Breakey, D., Meskell, C., Comparison of metrics for the evaluation of similarity in acoustic pressure signals, Journal of Sound and Vibration, 332, (15), 2013, p3605-3609
http://dx.doi.org/10.1016/j.jsv.2013.02.033

Comparison of metrics for the evaluation of similarity in acoustic pressure signals

David Breakey[a,*], Craig Meskell[a]

[a] *Department of Mechanical and Manufacturing Engineering, Trinity College, Dublin 2, Ireland.*

Abstract

Determining if aeroacoustic sound predictions are accurate is difficult because the question of how to define 'accurate' remains open. This communication evaluates four metrics for comparing time-domain pressure signals, each implying its own definition of 'accurate.' An adaptation of a Structural Similarity Metric (originating from image processing literature) to time-frequency representations of acoustic signals is shown to outperform typical metrics such as relative energy and mean square error.

Keywords: Aeroacoustics, Signal similarity, Sound prediction

1. Introduction

Recent techniques that allow comparison between time-resolved predicted and recorded sound fields include time-resolved comparisons of acoustic analogies with DNS calculations [1], semi-empirical sound calculations from Time-resolved PIV data [2, 3], Kirchhoff surface projection methods [4], and real-time projections to simplified source models with analytical solutions [5, 6]. Ultimately, the comparisons are made between either spectral or time-domain signal shapes, finding the maximum deviation or determining qualitatively if they are 'similar'. A quantitative similarity metric would be helpful in evaluating these techniques, and an optimal metric would be sensitive to errors indicating poor model agreement, while remaining relatively invariant to random noise and small errors arising from unmodelled phenomena.

*Corresponding author

Email addresses: breakeyd@tcd.ie (David Breakey), cmeskell@tcd.ie (Craig Meskell)

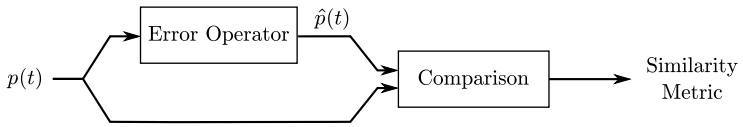

Figure 1: Process model for current study

This work considers four metrics from this perspective. A Structural Similarity Metric (SSIM) originating from image compression analysis has been successfully applied recently to time-frequency representations of biological neurogram signals [7] and speech signals [8], prompting the current study into its use for aeroacoustics. For images, the optimal compression stores an acceptable representation of the image as compactly as possible, ignoring extraneous detail and retaining only the information most important for human interpretation (*i.e.* structural information). The aeroacoustic analogue is that sound predictions cannot be expected to predict all the rich detail of the sound emissions, but they must capture the mechanics of the dominant sound emissions to be useful.

2. Analysis procedure

Fig. 1 shows the process model used in this study. A reference signal, $p(t)$, undergoes one of four different error operators (see section 2.1 and table 1) resulting in a corrupted signal, $\hat{p}(t)$, and the two signals are compared to give a single similarity metric. In this paper, the candidate metrics are compared as a function of the magnitude of the applied error.

2.1. Error Operators

Random noise Normally distributed random signal noise was added to the time series scaled by ϵ so that when $\epsilon = 1$, the energy of the signal noise is the same magnitude as the signal energy. This represents measurement errors and uncorrelated, unmodelled sound generation.

Time lag The time lag was applied by a frequency domain $e^{-\mathrm{i}\omega t_{\mathrm{lag}}}$ phase shift such that when $t_{\mathrm{lag}} f_c = 1$, the signal has been shifted by one period of the centre frequency. This represents uncertainty in the sound propagation path from aeroacoustic sources in turbulence.

Table 1: Error expressions for metric comparison

Name	Expression	Variable
Random noise[a]	$p(t) + \sqrt{\epsilon \langle p^2 \rangle} X(t)$	ϵ
Time lag	$p(t - t_{\text{lag}})$	$t_{\text{lag}} f_c$
Frequency shift	$p(t/(1 + \delta f_c / f_c))$	$\delta f_c / f_c$
Jittery phase[a]	$\frac{1}{N} \sum_i^N p(t + \kappa X(i)), N = 500$	κf_c
Quadratic noise	$\nu_2 p(t)^2 \sqrt{\langle p^2 \rangle / \langle p^4 \rangle} + (1 - \nu_2) p(t)$	ν_2
Cubic noise	$\nu_3 p(t)^3 \sqrt{\langle p^2 \rangle / \langle p^6 \rangle} + (1 - \nu_3) p(t)$	ν_3

[a] $X(\cdot)$ is a normally distributed random variable with unit standard deviation

Frequency shift A linearly interpolated resampling of the signal provided the frequency shift where the shifted centre frequency is $f_c + \delta f_c$, representing errors due to Doppler shift, associated with convecting sources.

