\lambda = 1$ and $M = 101$ are chosen. The choice of p is more complex. If the whistle calls were composed of R harmonics with stable, relatively equal amplitude contours, then the model order would be chosen as $p = 2R$. In reality, the higher harmonics are significantly weaker than the fundamental harmonic, and in regions where the whistle amplitude or frequency changes rapidly, the amplitudes of each harmonic fluctuate strongly. Due to the known harmonic structure of the whistles and the relatively weak amplitudes of higher harmonics, all harmonics are best estimated as multiples of the fundamental harmonic, f_1. A low model order of $p = 2$ is chosen, for which the frequency of the strongest harmonic is estimated, because of the occasional instability of the whistle harmonics. However, since higher harmonics are not accounted for in the model, the resulting fundamental frequency estimate has a higher error variance than if the data contained only the fundamental frequency contour. The solution is to apply a bandpass filter to isolate the fundamental harmonic from the higher harmonics before performing frequency estimation.

The wide frequency range of the bottlenose dolphin whistle calls require using two overlapping bandpass filters to isolate the fundamental frequency contour. The overlap region is chosen to be large enough to allow a smooth transition between the two frequency estimates. The Matlab command `filtfilt` [45] is used to perform zero-phasing filtering to ensure the resulting estimated frequency contours are correctly aligned in the time domain. Fig. 4-3 shows the frequency estimates for Whistle 1 obtained using a bandpass filter overlap region of 12-12.5 kHz and a transition time

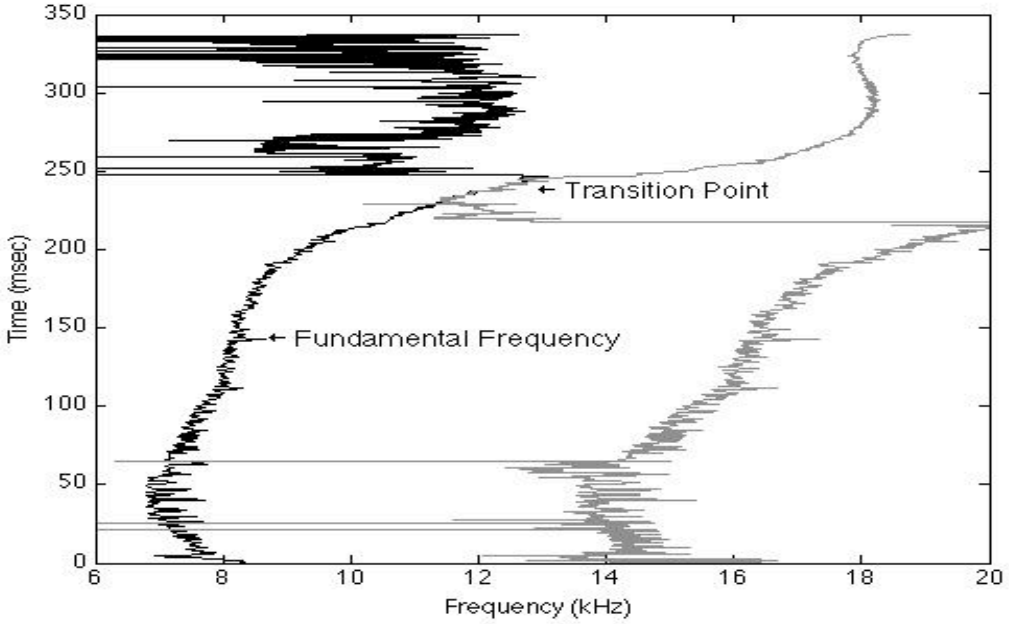

Figure 4-3: Fundamental frequency estimation of Whistle 1 in Fig. 4-1 using overlapping bandpass filters

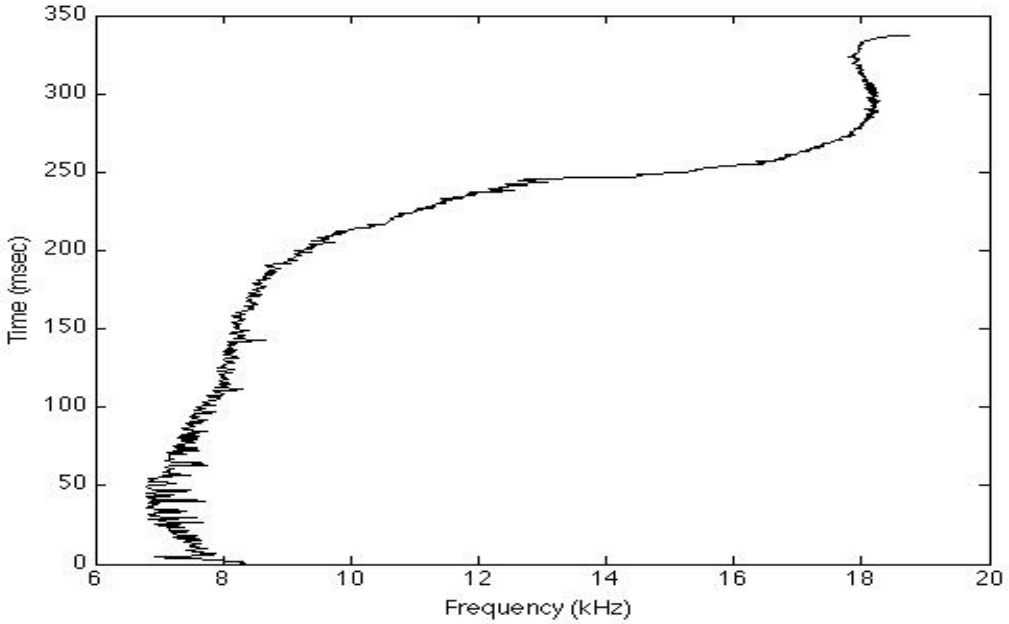

Figure 4-4: Fundamental frequency contour of Whistle 1 in Fig. 4-1

between frequency estimates of 237.65 msec. The resulting fundamental frequency contour is shown in Fig. 4-4. The frequency contours of the higher harmonics are $f_r[l] = r f_1[l]$ for $1 \leq l \leq L$, where L is the number of analysis blocks in the whistle.

The harmonic amplitude estimates are then found for each analysis block using an estimation width of $W = 20$ data points. For each data block, the number of harmonic amplitudes to be estimated is specified based on the frequency of the fundamental harmonic. For example, when the fundamental harmonic exceeds 10 kHz, there will be at most three harmonics present due to the frequency cutoff at 40 kHz. Overestimating the number of harmonics in the data gives spurious results. The estimated amplitude contours for Whistle 1 is shown in Fig. 4-5. It is important to keep in mind that the amplitude estimates are performed for the recorded whistle and are not necessarily representative of the actual whistle, since the higher harmonics are artificially cutoff by the recording equipment at frequencies greater than 40 kHz. The actual harmonic amplitudes most likely do not fluctuate as rapidly as seen in

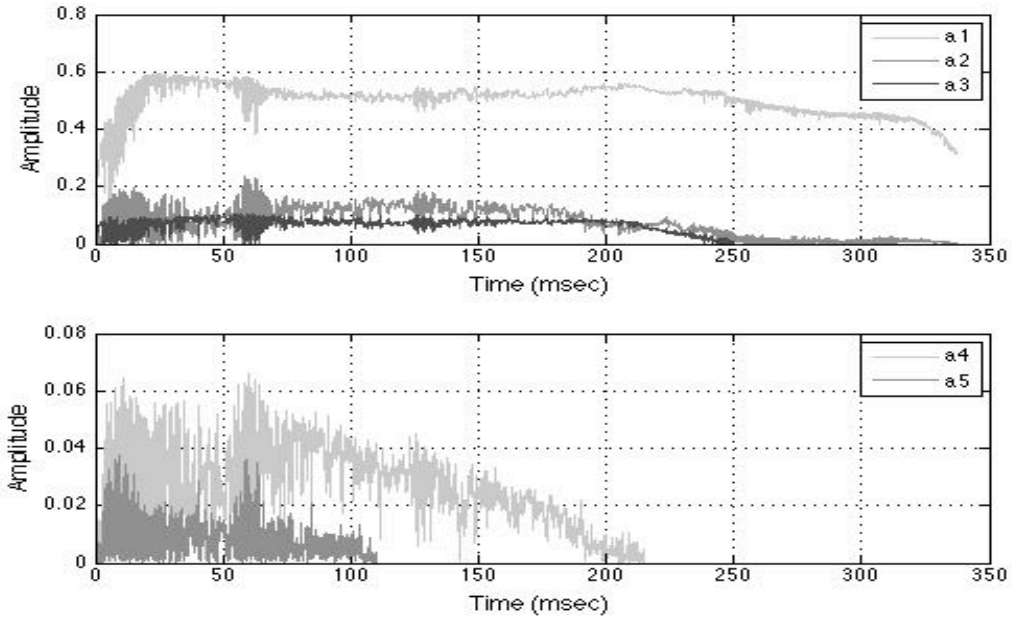

Figure 4-5: Estimated amplitude contours for Whistle 1 in Fig. 4-1

Fig. 4-5. The observed short-time variability in the amplitude contours accounts for model mismatch and frequency estimation error.

Fig. 4-6 shows the residual MSE for Whistle 1. The MSE is remarkably low in the middle of the whistle while the amplitude contours are relatively stable, indicating good frequency and amplitude estimation performance. In regions where the whistle is less stable, such as during the attack phase at the beginning of the whistle, the parameters vary more quickly, resulting in worse estimation performance. The synthetic whistle is then constructed from the harmonic frequency and amplitude contours according to the model in Eq. (3.1),

$$\hat{s}[l] = \sum_{r=1}^{R} \hat{a}_r[l] \cos\left(2\pi \sum_{i=1}^{l} \frac{r\hat{f}_1[i]}{f_s} + \hat{\theta}_r\right) \quad \text{for} \quad 1 \leq l \leq L, \tag{4.1}$$

where $\hat{a}_r$ are the harmonic amplitude contours, $\hat{\theta}_r$ are the initial phases of each harmonic, and $\hat{f}_1$ is the fundamental frequency contour. Since the human auditory system

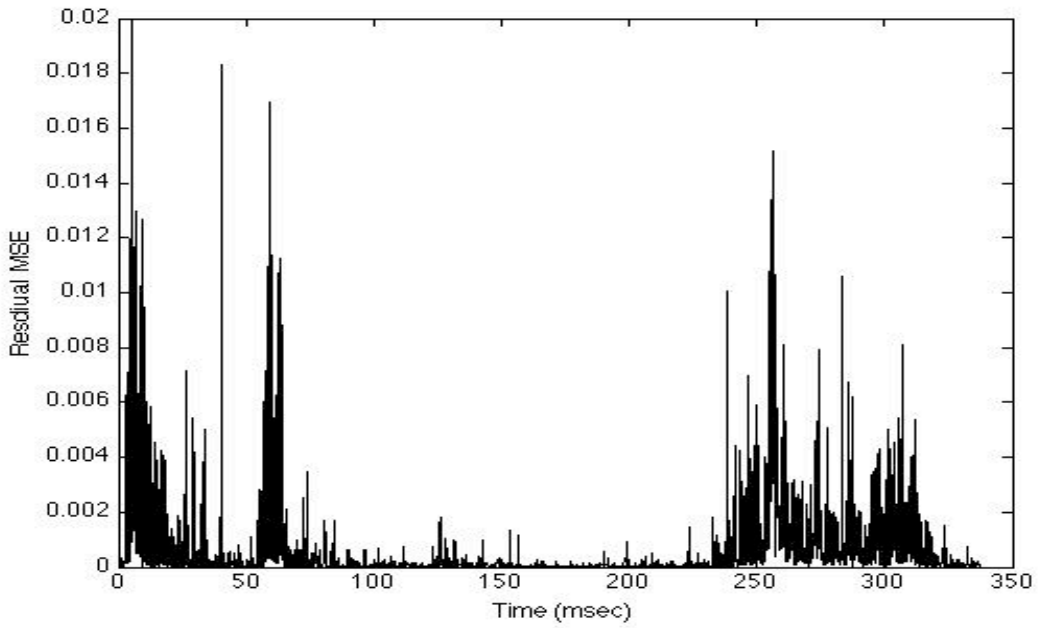

Figure 4-6: Residual MSE for Whistle 1 in Fig. 4-1

(HAS) is insensitive to the initial phase, the synthetic whistles could be constructed with $\hat{\theta}_r = 0$, but accounting for the initial phase difference between harmonics causes the synthetic whistle to more closely resemble the recorded whistle in the time domain.

