Accurate Fourier Coefficient and Spectrum Search Algorithm for Short Duration Signals

Mark E. M. Stewart, Member, IEEE

Abstract—The algorithm presented accurately finds the Fourier coefficients, or spectrum, of discretely sampled, short duration signals. It avoids the FFT’s leakage errors. The algorithm uses a classical search approach, namely, bracketing and searching for the frequency, magnitude, and phase of the Fourier modes that minimize signal power—an application of non-linear least squares. The robustness of the procedure is analyzed, showing that bracketing captures only the global minimum and the frequency interval of attraction increases with shorter sampling intervals. The algorithm’s operation count and cost are demonstrated. In noiseless test cases, the algorithm can routinely find signal modes to machine precision without leakage. It is validated and compared with the Burgess-Lanczos method. The algorithm is demonstrated by resolving a note from a string instrument and phonemes in human speech. The algorithm has superior accuracy for signals with many modes, and it is valuable for precisely finding the frequency spectrum during a short time interval within a non-stationary signal when the number of signal samples is limited.

Index Terms—Least Squares Methods, Parameter Estimation, Signal Analysis, Speech Processing, Frequency Estimation, Phase Estimation

Nomenclature

\( C_f \) Fourier coefficient (mode magnitude) at frequency \( f \)

\( f \) Frequency of a mode in the signal, Hz

\( f_s \) Uniform sampling frequency of signal, Hz

\( i \) Unit imaginary number

\( L \) Convergence metric, dB

\( M \) Number of cancelling modes

\( N \) Number of DFT samples in sampling interval, \( N = T f_s \)

\( N_f \) Number of DFT \( f \) samples at test frequency

\( p \) Integer power

\( P \) Signal sample value precision in bits

\( r(t) \) Gaussian random noise signal

\( s(t) \) Continuous function or signal (single channel)

\( S_k \) Sample value \( k \) from discretely sampling signal, \( s(t) \)

\( S_k^{\prime} \) Sample value \( k \) from residual signal, \( M \) modes removed

\( S_k^{\prime \prime} \) Sample value \( k \) of polishing signal for mode \( J \)

\( t \) Time, s, also \( t_s = k f_s \)

\( T \) Integration time or sampling interval duration, \( T = N f_s \), s

\( T^f \) Integration time at test frequency for DFT, s

\( \alpha \) Angular frequency, \( \alpha = 2 \pi f \), radians/s

\( \Delta t \) Time step between samples, \( \Delta t = 1/f_s \), s

\( \alpha, \delta \) Perturbation parameters for frequency, magnitude

\( \phi \) Perturbation parameter for phase

\( \Phi \) Function to be minimized

\( \theta_j \) Phase angle of mode that ensures \( C_f \) is real, radians

\( \lfloor x \rfloor \) Floor function, \( x \)

\( j \) Integer index of cancelling mode, from set of \( M \)

\( J \) Integer index of cancelling mode polishing focus

\( k \) Integer index of discrete time, \( k \Delta t \); also a general index

\( \ell \) Integer index of FFT mode, \( f_j = \ell f_s / N \)

\( ^{\wedge} \) Accent indicates complex value, otherwise real

I. INTRODUCTION

Finding the frequency and magnitude of a signal’s modes is a common problem in science and engineering. Currently, the Fast Fourier Transform (FFT) is a commonly used method. The FFT accurately represents the discrete signal and is very robust and efficient. Yet, in general, it does not provide precise spectrum information, including Fourier coefficients of the underlying signal modes.

The FFT suffers from ‘leakage’—additional modes in the spectrum beyond the underlying signal itself. For long duration, stationary signals, a sufficient number of signal samples exist so that FFT results can be averaged to better resolve frequencies. However, non-stationary signals do not have the time duration and number of samples, \( N \), to achieve frequency resolution with an FFT.

The current paper demonstrates a robust algorithm for accurately finding an auditory signal’s Fourier coefficients and spectrum. Leakage errors are avoided, and the method can be used in a short interval when an auditory signal is nearly stationary and the number of signal samples is limited. The algorithm is compared with the Burgess-Lanczos method [1].

The algorithm can be valuable beyond auditory signals, in particular, when the precise frequency, magnitude, or phase of signal modes is required.

Section I sets the background and notation for non-specialists. Section II outlines the algorithm. Section III is a perturbation analysis of bracketing. Section IV concerns important properties, and Section V provides examples.

A. Background

In an early example of spectrum analysis from the 1860’s, Helmholtz [2] used Helmholtz resonators of glass and later metal to detect individual tones in musical instruments, speech, and other vibrating objects. A Helmholtz resonator...
uses sympathetic resonance to detect vibration frequencies. Helmholtz explained the fundamental tone and ‘harmonic upper partial tones’ by Fourier series.

Fourier [3] used his namesake series to solve the differential equation for heat flow in objects. Fourier’s contribution was to recognize that a series of sine/cosine modes can represent almost all functions and signals. Further, these series form solutions to important differential equations, including the wave and heat equations.

Importantly, the wave equation models the underlying physics of many sound sources, plus sound transmission. A string, drum, or beam’s vibration are classic examples; sound propagation is another. Linearity of the equation and constant coefficients allow superposition of modes. Hence, the solutions to the wave equation are a sum of sine/cosine modes, and Fourier series represent the vibrations of many sound sources and their propagation. Yet, for those cases (fricative, plosive phonemes) where the physics of the sound source does not result in a wave equation model, Fourier series can still represent the signal.

Fourier series were first calculated manually, and a need has existed for accurate machines and computationally efficient algorithms that automate calculation. Michelson and Stratton [4] devised a mechanical harmonic analyzer and synthesizer that improved the accuracy of Lord Kelvin’s machine for analyzing tides [5].