Jittery phase This error simulates the effect of multiple paths from a turbulent sound source to an observer, a phenomenon attributable to turbulent fluctuations that cause differences in refraction during sound propagation, leading to random self-interference of the source. κ controls the standard deviation of the propagation time. When $\kappa f_c = 1$, the standard deviation is one period of the signal's centre frequency.

Quadratic and cubic noise These types of noise imitate non-linear interactions between acoustic sources. The noise is controlled as a ratio of linear and higher-order terms. When $\nu_i = 1$, with $i = 2$ or 3, the signal is fully quadratic or cubic, while $\nu_i = 0$ gives linear signals.

2.2. Metric Types

Four metrics are investigated in this paper: two classical metrics—relative energy (RE) and mean square error (MSE)—and two variations of the SSIM introduced by Wang et al. [9].

2.2.1. Relative energy (RE)

The RE metric simply compares the ratio of the total energy of two time signals. In functional form

$$\mathrm{RE} = \frac{\int \hat{p}(t)^2 dt}{\int p(t)^2 dt}. \tag{1}$$

For identical signals, RE = 1, which is the optimal value for similarity. RE is invariant under shifts in time and frequency provided the signal has a constant envelope. Different signals can have the same energy, leaving RE with little discriminative power. The spectral and directivity comparisons used frequently in aeroacoustics are a form of energy metric, where the comparison is made with subsets of the collected data.

2.2.2. Mean square error (MSE)

The MSE is a normalized measure of the mean deviation between two signals. It is expressed simply as

$$\mathrm{MSE} = \frac{\int (\hat{p}(t) - p(t))^2 dt}{\int p(t)^2 dt}. \tag{2}$$

Unlike RE, MSE is highly discriminatory between different signals, and MSE = 0 (the optimal value) only occurs when the signals are identical.

2.2.3. Structural Similarity Metric (SSIM)

The SSIM was developed for the comparison of the quality of images subject to various compression algorithms [9], and measures how similar images are with respect to their structure within a particular convolution window. In each convolution window and for each image (x, y), local means (μ_x, μ_y), standard deviations (σ_x, σ_y), and an auto-covariance (σ_{xy}) are calculated and compared. The SSIM is a local measure, which operates on a 2D scalar field, returning a 2D scalar field that can be averaged to find the mean SSIM (MSSIM) for a particular image pair. These values are calculated by

$$\mathrm{SSIM}(x, y) = l(x, y)c(x, y)s(x, y), \tag{3}$$

$$\mathrm{MSSIM} = \frac{1}{M} \sum_{j=1}^{M} \mathrm{SSIM}(x_j, y_j), \tag{4}$$

where

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{5}$$

$$C_1 = (K_1 L)^2, \qquad C_2 = (K_2 L)^2, \qquad C_3 = C_2/2. \tag{6}$$

C_1, C_2, and C_3 avoid singularities when the denominators would otherwise be small, where K_1 and K_2 are arbitrarily taken as 0.01 and 0.03, respectively, and L is the dynamic range of the image [9]. Regardless of the choice of constants, MSSIM = 1 occurs only if the images are identical. Following Wang et al. [9], a Gaussian-weighted convolution window was chosen.

Equation (4) implies that every part of the 2D field is equally important for analysing the structure, which is a poor assumption for acoustic signals. In light of this, a weighted SSIM (WSSIM) was used

$$\text{WSSIM} = \sum_{j=1}^{M} w(x_j)\text{SSIM}(x_j, y_j) \tag{7}$$

where $w(x_j)$ is a weighting function defined by the local magnitude of the reference signal.

1D SSIM ($WSSIM_{1D}$). Though the SSIM was developed for 2D signals (images), it can be applied to 1D time series by considering the signal as a $N_t \times 1$ pixel image and using a 1D convolution window. The formula is still given by Eq. (7). Computationally, this is preferable to the time-frequency case because it eliminates several steps from the calculation.

Time-frequency SSIM (WSSIM). Hines et al. [8], applied the SSIM to short-time Fourier transforms of acoustic signals, where the image 'intensity' was the modulus of the Fourier coefficients. In this work, a continuous wavelet transform (CWT) was used to take advantage of the higher temporal resolution for higher frequency signal components. The wavelet used here was a complex-valued Morlet wavelet with $\omega_0 = 6$ as it shares the characteristic shape of a modulated wavepacket, leading to efficient collapse of the investigated signals onto a few wavelet coefficients. The wavelet scales can be converted to pseudo-frequencies [10], and the resulting time-frequency 'image' is then $\tilde{p}(t, f_{\text{pseudo}}) = |\text{CWT}[p(t)]|$ for use in Eq. (7).