Fig. 4-7 compares the recorded and synthetic time domain representations for Whistle 1. Fig. 4-8 compares the spectrograms for the recorded and synthetic versions of Whistle 1. In-air playbacks using Matlab demonstrate that the synthetic whistle is almost indistinguishable from the recorded whistle. However, the sinusoidal model does not account for any stochastic 'noise-like' portions of the whistle, such as seen surrounding the fundamental frequency contour at the end of Whistle 1 in Fig. 4-8. Other dolphin whistles should be studied to determine whether this type of stochastic effect is actually produced by the dolphin.

Figs. 4-9 through 4-12 show the fundamental frequency and amplitude contours for Whistles 2 and 3 in Fig. 4-1. Each successive whistle has a longer duration and is characterized by increasingly stable frequency and amplitude contours. The residual MSE for Whistles 2 and 3 is shown in Fig. 4-13 and Fig. 4-14. Both Whistle 2 and 3 have a lower residual MSE than Whistle 1, as expected based on the stability of the frequency and amplitude contours. Each whistle has a higher residual MSE when the fundamental frequency is rapidly increasing toward the end of the whistle.

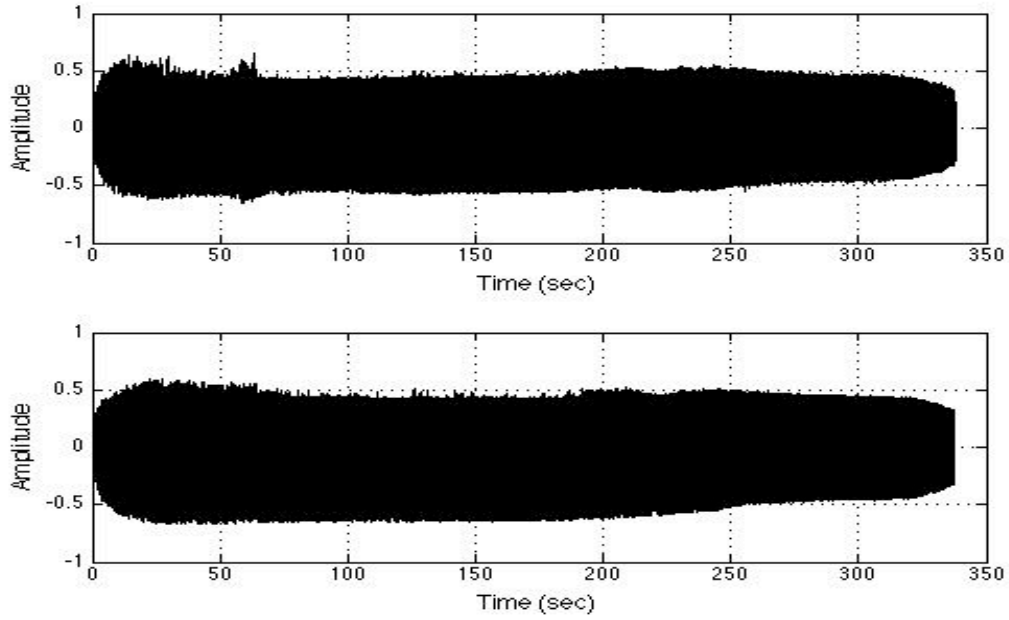

Figure 4-7: Recorded (top) vs. synthetic (bottom) versions of Whistle 1 in Fig. 4-1

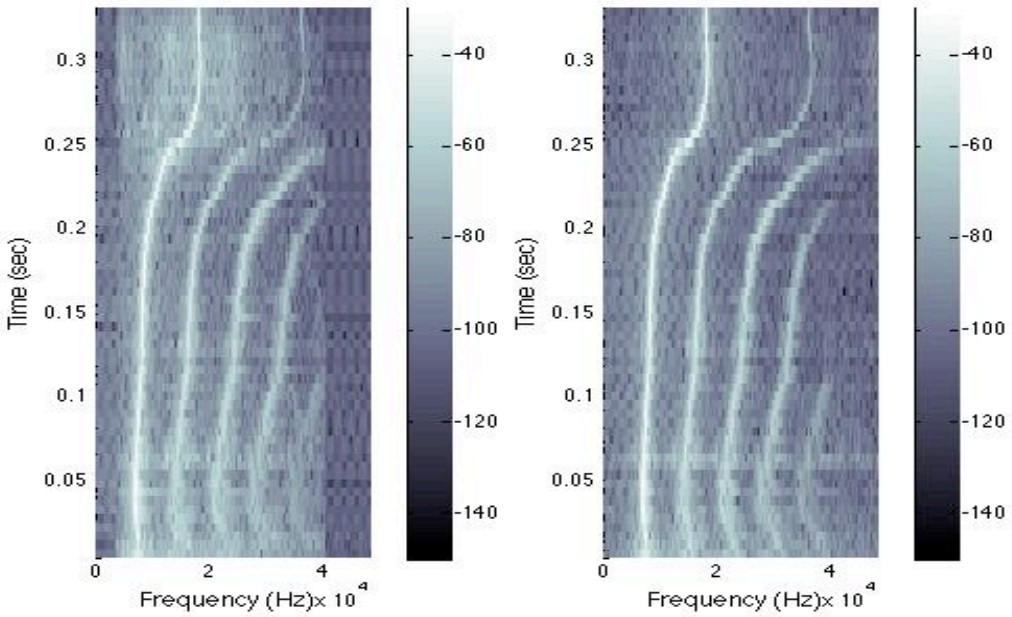

Figure 4-8: Spectrograms of recorded (left) vs. synthetic (right) versions of Whistle 1 in Fig. 4-1 (dB)

