ANALYSING NON-LINEARITY USING BISPECTRAL

Soroush Jamali

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano
Piazza Leonardo Da Vinci 32, 20122 Milano, Italy
soroush.jamali@mail.polimi.it

ABSTRACT

Bicoherence analysis is a well established method for identifying the quadratic nonlinearity of stationary processes. However, it is often applied without checking the basic assumptions of stationarity and convergence. The classic bicoherence, unfortunately, tends to give false positives – high bicoherence values without actual nonlinear coupling of different frequency components – for signals exhibiting rapidly changing amplitudes and limited length. The effect of false positive values can lead to misinterpretation of results, therefore a more prudent analysis is necessary in such cases. This paper analyses the properties of bispectrum and bicoherence in detail, generalizing these quantities to nonstationary processes. in addition. From speech to images, and videos, advances in machine learning have led to dramatic improvements in the details and realism of so-called AI-synthesized content. While there are many applications to this approach, this type of content can also create convincing and deceptive fakes. We seek to develop forensic techniques to distinguish sounds produced by the human speech system from a synthesized voice. We observed from speech synthesizer based on neural network introduce specific and unusual spectral correlations not typically found in human speech. Although not necessarily audible, these correlations can be measured using tools from bispectral analysis and used to distinguish human from synthesized speech.

Index Terms— Bispectral, Non-linearity, DeepFake

1. INTRODUCTION

The need to identify the existence (or lack) of second order nonlinear interactions in dynamical systems is a widespread problem in numerous disciplines. Examples can be found in population dynamics [1], geophysics [2], oceanology [3] or in plasma physics [4]. The basic idea is to investigate the threewave coupling in a signal, or more accurately, to characterize the fraction of the signal-energy of two waves that is quadratically phase coupled to a third wave at the sum frequency. The data analysis methods for stationary (on the scale of a single investigated time slice [5]) processes are well developed [6], [7]. However, more care is necessary when working with nonstationary processes and transient signals. There have been a number of previous approaches to extend the validity of the bicoherence method to rapidly changing systems. However, all of these required quasi-stationary behaviour on some short timescale. Good examples are the wavelet-bicoherence [8], the short-time Fourier transform based bicoherence [9], [10], or the method of appying adaptive windowing to select an optimal window length with maximized local coherence [11]. All of these, however, require quasi-stationarity on a short time scale, which is not always the case in various real-life systems. This paper introduces an approach to characterize nonlinearity of nonstationary systems based on the "classical" bicoherence technique [6], [7], complemented by statistical analysis. The goal is to estimate the significance level of the measured bicoherence for each frequency-frequency point, in order to filter likely false positives (type I error), i.e. values of high bicoherence despite the lack of actual phase coupling

Section II describes bicoherence calculation for stationary processes based on Kim et.al. [6], and exposes the problem with nonstationary systems. In section III detectio A validation using simple model systems follows

in Section IV. Finally, implications for lower and higher order coherence calculations are discussed in Section V.

2. BISPECTRAL ANALYSIS

For purposes of illustration consider a signal composed of a sum of two sinusoids:

$$x(n) = sin(\omega_1 + \phi_1) + sin(\omega_2 + \phi_2) \tag{1}$$

Consider now passing this signal through a simple non linearity:

$$y\left(n\right) = x^2\left(n\right) + x(n) \tag{2}$$

Expanding the polynomial and rewriting in terms of the harmonics using basic trigonometric identities gives:

$$\begin{aligned} y\left(n\right) = &\frac{1}{2}(1 + \sin\left(2\omega_1 n + 2\emptyset_1 - \frac{\pi}{2}\right))+ \\ &\frac{1}{2}(1 + \sin\left(2\omega_2 n + 2\emptyset_2 - \frac{\pi}{2}\right))+ \\ &2\sin\left((\omega_1 + \omega_2) + (\emptyset_1 + \emptyset_2)\right)+ \\ &2\sin\left((\omega_1 - \omega_2) + (\emptyset_1 - \emptyset_2)\right)+ \\ &\sin\left((\omega_1 n + \emptyset_1) + (\omega_2 n + \emptyset_2)\right) \end{aligned} \tag{3}$$

Notice that this simple non-linearity introduces new harmonics with frequencies and phases that are correlated (Figure 2). In addition, these correlations simply could not have been introduced through a linear transform. Our goal now is to try to detect these "un-natural" higher order correlations as a means of detecting the presence of a non-linearity.

