

IMPROVED MULTIPATH TIME DELAY ESTIMATION USING CEPSTRUM SUBTRACTION

Eric L. Ferguson*, Stefan B. Williams

Australian Centre for Field Robotics
The University of Sydney, Australia

Craig T. Jin

Computing and Audio Research Laboratory
The University of Sydney, Australia

ABSTRACT

When a motor-powered vessel travels past a fixed hydrophone in a multipath environment, a Lloyd's mirror constructive/destructive interference pattern is observed in the output spectrogram. The power cepstrum detects the periodic structure of the Lloyd's mirror pattern by generating a sequence of pulses (rahmonics) located at the fundamental quefrency (periodic time) and its multiples. This sequence is referred to here as the 'rahmonic component' of the power cepstrum. The fundamental quefrency, which is the reciprocal of the frequency difference between adjacent interference fringes, equates to the multipath time delay. The other component of the power cepstrum is the non-rahmonic (extraneous) component, which combines with the rahmonic component to form the (total) power cepstrum. A data processing technique, termed 'cepstrum subtraction', is described. This technique suppresses the extraneous component of the power cepstrum, leaving the rahmonic component that contains the desired multipath time delay information. This technique is applied to real acoustic recordings of motor-vessel transits in a shallow water environment, where the broadband noise radiated by the vessel arrives at the hydrophone via a direct ray path and a time-delayed multipath. The results show that cepstrum subtraction improves multipath time delay estimation by a factor of two for the at-sea experiment.

Index Terms— time delay estimation, underwater acoustics, cepstrum, source localization, autocorrelation

1. INTRODUCTION

In underwater acoustics, the Lloyd's mirror effect [1, 2] is observed as an interference pattern in the spectrogram (variation with time of the spectrum) of a receiving hydrophone. It occurs when the sinusoidal signal components that compose a source of broadband radiated noise arrive at the receiver via the direct propagation path and combine with amplitude-scaled time-delayed replicas arriving via indirect propagation paths (multipaths). If the two signal arrivals are *in phase*, they reinforce each other through coherent addition resulting in *constructive* interference; when they are in *antiphase*, they cancel and *destructive* interference occurs. As a result, the received intensities of a source's spectral lines are observed to undergo a periodic, or cyclic, variation with range [3]. The resultant intensity is given by

$$A_r^2 = A_d^2 + A_i^2 + 2A_dA_i \cos(\kappa\delta l), \quad (1)$$

where A_d and A_i are the amplitudes of the respective direct path and indirect path arrivals, $\kappa = 2\pi/\lambda$ (where λ denotes wavelength)

*The author gratefully acknowledges the contributions provided by the IEEE Oceanic Engineering Society Scholarships, the University of Sydney Electrical and Information Engineering PhD Scholarship, and the Defence Science and Technology Group Postgraduate Research Scholarship.

is the spatial frequency (or wavenumber), and δl is the path length difference between the two paths. The last term in (1) is oscillatory due to interference between the direct path and multipath arrivals, giving rise to a Lloyd's mirror interference pattern [4]. Destructive interference fringes occur when $\kappa\delta l = (2k - 1)\pi$, $k = 1, 2, 3, \dots$, where k is the fringe order (number). For adjacent destructive interference fringes

$$(\kappa_{k+1} - \kappa_k)\delta l = 2\pi. \quad (2)$$

Hence from (2), at any given instant in time, the destructive interference fringes have a uniform spacing in the direction of the frequency axis which is given by

$$(f_{k+1} - f_k) = \frac{c}{\delta l} = \frac{1}{\tau_{i,d}}, \quad (3)$$

where c is the isospeed of sound travel in the underwater medium and $\tau_{i,d}$ is the time difference in the arrival times of the signal propagating via the respective indirect and direct paths. Note that the temporal frequency difference $f_{k+1} - f_k$ equates to the reciprocal of the multipath time delay. Knowledge of the time delay between the direct propagation path and seafloor reflected multipath is useful, since it enables a source to be passively localized in range using a single sensor [5, 6, 7].