3. Results

3.1. Test signals

Two test signals were investigated: a simulated Gaussian-modulated sinusoid and a experimental far-field recording from a free jet flow [11]. Both signals were sampled at $f_s = 100$ kHz. The chosen centre frequency for the simulated signal was $f_c = 1/T_0 = 1225$ Hz, which corresponds to a Strouhal number $St = fD/U = 0.3$ for the experimental recording. In each case, the signals spanned $-300T_0 < t < 300T_0$, but only the centre third of the points were used in the comparison to avoid edge effects after applying the frequency and time shift errors. The simulated signal corresponds in shape to the solution of wavepacket-type models for jet noise [5]:

$$p(t) = -\exp(-t^2/T^2)\exp(\mathrm{i}2\pi f_c t), \tag{8}$$

where the characteristic width was set by $T = \sqrt{2}T_0$.

The experimental presure signal was the axisymmetric mode of a far-field microphone ring ($r/D = 47$ and $\theta = 20°$) of a $M = 0.6$, $D = 50$ mm jet in an anechoic jet facility at the *Centre d'Etudes Aérodynamiques et Thermiques* (CEAT), Institut Pprime, Poitiers, France. Further details of the jet facility are given by Cavalieri et al. [11]. For the WSSIM, the wavelet scales were chosen to correspond to 200 evenly-spaced pseudo-frequencies ranging from the experimental anechoic cut-off, 200 Hz, to 50 000 Hz ($\Delta f \approx 250$ Hz).

3.2. Metric peformance

An optimal metric would have steady, monotonic deviation from the optimal value as the error magnitude is increased and low sensitivity to random signal noise. The monotonic deviation ensures that the magnitude of the error operators can be detected if they are uncertain in the prediction model. Figs. 2 & 3 indicate the results of the metric computations for the simulated and experimental cases, respectively. The energy metric performs as expected with a monotonic sensitivity to signal noise and no discrimination under time and frequency shifts. The MSE shares this monotonic sensitivity to signal noise, but the bigger issue is the dependence of the response to a time lag on the autocorrelation of the input signal. For periodic and semi-periodic signals, the presence of several local minima and maxima could be misleading for prediction verification. The $\mathrm{WSSIM_{1D}}$ shares this problem under a time lag, though performs better with signal noise. Strikingly, the

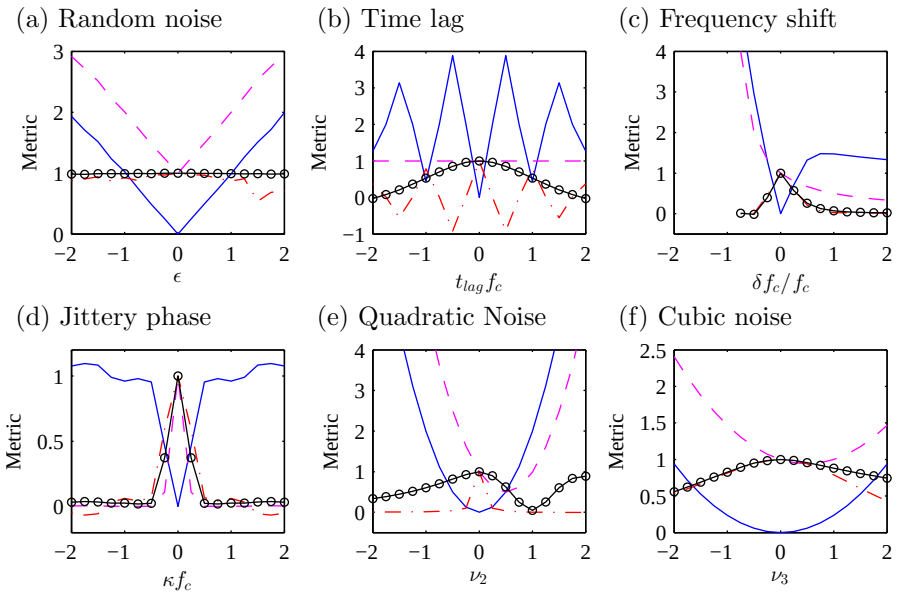

Figure 2: Results from metric tests with simulated data. – – – Relative energy (RE), ——— Mean square error (MSE), – · – · – 1D SSIM (WSSIM_{1D}), —○— Time-frequency SSIM (WSSIM).

WSSIM overcomes both of these issues, indicating the best performance. For jittery phase error, all the metrics perform similarly, and for the higher-order noise, though the shapes of each metric differ, none is obviously better than the others.

