

Results for International Student Challenge Problem in Acoustic Signal Processing 2014

Kay L. Gemba

Dep. of Ocean and Resources Engineering

University of Hawai'i at Mānoa

Honolulu, HI 96822, USA

gemba@hawaii.edu

1. Problem formulation

A truck with a 4-stroke diesel engine travels along a straight road with constant speed. Near the road is a microphone that senses the radiated acoustic noise from the truck during its passage past the microphone. The output of the microphone is sampled at the rate of 12,000 samples/second and 30 seconds of data are recorded during the trucks transit which can be found in the attached file: *truck30sec.wav*. During the recording of the data, the speed of sound propagation in air was a constant 347 m/s.

Problem: Assuming that the truck is a point source, (a) Plot the spectrogram of the acoustic data file *truck30sec.wav* and (b) Given that the strongest spectral line is the engine firing rate, calculate the trucks

1. engine firing rate (in Hz),
2. cylinder firing rate (in Hz),
3. number of cylinders,
4. tachometer reading (in revolutions/minute),
5. speedometer reading (speed in km/hour),
6. distance (in meters) of the closest point of approach of the truck to the sensor,
7. time (in seconds) at which the closest point of approach occurs.

2. Introduction

Discrete Fast-Fourier Transforms (FFT) can be used to estimate the frequency content of discrete, finite time signals. The frequency content for stationary signals is constant in time and the FFT is an appropriate analysis tool. However, the FFT is not localized in time which is a limitation when estimating the frequency content of nonstationary signals. A solution is to segment the original signal into smaller parts, compute the FFT for each part (usually with some overlap to smooth transitions between segments) and assemble the output in temporal order. This time-frequency (TF) representation is called a spectrogram and is an intuitive way to display a signal's frequency content in time. Time and frequency resolution are related by the uncertainty principle, an increase in frequency resolution reduces the time resolution and vice versa, e.g. estimating the instantaneous frequency (IF) of a source (such as a passing airplane, vessel or a truck) is a tradeoff between time and frequency resolution. Alternatively, IF estimates can be computed using the Wigner-Ville distribution. The following discussion will address how to estimate the IF of a passing truck which is subsequently compared to the output of a parametric model used to compute the parameters in the problem statement.

3. Analysis of Data

Spectrogram of passing truck

For a preliminary estimate of the truck's parameters, two spectrograms are computed (Fig. 1) using data obtained from [Ferguson and Culver \(2014\)](#). Figure 1(a) indicates that the the strongest spectral line (corresponding to the engine firing rate (EFR)) is close to 120 Hz. Several other spectral lines (and harmonics) can be observed as well as energy at higher frequencies when the truck is closer to the microphone. If energy from a moving

source is reflected from a boundary, its arrival pattern at the microphone corresponds to a Lloyd's mirror effect. No such effect is noticeable which might indicate that the recording was performed close to the ground and that most of the reflected energy was scattered or attenuated. Figure 1(b) shows a close up spectrogram of the strongest spectral line at approximately 120 Hz. A doppler shift over the 30 second recording can be observed.