In the earliest years of digital computers, when computers were at least three orders of magnitude slower than today, the Fast Fourier Transform (FFT) was devised to give the Fourier coefficients of a discretely sampled signal. While Cooley and Tukey [6] brought the FFT into wide use in software and hardware, others had independently developed and used similar techniques. Cooley [7] provides a history. Soon after Cooley and Tukey’s paper, FFT based spectrum analyzers replaced [8] frequency shifting a signal into a band-pass filter. Contemporary research includes improving discrete Fourier transform (DFT) performance using a reduced distribution of signal samples [9].

One problem in finding a signal’s spectrum is that DFTs usually introduce ‘leakage’—additional modes in the spectrum beyond the underlying signal modes. Fig. 1 shows a signal with three modes and its FFT. The mode at \( f = 172.2656 \) Hz corresponds to an FFT frequency, \( 4f/N \), and is found by the FFT without leakage. The mode magnitude is found exactly and the neighboring Fourier coefficients are at the background spectrum. The two other modes are not at FFT frequencies, and they show a local peak, but they also show leakage or non-zero Fourier coefficients across the spectrum. The current algorithm avoids the problem of leakage, since it can search at frequencies between the fixed FFT frequencies.

![Signal and FFT](image1)

**Fig. 1.** Noiseless signal (upper) and its FFT (lower) consists of three modes \( f = 172.2656 \) Hz, 635 Hz, and 4000 Hz; \( C_1 = 1.0, 0.05, 0.7; \theta_1 = 0.3\pi, 0.2\pi, 0.5\pi \). The mode at \( f = 172.2656 \) Hz corresponds to an FFT mode frequency and is found by the FFT without leakage. The two other modes, not at FFT mode frequencies, show leakage or non-zero Fourier coefficients spread across the spectrum (lower). Green line above is residual signal, \( \sum^n \); blue histogram lines below are the Fourier modes found by the current algorithm. Throughout the paper, unless stated otherwise, signals are sampled at \( f_s = 22.05 \text{ kHz}, N = 512, T = 23.22 \text{ ms}, P = 32\) bit resolution; signals are assumed to be band limited and without aliasing.

![Frequency Extension](image2)

**Fig. 2.** Periodic extension of a function (upper) and after Hanning filtering (lower). The function is periodic, but not of period \( T \). A discontinuity is created above, and the filter removes it by distorting the signal. A slope discontinuity can also be created in a periodic extension.
Leakage is not an error; the FFT does exactly what it should do. FFT modes are all periodic with period \( T \); consequently the series must also be periodic with period \( T \). However, if the sampled signal is not periodic with this period, a discontinuity is introduced (Fig 2). In particular, when the signal is periodically extended beyond the interval and has a discontinuity or slope discontinuity, the FFT places a discontinuity where intervals meet—a departure from the intended signal [1]. Data windowing functions reduce FFT leakage by reducing this discontinuity; the sample values are zeroed toward the interval ends, but the signal is distorted.

A statistical estimation approach to reducing the influence of leakage and noise, thereby improving an FFT spectrum, is to find an average value: divide a stationary, long duration signal into sub-intervals, apply an FFT to each sub-interval, and average the magnitude squared Fourier coefficients, \( |\hat{C}_f| \), at each frequency; in this manner, the variance of signal’s power spectrum (periodogram) is reduced [10] [11]. This procedure is often attributed to Bartlett [12]. For example, Lewicki [13] averaged fifty 1024-point FFTs to resolve the vibration spectrum of a helicopter transmission. For short duration signals, insufficient data exists for this approach.

The Burgess-Lanczos method [1] [14] takes a different approach to circumvent leakage. Instead of minimizing leakage, the method assumes an underlying mode exists and approximates it from adjacent FFT mode parameters.

Other methods exist, beyond the FFT. Instead of a Fourier transform, non-linear least squares (NLLS) and a minimization search can be used to find signal parameters. Friedlander and Francos [15] present theoretical analyses of using NLLS to estimate signal parameters. Angeby [16] applies NLLS and the Gauss-Newton method to polynomial phase signals (PPS), and notes that searching on small, sliding intervals eases numerical problems. For signal parameter estimation in the presence of noise, NLLS is considered statistically efficient. Further, it can be applied to non-uniformly sampled signals.

The all-poles method (maximum entropy method (MEM), autoregressive model (AR), linear predictive coding (LPC)) is another approach to finding spectral content. By using a rational function (1) with poles, sharp spectral features can occur at real frequencies and represent discrete modes in a signal. The method can produce remarkable results, but it is considered sensitive to the model’s order and captures noise.

In speech recognition [17] [18], the MEM is a popular method of representing a short duration signal due to its speed and simplicity.

\[
a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_p z^{-p} = 0.
\]

Since the 1980’s, wavelets [19] have been used to represent functions as a series, with a transform to find the series—not unlike Fourier series. Wavelets are appropriate for non-stationary and non-periodic signals, as well as functions with discontinuities and sharp peaks. A fast, recursive algorithm [20] exists for finding the discrete wavelet transform.

In electrical hardware, the Phase-Locked Loop (PLL) finds a frequency in a signal. In demodulating an FM radio signal, the frequency is found by a PLL or the Foster-Seeley discriminator.

The human cochlea is a spectrum analyzer for auditory signals; Helmholtz [2] recognized this. Von Békésy [21] demonstrated that sound frequencies resonate at different points along the cochlear partition since decreasing width changes its resonant frequency. Measurements indicate individual frequencies are not sharply resolved in the feline cochlea [22], although for some echo locating bats with long duration calls, resolution may be an order of magnitude sharper.