Fundamentally, the *cepstrum* is the spectrum of a logarithmic spectrum and is used in practice to detect *periodicity in a frequency spectrum*. The *quefrency* is a measure of the periodicity of the spectral ripple and it has the units of (periodic) time. In speech analysis, periodicity in the voice spectrum is observed as a regular spacing in frequency of the 'fundamental voice frequency' (pitch) and its harmonics [8, 9, 10]. If the spectrum of the logarithm of the speech power spectrum is computed, a peak which corresponds to the pitch period of the voiced-speech segment being analysed will appear [9].

For the present case, the transit of a broadband source past an acoustic sensor in an underwater acoustic multipath propagation environment, periodic structure is observed in the output spectrogram as an ordered arrangement of Lloyd's mirror interference fringes which form a regular periodic pattern. The *quefrency* can be considered as the periodic time associated with a series of interference fringes that are uniformly spaced in frequency. The power cepstrum is generated by taking the inverse Fourier transform of the logarithm of the power spectrum [11]. The first rahmonic is the fundamental period, which is equal to the reciprocal of the frequency interval between any two adjacent destructive interference fringes (or, equally, between any two adjacent constructive interference fringes). Hence, cepstrum analysis is suited to detecting and quantifying periodicity in a spectrogram. It is also suited to multipath time delay estimation [12, 13].

The original contribution of this research is a novel data processing technique that suppresses cepstrum components extraneous

to the rahmonics, thereby improving the estimation of the multipath time delay for broadband signals arriving at a receiving hydrophone. This improvement in time delay estimation is demonstrated using real data collected during an at-sea experiment.

2. UNDERWATER ACOUSTIC SENSOR OUTPUT MODEL

2.1. Direct path signal combined with indirect path signal

For the present case, the signal of interest is continuous broadband acoustic noise radiated by a small motor-powered boat underway. The output of the hydrophone, which is located above the seafloor, is modeled as a combination of: (a) an underwater acoustic signal propagating directly from the source to the sensor $s(t)$, where t denotes time; and (b) an amplitude-scaled time-delayed replica propagating via an indirect path that includes a seafloor boundary reflection $\alpha s(t - \tau_\beta)$, where α is the attenuation (or boundary reflection) coefficient ($0 < \alpha \leq 1$), and τ_β is the time delay. The combination of the signal with a single echo can be represented as [10]

$$x(t) = s(t) + \alpha s(t - \tau_\beta). \quad (4)$$

The time delay between the indirect propagation path and the direct propagation path results in a phase difference between the direct path and multipath signals equivalent to $2\pi f\tau_\beta$, where f is the signal frequency. Assuming that the transmission loss is the same for each propagation path, the instantaneous power of the resultant signal is given by [10]

$$|X(f)|^2 = |S(f)|^2 (1 + \alpha^2 + 2\alpha \cos(2\pi f\tau_\beta)), \quad (5)$$

where the power spectrum of the direct path signal $|S(f)|^2$ is modulated by a periodic function of the frequency. By taking the logarithm of the output power spectrum the *product* in (5) is converted to a *sum* of two components, i.e.

$$\log |X(f)|^2 = \log |S(f)|^2 + \log (1 + \alpha^2 + 2\alpha \cos(2\pi f\tau_\beta)), \quad (6)$$

where the additive periodic component has a period determined by the multipath delay τ_β .

2.2. Power cepstrum

For the periodic structure of a Lloyd's mirror interference pattern, the queffency (or periodic time) of the first harmonic is the reciprocal of the uniform frequency difference between adjacent destructive interference fringe minima. For the present case, Lloyd's mirror interference pattern formation results from the combination of the direct and indirect propagation path signals at the sensor and the queffency of the first rahmonic is the multipath time delay. In other words, cepstrum analysis of the hydrophone's noisy signal output provides an estimate of the multipath time delay τ_β , which is the difference in the arrival times of the direct path signal and the indirect path signal at the hydrophone. The definition of the *power cepstrum* $C_{XX}(\tau)$ can be expressed mathematically as [11]

$$C_{XX}(\tau) = \mathcal{F}^{-1} \{ \log |X(f)|^2 \}, \quad (7)$$

where τ is the queffency (or delay time), \mathcal{F}^{-1} denotes the inverse Fourier transform, and $|X(f)|^2$ is equivalent to the two-sided power spectrum of a received signal $x(t)$.