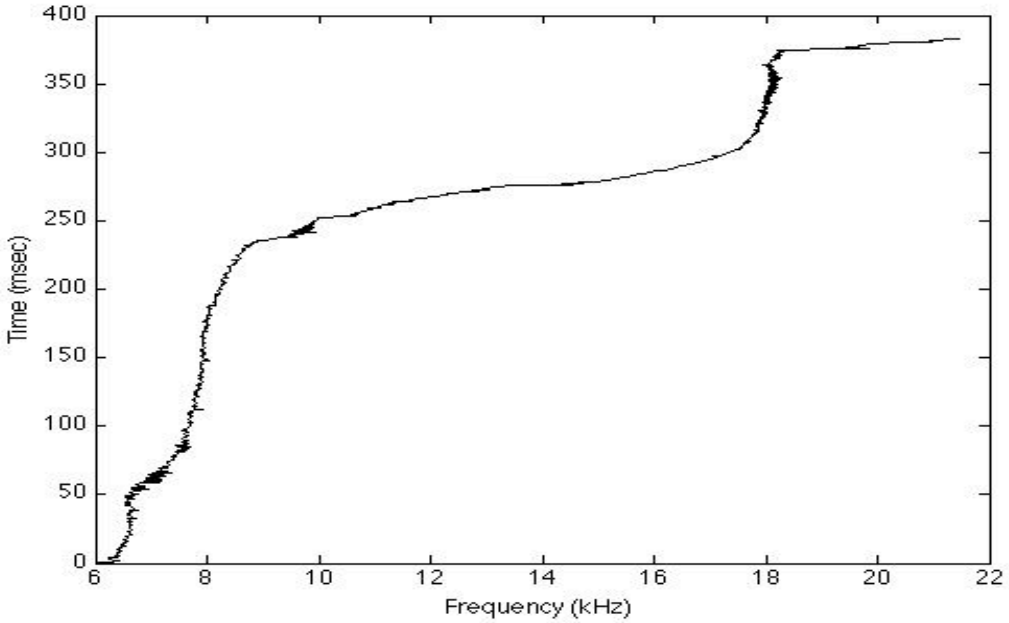

Figure 4-9: Fundamental frequency contour of Whistle 2 in Fig. 4-1

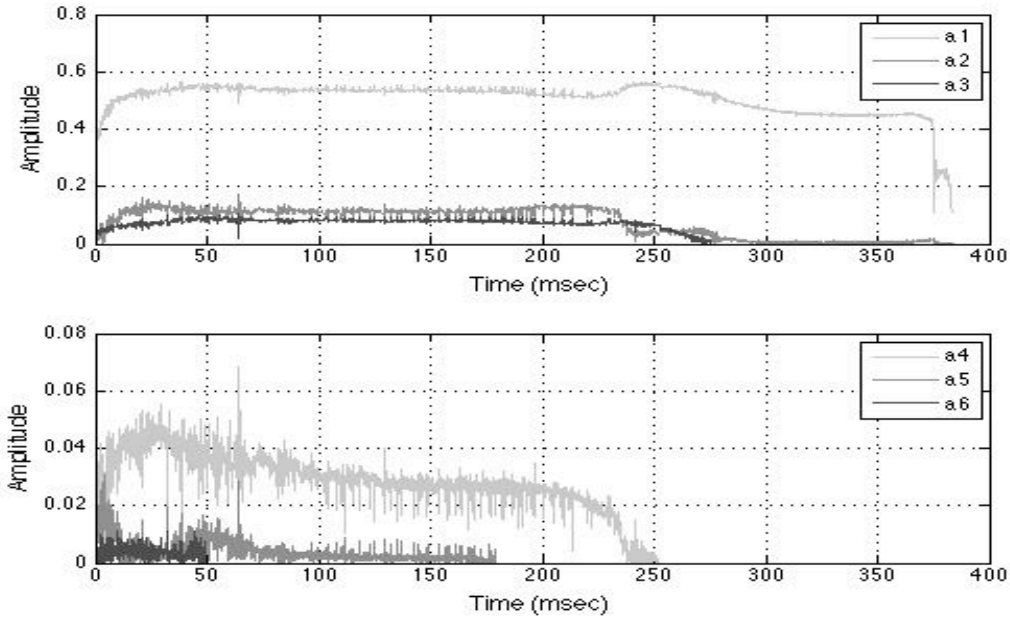

Figure 4-10: Estimated amplitude contours for Whistle 2 in Fig. 4-1

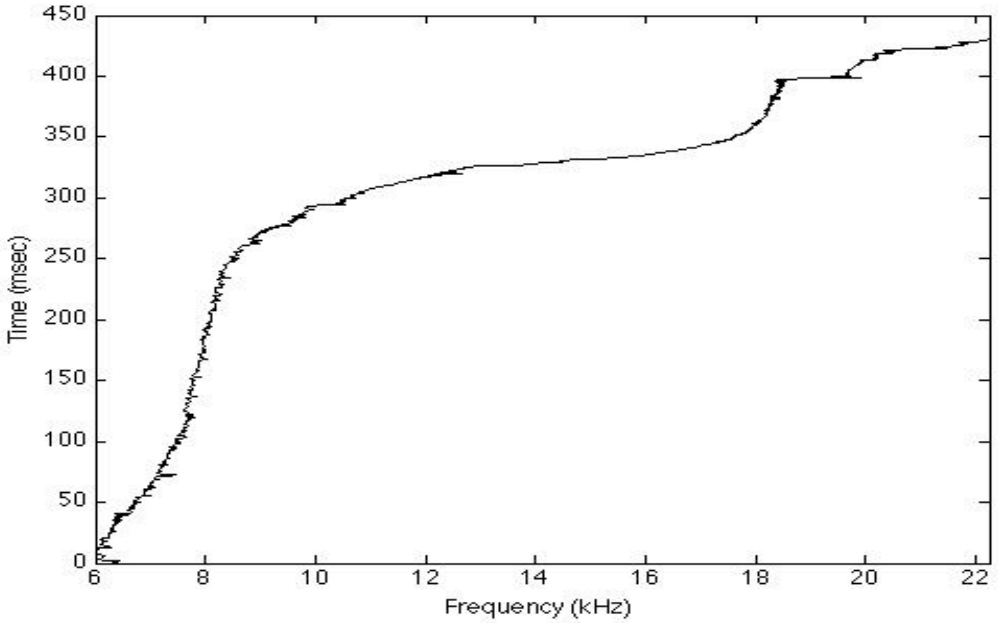

Figure 4-11: Fundamental frequency contour of Whistle 3 in Fig. 4-1

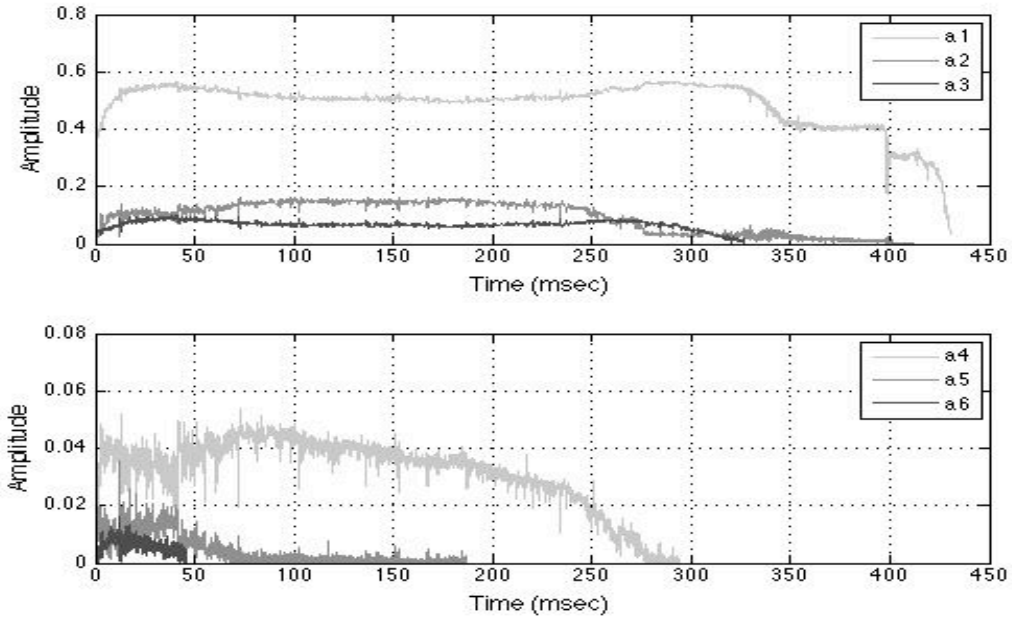

Figure 4-12: Estimated amplitude contours for Whistle 3 in Fig. 4-1

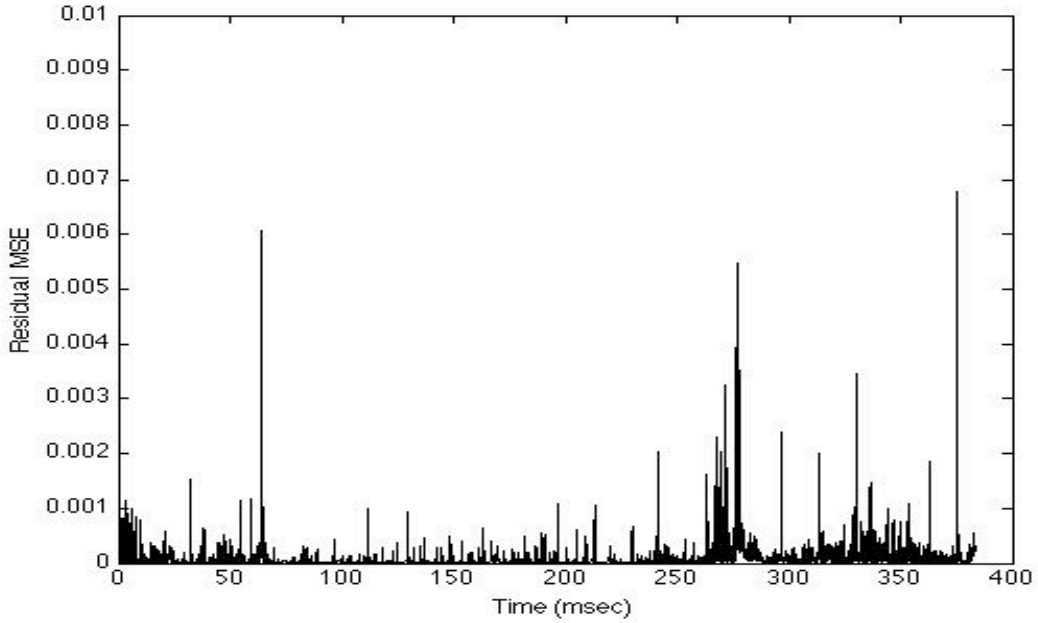

Figure 4-13: Residual MSE for Whistle 2 in Fig. 4-1

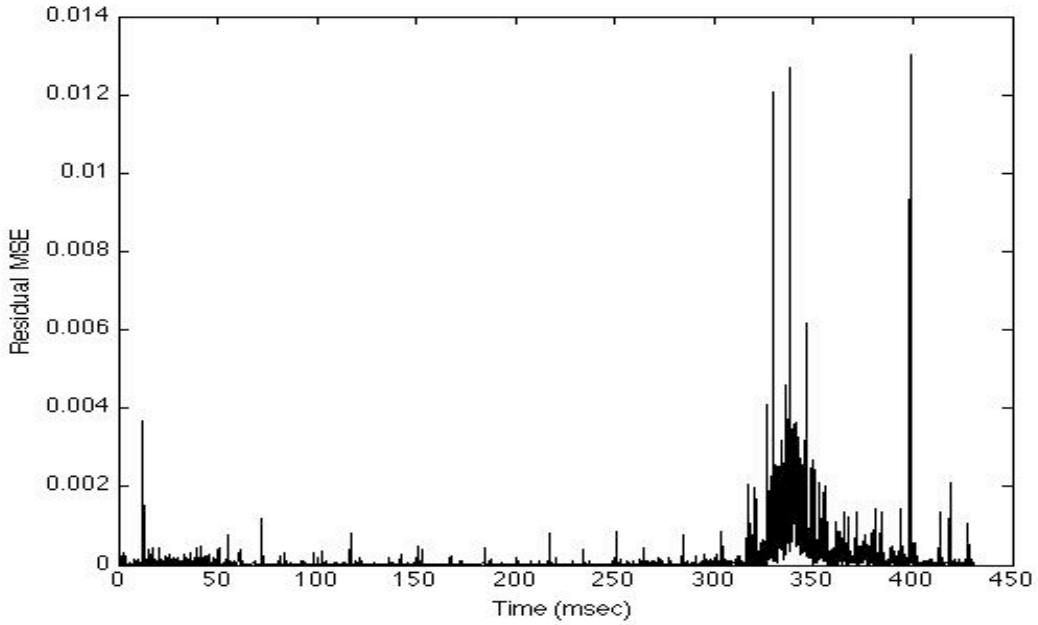

Figure 4-14: Residual MSE for Whistle 3 in Fig. 4-1

4.2 Watermarked Synthetic Whistle Calls

In a covert communications scenario, a blind watermark detection scheme is generally desired, in which the host signal is not needed for watermark retrieval. Due to the sensitivity of the HAS, a parametric watermark that produces a natural-sounding stego-signal provides the best opportunity for passing embedded information without alerting observers to the existence of the information. The harmonic frequency contours are chosen as the parameter set to be watermarked based on the strong performance of the sinusoidal model of Eq. (4.1) in representing recorded bottlenose dolphin whistle calls. In order to produce natural-sounding whistles using a retrievable watermark, the harmonic relationship between frequency contours should be maintained. Thus, different schemes for watermarking the fundamental frequency contour of a synthetic whistle should be considered in terms of the ease of watermark detection and retrieval and the naturalness of the resulting stego-signal.

The fundamental frequency contour regularly fluctuates about its instantaneous mean that can be described as a vibrato in the frequency contour. Instead of adding distortion on top of the observed vibrato, watermark retrieval can be enhanced by watermarking the instantaneous mean frequency (IMF) contour, f_{IMF}, which is assumed to be the original frequency contour if the vibrato effect did not occur. The vibrato can be thought of as a stochastic vibration or watermark f_W added to the smoothed frequency contour f_{IMF}, so that

$$f_1[l] = f_{IMF}[l] + f_W[l], \quad \text{for} \quad 1 \leq l \leq L. \tag{4.2}$$

Since the natural bottlenose dolphin whistles consist of distorted frequency contours, there is a good chance that robust watermarking methods can be utilized to produce natural-sounding synthetic whistles. However, if the watermark is too natural, it may be difficult to distinguish between natural and synthetic whistles.

The IMF contour is found as the weighted time-average of the fundamental fre-

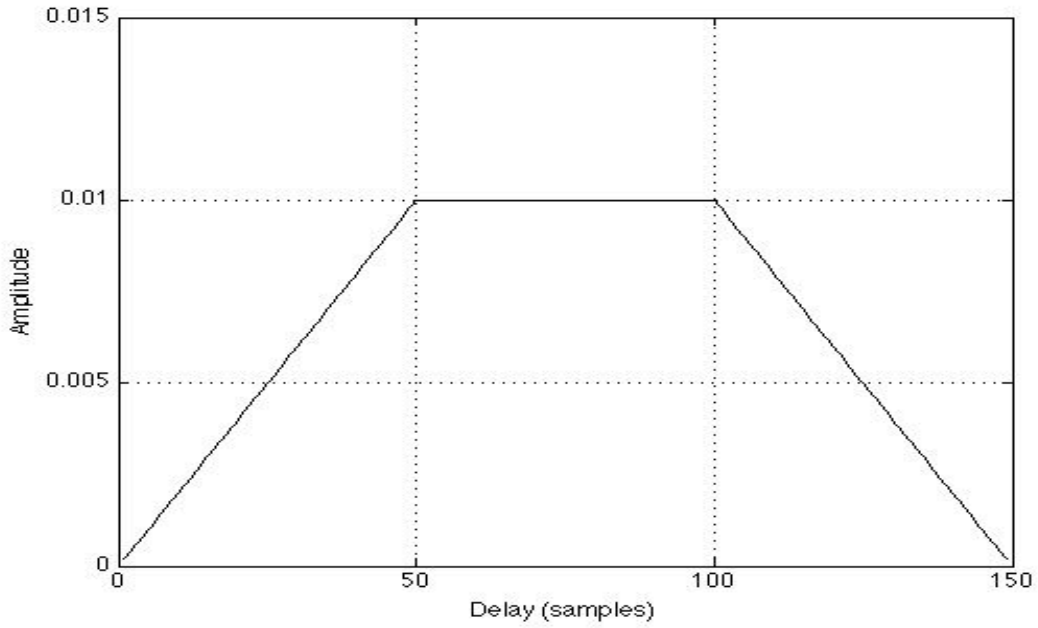

Figure 4-15: Impulse response of moving-average filter

quency contour using a moving-average filter with the impulse response shown in Fig. 4-15. The observed bottlenose dolphin whistle vibrato occurs with an average period of roughly 1 msec, so the effective impulse response length of the filter is chosen to be about 1 msec. The resulting moving-average filter gives equal weight to local frequency estimates while giving consideration to more distant values in order to smoothly estimate the IMF. The Matlab command `filtfilt` [45] is again used to perform zero-phase filtering. The fundamental frequency and IMF contours for a portion of Whistle 2 are shown in Fig. 4-16.