In this section, we describe the basic statistical tools used to analyze audio recordings. The bispectrum of a signal represents higher-order correlations in the Fourier domain. An audio signal y(k) is first decomposed according to the Fourier transform:

$$y\left(\omega\right) = \sum_{k=-\infty}^{\infty} y\left(k\right) e^{-ik\omega} \tag{4}$$

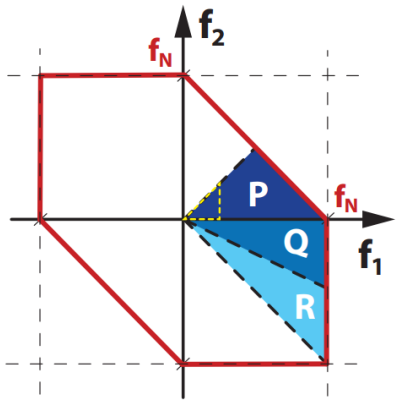

Figure 1: Bicoherence is defined on the f1; f2 plane. In case of finite signals it is bounded by the Nyquist-frequency, marked with red solid lines. Symmetries that arise from the definition reduce the plotting area to the ones marked with P, Q, R. Each of these contain equivalent information, therefore in the this paper we only plot the area marked with P

with $\omega \in [-\pi; \pi]$. It is common practice to use the power spectrum of the signal $P(\omega)$ to detect the presence of second-order correlations, which is defined as:

$$P(\omega) = Y(\omega) Y^*(\omega) \tag{5}$$

where $\star$ denotes complex conjugate. The power spectrum is blind to higher-order correlations, which are of primary interest to us. These correlations can, however, be detected by turning to higher-order spectral analysis [9]. The bispectrum, for example, is used to detect the presence of third-order correlations:

$$B(\omega_1, \omega_2) = Y(\omega_1) Y(\omega_2) Y^*(\omega_1 + \omega_2) \tag{6}$$

Unlike the power spectrum, the bispectral response reveals correlations between the triple of harmonics$[\omega_1, \omega_1, \omega_1+\omega_1]$,$[\omega_2, \omega_2, \omega_2+\omega_2]$,$[\omega_1, \omega_2, \omega_1+\omega_2]$ and $[\omega_1, -\omega_2, \omega_1-\omega_2]$. Note that, unlike the power spectrum, the bispectrum in (6) is a complex-valued quantity. From an interpretive stance it will be convenient to express the complex bispectrum with respect to its magnitude:

$$|B(\omega_1, \omega_2)| = |Y(\omega_1)|.|Y(\omega_2)|.|Y^*(\omega_1 + \omega_2)| \tag{7}$$

and phase:

$$\angle B(\omega_1, \omega_2) = \angle Y(\omega_1) + \angle (Y(\omega_2) - \angle Y^*(\omega_1 + \omega_2) \tag{8}$$

With the definition in eq. (7), $b(\omega_1; \omega_2) \in [0; 1]$, and can be interpreted similarly to classic coherence [6]. If (f1; f2) components are coupled (the conditions (1)-(2) are met), $b(\omega_1; \omega_2) \to 1$, else it will tend to zero as the number of averaging tends to infinity. In the case of signals with finite length the frequency domain of the bicoherence is bounded with the Nyquist frequency [13], which is shown as the red hexagonal frame in figure 2. Due to the symmetries of the bispectrum (detailed in appendix A) in the case of real signals it is enough to plot an even smaller region, which contains all the independent information (marked with "P" in Fig. 1). For further analysis we will only present results in this area.