B. Fourier Series, Coefficients, and Transform

The Fourier series and transform are outlined here for completeness and clarity of notation. The Fourier series is a representation of a continuous function or signal, \( \hat{s}(t) \). The Fourier series is expressed in terms of Fourier coefficients, \( \hat{C}_f \), and modes, \( e^{2\pi if_1 t} = \cos(2\pi f_1 t) + i\sin(2\pi f_1 t) \), on \([0,T]\) as,

\[
\hat{s}(t) \approx \sum \hat{C}_f e^{2\pi if_1 t}.
\]Modes come in pairs, \( \pm f_1 \). Here we focus on real signals, \( s(t) \), and the real part of the series, namely

\[
s(t) \approx \sum C_f \cos(2\pi f_1 t + \theta_f) \text{, } C_f = 2|\hat{C}_f|, \quad f_1 \geq 0 \quad (2b)
\]

A constant, real phase, \( \theta_f \), is introduced to make magnitude, \( C_f \), real.

By the Fourier Expansion Theorem [23], the series in (2) will converge to \( s(t) \) in an interval, if \( s(t) \) is piecewise continuous in the interval with a finite number of corners, that is, slope discontinuities. \( s(t) \) must also be periodic outside the interval (a sufficient condition, but not necessary). The coefficients \( \hat{C}_f \) are given by (3) in the interval \([0, T]\).

The wave equation describes the vibration of stringed instruments and drums and is solved mathematically with a Fourier series. Consequently, a few harmonic modes describe the sound. Similarly, the vibrating human vocal folds generate harmonic modes. Many objects and machines have natural frequencies—due to their shape, restraint, and elastic properties—which define their resonance and vibration in terms of Fourier modes.

In contrast, the fricative phonemes of human speech (voiceless ‘s’ and ‘f’) are created by turbulent airflow between the tongue and teeth; there are multiple flowpath locations where flow separation can occur. Turbulent airflow involves transient eddies of varying sizes that produce sounds at random frequencies within a frequency range. The mathematical analysis of this sound source does not yield Fourier series solutions, but with enough modes and the Fourier Expansion Theorem, sound representation is possible. Similarly, Fourier series represent the plosive phonemes of human speech (‘b’, ‘p’, ‘t’) and their abruptly starting air flow
and pressure wave. Section V gives examples.

The Fourier Transform is a way of finding Fourier coefficients and the series. Under appropriate conditions, the Fourier coefficients, \( \hat{C}_f \), of a signal, \( s(t) \), are given by,

\[
\hat{C}_f = \frac{1}{T} \int_0^T e^{i2\pi ft} s(t) \, dt
\]

where \( T \) is an integration time that must be an integer number of periods at the frequency, \( f \). Further, \( s(t) \) must be periodic in the interval \([0,T]\) or zero outside of it—otherwise leakage can result.

The FFT is a practical method of efficiently finding Fourier coefficients using a discrete approximation to the Fourier Transform (3).

\[
\hat{C}_f \approx \frac{1}{N} \sum_{k=0}^{N-1} e^{2\pi i f t_k} s(t_k)
\]

where \( t_k = \frac{k}{N} T, \ell = 0, \pm \frac{N}{2} \), and \( N = 2^p \).

The function, \( s(t) \), is uniformly sampled at a frequency, \( f_s \). The FFT frequencies, \( f^\prime \), are fixed by the sampling frequency, \( f_s \), and the number of samples, \( N \); in particular, \( f^\prime = f_s, f^\prime = f_s/2 \), plus \( \frac{N}{2} - 1 \) harmonics. Importantly, this choice of \( f^\prime \) (with \( N = 2^p \)) ensures an integral number of periods of the FFT mode within the sampled data.

Being constrained to these frequencies allows efficient calculation of coefficients for large \( N \). The FFT is efficient because it takes only \( O(N \log_2 N) \) operations to calculate these Fourier coefficients at \( N/2 + 1 \) frequencies. In particular, repeated evaluations of \( e^{2\pi i ft} s(t) \) occur for different frequencies, but the algorithm elegantly orders operations to eliminate these repetitions. Further, evaluation of the trigonometric functions in \( \exp(2\pi i ft/N) \) is simplified.

### C. FFT Leakage and the Burgess-Lanczos Method

Yet another way to view leakage is to calculate the Fourier coefficients, \( C_i \), of a single mode, \( \cos 2\pi f \mu t \), where \( f \mu = \mu f_s, n \). Using (3), leakage in the continuous Fourier transform can be expressed in closed form as,

\[
C_i = 2 \frac{1}{T} \cos 2\pi f \mu t \cos 2\pi f t \, dt ; \mu \neq \ell
\]

For integral \( \mu \), the mode \( \cos 2\pi f \mu t \) is periodic with period \( T \), and \( C_i = \delta_{\mu \ell} \), that is, there is no leakage (\( \delta \) is the Kronecker delta). For real \( \mu \) between \( \ell \) and \( \ell + 1 \), that is, \( \mu = \ell + \epsilon \), for a frequency perturbation \( \epsilon \), the mode is not periodic with period \( T \). Non-zero Fourier coefficients occur across the spectrum. Fig. 3 graphs (5) for the modes in Fig. 1 and compares with FFT results.

The Burgess-Lanczos method circumvents leakage by estimating a mode from adjacent FFT modes. It assumes the signal contains a single mode, \( \cos 2\pi f \mu t \), and uses (5) to approximate the unknown mode’s frequency, magnitude, and phase. In particular, the adjacent Fourier (or FFT) coefficients \( \hat{C}_i \) and \( \hat{C}_{\ell+1} \) are assumed to dominate (\( \ell \leq \mu < \ell + 1 \)), as the other coefficients quickly become small. From FFT results, the mode may be approximated [1] by,

\[
\varepsilon = \left| \frac{\hat{C}_i}{\hat{C}_i + \hat{C}_{\ell+1}} \right| ; \quad f_\mu = (\ell + \varepsilon) \frac{f_s}{N}
\]

\[
\hat{C}_i = \frac{2\sin(\pi \ell/N)}{A_\varepsilon \sin \pi \varepsilon} \hat{C}_{\ell+1} ; \quad \theta_\mu = \theta_{\ell+1} + \frac{\pi(N-1)\varepsilon}{N}
\]

where \( A_\varepsilon \) is a normalization, typically \( 1/N \), sometimes 1. Equations 6 are appropriate for \( 0 \leq \varepsilon \leq 0.5 \), but, for \( 0.5 < \varepsilon < 1 \), better estimates of \( C_\mu \) and \( \theta_\mu \) are:

\[
\hat{C}_i = \frac{2\sin(\pi(1-\varepsilon)/N)}{A_\varepsilon \sin \pi (1-\varepsilon)} \hat{C}_{\ell+1} ; \quad \theta_\mu = \theta_{\ell+1} - \frac{\pi(N-1)(1-\varepsilon)}{N}
\]