In the present case, the power cepstrum is given by the inverse Fourier transform \mathcal{F}^{-1} of (6), which can be written as

$$C(\tau) = \mathcal{F}^{-1} (\log |S(f)|^2) + \sum_{n=1}^{\infty} a_n \delta(\tau - n\tau_\beta), \quad (8)$$

where a_n is the strength of the n^{th} rahmonic and is given by

$$a_n = \frac{2}{n} (-1)^{n+1} \alpha^n \text{ for } n = 1, 2, 3, \dots \quad (9)$$

The rahmonic strengths are observed to alternate in sign and decrease as the reciprocal of the rahmonic number n , in the same way as the Mercator series. Hence, the power cepstrum of the received signal consists of the power cepstrum of the direct path signal $S(f)$ and a sequence of impulse functions (or rahmonics) of strength a_n located at queffency τ_β and its multiples. Figure 1(a) shows the computed power cepstrum $C(\tau)$ when a broadband direct-path signal is combined with a time-delayed replica ($\alpha = 1$, $\tau_\beta = 224 \mu\text{s}$). Figures 1(b) and 1(c) show the variation with queffency of the respective power cepstrum components; the non-rahmonic component is given by $C_1(\tau) = \mathcal{F}^{-1} (\log |S(f)|^2)$, and the rahmonic impulse component is given by $C_2(\tau) = \sum_{n=1}^{\infty} a_n \delta(\tau - n\tau_\beta)$. Figure 1(d) shows the autocorrelation as a function of positive lag, with the maximum value of the echo's sinc function corresponding to a lag (i.e. the echo time delay) of $224 \mu\text{s}$. In practice, additive noise biases the autocorrelation function and may obscure the peaks corresponding to propagation time delay difference [14]. Also, when there is more than one indirect propagation path, the identification of individual propagation paths using autocorrelation analysis becomes more difficult. Cepstrum analysis is thus preferred because each propagation path has a series of more readily identifiable cepstrum peaks located at the rahmonic queffencies that characterize a specific propagation path [14].

2.3. Cepstrum subtraction

A novel data processing technique is presented that suppresses cepstrum components extraneous to the rahmonics. From (8), the power cepstrum of the received signal $C(\tau)$ consists of the non-rahmonic components $C_1(\tau)$ derived from stationary process $S(f)$ (which remains stationary after inverse Fourier transformation [15]), and a sequence of rahmonic impulse functions $C_2(\tau)$ for another (independent) stationary process [16]. Note that $C_1(\tau)$ is independent of the time delay variable τ_β , whereas $C_2(\tau)$ is dependent, with impulses located at queffency τ_β and its multiples.

The ensemble average of the power cepstrum over M realizations may be given by

$$\frac{1}{M} \sum_{m=1}^M C_m(k) = \frac{1}{M} \sum_{m=1}^M C_{1,m}(k) + \frac{1}{M} \sum_{m=1}^M C_{2,m}(k), \quad (10)$$

where $C_{1,m}(k)$ and $C_{2,m}(k)$ are the m^{th} realizations at discrete queffency k for the non-rahmonic and rahmonic components, respectively. Assuming τ_β varies over the M realizations, the expected value of $C_2(k)$ at any queffency is zero [17], i.e.

$$\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M C_{2,m}(k) = 0. \quad (11)$$