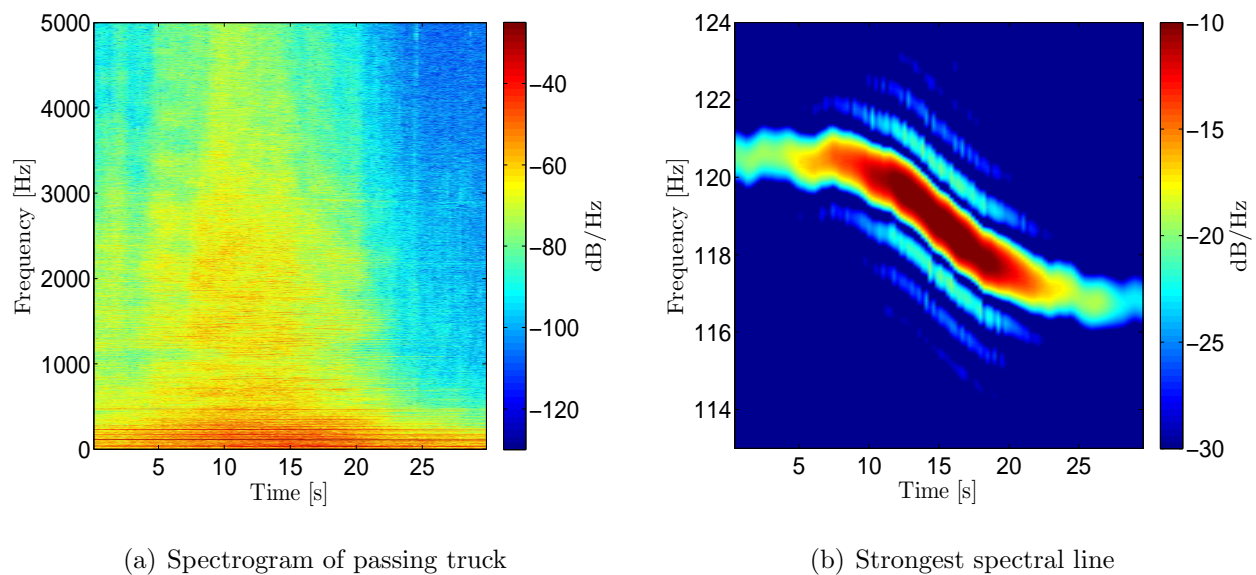


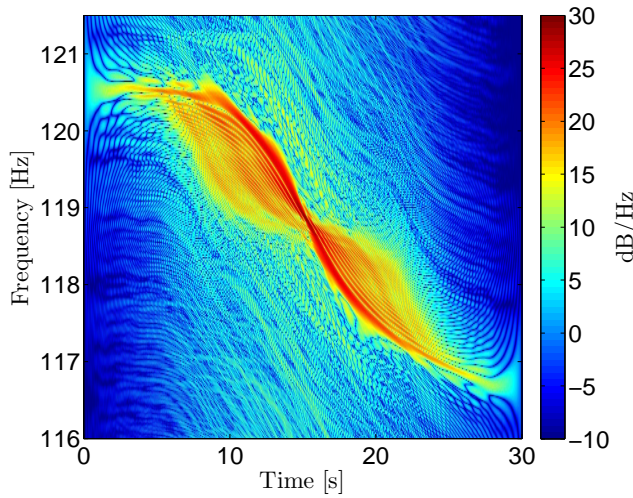
Figure 1: (a) Spectrogram of truck30sec.wav with several spectral lines below 500 Hz. The strongest spectral line (corresponding to the EFR) is at nominally 120 Hz. (b) Zoomed in plot of spectrogram showing doppler shift of the EFR (frequency resolution 0.03 Hz, time resolution 0.24 s, 75% overlap).

Parameter estimation

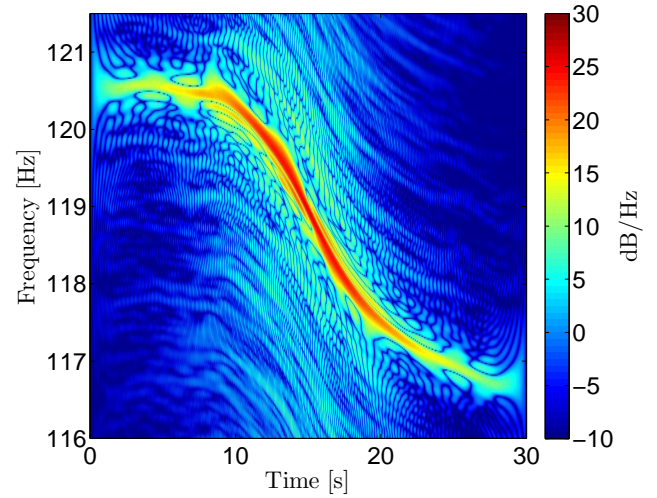
Vehicle parameters are estimated using a parametric model (Ferguson (1994)). Here, this model approximates the instantaneous frequency of a constant tone emitted by the passing truck of constant speed along a straight road. Four model parameters (EFR, speed of the vehicle, distance from the microphone and time of closest approach) are varied and results

are compared to the IF extracted from a TF plot using an acoustic recording (referred to as data). Optimum parameters are found by minimizing the squared difference between model output and data. The next two paragraphs will discuss how the TF plots are computed and how the parametric model is implemented.