Burgess [1] also indicated that Hanning filtering and larger sample sizes, \( N \), would give improved estimates, and he provided equations and error estimates. This method is very efficient and significantly improves the accuracy of FFT mode estimates for low noise signals. Prediction accuracy is degraded by the leakage of nearby modes and by noise. Fig. 3 suggests accuracy limitations, since the closed form and FFT results differ, as do FFTs of different lengths, \( N \); further these differences are comparable in size to the small mode.

### II. ALGORITHM OUTLINE

To accurately determine Fourier coefficients for a short duration signal, the current algorithm uses NLLS and a classical search strategy, namely bracketing and search, to minimize a function and find a mode. A third step, polishing, is also required. The function to be minimized, \( \Phi(S_{\ell+1}) \) (8), is the power of a reduced signal, namely, the signal, \( S_{\ell+1} \), minus a cancelling mode. The search space includes frequency, \( f_{\text{test}} \), magnitude, \( C_{\text{test}} \), and phase, \( \theta_{\text{test}} \).

\[
\Phi(S_{\ell+1}) = \frac{1}{N} \sum_{k=0}^{N-1} \left| S_k - C_{\text{test}} \cos(2\pi f_{\text{test}} k \Delta t + \theta_{\text{test}}) \right|^2
\]

Bracketing must exclude local minima from the search.
Powell’s method is the most efficient and robust search method tested. Once $\Phi(S_j)$ is minimized, the cancelling mode is subtracted from the signal, $S_k$, to form a new residual signal, $S_k^M$, (9), and the bracketing and search is repeated with $\Phi(S_k^M)$ to find the next largest mode. These steps are shown in Table I.

$$S_k^M = S_k - \sum_{j=1}^{M} C_j \cos(2\pi f_j k \Delta t + \theta_j) \quad k = 0, N-1$$ (9)

Just as the Burgess-Lanczos estimate has errors due to interaction with other modes, the search result is not exactly the mode of the signal. In particular, the minimum of $\Phi(S_{kl})$ is shifted away from the exact result by the presence of other modes in the signal, hence the frequency, magnitude, and phase found for the cancelling mode are well approximated, but not exact. To circumvent this problem, as many of these offending modes as possible are removed while leaving the mode to be refined. In particular, the polishing function $S_{kl}$ (10) isolates mode $J$ by subtracting all other approximated modes from the original signal leaving only the residual signal and the $J$th mode to be polished. The search method is also applied to this isolated mode, $\Phi(S_{kl})$.

$$S_{kl} = S_k - \sum_{j \neq J} C_j \cos(2\pi f_j k \Delta t + \theta_j)$$

for $k = 0, N-1$, $j = 1, M$, $j \neq J$. (10)

The algorithm continues until the signal is sufficiently well represented. The algorithm is implemented in single precision, $P=32$. Table II demonstrates results and compares with the Burgess-Lanczos method.

III. ANALYSIS OF BracketING, Search METHODS, and POLISHING

This section provides more detailed explanations and analysis of bracketing, search, and polishing. Two important properties must be guaranteed. First, ensure generally that bracketing includes only the region of attraction of a global minimum—no local minima. Second, ensure that polishing works generally, and the global minimum converges to the desired mode. For a first reading, scrutiny of all details may not be necessary.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>OUTLINE OF ALGORITHM AND PERFORMANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form discrete signal, $S_i = s(kf_j), k=0,N-1$</td>
<td>% CPU Time</td>
</tr>
<tr>
<td>For each cancelling mode $j$, $j = 1, M$</td>
<td></td>
</tr>
<tr>
<td>FFT of signal, $S_i^{k-1}$, for Initial Bracketing</td>
<td>2.1</td>
</tr>
<tr>
<td>For each test frequency, $f_i$, $i = 1, B_i$</td>
<td></td>
</tr>
<tr>
<td>{ DFTs ($f_i$) } * final bracketing */</td>
<td>4.5</td>
</tr>
<tr>
<td>Refinement search by Powell’s method $\Phi(S_i^{k+1})$</td>
<td>28.5</td>
</tr>
<tr>
<td>Remove Approx Mode, Reduced Signal, $S_i^{k'}$</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>/* Polishing */</td>
<td></td>
</tr>
<tr>
<td>For each mode found, mode $J_i$, $J = 1, j$</td>
<td></td>
</tr>
<tr>
<td>Form Polishing Signal, $S_{kl}$, for mode $J_i$</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Refinement search by Powell’s Method $\Phi(S_{kl})$</td>
<td>64.8</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>% CPU Time is based on signal in Fig 1; $N=512$, $M=3$, $B_i=21$.</td>
<td></td>
</tr>
</tbody>
</table>

A. Analysis of Bracketing

With an oscillatory signal and cancelling modes, many local minima are observed for $\Phi$. Bracketing must exclude local minima and find a starting point within the region of attraction of the global minimum. However, perturbation analysis shows that an FFT helps find the bracketing interval.

1) Perturbation Analysis of the Global Minimum

Perturbation analysis of the bracketing requirements considers a simplified model problem, namely, a single mode, $\cos(2\pi ft)$, representing the discrete target signal, $S_k = \cos(2\pi kf_j)$. Further, the analyses involve an approximate cancelling mode with a small error in frequency, magnitude, or phase. Table III shows these cancelling modes and error parameters. The minimization function is the substitution of the discrete target signal and cancelling mode into $\Phi$, (8). Note that the summation in $\Phi$ is approximated by the integral $f_i \int_{\Delta t} dt$.