In a covert communications scenario, it would be desirable to be able to retrieve the watermark under relatively low SNR conditions, such as SNR = 5 dB. This requires a relatively robust watermarking scheme that facilitates watermark retrieval even when frequency estimation performance is relatively bad. Liu's F-QIM watermarking scheme [38], which is based on detecting the difference between separate frequency quantizers, would require either large frequency deviations between quantizer levels

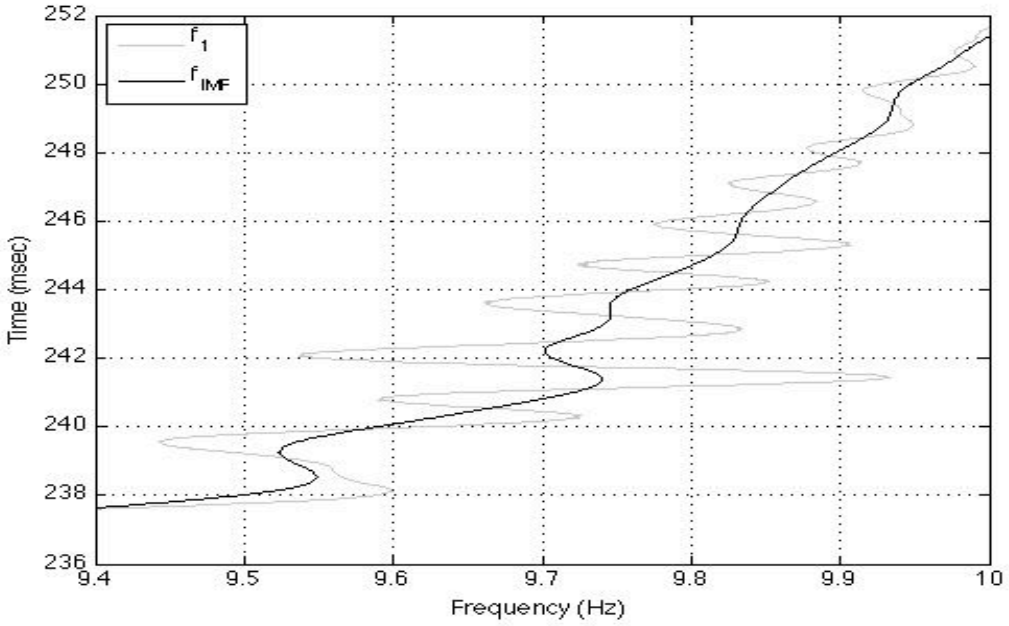

Figure 4-16: Fundamental frequency and IMF contours for a portion of Whistle 2 in Fig. 4-1

or high SNR to ensure robust watermark retrieval due to the frequency estimation performance. The remainder of this chapter considers two watermarking schemes that are relatively robust for a range of SNR. The first scheme constructs a watermark composed of linear chirp segments separated by an abrupt frequency shift. The second scheme constructs a watermark that simulates the natural vibrato of the fundamental frequency using continuous-phase modulation (CPM).

4.2.1 Linear Chirp Segments With Abrupt Frequency Shifts

The goal of most communications systems is to maximize the achievable data rate for which transmitted information can be reliably decoded. This implies that each information bit will correspond to a minimal number of samples in the transmitted signal. Thus, from the perspective of data rate, an optimal frequency watermarking scheme will have a relatively low number of samples per information bit available for

frequency estimation. Increasing the sample rate at the receiver will also generally improve frequency estimation performance by providing more samples per information bit, but it is assumed fixed when choosing a watermarking scheme. At low SNR, small changes in the frequency contour may be obscured by the increased frequency estimation variance, making robust QIM-based watermarking schemes unattractive in terms of perceptual distortion of the host signal. To improve frequency estimation performance and limit perceptual distortion, the IMF contour should be watermarked with a generally smoothly-varying signal that can be tracked over time using the HTLS frequency estimator or other frequency estimators.

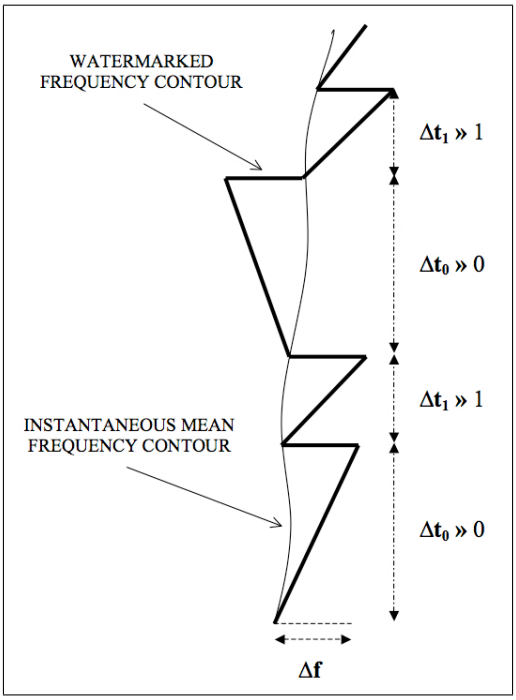

Figure 4-17: Watermarking scheme based on linear chirp segments with abrupt frequency shifts

A potential watermarking scheme, portrayed in Fig. 4-17, approximates the IMF contour using linear chirp segments with abrupt frequency shifts Δf. The watermarked information is encoded in the amount of time between abrupt frequency shifts, Δt_0 and Δt_1. The slope of each linear chirp segment is chosen to achieve a fre-

quency separation of Δf from the IMF contour after a duration Δt_0 or Δt_1 specified by each information bit. The synthetic stego-signal is then constructed according to Eq. (4.1) using the watermarked fundamental frequency contour and the amplitude contours estimated using the original fundamental frequency contour estimate. An alternative to the watermarking scheme in Fig. 4-17 is to tag the midpoint instead of the initial point of each linear chirp segment to the IMF contour.

Fig. 4-18 shows the linear chirp watermarked fundamental frequency contour based on Whistle 2 of Fig. 4-1. The watermarked contour was constructed using a random information bit stream and the parameters $\Delta f = 150$ Hz, $\Delta t_0 = 1$ msec and $\Delta t_1 = 2$ msec. In-air playbacks using Matlab demonstrate that there is a small perceptible difference between the recorded and watermarked synthetic whistles. The parameter that most effects the perceptible distortion of the host signal is the frequency shift, Δf. At relatively high SNR, the frequency estimation performance will be improved, and thus require a smaller Δf for reliable watermark retrieval. As SNR decreases,

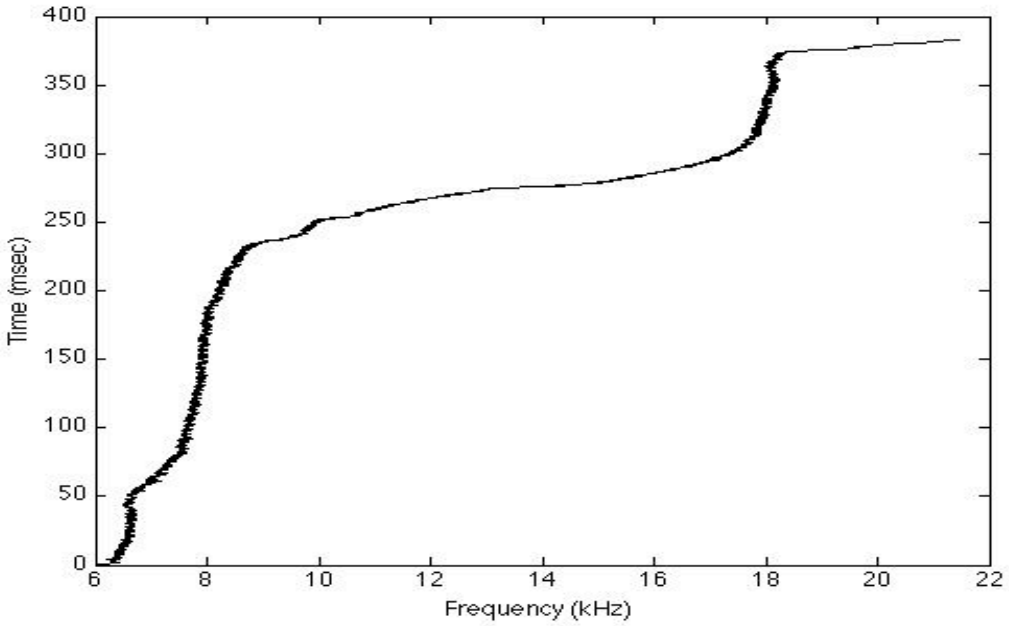

Figure 4-18: Linear chirp watermarked frequency contour of Whistle 2 in Fig. 4-1

the frequency estimation variance increases, and a larger Δf is needed to differentiate between an actual frequency shift and estimation error. Watermark retrieval is performed by detecting abrupt frequency shifts in the fundamental frequency contour of the received whistle.

4.2.2 Continuous Phase Modulation

Due to the inherent vibrato observed in the bottlenose dolphin whistle calls, an alternative to the linear chirp watermarking scheme is to embed information in a synthetic vibrato using continuous phase modulation (CPM) as shown in Fig. 4-19. CPM signals [52] have a continuous carrier phase

$$\phi(t;\mathbf{I}) = 2\pi \sum_{k=-\infty}^{n} I_k h_k q(t-kT), \quad nT \leq t \leq (n+1)T \tag{4.3}$$

where $\{I_k\}$ is a sequence of M-ary information symbols selected from the alphabet $\pm 1, \pm 3, \ldots, \pm(M-1)$, $\{h_k\}$ is a sequence of modulation indices, and $q(t)$ is some normalized waveform shape. While many types of CPM could be used to construct a synthetic whistle vibrato, a simple type called minimum-shift keying (MSK) can be used to illustrate a watermarking scheme using CPM.

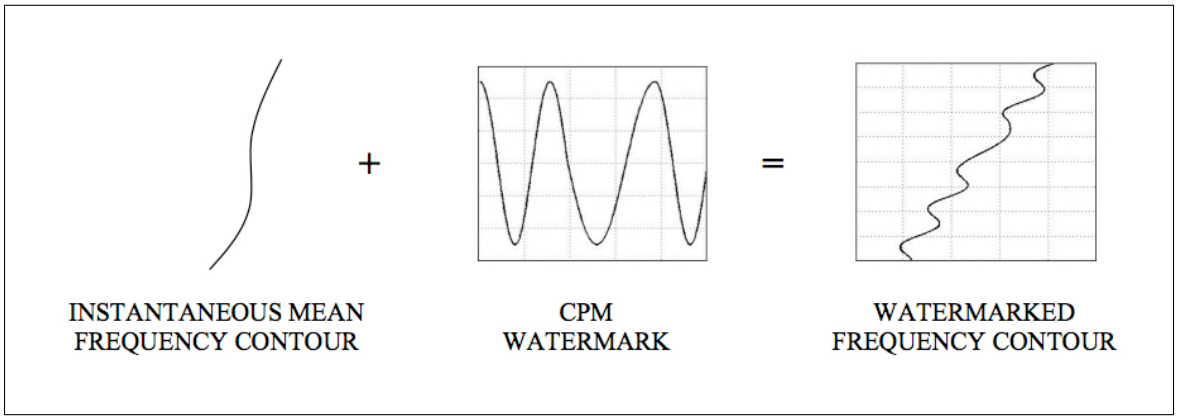

Figure 4-19: Watermarking scheme based on CPM perturbation of the IMF contour

MSK is a special form of binary CPM in which the modulation index $h = \frac{1}{2}$ and normalized waveform shape

$$q(t) = \begin{cases} 0 & (t < 0) \\ t/2T & (0 \le t \le T) \\ 1/2 & (t > T) \end{cases}. \tag{4.4}$$

The phase of the MSK carrier in the interval $nT \le t \le (n+1)T$ is

$$\begin{aligned} \phi(t; \mathbf{I}) &= \frac{1}{2}\pi \sum_{k=-\infty}^{n-1} I_k + \pi I_n q(t - nT) \\ &= \theta_n + \frac{1}{2}\pi I_n \left(\frac{t - nT}{T} \right), \quad nT \le t \le (n+1)T, \end{aligned} \tag{4.5}$$

where

$$\theta_n = \frac{1}{2}\pi \sum_{k=-\infty}^{n-1} I_k \quad . \tag{4.6}$$

The modulated MSK carrier signal with amplitude A and carrier frequency f_c is

$$\begin{aligned} s(t) &= A \cos \left[2\pi f_c t + \theta_n + \frac{1}{2}\pi I_n \left(\frac{t - nT}{T} \right) \right] \\ &= A \cos \left[2\pi \left(f_c + \frac{1}{4T} I_n \right) t - \frac{1}{2} n\pi I_n + \theta_n \right], \quad nT \le t \le (n+1)T. \end{aligned} \tag{4.7}$$

From Eq. (4.7), it can be seen that for each interval $nT \le t \le (n+1)T$, MSK can be thought of as having one of two frequencies,

$$\begin{aligned} f_0 &= f_c - \frac{1}{4T} \\ f_1 &= f_c + \frac{1}{4T}, \end{aligned} \tag{4.8}$$

with an adjusted phase to achieve a continuous phase across all intervals.