Here, the physical significance of a varying τ_β is a moving broadband source, such that the queffency of the rahmonic component changes throughout the ensemble average time. Therefore, the non-rahmonic component $C_1(k)$ can be estimated by taking the average of the received power cepstrum $\bar{C}_1(k)$ over a sufficiently long time (i.e. for a sufficiently large M), where

$$\bar{C}_1(k) = \frac{1}{M} \sum_{l=1}^M C_m(k). \quad (12)$$

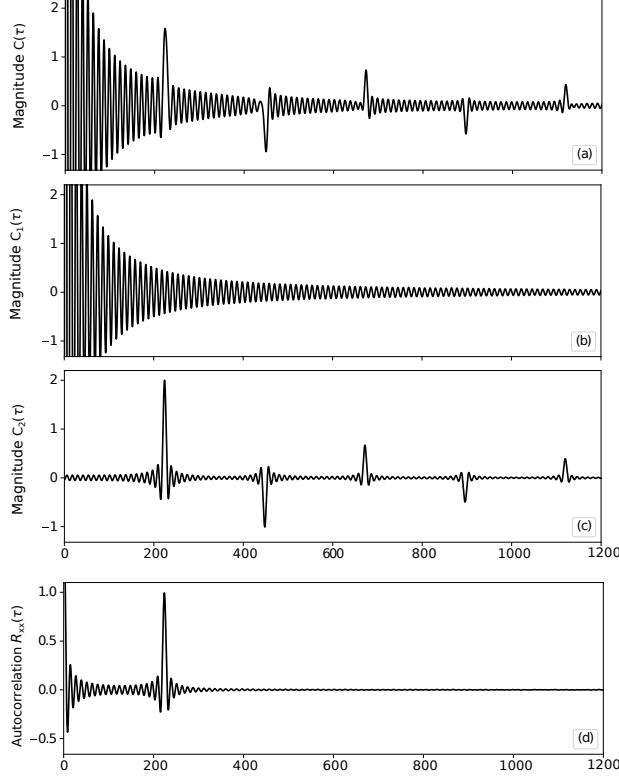


Fig. 1. (a) Variation with quefrency of the power cepstrum $C(\tau)$ of a direct-path signal combined with an indirect-path echo $\alpha = 1$, $\tau_\beta = 224 \mu\text{s}$. (b) Variation with quefrency of the power cepstrum component $C_1(\tau)$. (c) Similar to (b), but for the rahmonic impulse component $C_2(\tau)$. (d) Variation with lag of the autocorrelation function for the direct path signal combined with the multipath echo.

Note that the long time-averaged rahmonic contribution at quefrency k is negligible for a sufficiently large M . This approach is consistent with other ensemble averaging methods [18].

The rahmonic component for realization m can be estimated by subtracting the estimate of the non-rahmonic component from the received power cepstrum. The cepstrum subtraction process is described mathematically by,

$$\hat{C}_{2,m}(k) = C_m(k) - a(k)\bar{C}_1(k), \quad (13)$$

where $\hat{C}_{2,m}(k)$ is an estimate of the rahmonic component for the m^{th} frame at discrete quefrency k , $C_m(k)$ is the received power cepstrum for the m^{th} frame at quefrency k , and $\bar{C}_1(k)$ is the long time-averaged received cepstrum given by (12). The parameter $a(k)$ controls the amount of the non-rahmonic component that is subtracted from the power cepstrum. For full non-rahmonic component subtraction $a(k) = 1$, and for over-subtraction $a(k) > 1$, analogous to spectral subtraction [19]. The cepstrum subtraction process is visualized in Figure 1, where the non-rahmonic contribution in (b) is subtracted from the received power cepstrum in (a), leaving only the rahmonic component in (c).

Note that the cepstrum subtraction method has the same formulation as cepstrum mean normalization (which is used in speech processing) but with the addition of factor $a(k)$ [8, 20, 21]. Rather than removing channel-effects (distortions) for Mel-frequency cepstrum coefficients, here cepstrum components extraneous to the rahmonics