Table IV and Fig. 5 show these perturbation expansions in frequency, magnitude, and phase neighborhoods of the global minimum. A spreadsheet calculation, with the signal and $\Phi$ (8) was used to validate the perturbation analysis, and it is also plotted in the figures.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>PERFORMANCE COMPARISON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>$c_{f_1}$</td>
</tr>
<tr>
<td>Exact</td>
<td>172.2656</td>
</tr>
<tr>
<td>Current</td>
<td>172.2659</td>
</tr>
<tr>
<td>BL</td>
<td>172.3051</td>
</tr>
<tr>
<td>BL Hanning</td>
<td>172.3492</td>
</tr>
</tbody>
</table>

Performance comparison of the current search algorithm and the Burgess-Lanczos (BL) method with and without Hanning filtering on the noiseless, three-mode signal from Fig. 1. Burgess-Lanczos method modes successively removed, largest to smallest, to reduce interaction error; polishing was not advantageous. $P = 32$ bits.
As a guide to bracketing the search, \( \Phi \) is quadratic in a neighborhood of the global minimum for errors in frequency, \( \varepsilon \), magnitude, \( \delta \), and phase, \( \phi \). Further, the region of attraction for phase and magnitude is independent of sampling interval parameters \( T \) and \( N \). The bracketing interval for frequency depends on sampling interval period, \( T = 2\pi f_s \), the number of mode periods in the sampling interval, times \( 2\pi \).

The bracketing interval width for frequency depends on sampling interval period, \( T \), and Fig. 4 shows two different periods, \( T = 23.22 \text{ ms} \) and \( T = 11.61 \text{ ms} \). In particular, replacing \( \varepsilon \) in Table III with frequency error, \( \Delta f \), the dominant quadratic term is \( \Phi(\Delta f) \propto \Delta f^2 T^2 \). Hence for shorter periods, \( T, \Phi \) has a wider parabola near the minimum, and the frequency interval of attraction is larger.

This interval of attraction analysis is consistent with Angeby’s observation [16] and implementation where short sliding windows significantly improve numerical properties.

2) Initial Bracketing by FFT Periodogram

The discrete modes of an FFT are separated by frequency intervals of \( f_s/N \). The first bracketing step is to use an FFT periodogram to narrow the search. In particular, the largest FFT coefficient, \( |C_f| \), is found, where \( f = k f_s/N \), and a conservative frequency bracketing interval is \( [(k-1)f_s/N,(k+1)f_s/N) \)—a reduction to \( 4/N \) of the FFT’s spectrum. A second bracketing step provides further reduction to a starting point well within the global minimum’s region of attraction.

3) Final Bracketing by Individual Mode DFT

The Burgess-Lanczos method could be used for final bracketing, but it is used for validation instead. An alternate method is introduced here, designated \( DFT_F \).

The FFT is constrained to find Fourier coefficients at frequencies separated by \( f_s/N \), and this separation is determined by the sampling frequency, \( f_s \), sample size, \( N \), and the requirement for an integral number of periods of the mode. Yet, both testing between the FFT frequencies, \( f_{\text{test}} \), and successful bracketing is possible by using less that the full data set \( S_k, k=0, N-1 \), (Fig. 6) and by using an approximation to the Fourier Transform integral (4).

For a signal, \( s(t) \), where the FFT suggests a spectral maximum near \( k f_s/N \), consider a test mode at the frequency \( f_{\text{test}} \in [(k-1)f_s/N,(k+1)f_s/N] \). The largest number of complete periods of this mode within the available \( N \) samples is \( f_{\text{test}} T \), the number of data points inside this interval is \( N = \lfloor f_{\text{test}} T \rfloor \), and the interval length is approximately \( T = \frac{\lfloor f_{\text{test}} T \rfloor}{f_{\text{test}}} T \) in

\[
\hat{C}_{f_{\text{test}}} \approx \frac{2}{T} \sum_{k=0}^{N-1} e^{2\pi i f_{\text{test}} t_k} s(t_k), \text{ where } t_k = \frac{k}{f_s}. \tag{11}
\]

Since less than a test mode’s period is lost from each interval, \([0, T]\), high frequencies, \( f_{\text{test}} \), have much less sample loss than low frequencies. Complete loss occurs and \( \tilde{T} \) is zero for \( f_{\text{test}} < f_s/N \); a brute force method is used for bracketing at these low frequencies.
The computational cost is estimated as $10N$ for each mode tested, and DFT relative cost is shown in Table I. The DFT shows leakage, but does not show minimum shifting.

To bracket the global minimum of $\Phi$, the strategy is to evaluate the DFT at test frequencies, $f_{\text{test}}$, in the interval $[(k-1)f_s/N, (k+1)f_s/N]$. The DFT gives magnitude, $C_{f_{\text{test}}}$, and phase, $\theta_f$. Fig. 7 shows an example. The largest DFT coefficient, $|C_{f_{\text{test}}}|$, approximates the center of the interval of attraction, and is used as the starting point for multidimensional minimization.

How close is close enough? Could the presence of other signal modes reduce the interval of attraction from the single mode estimates of Table III? Experimentally, no issues have been detected in a large number of signals after searching to a resolution of $0.1f_s/N$. Theoretically, the multimode perturbation analysis of section C shows that other modes do not change the quadratic expansion term—only the minimum’s location is shifted.

B. Multidimensional Minimization by Powell’s Method

The search algorithm must efficiently and exactly perform multidimensional minimization on $\Phi(S_k)$, that is, find the frequency, $f$, magnitude, $C_f$, and phase, $\theta_f$ of a mode that best cancels a mode in the signal, $S_k$. Other search algorithms have been used to verify the code, yet Powell’s method is robust, can provide quadratic convergence (the current implementation [11] does not include the necessary extensions), and requires only function evaluations—not derivative evaluations.