The synthetic vibrato signal $f_W[l]$ can be constructed by sampling Eq. (4.7) at

the points $t = l/f_s$ with a carrier rate of $f_c = 1/T$,

$$f_W[l] = A\cos\left[\frac{2\pi l}{Tf_s}\left(1+\frac{1}{4}I_n\right) - \frac{1}{2}n\pi I_n + \theta_n\right], \quad nTf_s \le l \le (n+1)Tf_s, \tag{4.9}$$

where $\{I_n\}$ is a sequence of binary information symbols ± 1. The CPM watermarked fundamental frequency contour is

$$f_{CPM}[l] = f_{IMF}[l] + f_W[l], \quad 1 \le l \le L. \tag{4.10}$$

Fig. 4-20 shows both the unmodified and CPM-watermarked fundamental frequency contours for a portion of Whistle 2 in Fig. 4-1, where f_{CPM} is constructed using the parameters $A = 50$ Hz and $T = 1$ msec. The main distinguishing feature between the two fundamental frequency contours is that the watermarked contour vibrato has a constant amplitude as opposed to the variable strength vibrato in the

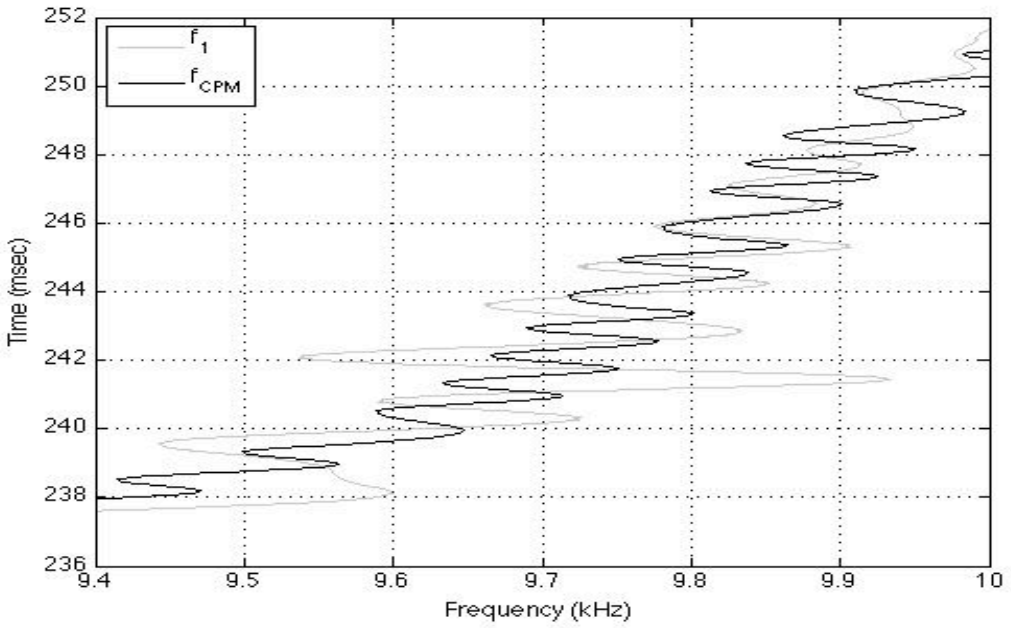

Figure 4-20: Unmodified and CPM-watermarked frequency contours for a portion of Whistle 2 in Fig. 4-1

recorded whistle. In-air playbacks using Matlab demonstrate that the watermarked whistle, constructed from Eq. (4.1) using the unmodified amplitude contour estimates, is essentially imperceptible from the recorded whistle, with the exception of slight background noise in the recorded whistle. Proakis [52] covers CPM demodulation methods that can be used for watermark retrieval after estimating the fundamental frequency contour of the received whistle.

Chapter 5

Experimental Results

This chapter presents results from the Rescheduled Acoustic Communications Experiment (RACE08) conducted in Narragansett Bay during March 2008. Synthetic whistle calls based on the bottlenose dolphin whistle call in Fig. 4-1 were transmitted throughout the experiment. The frequency estimation performance of the HTLS algorithm is demonstrated for both natural and watermarked frequency contours.

5.1 RACE08 Description

RACE08 was conducted at the University of Rhode Island's Narragansett Bay Campus, shown in Fig. 5-1, from March 1st through March 25th. Acoustic signals were transmitted from a stationary tripod located roughly 100 meters from shore in water depth of 9 meters. The primary source transducer, an ITC-1007 spherical transducer with resonant frequency of approximately 11kHz, was located about 4 meters from the sea floor. A source array composed of three Datasonics AT-12ET transducers, located beneath the ITC-1007, was not used for transmitting synthetic whistles. Three main receiver array tripods were located roughly 400 meters to the East, 400 meters to the North, and 1000 meters to the North of the source array tripod. The 400 meter receiver arrays were composed of 24 elements with 5 cm spacing. The 1000 meter

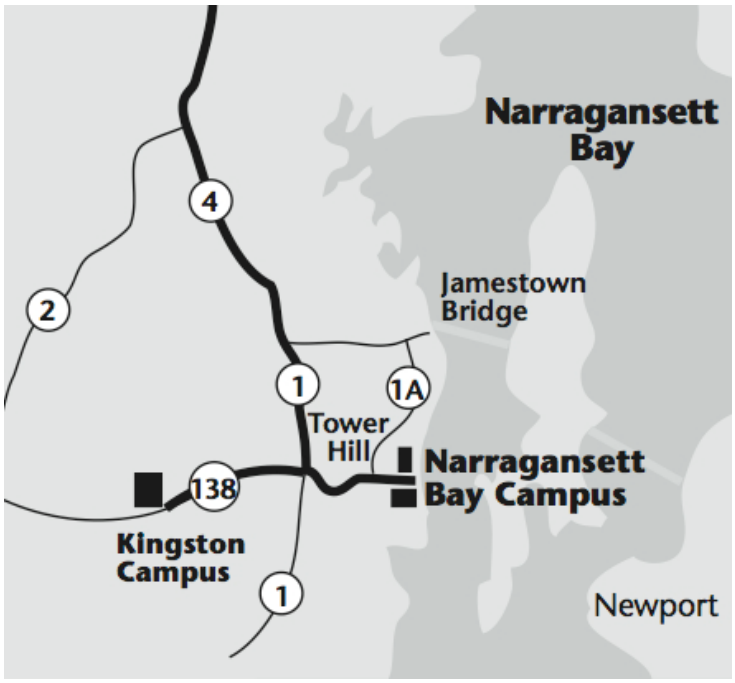

Figure 5-1: University of Rhode Island's Narragansett Bay Campus

receiver array was composed of 12 elements with 12 cm spacing. The bottom element of each receiver array was located 2 meters above the sea floor. The water depths between source and receiver arrays ranged from 9 to 14 meters. A reference ITC-100 hydrophone was mounted 1 meter from the ITC-1007 source transducer. The sample rate of the transmitter and all receivers was 39062.5 Hz (1e7/256).

5.2 RACE08 Results

Synthetic whistle calls, based on the bottlenose dolphin whistle call in Fig. 4-1, were transmitted on the ITC-1007 source transducer at two hour intervals throughout the RACE08 experiment. The results presented here, taken from the 8:00 P.M. EDT transmission on March 23rd, were chosen for relatively calm environmental conditions in Narragansett Bay.

Fig. 5-2 compares spectrograms of unmodified synthetic whistle calls received at

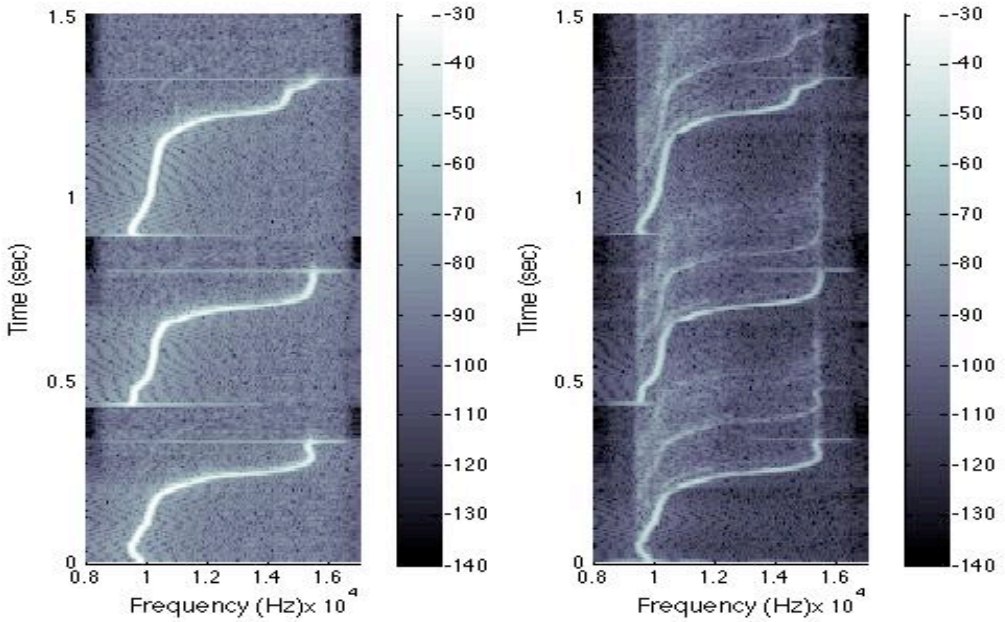

Figure 5-2: Spectrograms of unmodified synthetic whistle calls received at the reference (left) and N1000 (right) hydrophones (dB)

the reference and North 1000 meter (N1000) hydrophones. The reference hydrophone records the whistle call without multipath or intersymbol interference (ISI), while the N1000 hydrophone sees an impulse response of length greater than 0.5 seconds. The relatively long impulse response is due to strong reflections from shore in the narrow channel.

Fig. 5-3 compares spectrograms of watermarked synthetic whistle calls received at the reference and N1000 hydrophones. The watermarking scheme was similar to that portrayed in Fig. 4-17, except that the frequency was held constant for each information bit, resulting in a variable abrupt frequency shift Δf. The parameters $\Delta t_0 = 10.2$ msec and $\Delta t_1 = 20.4$ msec were chosen for initial testing to ensure frequency estimation and watermark retrieval could be demonstrated. The presence of the watermark is clearly seen at the N1000 hydrophone in Fig. 5-3, since the multipath energy only appears at discrete frequencies determined by the watermarking scheme.