Real-life digital signals sampled at a given f_s sampling frequency have finite length (with M discrete elements). The calculation of the averages in (6)(7), based on the definition in (5) requires the signal to be divided it into N parts with equal $T = \frac{\frac{M}{N}}{f_s}$ length. The T length of a single part and the f_s sampling frequency determine the f_N Nyquist frequency and Δf frequency resolution of the calculations:

$$f_N = \frac{f_s}{2} \tag{9}$$

$$\Delta f = f_s \frac{N}{M} \tag{10}$$

It means that increasing the N number of averages can only be achieved at the cost of degrading the frequency resolution, making a trade-off necessary for any given phenomenon under investigation. From a practical point of view, first we should consider the minimal Δf frequency resolution required for the investigated phenomena, which will then set the maximum N number of averages. Splitting the signal into parts (practically windowing with a boxcar window function) leads to sidelobes and other undesired features in the Discrete Fourier Transform. For the suppression of these effects it is beneficial to apply window functions to the smaller parts. Throughout the paper we utilize the Hann window with 50 % overlap [14], where the window function is defined as

$$\omega_{Hann}(t) = sin^2(\frac{\pi}{T}t) \tag{11}$$

extending for a single period between two zero function values.

Let us now introduce the $X^{(i)}$ Discreet Fourier Transformed (DFT) complex vector of the signal's i_{th} part, which has $2n = M = N$ elements, and its frequency resolution is defined by (8). A single element of this vector will be marked as $X_k^{(i)}$, where $k \in (1, ...n)$ and the k_{th} element corresponds to the $k \cdot \Delta f$ frequency component. (Note that, as far as we are analysing real signals, we consider only the positive frequencies of the transform).We define the $S^{(i)}$ cyclically shifted frequency matrix of the $X_j^{(i)*}$ complex conjugate vector elements :

$$S_{kl}^{(i)} = \begin{cases} X_{k+l}^{(i)*}, \; if \; k \leq land \, k + l \leq \; n \\ 0, otherwise \end{cases} \tag{12}$$

Also from an interpretive stance it is helpful to work with the normalized bispectrum [2], the bicoherence:

$$B_c(\omega_1, \omega_2) = \frac{Y(\omega_1) Y(\omega_2) Y^*(\omega_1 + \omega_2)}{\sqrt{|Y(\omega_1) \; Y(\omega_2)|^2 |Y(\omega_1 + \omega_2)|^2}} \tag{13}$$

This normalized bispectrum yields magnitudes in the range [0; 1]. Throughout, we compute the bicoherence with a segment length of N = 64 with an overlap of 32 samples. In the absence of noise, the bicoherence can be estimated from a single realization as in 6. However in the presence of noise some form of averaging is required to ensure stable estimates. A common form of averaging is to divide the signal into multiple segments. For example the signal y(n) with $n \in [1, N]$ can be divided into K segments of length $M = \frac{N}{k}$, or K overlapping segments with$M > \frac{N}{k}$. The bicoherence is then estimated from the average of each segment's bicoherence spectrum:

$$\hat{B}_c(\omega_1, \omega_2) = \frac{\frac{1}{k}\sum_k Y(\omega_1) Y(\omega_2) Y^*(\omega_1 + \omega_2)}{\sqrt{\frac{1}{k}\sum_k |Y(\omega_1) \; Y(\omega_2)|^2 \frac{1}{k}\sum_k |Y(\omega_1 + \omega_2)|^2}} \tag{14}$$

The bicoherence includes a normalization factor that removes the magnitude dependence. There is some inconsistency with the definition of the bicoherence normalization constant. Some of the definitions that have been used are

$$\hat{B}\left(\omega_1,\omega_2\right)=\frac{\left|\sum_k Y\left(\omega_1\right)Y\left(\omega_2\right)Y^*\left(\omega_1+\omega_2\right)\right|}{\sqrt{\sum_k\left|Y\left(\omega_1\right)\right|^2\left|Y\left(\omega_2\right)\right|^2\left|Y^*\left(\omega_1+\omega_2\right)\right|^2}} \tag{15}$$

which was provided in Sigl and Chamoun 1994, but does not appear to be correctly normalized. Alternatively, plasma physics typically uses