C. Perturbation Analysis of Polishing Requirements

Perturbation analysis shows that while searching for a cancelling mode by minimizing $\Phi(S_k)$, the minimum is shifted away from the correct cancelling mode by the presence of other modes in the signal. Hence the frequency, magnitude, and phase of the cancelling mode are well approximated by searching for the minimum of $\Phi$, but polishing is required for exact results.
The extended perturbation analysis uses two modes, \( \cos(2\pi f_1 t) + a \cos(2\pi f_2 t) \), to represent the discrete target signal, \( S_k \). The parameterization of the cancelling mode is as before. Unfortunately, \( \Phi \) is not of the form \( A\varepsilon^2 \) near the minimum as for a single mode. Instead, \( \Phi \) is of the form \( A\varepsilon^2 + B\varepsilon + C \) in a neighborhood of its minimum at \( \varepsilon = -B/(2A) \) where \( A, B, \) and \( C \) are expressions that are constant in a neighborhood of the minimum. In particular, the presence of a second mode shifts the minimum of \( \Phi \) away from \( \varepsilon = 0 \) and the desired result. Similarly for \( \delta \) and \( \phi \).

These errors are small. Table IV and (12) show the error in the minimum when the signal and cancelling mode are substituted into \( \Phi \). For the signal in Fig. 1 without polishing, the frequencies are correct to 0.16 Hz, the magnitudes to 0.4%, and the phases to within 0.2 degrees.

\[
E_i = \frac{\alpha}{\alpha + \beta} \cos(\alpha + \beta)T + \frac{\alpha}{\alpha - \beta} \cos((\alpha - \beta)T)
\]

Equation (12) gives

\[
E_i = \frac{\alpha}{\alpha + \beta} \cos((\alpha + \beta)T - 1) + \frac{\alpha}{\alpha - \beta} \cos((\alpha - \beta)T - 1)
\]

To circumvent this problem, a discrete polishing signal, \( S_{ij} \), (10) is formed where all approximated cancelling modes are removed—except for the focus of the polishing, the \( j^{th} \) cancelling mode. By removing the best estimates of the other modes in \( S_{ij} \), this interference is reduced. By repeated polishing, the coefficients, \( C_j \), of the cancelling modes converge to the best representation of the discrete signal.

IV. Important Algorithm Properties

Several algorithm properties are important and are compared in this section, namely, convergence, computational complexity, convergence metrics, tolerance of signal noise, and resolution of two modes with nearly identical frequencies.

A. Convergence to the Signal

Fig. 8 shows convergence versus computation time, that is, the reduction in error with additional Powell search iterations. The graph indicates linear convergence. Powell’s method can give quadratic convergence, but the implementation used here [11], does not include the necessary extensions.

B. Computational Complexity

The current algorithm must be computationally efficient for all plausible input. Table I gives the observed relative CPU time for the signal in Fig. 1, plus the operation count for each algorithm step. The Powell’s method searches and repeated evaluations of \( \Phi \) are the most time consuming steps. The current algorithm is much more time consuming than a single FFT—50 to 100 times more in this example.

However, comparisons can be difficult. When performing an FFT, the temptation exists to increase the FFT’s sample size, \( N \), to increase the frequency resolution. Further, an FFT may need to be averaged over many sub-intervals [13] which further increases the computational cost. In contrast, the current algorithm is unencumbered by fixed FFT sampling frequencies and is accurate on a single interval; the temptation is to reduce sample size, \( N \), and reduce computational cost. If

![Fig. 8](image-url)
the Burgess-Lanczos method can successfully find signal modes, its computational efficiency is exceptionally good.

If noise or a fricative is present, if nearly identical frequencies are present, then the signal is not well captured by a few modes. Additional computations are required to find these additional modes. Termination conditions can help.

C. Precision and Convergence Metrics

The precision, \( P \), of signal sample values, \( S_k \), places limits on the smallest modes that should be searched for. In particular, audio formats often decode signal values to a precision \( P = 8 \) bits, and for a signal scaled so as not to exceed a magnitude of one, signal details smaller than \( 2^{1-P} \) are lost by rounding to the nearest bit. The search algorithm should not search for modes smaller than this precision threshold. Similarly, signal peak magnitudes exceeding 1.0 are rounded to the nearest bit. The search algorithm should not consider good modes smaller than \( 2^{1-P} \) precision.

Note, signals that do not utilize the dynamic range of the precision, \( P \), to the precision threshold. Values less than 10 dB are considered good.

A second metric, \( L_2 \), compares the residual signal’s power to the precision threshold.

\[
L_2 = 20 \log \left( \frac{1}{N} \sum_{k=0}^{N-1} (S_k^M)^2 / 2^{1-P} \right) \tag{14}\]

A third metric, \( L_3 \), measures the reduction in signal power due to removing cancelling modes.

\[
L_3 = 10 \log \left( \sum_{k=0}^{N-1} (S_k^M)^2 / \sum_{k=0}^{N-1} (S_k^0)^2 \right) \tag{15}\]

Note, signals that do not utilize the dynamic range of the precision will have poor \( L_2 \) performance. Further note that for each metric, larger negative values indicate better performance. Table V shows metric values for each test case.

For storage, audio formats may use signal compression to increase signal precision, \( P \), and/or store in fewer bits. For example, current telephone signal compression codes 12 or 13 bits into an 8-bit value, and current MP3 and Ogg Vorbis formats store in fewer than \( P=8 \) bits. CD quality sound is \( P=16 \) bits.

D. Effect of Signal Noise on Bisection, Search, and Results

Noise can introduce random errors into the predictions of mode frequency, magnitude, and phase. Further, noise places limits on the smallest modes that should be searched for.