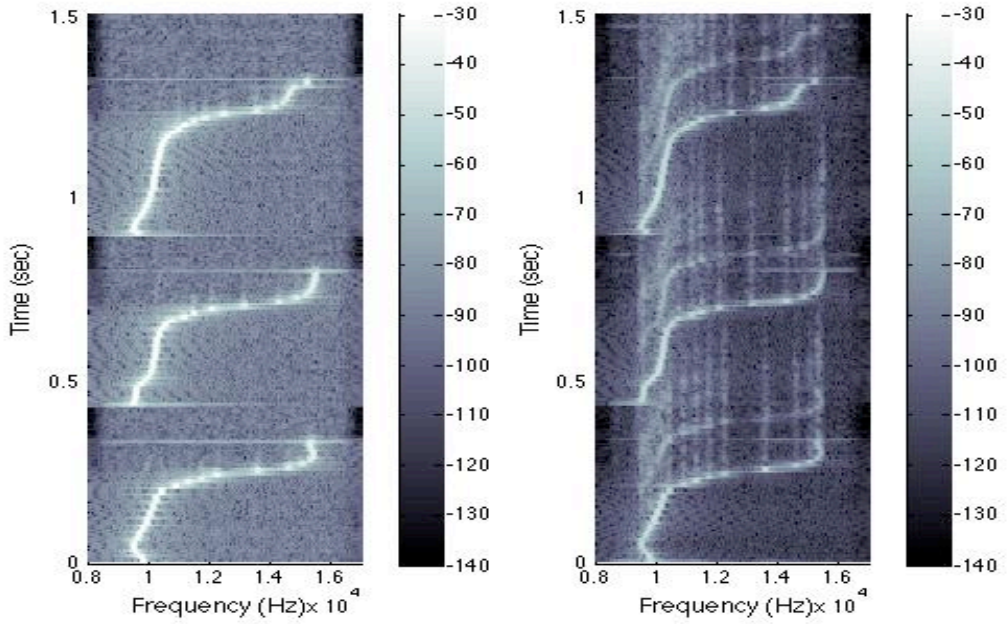

Figure 5-3: Spectrograms of watermarked synthetic whistle calls received at the reference (left) and N1000 (right) hydrophones (dB)

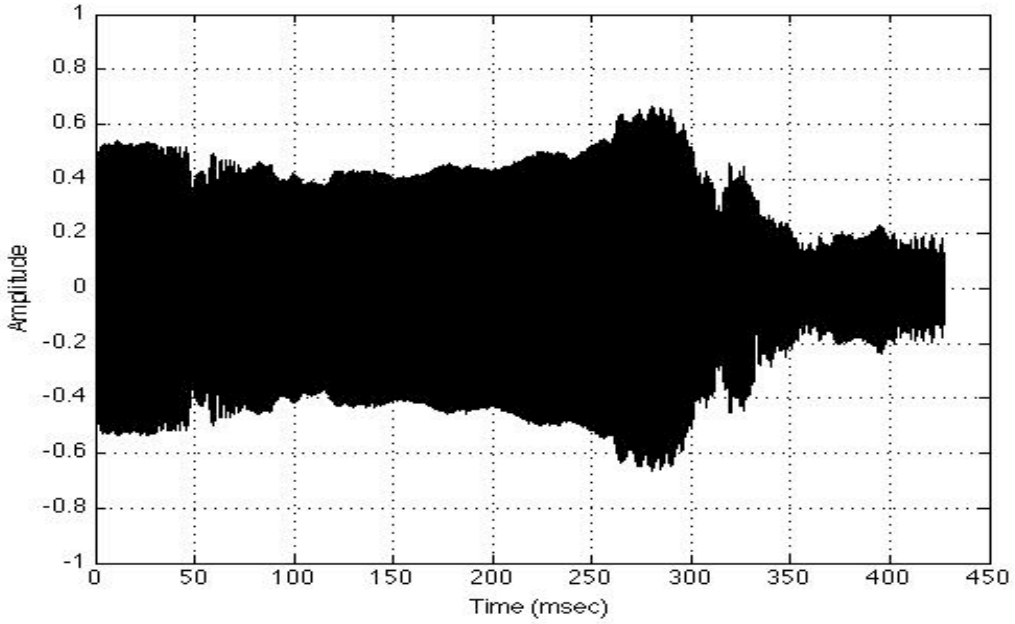

Figure 5-4: Reference hydrophone recording of unmodified Whistle 3 in Fig. 5-2

Fig. 5-4 shows the third whistle (Whistle 3) from Fig. 5-2 as recorded by the reference hydrophone. Due to the frequency response of the ITC-1007 transducer, the amplitude of Whistle 3 varies in time as the frequency changes. The rest of this chapter examines the frequency estimation performance of the HTLS algorithm for both unmodified and watermarked versions of Whistle 3.

Fig. 5-5 compares the frequency estimation performance for both unmodified and watermarked versions of Whistle 3 received by the reference hydrophone, using the parameters $\lambda = 1$, $M = 101$, and $p = 2$. A major drawback of this watermarking scheme is that when the unmodified frequency contour is relatively constant, there is little frequency separation between information bits, and watermark retrieval requires excellent frequency estimation. By using linear chirp segments with abrupt frequency shifts Δf, robust watermark retrieval is possible independent of the unmodified frequency contour.

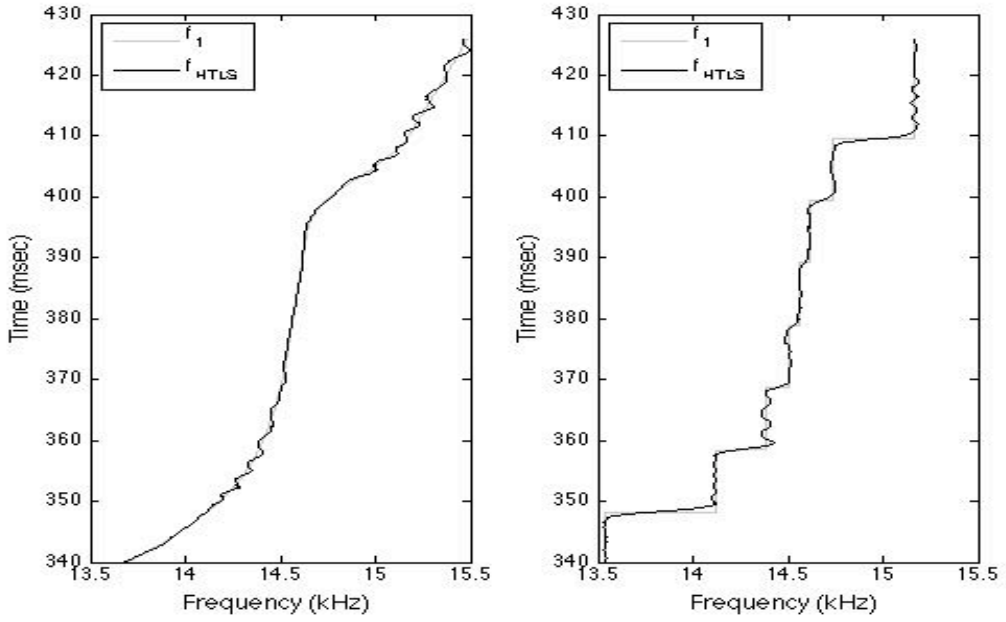

Figure 5-5: Frequency estimation performance for unmodified (left) and watermarked (right) whistle contours received at reference hydrophone

Fig. 5-6 compares the frequency estimation performance for both unmodified and watermarked versions of Whistle 3 received by the N1000 hydrophone, using the parameters $\lambda = 1$, $M = 101$, and $p = 2R$ with up to 3 harmonics. The effect of ISI is combatted by increasing the model order to account for major peaks in the impulse response, yielding good frequency estimation of the transmitted contour. However,

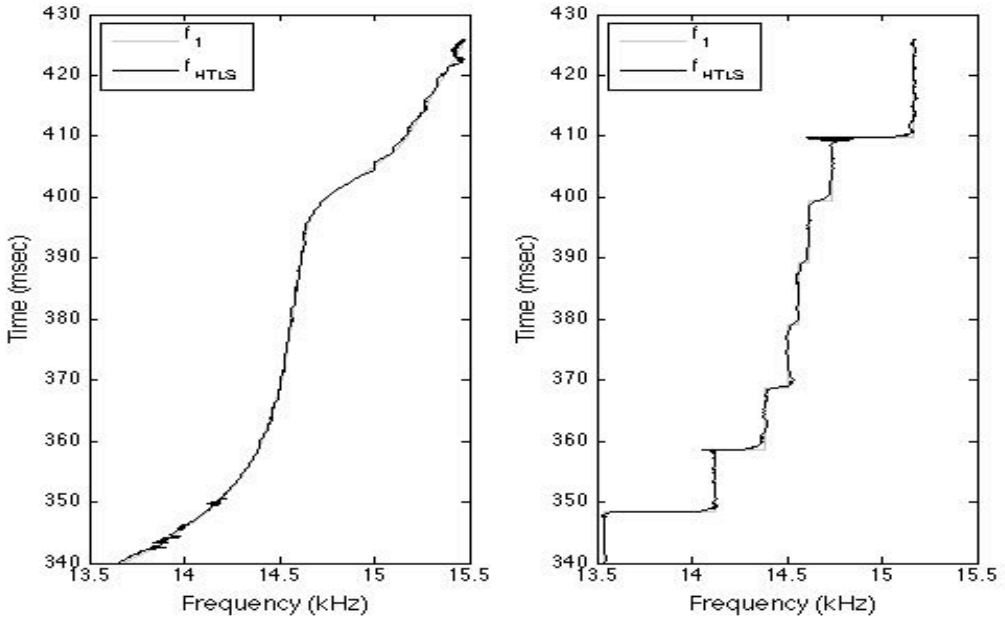

Figure 5-6: Frequency estimation performance for unmodified (left) and watermarked (right) whistle contours received at N1000 hydrophone

overestimating the model order harms the frequency estimation performance, so p was manually adjusted to account for the onset of strong multipath arrivals.

Fig. 5-7 and Fig. 5-8 show the complete estimated frequency contours for unmodified and watermarked versions of Whistle 3 received by the N1000 hydrophone. As seen in Fig. 5-7, ISI can cause sudden spurious frequency estimation results. Discounting the outliers in Fig. 5-7, the standard deviation of the unmodified whistle frequency estimate is 21.6 Hz, while the standard deviation of the watermarked frequency estimate is 20.8 Hz.

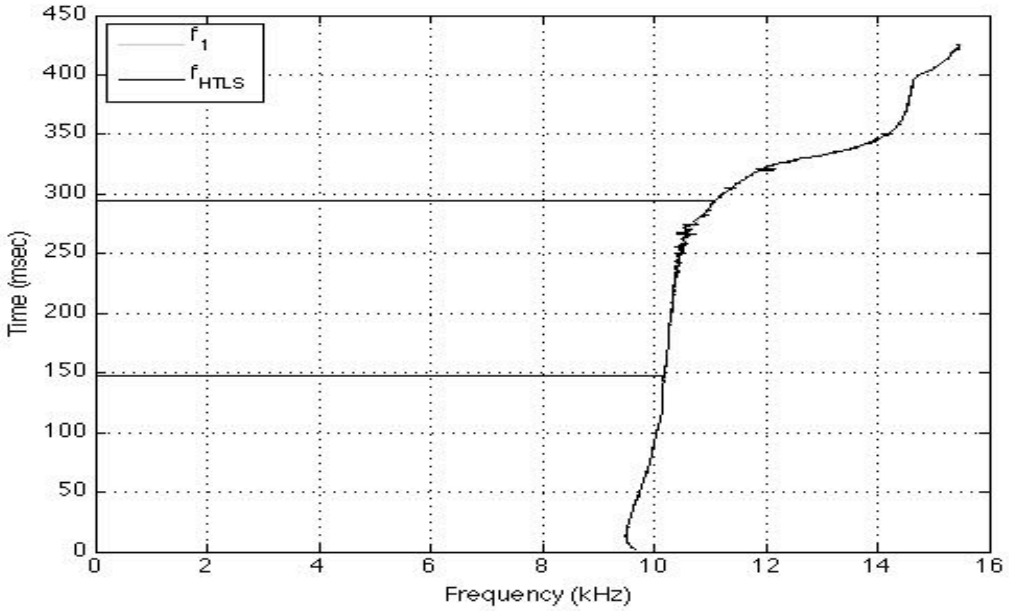

Figure 5-7: Frequency estimation performance for unmodified whistle contour received at N1000 hydrophone