4. Conclusions

This work compared the performance of four similarity metrics for time-domain pressure signals. The relative energy metric exhibits poor discrimination between signals and high sensitivity to random noise. The mean square error metric is also highly sensitive to random noise and phase error. The Weighted Structural Similarity Metric using a time-frequency signal representation shows promise because it is resistant to contamination by random noise while still providing clear discrimination between time lags and frequency shifts, information pertinent to noise source model validation.

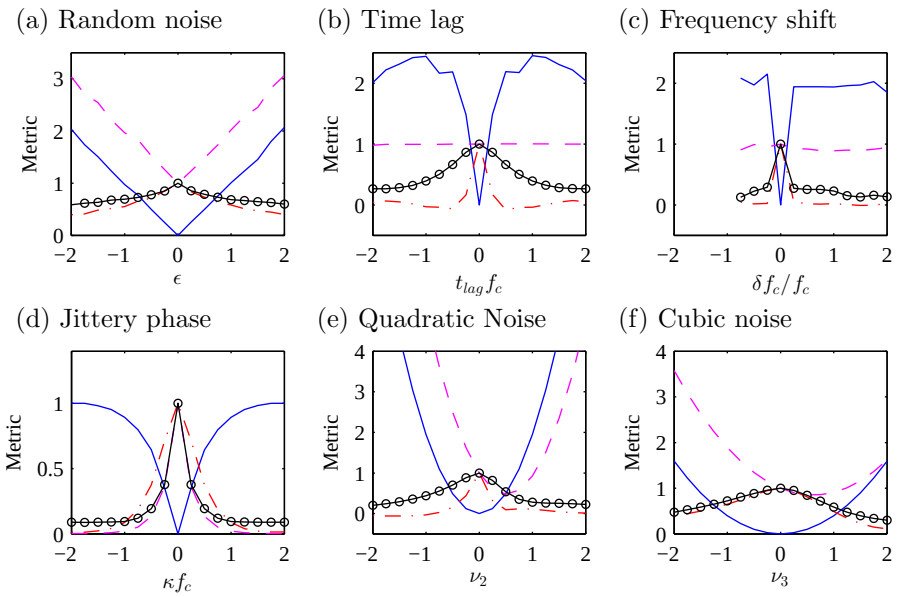

Figure 3: Results from metric tests with experimental far-field data. – – – Relative energy (RE), ——— Mean square error (MSE), – · – · – 1D SSIM (WSSIM_{1D}), —o— Time-frequency SSIM (WSSIM).

Acknowledgements

The authors thank the late Prof. John Fitzpatrick, who originated this project. D. B. also thanks Drs. Peter Jordan and André Cavalieri for their relevant discussions and help obtaining the experimental data. Science Foundation Ireland supported this work under contract 09/RFP/ENM2469.

References

[1] A. Samanta, J. B. Freund, M. Wei, S. K. Lele, Robustness of acoustic analogies for predicting mixing-layer noise, AIAA Journal 44 (2006) 2780–2786.

[2] V. Koschatzky, P. Moore, J. Westerweel, F. Scarano, B. Boersma, High speed PIV applied to aerodynamic noise investigation, Experiments in Fluids 50 (2011) 863–876.

[3] C. Haigermoser, Application of an acoustic analogy to PIV data from rectangular cavity flows, Experiments in Fluids 47 (2009) 145–157.

[4] A. S. Lyrintzis, Review: the use of Kirchhoff's method in computational aeroacoustics, Journal of Fluids Engineering 116 (1994) 665–676.

[5] A. V. G. Cavalieri, P. Jordan, A. Agarwal, Y. Gervais, Jittering wave-packet models for subsonic jet noise, Journal of Sound and Vibration 330 (2011) 4474–4492.

[6] F. Kerhervé, P. Jordan, A. V. G. Cavalieri, J. Delville, C. Bogey, D. Juvé, Educing the source mechanism associated with downstream radiation in subsonic jets, Journal of Fluid Mechanics 710 (2012) 606–640.

[7] A. Hines, N. Harte, Speech intelligibility prediction using a neurogram similarity index measure, Speech Communication 54 (2012) 306–320.

[8] A. Hines, J. Skoglund, A. Kokaram, N. Harte, ViSQOL: the virtual speech quality objective listener, in: Proceedings of the International Workshop on Acoustic Signal Enhancement, Aachen, Germany, 2012, pp. 1–4.

[9] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (2004) 600–612.

[10] C. Torrence, G. P. Compo, A practical guide to wavelet analysis, Bulletin of the American Meteorological Society 79 (1998) 61–78.

[11] A. V. G. Cavalieri, P. Jordan, T. Colonius, Y. Gervais, Axisymmetric superdirectivity in subsonic jets, Journal of Fluid Mechanics 704 (2012) 388–420.