$$\hat{B}_c\left(\omega_1,\omega_2\right)=\frac{\left|Y\left(\omega_1\right)Y\left(\omega_2\right)Y^*\left(\omega_1+\omega_2\right)\right|}{\langle\left|Y\left(\omega_1\right)Y\left(\omega_2\right)\right|\rangle\langle\left|Y^*\left(\omega_1+\omega_2\right)\right|\rangle} \tag{16}$$

where the angle brackets denote averaging. Note that this is the same as using a sum, because k is the same in the numerator and the denominator. This definition is directly from Nagashima 2006, and is also referred to in He 2009 and Maccarone 2005. Finally, one of the most intuitive definitions comes from Hagihira 2001 and Hayashi 2007, which is

$$\hat{B}_c\left(\omega_1,\omega_2\right)=\frac{\left|\sum_k Y\left(\omega_1\right)Y\left(\omega_2\right)Y^*\left(\omega_1+\omega_2\right)\right|}{\sum_k\left|Y\left(\omega_1\right)Y\left(\omega_2\right)Y^*\left(\omega_1+\omega_2\right)\right|} \tag{17}$$

The numerator contains the magnitude of the bispectrum summed over all of the time series segments. This quantity is large if there is phase coupling, and approaches 0 in the limit of random phases. The denominator, which normalizes the bispectrum, is given by calculating the bispectrum after setting all of the phases to 0. This corresponds to the case where there is perfect phase coupling, because all of the samples have zero phase. Therefore, the bicoherence has a value between 0 (random phases) and 1 (total phase coupling).

3. DETECTION NON-LINEARITY

In the previous section we saw that a simple squaring non-linearity applied to the sum of two sinusoids introduces new harmonics with distinct higher-order correlations between the individual harmonics, 3. We also saw how the bispectrum/bicoherence can be used to detect these correlations, 6. If such higherorder correlations are weak in "natural" signals then their presence can be used as an indication of tampering thus casting the authenticity of the signal into a suspicious light. In the next section we provide some empirical evidence that for human speech, these higher-order correlations are in fact relatively weak. But first we need to show why we might expect increased activity in the bicoherence when an arbitrary nonlinearity is applied to an arbitrary signal. To this end consider first an arbitrary function $f(\cdot)$:

$$y(n)=f(x(n)) \tag{18}$$

For notational convenience this system is expressed in an equivalent vector form as $\vec{y}=f(\vec{x})$. The vector valued function$f(\cdot)$: can be expressed in terms of scalar valued functions

$$f(\vec{x})=(f_1((\vec{x}))+f_2((\vec{x}))...f_N((\vec{x}))) \tag{19}$$

Each of these scalar valued functions $f_i(\cdot)$ can subsequently be expressed in terms of its Taylor series expansion about a point $\vec{P}$:

$$f_i(\vec{x})=f_i(\vec{P})+\sum_j\frac{\partial f_i}{\partial x_i}x_i\ +\ \frac{1}{2}\sum_{j,k}\frac{\partial^2 f_i}{\partial x_j\partial x_k}x_ix_k+\ ... \tag{20}$$

For simplicity this expansion is truncated after the secondorder term and rewritten in vector/matrix form as :

$$f_i(\vec{x})\approx\ a\ +\ \vec{b}\vec{x}\ +\frac{1}{2}\vec{x}C\vec{x} \tag{21}$$

with the scalar $a=f_i(\vec{p})$, the vector $\vec{b}_j=\frac{\partial f_i}{\partial x_i}|\vec{p}_j$ and the matrix $[C]_{jk}=\frac{\partial^2 f_i}{\partial x_i}\partial x_k\,|\vec{p}_j$, with $i,j,k\in[1,N]$.The point of all of this is simply that the application of an arbitrary non-linearity $f(\vec{x})$ results in an output that contains a sum of a linear term $\vec{b}\vec{x}$. and a quadratic term $\vec{x}C\vec{x}$. Note the similarity with the original example in the previous section, 2. Consider now the input signal expressed in terms of its Fourier series:

$$x(n)\ =\ \sum_{k=0}^{\infty}a_k\sin(kn\ +\ \emptyset_k) \tag{22}$$

The situation is now very similar to that of the previous section, 1 and 2. That is, the application of an arbitrary non-linearity $f(\cdot)$ to an arbitrary signal $x(n)$ will result in higher-order correlations between various harmonics, and hence an increased bicoherence activity. fact, even more specific predictions can be made by looking closely at how a non-linearity affects the bicoherence magnitude and phase. Recall that a non-linearity introduces new harmonics that are correlated with the original harmonics, Figure 2. For example, given a pair of harmonics and ω_1 and ω_2 , a non-linearity produces a new harmonic $\omega_1+\omega_2$whose amplitude is correlated to ω_1 and ω_2 . To the extent that such correlations do not occur naturally, these correlations will result in a larger response in the bicoherence magnitude, 7. In addition, if the initial harmonics have phases $\emptyset_1$ and $\emptyset_2$, then the phase of the newly introduced harmonic is$\emptyset_1+\emptyset_2$. It is then easy to see that the bicoherence phase for this pair of harmonics will be 0, 8. Again, to the extent that these higher-order correlations do not occur naturally, the correlations will result in a bias of the bicoherence phase towards 0. A phase bias towards $\frac{\pi}{2}$ may also occur due to correlations of any harmonic ω_1 with itself. That is, the bicoherence phase $\angle B(\omega_1+\omega_2)=\emptyset_1+\emptyset_2-(2\emptyset_1-\frac{\pi}{2})=\frac{\pi}{2}$ We predict that a non-linearity will manifest itself in the bicoherence in two ways: for certain harmonics an increase in the magnitude, and for those harmonics a phase bias towards 0 and/or $\frac{\pi}{2}$. In the following section we will show how these principles can be used in the detection of digital forgeries.

4. RESULTS

We first show how the bicoherence can detect the presence of global and local non-linearities of the form described in Section 2. We then show how these ideas generalize to a broader class of nonlinearities that might be involved in the creation of digital forgeries Shown in Figure 1 is the bicoherent magnitude and phase for three different human speakers. Shown in the second to the sixth rows are the bicoherent magnitude and phase for five different synthesized voices, as described in Section 2.1. Each bicoherent magnitude and phase panel are displayed on the same intensity scale. There are also strong differences in the bicoherent phases across all

synthesized speech. The bicohernece, 7, is computed for each human and synthesized speeches, from which the bicoherence magnitude and phase are computed. These two-dimensional quantities are normalized such that the magnitude and phase for each frequency ω_1 are normalized into the range $[0, 1]$ by subtracting the minimum value and dividing by the resulting maximum value. The normalized magnitude and phase are each characterized using the first four statistical moments. Let the random variable M and P denote the underlying distribution for the bicoherence magnitude and phase. The first four statistical moments are given by:

where $E_X[\cdot]$ is the expected-value operator with regards to random variable X. From the magnitude $X = M$ and phase $X = P$, these four moments are estimated by replacing the We collected a data set consisting of 5 human and synthesized speech recordings. The human speech are obtained from 5 people (three male and two female). These recordings were extracted from various high-quality podcasts. Each recording averaged 3:5 seconds in length. We also include samples generated using the nonlinear transformation function in order to control the level of non-linearity. non-linearity made in the audio signal with various type of transformation including following equations:

$$y(n) = \frac{x(n)}{\sqrt{1 + x^2(n)}} \tag{23}$$

$$y(n) = x^2(n) + x(n) \tag{24}$$

$$y(n) = tanh(x(n)) = \frac{e^{x(n)} - e^{-x(n)}}{e^{x(n)} + e^{-x(n)}} \tag{25}$$