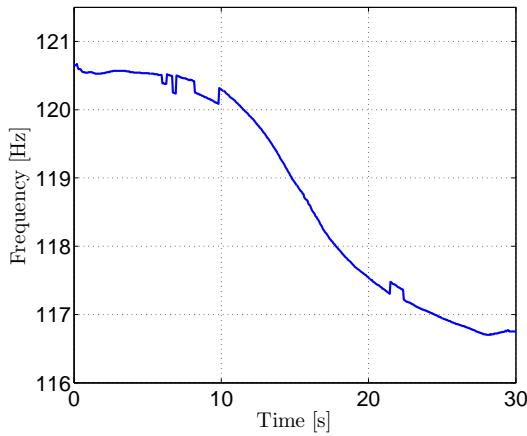
To estimate the IF of the EFR from the data, TF representations are computed using the Short Time Fourier Transform (STFT) and the Pseudo-Wigner-Ville distribution (PWVD). For both methods, the main variables are the sampling rate and window type. First, the signal is filtered and downsampled to 6000 samples/sec to reduce computational load: the lowpass filter has a pass-band edge at 250 Hz (0.1 dB ripple ratio in the pass-band) and -60 dB stop-band attenuation at 500 Hz. Second, an appropriate window is selected. By definition, instantaneous frequency equals the derivative of the phase and [Cohen and Lee \(1989\)](#) showed that this is the case (for continuous time) as the window approaches a delta function (e.g. the window vanishes). For discrete data, this is not possible and at least a rectangular window is used to segment the data. In particular, for the Wigner-Ville distribution, cross-terms around the main frequency component (Fig. [2\(a\)](#)) cause discontinuities in the IF estimate (Fig. [2\(c\)](#)) and a window is required to smooth out the discontinuities. By the conventional measure of TF resolution (the product of time and frequency 'bandwidths'), a Gaussian window is optimal. By the measure of energy concentration, the prolate spheroidal function (which the Kaiser window ([Kaiser \(1974\)](#)) approximates) is the better option. Therefore, a Kaiser window (beta=4.5) is used to compute both STFT and PWVD plots (an example using PWVD is show in Fig. [2\(b\)](#)). The beta value is a trade off parameter between main-lobe and side-lobe width, the value of 4.5 is selected by minimizing discontinuities in the peak detection algorithm (Fig. [2\(d\)](#)). The peak detection algorithm ([Boashash \(2003\)](#)) operates



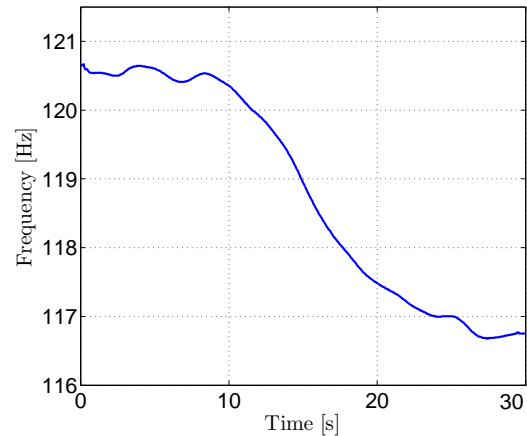
(a) Wigner-Ville Distribution



(b) Pseudo Wigner-Ville Distribution (Kaiser window)



(c) Extracted IF (no window)



(d) Extracted IF (Kaiser window)

Figure 2: Instantaneous frequency estimates using Wigner-Ville and Pseudo Wigner-Ville distribution. The plot in (a) is computed using no window. Notice the narrow doppler shift and significant cross-terms. The plot in (b) is computed using a Kaiser window (beta=4.5). Plots in (c) and (d) show the output of the peak detection algorithm for (a) and (b), respectively. The window eliminates discontinuities but increases oscillatory behavior at the boundaries. Note the increased time-frequency resolution in (b) in comparison to the STFT in Fig. 1(b).

over the range of the doppler shift (115 - 122 Hz) at each time step and selects the maximum value. For IF estimates using the STFT, the power spectrum of each window is computed before peak detection. The algorithm operates directly on the PWVD output. While it is possible to further increase the IF accuracy for a nonstationary signal using an adaptive window length (Lerga and Sucic (2009)), the current choice is deemed sufficient.