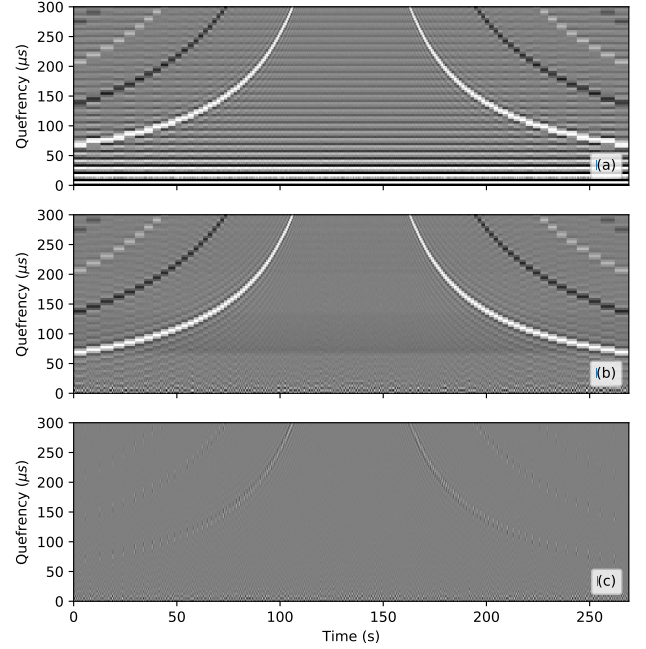


Fig. 2. Various power cepstrograms for simulated acoustic data, with quefrencies up to $300 \mu\text{s}$. (a) Variation with time of the power cepstrum, (b) Similar to (a), but with cepstrum subtraction where the non-rahmonic component is estimated from the entire transit ($M = 2700$), and (c) Similar to (b) but the non-rahmonic component is estimated from the last 2 seconds ($M = 20$) of received cepstra, which have significant rahmonic contributions that bias the estimate.

are suppressed. Also the two methods are applied to the outputs of different signal models.

Figure 2 shows various power cepstrograms for a simulated motor-boat transit using synthetic broadband radiated noise. The source/sensor geometry is the same for the at-sea experiment discussed below in subsection 3.1. Figure 2(a) shows the power cepstrum without cepstrum subtraction. The horizontal banding associated with the non-rahmonic component is prominent, which can be seen to interfere with the rahmonics. Figure 2(b) shows that the effect of cepstrum subtraction is to suppress the non-rahmonic component, thus restoring the fidelity of the rahmonic component. Here, $M = 2700$, which is sufficiently large such that the contribution of the rahmonic component to the time averaged-power cepstrum is negligible. The strengths of the rahmonics relative to the non-rahmonics have increased. In contrast, Figure 2(c) shows that the cepstrum subtraction process has suppressed both the rahmonic and non-rahmonic components because $M = 20$ is small, i.e. the contribution of the rahmonic component to the time-average power cepstrum is significant. Cepstrum subtraction requires then multipath time delay τ_β to be a rapidly changing over M . Otherwise, the rahmonic strengths bias the estimate of the non-rahmonic component of the power cepstrum, violating (11).

3. EXPERIMENTAL RESULTS

The multipath time delay, or the time difference in the arrivals of the signal via the direct propagation path and the signal via the indi-

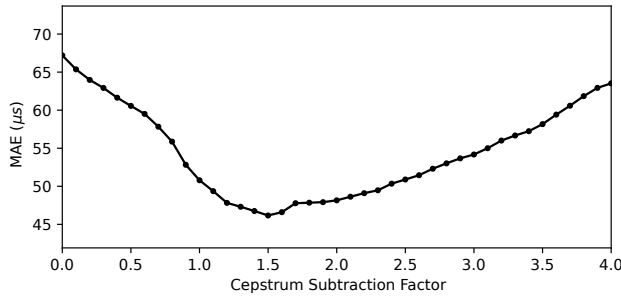


Fig. 3. Variation with the cepstrum subtraction factor $a(k)$ of the mean absolute error for multipath time delay estimation.

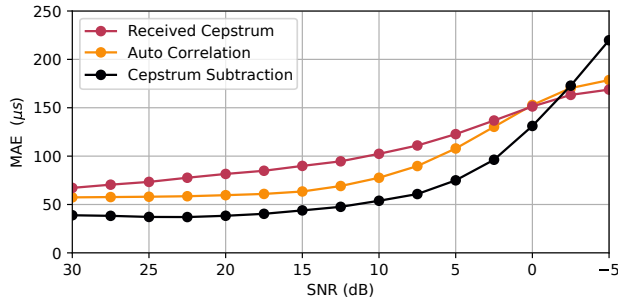


Fig. 4. Variation with signal-to-noise ratio of the mean absolute error for multipath time delay estimation.

rect seafloor-reflected multipath, is estimated using various cepstrum methods and the autocorrelation function. The performance of the time delay estimation methods is compared using acoustic data collected during an at-sea experiment.