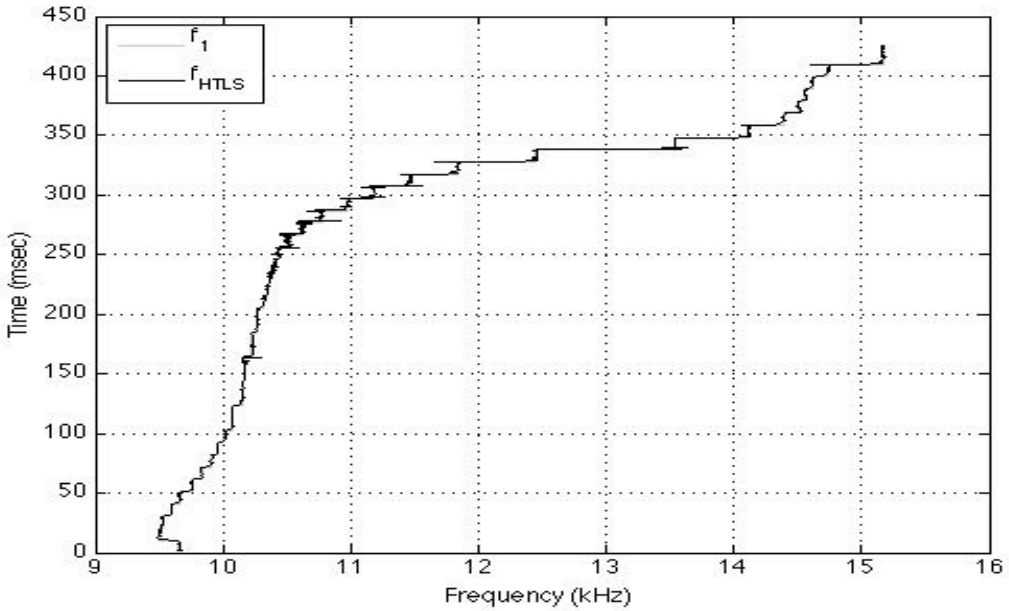

Figure 5-8: Frequency estimation performance for watermarked whistle contour received at N1000 hydrophone

Although the distortion due to ISI presents a challenge to watermark retrieval, it can be overcome in mild environmental conditions with clearly defined multipath arrivals by appropriately increasing the model order used in frequency estimation. In severe environmental conditions, where the multipath arrivals reflected off surface waves are less clearly defined, the frequency estimation performance will degrade. Further testing with the watermarking schemes presented in Section 4.2 should be performed in various environmental and bathymetric conditions to establish the operational limits on robust watermark retrieval.

Chapter 6

Conclusions and Future Directions

The work presented in this thesis develops a method for high-resolution modeling of marine mammal whistle calls that can be used to generate natural sounding synthetic whistles for biological research or covert communications. Although McAulay and Quatieri [46] reported good results in applying their human speech processing sinusoidal model to the synthesis of whale sounds, their technique was based on a block-by-block estimation of slowly-varying parameters. By applying a relatively short sliding window with hop size of $H = 1$, the quickly-varying parameters of chirp signals can be accurately estimated. Essentially, higher resolution estimates are found for the fundamental frequency and amplitude contours used by Buck *et al.* [5] in the modification and synthesis of bottlenose dolphin whistle calls. Due to the sensitivity of the HAS, the optimal scheme for watermarking marine mammal whistle calls is based on slight imperceptible modifications of the fundamental frequency contour. High-resolution frequency estimation is essential for producing natural sounding stego-signals that are robust to channel-induced signal distortion and additive ambient noise.

An interesting result, previously unknown due to the lower resolution of other techniques, is that the bottlenose dolphin whistles exhibit an inherent fluctuating vibrato of the fundamental frequency contour, presumably due to the physical mechanism

for generating whistles. A typical vibrato of the bottlenose dolphin fundamental frequency, ranging from 6 to 22 kHz, has a period of 1 msec with a magnitude from 50 to 100 Hz. The presence and resolvability of the inherent vibrato naturally lead to watermarking the instantaneous mean fundamental frequency contour with a synthetic vibrato using CPM signals.

Directions for future work can be divided into two categories: updating the existing model to better describe marine mammal whistle generation and addressing operational aspects of a covert communications system. The major distinction between these categories is that modeling can performed offline at ideal SNRs, while a covert communications system will optimally operate online at degraded SNRs.

Accurate modeling of marine mammal whistle calls requires high-quality recordings with a high SNR and sufficient sample rate to capture the desired harmonics without aliasing. The custom built suction cup hydrophone, used in the Sarasota Bottlenose Dolphin Whistle Catalog to record whistles during brief capture-release events, provides recordings with excellent SNR. For the whistle recording studied in this thesis, the high frequency harmonics are cutoff above 40 kHz. Optimal recordings should use a high enough sample rate to resolve the desired harmonics and employ anti-aliasing filters to limit whistle distortion. A large number of bottlenose dolphin whistle calls should be analyzed to determine characteristic modulations of the frequency and amplitude contours. If these characteristics can be accurately modeled, natural sounding whistles can be generated from scratch, without requiring a whistle recording to develop frequency and/or amplitude contours. The existing sinusoidal model could be updated to include components of the whistles that are not confined to narrow band harmonics. The apparent stochastic effects of the whistles, such as during the attack or final phases of the whistles, could be modeled in a similar fashion as Levine's sinusoid+noise+transient model [36]. Finally, the bottlenose dolphin vocal tract could be modeled to improve the sinusoidal synthesis model, as shown in Fig. 1-1.

One of the drawbacks for using the HTLS algorithm to track fundamental frequency contours in a covert communications system is the high computational load required to obtain a frequency estimate for each sample. A recursive implementation of the weighted HTLS algorithm, using an appropriate forgetting factor λ to discard old data, would greatly improve the algorithm's computational load for real time applications. Liang [37] discusses using the SVD-update algorithm of Bunch and Nielsen [7] after calculating the initial SVD to reduce the computational loading of sequential chirp parameter estimation. Taking advantage of the state-space model utilized in the HTLS algorithm, an extended Kalman filter [22] could be developed to track parameter changes throughout a whistle call. It would be beneficial to develop a more robust way to deal with channel-induced ISI, such as using the Expectation-Maximization (EM) algorithm [48] to estimate channel conditions and performing channel equalization prior to frequency estimation. It could also turn out that other frequency estimators, such as Quinn's FTI frequency estimator, are a better choice than the HTLS algorithm for watermark detection and retrieval. Quinn [55] combines FTI frequency estimation with a Hidden Markov Model (HMM) to track slowly varying frequencies at low SNR. HMMs could be developed to improve freequency tracking of marine mammal whistle calls at low SNR.

Different watermarking schemes should be tested and compared in terms of their ability to produce natural sounding synthetic stego-signals, potential achievable data rates, and watermark detection and retrieval performance. While this thesis focused on the frequency of estimation of a single marine mammal whistle call, an operational environment at sea will often include actual marine mammal whistle calls in addition to the synthetic stego-signal. Sturtivant and Datta [67] have looked at extracting whistle contours from recordings of several dolphins. An eventual covert communications system will most likely need to be able to overcome acoustic interference from biologics that respond to the natural sounding stego-signals.

Appendix A

Prony's Derivation of the Linear Prediction Equations

Prony demonstrated that the nonlinear aspects of Eq. (2.3),

$$x[n] = \sum_{k=1}^{p} h_k z_k^{n-1} \quad , \tag{A.1}$$

can be embedded into a polynomial factorization problem [43]. He showed that the poles z_k can be resolved separately from the parameters h_k, which can then be found by solving Eq. (2.6). The key to the separation is to recognize that Eq. (A.1) is the solution to a homogeneous linear constant-coefficient difference equation. In order to find the form of this difference equation, first define the polynomial $\phi(z)$ that has the poles z_k as its roots,

$$\phi(z) = \prod_{k=1}^{p} (z - z_k) \quad . \tag{A.2}$$

If the products of Eq. (A.2) are expanded into a power series, the polynomial may be represented as the summation,

$$\phi(z) = \sum_{m=0}^{p} w[m] z^{p-m} \quad , \tag{A.3}$$

with complex coefficients $w[m]$ such that $w[0] = 1$. Shifting the index in Eq. (A.1) from n to $n - m$ and multiplying by the parameter $w[m]$ yields

$$w[m]x[n-m] = w[m]\sum_{k=1}^{p} h_k z_k^{n-m-1} \quad . \tag{A.4}$$

Forming similar products $w[0]x[n], \ldots, w[p]x[n-p]$ and summing produces

$$\begin{aligned}\sum_{m=0}^{p} w[m]x[n-m] &= \sum_{m=0}^{p} w[m]\sum_{k=1}^{p} h_k z_k^{n-m-1} \\ &= \sum_{k=1}^{p} h_k \sum_{m=0}^{p} w[m] z_k^{n-m-1} \quad ,\end{aligned} \tag{A.5}$$

which is valid for $p+1 \leq n \leq 2p$. Making the substitution $z_k^{n-m-1} = z_k^{n-p-1} z_k^{p-m}$,

$$\begin{aligned}\sum_{m=0}^{p} w[m]x[n-m] &= \sum_{k=1}^{p} h_k z_k^{n-p-1} \sum_{m=0}^{p} w[m] z_k^{p-m} \\ &= \sum_{k=1}^{p} h_k z_k^{n-p-1} \phi(z)\Big|_{z=z_k} = 0 \quad .\end{aligned} \tag{A.6}$$

Eq. (A.6) is the linear difference equation whose homogeneous solution is given by Eq. (A.1). Eq. (A.3) is the *characteristic equation* associated with this linear difference equation. The set of valid linear prediction equations is expressed as

$$\begin{bmatrix} x[p] & x[p-1] & \ldots & x[1] \\ x[p+1] & x[p] & \ldots & x[2] \\ \vdots & \vdots & \ddots & \vdots \\ x[2p-1] & x[2p-2] & \ldots & x[p] \end{bmatrix} \begin{bmatrix} w[1] \\ w[2] \\ \vdots \\ w[p] \end{bmatrix} = - \begin{bmatrix} x[p+1] \\ x[p+2] \\ \vdots \\ x[2p] \end{bmatrix} . \tag{A.7}$$

Although it is derived from different assumptions, the modern Prony's method, which accounts for error in the model, is equivalent to the covariance method of linear prediction [41].

Bibliography

[1] Theagenis J. Abatzoglou, Jerry M. Mendel, and Gail A. Harada. The constrained total least squares technique and its applications to harmonic superresolution. *IEEE Transactions on Signal Processing*, 39(5):1070:1086, May 1991.

[2] Luis B. Almeida and Fernando M. Silva. Variable-frequency synthesis: An improved harmonic coding scheme. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 27.5.1–4, 1984.

[3] K. Beeman. Digital signal analysis, editing, and synthesis. In S. L. Hopp, M. J. Owren, and C. S. Evans, editors, *Animal Acoustic Communication*, pages 59–103. Springer, Berlin, 1998.

[4] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35(3&4):313–336, 1996.

[5] John R. Buck, Hugh B. Morgenbasser, and Peter L. Tyack. Synthesis and modification of the whistles of the bottlenose dolphin, *Tursiops truncatus*. *Journal of the Acoustical Society of America*, 108(1):407–416, July 2000.

[6] John R. Buck and Peter L. Tyack. A quantitative measure of similarity for *Tursiops truncatus* signature whistles. *Journal of the Acoustical Society of America*, 94(5):2497–2506, November 1993.

[7] J. R. Bunch and C. P. Nielsen. Updating the singular value decomposition. *Numerische Mathematik*, 31:111–129, 1978.

[8] Brian Chen and Gregory W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transaction on Information Theory*, 47(4):1423–1443, May 2001.

[9] Kevin G. Christian. *Generic Compression and Recall of Signals with Application to Dolphin Whistles*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, September 1993.