Estimating the model parameters is an optimization problem. The derivation of the model is given in Ferguson (1994) and only a short introduction will be given here. The goal is to minimize the squared difference between the data g_{t_j} and the model f_{t_j} :

$$\underset{f_0, v, h, t_c}{\text{minimize}} \sum_{j=1}^N (g_{t_j} - f_{t_j})^2 = \underset{f_0, v, h, t_c}{\text{minimize}} [(g_{t_1} - f_{t_1})^2 + (g_{t_2} - f_{t_2})^2 + \dots + (g_{t_N} - f_{t_N})^2], \quad (1)$$

where

$$f_{t_j} = \frac{f_0 c^2}{c^2 - v^2} \left(1 - \frac{v^2 (t_j - t_c)}{[h^2 (c^2 - v^2) + v^2 c^2 (t_j - t_c)^2]^{1/2}} \right). \quad (2)$$

The optimization parameters EFR (f_0), speed of the vehicle (v), distance from the microphone (h) and time of closest approach (t_c) are input to the model in Eq. (2). The speed of sound c is a constant and the instantaneous frequency f at time t_j is the output of the model. The model output has the same temporal resolution (index j) as the data. While the original paper suggests using the Gauss-Newton algorithm (GNA) on the partial derivatives (w.r.t. the optimization parameters) in Eq. (2) to find an optimum solution, a different implementation is selected here. A nonlinear least-squares optimization problem for vector-valued functions (the right side of Eq. (1)) can be solved using a trust region reflective algorithm (Levenberg-Marquardt Algorithm, LMA). The algorithm minimizes Eq. (1) using the residual vector $(g_{t_j} - f_{t_j})^2$ (of length N) and not its final sum over the index j . The LMA is an efficient implementation for a vector valued optimization problem because, unlike using GNA, explicit knowledge of the partial derivatives of Eq. (2) is not required. The LMA is

essentially a hybrid approach between the GNA and the method of steepest decent. Unlike the GNA, the LMA does not require a good first estimate of the solution but might need more iterations to converge. The possibility of local minima (for both GNA and LMA) can be avoided by selecting different initialization values, i.e. parameter boundary values or the set of parameters corresponding to the latest computed minimum.

Table 1 shows an overview of the probable range (lower and upper boundary) of parameters and initialization values (best guess) for the model. The probable range is determined from the spectrogram in Fig. 1(b) and literature. Both EFR and time of closest approach are estimated from the spectrogram: the EFR is expected to be between 116-121 Hz and the time of closest approach occurs in the center of the doppler shift (between 10-22 s). The upper range of the speed of the vehicle is limited by a scaling ambiguity around the true speed (Cevher et al. (2009), given at approx. 72 km/h or greater), which can only be resolved by providing the recording distance. The upper bound for the recording distance is unknown and is assumed to be less than 100 m. Initialization values are selected after a coarse global grid search using the standard spectrogram function in MATLAB.

To estimate the cylinder firing rate (CFR), real cepstra of the data are computed (see Oppenheim and Schaffer (2004) for an introduction to concepts and cepstral glossary; see Thomas and Wilkins (1970) for a discussion using cepstral analysis and autocorrelation to find the CFR). The spectrum of the log of the spectrum of a time waveform is referred to as the cepstrum, the independent variable is called quefrency and harmonics are called rahmonics. The motivation for cepstral analysis is to separate periodic components in the time signal.

Ideally, cylinder firing is periodic and should correspond to a clear peak in the cepstrum.

However, this might not be true for trucks with computer controlled injection systems. Here, it is assumed that this variation is small (or that the CFR is pseudo-periodic due to constant speed of the truck and even terrain). The frequency of the CFR can be computed by identifying the delay of the first rahmonic in the cepstrum and dividing the sampling rate by the delay. Once the CFR is calculated, the number of cylinders p can be estimated (Cevher et al. (2009)) using

$$p = \frac{EFR}{CFR}. \quad (3)$$

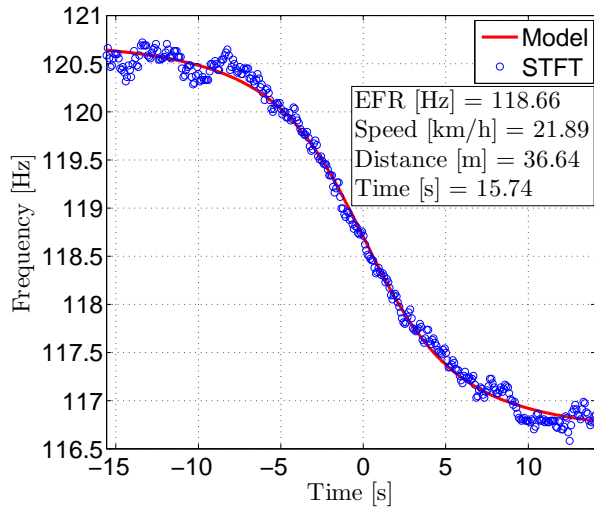
In a 4 cycle engine, every piston has to go up and down twice to fire once and the tachometer reading χ in rpm can be estimated as follows:

$$\chi = CFR \times 60 \times 2. \quad (4)$$

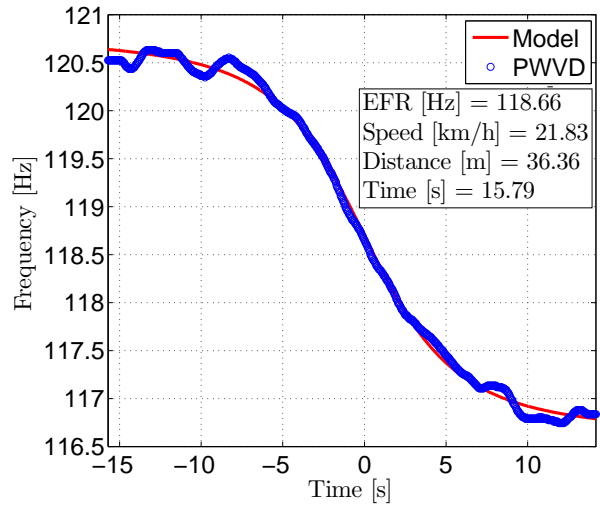
4. Results

Results for each TF method are shown in Fig. 3. The model approximates the data well and results using the STFT as well as the PWVD are very similar. The LMA converges to the optimum solution in all runs after roughly 40 iterations (less than one second). The tolerance is set to 10^{-20} and gradients are computed using central difference scheme to increase accuracy. The recording distance at the closest point is estimated to be 36.5 m and speed of the vehicle is estimated just below 22 km/h. The EFR is consistent at 118.7 Hz and the time of closest approach is at 15.8 s. Results for each method as well as final results are summarized in Table 1.

Figure 4 shows a filtered and downsampled (1000 samples/sec) real cepstrum. The signal is filtered for illustration purposes to eliminate tire and other noise sources at higher frequencies. The first rahmonic at a delay of 51 samples corresponds to a CFR of 19.6 Hz. To



(a) TF plot using STFT with fitted model



(b) TF plot using PWVD with fitted model

Figure 3: Parameter estimation results. (a) Time-frequency plot using STFT with frequency resolution of 0.015 Hz and time resolution of 0.08 s (75% window overlap). (b) Time-frequency plot using PWVD with twice the resolution of (a) using a smoothing window of 8999 points.

Table 1: Parameter overview, input range to the model and results. The last column shows final results.

Parameter	Probable range	Initialization	STFT	PWVD	Final results
EFR [Hz]	116 - 121	118.9	118.66	118.66	118.66
CFR [Hz]	-	-	-	-	19.77
Cylinders	-	-	-	-	6
Tachometer [rpm]	-	-	-	-	2372
Speed [km/hr]	0.1 - 72	19.4	21.89	21.83	21.85
Distance [m]	0.1 - 100	29.5	36.6	36.4	36.5
Time [s]	10 - 22	16	15.74	15.79	15.77

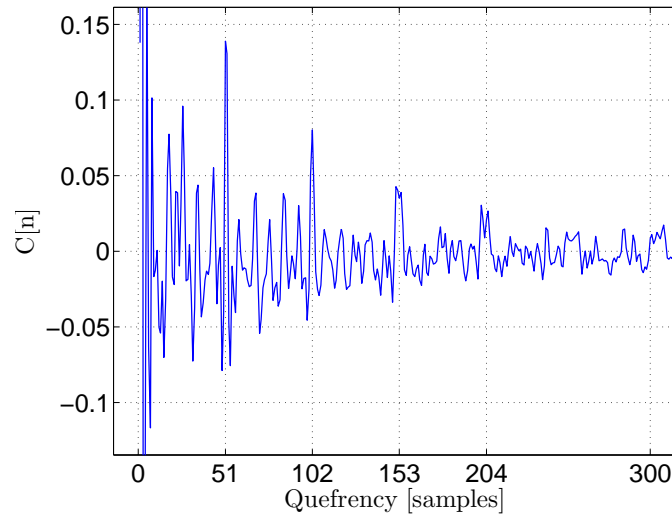


Figure 4: Zoomed in real cepstrum of truck30sec.wav (sampling rate of 1000 samples/sec). The first rahmonic peak at nominally 51 samples corresponds to a CFR of 19.6 Hz.