3.1. Dataset

An experiment was conducted where a propeller-driven boat, which radiated broadband underwater acoustic noise, transited past a hydrophone located 1 m above the seafloor in very shallow water (20 m deep). Fourteen transits were recorded over a two day period, where the vessel approached the sensor from different directions. A spectrogram of the hydrophone's output was observed to display a Lloyd's mirror interference pattern over the entire frequency band (≤ 90 kHz) of the receiver, which was sampled at 250 kHz. Recording commenced when the vessel was inbound at a ground range of 200 m. The vessel transited past the sensor, and recording was terminated when the vessel was 200 m outbound. The vessel's position was logged at 0.1 s intervals using a RF tracking device.

The multipath time delay is predicted as a function of the vessel's ground range, assuming an isovelocity (1520 ms^{-1}) sound propagation medium and specular reflection from a flat seafloor [22]. Acoustic recordings are divided into 0.1 s duration sound clips. The power cepstrum, power cepstrum with cepstrum subtraction, and autocorrelation function, are computed for each sound clip. The estimate of the multipath time delay is the quefrency (or lag) corresponding to the peak value in each of these functions. The estimated multipath time delay is compared with the predicted time delay. Six minutes of background recordings (with no source present) was also collected in order to measure the ambient noise spectrum, which was

used to produce various levels of additive noise for signal-to-noise ratio (SNR) experiments.

3.2. Comparison of time delay estimation methods

Figure 3 shows the mean absolute error (MAE) for multipath time delay estimation with cepstrum subtraction, as a function of the subtraction factor $a(k)$ in (13). When $a(k) = 0$, there is no cepstrum subtraction and the time delay is estimated from the received power cepstrum. When $a(k) = 1$, full cepstrum subtraction occurs. When $a(k) > 1$, there is over-subtraction. Non-rahmonic components are estimated by taking the mean of the received power cepstra for the entire acoustic dataset. The MAE decreases with cepstrum subtraction, which demonstrates the benefit of non-rahmonic component suppression in the received power cepstrum. The minimum MAE occurs when $a(k) = 1.5$. This level of over-subtraction reduces the MAE by $20 \mu\text{s}$ when compared with nil subtraction i.e. $a(k) = 0$.

The multipath time delay is estimated from (a) the power cepstrum, (b) the power cepstrum with cepstrum (over) subtraction, and (c) the autocorrelation function. The estimate of the multipath time delay is the quefrency (or lag) corresponding to the peak value in each of these functions. Colored noise with the same power spectral density as background noise recordings (without the source) is added to recordings in varying amounts in order to change the SNR. This enables time delay estimation performance to be tested as a function of SNR. Figure 4 shows that cepstrum subtraction provides the best overall performance, confirming that suppression of non-rahmonic components improves time delay estimation performance.

4. CONCLUSION

Cepstrum analysis enables estimation of the multipath time delay for broadband signals arriving via direct and indirect signal propagation paths at a receiving hydrophone. A cepstrum subtraction method suppresses the cepstrum components extraneous to the rhamonics. As a result, the magnitude of the rahmonics relative to the non-rahmonics are increased. Also, there is an improvement in multipath time delay estimation, which is demonstrated using real data collected during an at-sea experiment. The experimental results showed improved multipath time delay estimation performance in a very shallow water environment when compared with other time delay estimation methods, and robustness to decreasing signal-to-noise ratios. Multipath time delay estimation using cepstrum subtraction is superior to using the autocorrelation function when the $\text{SNR} \geq 0$. However, without cepstrum subtraction, the autocorrelation function is better than the power cepstrum for multipath time delay estimation in the present application. Cepstrum over-subtraction improves multipath time delay estimation.