Throughout these examples the input is a 1-D fractal signal $x(n)$ with 2048 samples. In this first set of experiments a signal is subjected to a global non-linearity of the form:

$$y(n) = \alpha x^2(n) + x(n) \tag{26}$$

with $\alpha \in [0, 1]$. The bicoherence is computed by dividing the signal into overlapping segments of length 32 with an overlap of 16. A 1024-point windowed DFT, $Y_k(\omega)$, is estimated for each segment from which the bicoherence is estimated $\hat{B}_c = (\omega_1, \omega_2)$, (14). Shown in the first column of Figure 3 are several signals with varying amounts of non-linearity, i.e., increasing values of in 26. The signals are all plotted on the same scale of $[0, 1]$. Shown within each row is a small portion of the input signal $y(n)$, the normalized power spectrum, the bicoherence magnitude $|\hat{B}_c = (\omega_1, \omega_2)|$, and a histogram of the bicoherence phase $\angle\hat{B}_c = (\omega_1, \omega_2)$ for those frequencies with $|\hat{B}_c = (\omega_1, \omega_2)| > 0.2$. With increasing amounts of non-linearity there is an increase in the magnitude and phase bias towards zero. At the same time the normalized power spectrum remains virtually unchanged - it is blind to the higher-order correlations introduced by the non-linearity. The increase in bicoherence activity can be quantified by measuring the mean magnitude and the phase bias. To measure the phase bias away from uniform we compute the variance across the bin counts from a discrete histogram of the phases (i.e., deviations from a uniform distribution will have a large variance). To show the general robustness of this measure the bicoherence is measured for one hundred independent fractal signals with varying levels of non-linearity. Shown in Figure 4 are the mean magnitude and phase bias plotted as a function of increasing amounts of non-linearity. Each data point corresponds to the average response over the one hundred trials, the error bars correspond to the standard error. With increasing amounts of non-linearity the

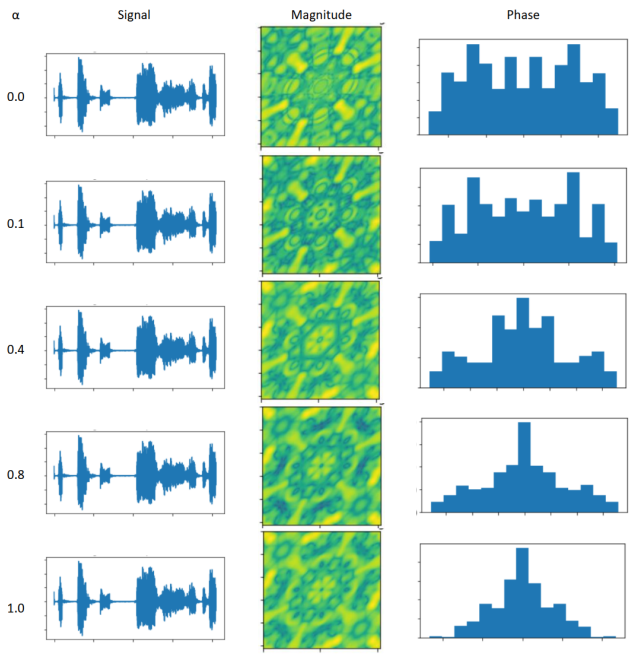

Figure 2: Detecting non-linearities. Shown from left to right is a portion of the input signal $y(n) = \alpha x^2(n) + x(n)$, its normalized power spectrum $P(\omega)$ with $\omega \in [-\pi, \pi]$, its bicoherence $\hat{B}_c(\omega_1, \omega_2)$ with $\omega_1, \omega_2 \in [-\pi, \pi]$, and the bicoherence phase histogram plotted from $[-\pi, \pi]$. As the non-linearity increases there is no change in the power spectrum, but a significant increase in the overall magnitude and phase bias (see also Figure 4)

mean magnitude and phase bias increase by 35 % and 165%, respectively.

for an other experiment we studied the effect of different non-linearity transformation 27 and 29 and we repeat the experiment as the previous section. hence we want examine the performance of bicoherence we limited the equation with only one degree of non-linearity. in the figure 4 you can see the effect of non-linearity in increasing both Bicoherence variance and standard deviation. As it illustrated in the figure 4 there are some points that are coordinated with others. this phenomena is mentioned in other papers as false positive. The effect of false positive values can lead to misinterpretation of results, therefore a more prudent analysis is necessary in such cases.