increase the sample resolution, a real cepstrum is computed with the original sampling rate. Erratic peaks due to the increased resolution are smoothed by applying a window. Results are similar to Fig. 4, a distinct peak at 607 samples corresponds to a CFR of 19.77 Hz. The number of cylinders (Eq. (3)) is found to be $6.002 \approx 6$. Values for 4,8,10, and 12 cylinders correspond to theoretical delays of 33, 66, 83, and 100, respectively. Examining Fig. 4, the closest peak to a delay of 33 samples is found at 26 samples (no match). There are no peaks close to 66 and 83 samples and the peak at 100 samples (possible match) is likely a rahmonic of 51 samples (a V12 should be a multiple of a V6). In addition, the lowest deterministic tone (in the spectrum) is called the cylinder fire rate (Cevher et al. (2009)) and is found to be nominally 20 Hz by visual inspection. This result is close to the above 19.77 Hz which suggests that the vehicle has 6 cylinders. An estimate for the tachometer reading (Eq. (4)) is computed to be 2372 rpm. Final results are shown in the last column of Table 1.

5. Discussion

Model output is fairly constant for most parameters. EFR and time of closest approach shift the model output vertically and horizontally while both speed and distance dictate the shape of the doppler curve. It seems plausible that the tradeoff between time and frequency resolution has less effect on the center frequency and time of closest approach for high SNR data. However, the distance of the microphone to the truck is sensitive to the TF resolution of the data and results from the original paper ([Ferguson \(1994\)](#)) indicate that the model overestimates this parameter for shorter ranges. It is found that an increased temporal resolution corresponded to a lower range estimate, which is selected in the analysis. In the absence of a ground truth (a preliminary recording of a 4-cylinder truck did not yield acceptable SNR), calibration parameters are closely matched to values cited in [Ferguson \(1994\)](#). While both rectangular and Kaiser windows were used to compute preliminary estimates, final results are largely invariant to the choice of the window.

The current approach might be improved by treating the TF plot as an image, and enhance the 'ridge' using edge enhancement/detection (Sobel operators or Canny edge enhancement). The output from the model can be transformed to a 2D image which would render the peak detection algorithm obsolete.

References

- Boashash, B., editor (2003). *Time Frequency Signal Analysis and Processing*. Elsevier Science, Oxford, UK, first edition.
- Cevher, V., Chellappa, R., and McClellan, J. (2009). Vehicle Speed Estimation Using Acoustic Wave Patterns. *IEEE Transactions on Signal Processing*, 57(1):30–47.
- Cohen, L. and Lee, C. (1989). Instantaneous frequency and time-frequency distributions. In *IEEE International Symposium on Circuits and Systems*, number 217, pages 1231–1234. IEEE.
- Ferguson, B. G. (1994). Application of the short-time Fourier transform and the WignerVille distribution to the acoustic localization of aircraft. *J. Acoust. Soc. Am.*, 96(2):821–827.
- Ferguson, B. G. and Culver, R. L. (2014). Acoustics Today. *International Student Challenge Problem in Acoustic Signal Processing*, 10(2):26–29.
- Kaiser, J. F. (1974). Nonrecursive Digital Filter Design Using the Io-sinh Window Function. In *IEEE Int. Symp. Circuits and Systems (ISCAS74)*, pages 20–23, San Francisco.
- Lerga, J. and Sucic, V. (2009). Nonlinear IF Estimation Based on the Pseudo WVD Adapted Using the Improved Sliding Pairwise ICI Rule. *IEEE Signal Processing Letters*, 16(11):953–956.
- Oppenheim, A. and Schafer, R. (2004). Dsp history - From frequency to quefrequency: a history of the cepstrum. *IEEE Signal Processing Magazine*, 21(5):95–106.
- Thomas, D. and Wilkins, B. (1970). Determination of engine firing rate from the acoustic waveform. *Electronics Letters*, 6(7):193–194.