5. REFERENCES

- [1] Robert J Urick, *Principles of Underwater Sound*, pp. 123–125, McGraw-Hill, New York, 2nd edition, 1975.
- [2] William M Carey, "Lloyd's mirror-image interference effects," *Acoustics Today*, vol. 5, no. 2, pp. 14–20, 2009.
- [3] Daphne Kapolka, Jason K Wilson, Joseph A Rice, and Paul Hursky, "Equivalence of the waveguide invariant and two path ray theory methods for range prediction based on lloyd's mirror patterns," in *Proceedings of Meetings on Acoustics 155*. ASA, 2008, vol. 4, p. 070002.

- [4] Adrian D Jones, David W Bartel, and Paul A Clarke, "An application of range-frequency striations to seafloor inversion in shallow oceans," in *Proceedings of Acoustics 2012-Fremantle*, 2012, pp. 1–7.
- [5] Sérgio M Jesus, Michael B Porter, Yann Stéphan, Emanuel Coelho, and Xavier Démoulin, "Broadband source localization with a single hydrophone," in *OCEANS'98 Conference Proceedings*. IEEE, 1998, vol. 2, pp. 1078–1082.
- [6] Sérgio M Jesus, Michael B Porter, Yann Stéphan, Xavier Démoulin, Orlando C Rodríguez, and Emanuel MM Ferreira Coelho, "Single hydrophone source localization," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 3, pp. 337–346, 2000.
- [7] Kam W Lo, Brian G Ferguson, Yujin Gao, and Alain Maguer, "Aircraft flight parameter estimation using acoustic multipath delays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 1, pp. 259–268, 2003.
- [8] A Michael Noll, "Short-time spectrum and "cepstrum" techniques for vocal-pitch detection," *The Journal of the Acoustical Society of America*, vol. 36, no. 2, pp. 296–302, 1964.
- [9] A Michael Noll, "Cepstrum pitch determination," *The Journal of the Acoustical Society of America*, vol. 41, no. 2, pp. 293–309, 1967.
- [10] Alan V Oppenheim and Ronald W Schafer, "From frequency to quefrency: A history of the cepstrum," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 95–106, 2004.
- [11] Robert B Randall, "A history of cepstrum analysis and its application to mechanical problems," *Mechanical Systems and Signal Processing*, vol. 97, pp. 3–19, 2017.
- [12] Eric L Ferguson, Rishi Ramakrishnan, Stefan B Williams, and Craig T Jin, "Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2657–2661.
- [13] Eric L Ferguson, Stefan B Williams, and Craig T Jin, "Sound source localization in a multipath environment using convolutional neural networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2018 IEEE International Conference on*. IEEE, 2018, pp. 2386–2390.
- [14] Julius S Bendat and Allan G Piersol, "Engineering applications of correlation and spectral analysis," *John Wiley & Sons*, pp. Sect. 6.2, 145–153, 1993.
- [15] Julius S Bendat and Allan G Piersol, *Random Data: Analysis and Measurement Procedures*, chapter 3, pp. 56–98, John Wiley & Sons, New York, 1st edition, 1971.
- [16] Julius S Bendat and Allan G Piersol, *Random Data: Analysis and Measurement Procedures*, chapter 1, pp. 1–36, John Wiley & Sons, New York, 1st edition, 1971.
- [17] Timo Gerkmann and Rainer Martin, "On the statistics of spectral amplitudes after variance reduction by temporal cepstrum smoothing and cepstral nulling," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4165–4174, 2009.
- [18] Julius S Bendat and Allan G Piersol, *Random Data: Analysis and Measurement Procedures*, pp. 344–380, John Wiley & Sons, New York, 1st edition, 1971.
- [19] S V Vaseghi, *Advanced Digital Signal Processing*, chapter 12, pp. 321–338, John Wiley & Sons, West Sussex, UK, 4th edition, 2008.
- [20] Philip N Garner, "Cepstral normalisation and the signal to noise ratio spectrum in automatic speech recognition," *Speech Communication*, vol. 53, no. 8, pp. 991–1001, 2011.
- [21] Sadaoki Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 2, pp. 254–272, 1981.
- [22] Eric L Ferguson, Rishi Ramakrishnan, Stefan B Williams, and Craig T Jin, "Deep learning approach to passive monitoring of the underwater acoustic environment," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 3351–3351, 2016.