[10] Leon Cohen. *Time-Frequency Analysis*. Prentice Hall Signal Processing Series. Prentice Hall PTR, Englewood Cliffs, NJ, 1995.

[11] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, Jessica Fridrich, and Ton Kalker. *Digital Watermaking and Steganography*. The Morgan Kaufmann Series in Multimedia Information and Systems. Morgan Kaufmann Publishers, Burlington, MA, second edition, 2008.

[12] Nedeljko Cvejic and Tapio Seppänen, editors. *Digital Audio Watermarking Techniques and Technologies : Applications and Benchmarks*. Information Science Reference, Hershey, PA, 2008.

[13] Bart De Moor. Total least squares for affinely structured matrices and the noisy realization problem. *IEEE Transactions on Signal Processing*, 42(11):3104–3113, November 1994.

[14] Baron (Gaspard Riche) de Prony. Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures. *Journal de l'Ecole Polytechnique*, 1(2):24–76, 1795.

[15] John R. Deller, Jr., John G. Proakis, and John H. L. Hansen. *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company, New York, NY, 1993.

[16] Petar M. Djurić and Steven M. Kay. Parameter estimation of chirp signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(12):2118–2126, December 1990.

[17] Paulo A. A. Esquef and Luiz W. P. Biscainho. Spectral based analysis and synthesis of audio signals. In Hector Perez-Meana, editor, *Advances in Audio and Speech Signal Processing: Technologies and Applications*, chapter 3, pages 56–92. Idea Group Publishing, Hershey, PA, 2007.

[18] Gene H. Golub and Charles F. Van Loan. An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17(6):883–893, December 1980.

[19] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins Series in the Mathematical Sciences. The John Hopkins University Press, Baltimore, MD, third edition, 1996.

[20] Fredric J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, January 1978.

[21] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., 1996.

[22] Simon Haykin. *Adaptive Filter Theory*. Prentice Hall Information and System Sciences Series. Prentice-Hall, Inc., Upper Saddle River, NJ, fourth edition, 2002.

[23] Xing He and Michael Scordilis. Spread spectrum for digital audio watermarking. In Nedeljko Cvejic and Tapio Seppänen, editors, *Digital Audio Watermarking Techniques and Technologies : Applications and Benchmarks*, chapter 2, pages 11–49. Information Science Reference, Hershey, PA, 2008.

[24] Xiaozhou Huang. Autoregressive synthesis of bottlenose dolphin (*Tursiops Truncatus*) whistles. Master's thesis, University of Massachusetts Dartmouth, Dartmouth, MA, July 2002.

[25] M. H. Kahn, M. S. Mackisack, M. R. Osborne, and G. K. Smyth. On the consistency of Prony's method and related algorithms. *Journal of Computational and Graphical Statistics*, 1(4):329–349, December 1992.

[26] Stefan Katzenbeisser. Principles of steganography. In Stefan Katzenbeisser and Fabien A.P. Petitcolas, editors, *Information Hiding Techniques for Steganography and Digital Watermarking*, chapter 2, pages 17–41. Artech House, Inc., Norwood, MA, 2000.

[27] Stefan Katzenbeisser and Fabien A.P. Petitcolas, editors. *Information Hiding Techniques for Steganography and Digital Watermarking*. Artech House Computer Security Series. Artech House, Inc., Norwood, MA, 2000.

[28] Steven M. Kay. *Modern Spectral Estimation*. Prentice-Hall Signal Processing Series. Prentice-Hall, Inc., Upper Saddle River, NJ, 1988.

[29] Daniel B. Kilfoyle and Arthur B. Baggeroer. The state of the art in underwater acoustic telemetry. *IEEE Journal of Oceanic Engineering*, 25(1):4–27, January 2000.

[30] Sridhar Krishnan, Behnaz Ghoraani, and Serhat Erkucuk. Time-frequency analysis of digital audio watermarking. In Nedeljko Cvejic and Tapio Seppänen, editors, *Digital Audio Watermarking Techniques and Technologies : Applications and Benchmarks*, chapter 9, pages 187–204. Information Science Reference, Hershey, PA, 2008.

[31] Ramdas Kumaresan, Donald W. Tufts, and Louis L. Scharf. A Prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models. *Proceedings of the IEEE*, 72(2):230–233, February 1984.

[32] Martin Kutter and Frank Hartung. Introduction to watermarking techniques. In Stefan Katzenbeisser and Fabien A.P. Petitcolas, editors, *Information Hiding Techniques for Steganography and Digital Watermarking*, chapter 5, pages 97–120. Artech House, Inc., Norwood, MA, 2000.

[33] Philippe Lemmerling, Ioannis Dologlou, and Sabine van Huffel. Speech compression based on exact modeling and structured total least norm optimization. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 353–356, Seattle, WA, May 1998.

[34] Philippe Lemmerling and Sabine van Huffel. Analysis of the structured total least squares problem for hankel/toeplitz matrices. *Numerical Algorithms*, 27:89–114, 2001.

[35] Philippe Lemmerling and Sabine van Huffel. Structured total least squares: Analysis, algorithms, and applications. In Sabine van Huffel and Philippe Lemmerling, editors, *Total Least Squares and Errors-In-Variables Modeling*, pages 79–91. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[36] Scott Nathan Levine. *Audio representations for data compression and compressed domain processing*. PhD thesis, Stanford University, Stanford, CA, 1998.

[37] R. M. Liang and K. S. Arun. Parameter estimation for superimposed chirp signals. In *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 273–276, 1992.

[38] Yi-Wen Liu. Audio watermarking through parametric synthesis models. In Nedeljko Cvejic and Tapio Seppänen, editors, *Digital Audio Watermarking Techniques and Technologies : Applications and Benchmarks*, chapter 3, pages 50–81. Information Science Reference, Hershey, PA, 2008.

[39] M. S. Mackisack, M. R. Osborne, and G. K. Smyth. A modified Prony algorithm for estimating sinusoidal frequencies. *Journal of Statistical Computation and Simulation*, 49:111–124, 1994.

[40] Malcolm D. Macleod. Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones. *IEEE Transactions on Signal Processing*, 46(1):141–148, January 1998.

[41] John D. Markel and Augustine H. Gray, Jr. *Linear Prediction of Speech*, volume 12 of *Communication and Cybernetics*. Springer-Verlag, Germany, 1976.

[42] John Markhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, April 1975.

[43] S. Lawrence Marple, Jr. *Digital Spectral Analysis: with applications*. Prentice-Hall Signal Processing Series. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1987.

[44] Nicola Mastronardi, Philippe Lemmerling, and Sabine van Huffel. Fast structured total least squares algorithms via exploitation of the displacement structure. In Sabine van Huffel and Philippe Lemmerling, editors, *Total Least Squares*

and Errors-In-Variables Modeling, pages 93–106. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[45] MATLAB. *Signal Processing Toolbox User's Guide*. The MathWorks, Inc., 2008.

[46] Robert J. McAulay and Thomas F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-34(4):744–754, August 1986.

[47] Maciej Niedźwiecki. *Identification of Time-Varying Processes*. John Wiley & Sons, Ltd, West Sussex, England, 2000.

[48] Mauri Nissilä and Subbarayan Pasupathy. Adaptive bayesian and EM-based detectors for frequency-selective fading channels. *IEEE Transactions on Communications*, 51(8):1325–1336, August 2003.

[49] Alan V. Oppenheim and Ronald W. Schafer with John R. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, second edition, 1999.

[50] Fabien A.P. Petitcolas. Introduction to information hiding. In Stefan Katzenbeisser and Fabien A.P. Petitcolas, editors, *Information Hiding Techniques for Steganography and Digital Watermarking*, chapter 1, pages 1–14. Artech House, Inc., Norwood, MA, 2000.

[51] James C. Preisig. Acoustic propagation considerations for underwater acoustic communications network development. *SIGMOBILE Mobile Computing and Communications Review*, 11(4):2–10, 2007.

[52] John G. Proakis. *Digital Communications*. Irwin/McGraw-Hill, fourth edition, 2000.

[53] Thomas F. Quatieri and Robert J. McAulay. Speech transformations based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-34(6):1449–1464, December 1986.

[54] B. G. Quinn and J.M. Fernandes. A fast efficient technique for the estimation of frequency. *Biometrika*, 78:489–498, 1991.

[55] B. G. Quinn and E. J. Hannan. *The Estimation and Tracking of Frequency*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, United Kingdom, 2001.

[56] Md. Anisur Rahman and Kai-Bor Yu. Total least squares approach for frequency estimation using linear prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(10):1440–1454, October 1987.

[57] David C. Rife and Robert R. Boorstyn. Single-tone parameter estimation from discrete-time observations. *IEEE Transactions on Information Theory*, IT-20(5):591–598, September 1974.

[58] David C. Rife and Robert R. Boorstyn. Multiple tone parameter estimation from discrete-time observations. *The Bell System Technical Journal*, 55(9):1389–1410, November 1976.

[59] David C. Rife and G. A. Vincent. Use of the discrete fourier transform in the measurement of frequencies and levels of tones. *Bell System Technical Journal*, 49:197–228, 1970.

[60] J. Ben Rosen, Haesun Park, and John Glick. Total least norm formulation and solution for structured problems. *SIAM Journal on Matrix Analysis and Applications*, 17(1):110–126, January 1996.

[61] Supratim Saha and Steven M. Kay. Maximum likelihood parameter estimation of superimposed chirps using Monte Carlo importance sampling. *IEEE Transactions on Signal Processing*, 50(2):224–230, February 2002.

[62] Laela S. Sayigh, H. Carter Esch, Randall S. Wells, and Vinvent M. Janik. Facts about signature whistles of bottlenose dolphins, *Tursiops truncatus*. *Animal Behaviour*, 74(6):1631–1642, December 2007.

[63] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, AP-34(3):276–280, March 1986.

[64] Xavier Serra and Julius Smith, III. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4):12–24, Winter 1990.

[65] Julius Smith, III and Xavier Serra. PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation. In *Proceedings of the 1987 Computer Music Conference*, San Francisco, CA, 1987. Computer Music Association.

[66] Gordon K. Smyth and Douglas M. Hawkins. Robust frequency estimation using elemental sets. *Journal of Computational and Graphical Statistics*, 9(1):196–214, March 2000.

[67] C. Sturtivant and S. Datta. Techniques to isolate dolphin whistle and other tonal sounds from background noise. *Acoustic Letters*, 18(10):189–193, 1995.

[68] Donald W. Tufts and Ramdas Kumaresan. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. *Proceedings of the IEEE*, 70(9):975–989, September 1982.

[69] Donald W. Tufts and Ramdas Kumaresan. Singular value decomposition and improved frequency estimation using linear prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-30(4):671–675, August 1982.

[70] Sabine van Huffel, L. Aerts, J. Bervoets, J. Vandewalle, C. Decanniere, and P. van Hecke. Improved quantitative time-domain analysis of NMR data by total least squares. In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, *Signal Processing VI: Theories and Applications*, volume III, pages 1721–1724. Elsevier Science Publishers, Amsterdam, The Netherlands, 1992.

[71] Sabine van Huffel, Hua Chen, Caroline Decanniere, and Paul van Hecke. Algorithm for time-domain NMR data fitting based on total least squares. *Journal of Magnetic Resonance*, A 110:228–237, 1994.

[72] Sabine van Huffel, Haesun Park, and J. Ben Rosen. Formulation and solution of structured total least norm problems for parameter estimation. *IEEE Transactions on Signal Processing*, 44(10):2464–2474, October 1996.

[73] Sabine van Huffel and Joos Vandewalle. *The Total Least Squares Problem : Computational Aspects and Analysis*, volume 9 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1991.

[74] Norbert Wiener. *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. M.I.T. Press, Cambridge, MA, 1949.

[75] Arie Yeredor. The extended STLS algorithm for minimizing the extended LS criterion. In Sabine van Huffel and Philippe Lemmerling, editors, *Total Least Squares and Errors-In-Variables Modeling*, pages 101–117. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.