Fig. 9 shows the three-mode signal, from Fig. 1, with and without Gaussian random noise, \( n(t) \). In both cases, the three modes are found, however the predicted parameters have small errors. Further, these errors are random since mode searches in different sampling intervals will have different noise signals. Lastly, the noise is large enough—relative to the convergence criteria—that the algorithm adds cancelling modes in an effort to capture this noise. The convergence criteria must consider the noise level to avoid this undesirable effect.

Fig. 10 shows the prediction errors at various noise levels for the current and Burgess-Lanczos methods.

E. Resolution of Two Modes with Nearly Identical Frequencies

Can the algorithm resolve modes with nearly the same frequency? Any spectral resolution algorithm must deal with the fundamental problem that: two modes at nearly the same frequency on a short interval look like a single mode at the mean frequency, if \( O(2\pi f_0 T) < 1 \).

Consider two modes of frequency \( f_0 + f \Delta f \) and \( f_0 - f \Delta f \), separated by a small frequency spacing, \( \lambda = \Delta f T \), and a phase offset \( \phi \in [0, 2\pi] \). Then, for any metric/algorithm, the combined signal is:

\[
\sin(2\pi f(1+\lambda) t + \phi) + \sin(2\pi f(1-\lambda) t - \phi) = 2 \sin(2\pi f t) \cos(2\pi f \lambda t + \phi) \approx 2 \sin(2\pi f t) \{ \cos(\phi) + \sin(\phi) - 2\pi f \lambda t + \cos(\phi) \} 2\pi f \lambda t^2 + \ldots \} \tag{16}\]

For short duration signals, the sampling interval, \( [0,T] \), is small, and these two modes approach a single mode when \( O(2\pi f_0 T) < 1 \). Note that this general analysis applies to any algorithm designed to resolve frequency components. Since \( T \)

---

### Table V: Convergence of Example Signals

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Precision, ( P )</th>
<th>Current Method</th>
<th>Burgess-Lanczos</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( L_1 ) (dB)</td>
<td>( L_2 ) (dB)</td>
<td>( L_3 ) (dB)</td>
</tr>
<tr>
<td>No Noise, 3 Mode Case, Fig. 1</td>
<td>32</td>
<td>8.1</td>
<td>-5.0</td>
</tr>
<tr>
<td>Noise, 3 Mode Case, Fig. 9</td>
<td>32</td>
<td>59.1</td>
<td>45.8</td>
</tr>
<tr>
<td>Two Modes ( \Delta f=20Hz ), Fig. 11</td>
<td>32</td>
<td>52.2</td>
<td>26.4</td>
</tr>
<tr>
<td>Two Modes ( \Delta f=50Hz ), Fig. 12</td>
<td>32</td>
<td>22.2</td>
<td>0.9</td>
</tr>
<tr>
<td>Guitar Open e, Fig. 13</td>
<td>8</td>
<td>0.0</td>
<td>-14</td>
</tr>
<tr>
<td>Voiced Phoneme 'oo', Fig. 14</td>
<td>8</td>
<td>-0.1</td>
<td>-11.9</td>
</tr>
<tr>
<td>Fricative 's', Fig. 15</td>
<td>8</td>
<td>6.5</td>
<td>-5.2</td>
</tr>
<tr>
<td>Plosive 'p', Fig. 16</td>
<td>8</td>
<td>19.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Plosive 'b', Fig. 17</td>
<td>8</td>
<td>20.7</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Convergence of residual signal, \( S_k^M \), in each test case for each metric. Both the current method and the Burgess-Lanczos method are compared. Larger negative metric values indicate better performance.
Fig. 9. Three mode signal (from Fig. 1) with Gaussian random noise (upper) and without (lower); SNR = 40 dB. The three modes are found, but additional cancelling modes occur. Leakage is still present for modes not at FFT frequencies.

= \frac{N}{f}, increasing the number of samples, N, or decreasing the sampling frequency, f, will improve resolution.

How does the current algorithm perform? Fig. 11 shows a three mode signal with two modes near 4500 Hz separated by 20 Hz, and \(2\pi\Delta f T=2.9\). The two modes are not well resolved, and additional spurious modes are found. In contrast, Fig. 12 shows the identical signal except the two modes near 4500 Hz are separated by 50 Hz, and \(2\pi\Delta f T=7.3\).

V. PRACTICAL EXAMPLES

Four auditory signals are presented in this section, and they range from a guitar note dominated by harmonics to voiced, fricative, and plosive phonemes from human speech. For each signal, results from the current algorithm are shown.

A. Guitar Note

The notes of string and woodwind instruments would be dominated by harmonics, like the guitar note in Fig. 13, due to the physics and mathematics of vibrating strings and air columns. Table VI shows modes found by the current method: fundamental frequency and eight harmonics plus two small, low frequency modes which are not harmonic. The residual signal, \(S_k^M\), is below the precision bounds of the signal. Additional harmonics exist that are small and do not significantly reduce \(L_J\). The Burgess-Lanczos method finds only the four largest modes, as shown in Table VI. The current method finds signal modes in this steady signal (Fig. 13) without spurious modes.

B. Voiced Phoneme ‘oo’

A voiced phoneme ‘oo’ (vocal folds vibrating) from a male speaker is shown in Fig. 14. Table VII shows 5 harmonics of a fundamental frequency of 83.55 Hz, plus approximate...
harmonics, found with the current method. The residual signal is below the precision bounds of the signal. The Burgess-Lanczos method finds only a single mode, as shown in Table VII.

The human vocal folds are a pair of semi-circular membranes restrained along their circular arcs; harmonics are known to occur. The shape of the human vocal tract is known to shift formant frequencies [24]; the approximate harmonics above 550 Hz in Table VII may be due to this effect. Modes are separated by more than 50 Hz except for three small modes.