5. REFERENCES

[1] H. Ombao and S. Van Bellegem. Evolutionary coherence of nonstationary signals. IEEE Transactions on Signal Processing, June 2008 56(6):2259–2266,

[2] *L. Dyrud, B. Krane, M. Oppenheim, H. L. Pecseli, K. Schlegel, ´ J. Trulsen, and A. W. Wernik. Low-frequency electrostatic waves in the ionospheric e-region: a comparison of rocket observations and numerical simulations. Annales Geophysicae, 24(11):2959–2979, 2006.*

[3] *X. Xie, X. Shang, and G. Chen. Nonlinear interactions among internal tidal waves in the northeastern South China Sea. Chi-*

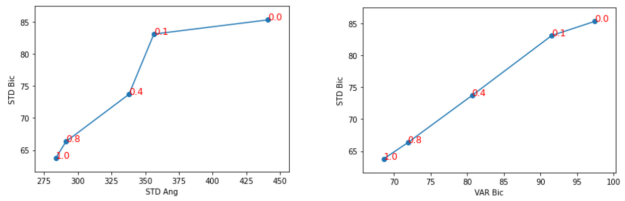

Figure 3: Bicoherence averaged over one-hundred independent signals of the form $\alpha x^2\,(n) + x(n)$ with $\alpha \in [0.0, 1.0]$. The error bars correspond to the standard error. As the degree of non-linearity increases both the magnitude and phase bias increase (see also Figure 3).

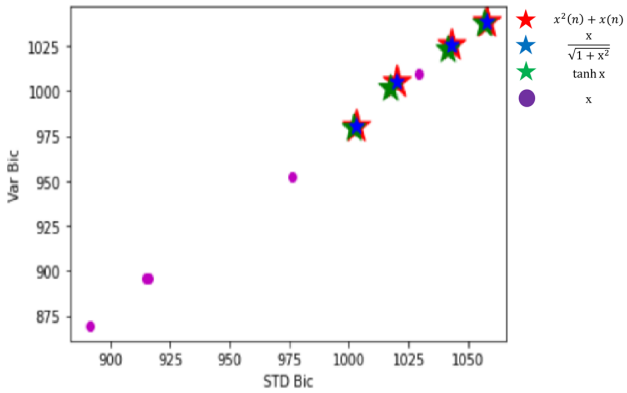

Figure 4: the value above are scaled in dB. the difference between the signals went through a non-linear transformation is obvious. values represented as "⋆" are signal with increases in harmonicas. ○ are representing the original sound signals

nese Journal of Oceanology and Limnology, 28:996–1001, September 2010.

[4] *P. Manz, P. Lauber, V. E. Nikolaeva, T. Happel, F. Ryter, G. Birkenmeier, A. Bogomolov, G. D. Conway, M. E. Manso, M. Maraschek, D. Prisiazhniuk, and E. Viezzer. Geodesic oscillations and the weakly coherent mode in the i-mode of asdex upgrade. Nuclear Fusion, 55(8):083004, 2015.*

[5] *Julius S. Bendat and Allan G. Piersol. Random data: analysis and measurement procedures, volume 729. John Wiley and Sons, 2011*

[6] *Y. C. Kim and E. J. Powers. Digital bispectral analysis and its applications to nonlinear wave interactions. IEEE Transactions on Plasma Science, 7(2):120–131, June 1979.*

[7] *Y. C. Kim, J. M. Beall, E. J. Powers, and R. W. Miksad. Bispectrum and nonlinear wave coupling. The Physics of Fluids, 23(2):258–263, 1980*

[8] *B. Ph. van Milligen, C. Hidalgo, and E. Sanchez. Nonlinear phenomena ´ and intermittency in plasma turbulence. Phys. Rev. Lett., 74:395–398, Jan 1995*

[9] *Vinod Chandran. Time-varying bispectral analysis of visually evoked multi-channel eeg. EURASIP Journal on Advances in Signal Processing, 2012(1):140, 2012.*

[10] *G. I. Pokol, L. Horvath, N. Lazanyi, G. Papp, G. Por, V. Igochine, et al. Continuous linear time-frequency transforms in the analysis of fusion plasma transients. In Proc. of the 40th EPS Plasma Physics Conf., 2013.*