C. Fricative Phoneme ‘S’

A fricative phoneme, ‘S’, from a male speaker is shown in Fig. 15. Signal frequencies are centered near 4500 Hz. The weak signal is near the precision bounds of the signal. The modes found are typically separated by 50 Hz or more; only one separation of ~30 Hz exists. Increasing the number of cancelling modes, M, from 20 to 30 reduces L₁, L₂, and L₃ by only about 3 dB.

D. Plosive Phoneme ‘P’

A plosive phoneme, ‘P’, from a male speaker is shown in Fig. 16. The stop (closed airway—no airflow or sound) is

<table>
<thead>
<tr>
<th>Mode Freq (Hz)</th>
<th>Harmonic</th>
<th>Magnitude</th>
<th>Relative Mag (%)</th>
<th>Phase (π Rad)</th>
<th>Mode Freq (Hz)</th>
<th>Magnitude</th>
<th>Phase (π Rad)</th>
</tr>
</thead>
<tbody>
<tr>
<td>46.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>121.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>331.05</td>
<td>1</td>
<td>0.0496</td>
<td>100</td>
<td>1.4977</td>
<td>331.28</td>
<td>.0498</td>
<td>1.4916</td>
</tr>
<tr>
<td>661.43</td>
<td></td>
<td>0.0178</td>
<td>35.9</td>
<td>0.5794</td>
<td>662.86</td>
<td>.0178</td>
<td>0.5472</td>
</tr>
<tr>
<td>992.10</td>
<td>2</td>
<td>0.0159</td>
<td>32.1</td>
<td>1.5536</td>
<td>992.02</td>
<td>.0157</td>
<td>1.5572</td>
</tr>
<tr>
<td>1323.95</td>
<td>3</td>
<td>0.0024</td>
<td>5.0</td>
<td>0.3348</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1983.22</td>
<td>5</td>
<td>0.0119</td>
<td>24.0</td>
<td>1.8695</td>
<td>1982.47</td>
<td>.0118</td>
<td>1.8887</td>
</tr>
<tr>
<td>2314.22</td>
<td>6</td>
<td>0.0067</td>
<td>13.5</td>
<td>1.9785</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2645.17</td>
<td>7</td>
<td>0.0023</td>
<td>4.6</td>
<td>1.0998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3972.34</td>
<td>11</td>
<td>0.0019</td>
<td>3.8</td>
<td>1.7988</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4303.84</td>
<td>13</td>
<td>0.0023</td>
<td>4.6</td>
<td>1.3622</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Modes found by the two methods for the guitar open 'e' note of Fig. 13.
shown before 7605 ms. The strong pressure wave resulting from resumed airflow is clearly visible in the signal—and the frequency spectrum has large, low frequency modes. The fricative modes found near 2000 Hz are typically separated by about 50 Hz; three spacings are of ~30 Hz.

All signals were recorded in Ogg Vorbis format [25] using Audacity [26] on a PC with a Gigaware desktop microphone. The human speech samples are of a male speaker. All searches are limited to twenty modes, M=20. Signals are sampled at \( f_s = 22.05 \text{ kHz} \) (recorded at 44.1 kHz and desampled), \( N = 512, T = 23.22 \text{ ms} \), and \( P = 8 \text{ bit} \) resolution. Signals are assumed to be bandlimited even though the Nyquist frequency is 11 kHz and the frequency limit of human auditory sensitivity is 20 kHz.

VI. CONCLUSION

For the perception of tone in human speech and music, capturing the subtle variations in modal frequencies and magnitudes is important. The current method can accurately find the spectrum of these short duration signals.

Each of the algorithms considered here are useful. The FFT continues to be a pillar of signal processing. If better accuracy is required in a low-noise signal with a few modes, the Burgess-Lanczos algorithm is efficient and appropriate. Yet, if

### TABLE VII

<table>
<thead>
<tr>
<th>VOICED PHONEME ’OO’ MODES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mode Freq (Hz)</strong></td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>83.55</td>
</tr>
<tr>
<td>168.38</td>
</tr>
<tr>
<td>253.88</td>
</tr>
<tr>
<td>339.30</td>
</tr>
<tr>
<td>418.73</td>
</tr>
<tr>
<td>508.28</td>
</tr>
<tr>
<td>551.64</td>
</tr>
<tr>
<td>599.09</td>
</tr>
<tr>
<td>680.28</td>
</tr>
<tr>
<td>715.84</td>
</tr>
<tr>
<td>762.13</td>
</tr>
<tr>
<td>847.26</td>
</tr>
<tr>
<td>2373.52</td>
</tr>
<tr>
<td>3057.92</td>
</tr>
</tbody>
</table>

Modes found by the two methods for the voiced phoneme ‘oo’ of Fig. 14.
high accuracy Fourier coefficients are required in a short duration signal with many modes, the current algorithm would be more appropriate.

ACKNOWLEDGMENT

Discussions with Donald Braun provided valuable insights into the Burgess-Lanczos method. Ambady Suresh provided very helpful comments on the paper.

REFERENCES


Mark E. M. Stewart (M’2012) received the B. Math degree in applied mathematics from the University of Waterloo, Waterloo, Canada in 1985 and the M.A. and Ph.D. degrees in applied and computational mathematics from Princeton University, Princeton, NJ in 1987 and 1990.

Currently, he is a senior research engineer working at NASA Glenn Research Center in Cleveland, Ohio. His NASA work involves high-fidelity, scientific computing simulations of propulsion systems, including multidisciplinary fluid/thermal/structural analysis of nuclear thermal propulsion fuel elements, microgravity fluid flow, turbofan engines, hypersonic air-breathing engines, external aerodynamics, combustion, multidisciplinary fluid-combustion-thermal-structural analysis, and engineering design. His outside interests include semantic analysis of scientific programs and spreadsheets. The problem in the current paper arose in attempts to simulate neural processing of semantics in the human auditory system.

Dr. Stewart is a member of the IEEE and the AIAA.