COHERENSI: A New Full-Reference IQA Index Using Error Spectrum Chaos

Tamir Hegazy and Ghassan AlRegib
Center for Signal and Information Processing (CSIP)
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA , USA
tamir.hegazy@ece.gatech.edu, alregib@gatech.edu

***Abstract*—We introduce a new family of quality indices that are based on the reported findings that biological visual systems exhibit interference in spatial harmonic distortion, and that local phase distortion is more perceived than global phase distortion. This paper specifically addresses a full-reference version of the proposed indices. The proposed quality index has three components: harmonic chaos analysis, phase chaos analysis, and pooling. The first two parts produce two maps that are weighted and summed to produce a single map. Then, we use the mean of the logarithmically scaled combined map as the quality index. We further propose a multi-scale variation of the index. Experimental evaluations were conducted on more than 1,300 images from the `TID2013` image database. We specifically test the quality index on three distortion categories: `Simple`, `Actual`, and `Noise`. Then, we cross-validate the results with the `LIVE` database. The results show that the proposed index outperforms all known indices based on correlation with subjective evaluation.**

***Index Terms*—Full-Reference, Image Quality, Perception**

I. INTRODUCTION

Image quality assessment (IQA) is becoming more important in the light of recent advancements in imaging technologies, the considerable increase in Internet visual data size, and the continuation of the network to serve as the bottleneck to access high quality visual data. Image compression, communication errors, and processing, are among several sources of image distortion. While subjective IQA is the natural way to perform perceptual IQA, it is time-consuming and expensive because of human involvement and the test setup time/cost, let alone the need for tens or hundreds of subjects. Automating IQA entails the creation of objective indices that are consistent with subjective evaluation. The available objective indices can generally be categorized as full-reference (FR), reduced-reference (RR), or no-reference (NR) IQA. We propose a new and novel family of image quality indices based on the harmonic content and the phase disorder in the error between the reference error-free image and the distorted image. In this paper, we focus on the FR version of the proposed index.

The human visual system (HVS) has 3 major properties: (*i*) frequency sensitivity, (*ii*) luminance sensitivity, and (*iii*) several masking effects [1] (contrast sensitivity, frequency masking, etc.). Frequency sensitivity determines the HVS sensitivity to various spatial frequencies. Luminance sensitivity measures the just-noticeable level of noise on a constant background. Masking refers to the effect of decreasing visibility of one signal in the presence of another.

In our methodology, we argue that the behaviors of harmonics and phase produced by distortion are sufficient for analyzing distortion perception. Furthermore, the error between the reference error-free image and the distorted image contains all information needed to measure the 3 properties above. We show that chaotic behaviors of error harmonics and phase lead to more distortion perception. As a result, the proposed quality index analyzes the harmonics and phase content in the error signal and tries to collectively quantify the 3 HVS properties. For the rest of the paper, we refer to the proposed quality index as COHERENSI, which is derived from Chaos of Harmonics/phase in Error as a REference-based Novel Similarity Index. Experimental evaluations reveal that COHERENSI is more correlated with subjective scores than the best known counterparts for a wide range of distortion types. In this paper, we consider the `Simple`, `Actual`, and `Noise` categories of the `TID2013` image database [2]. The 3 categories comprise 11 of the most real-life distortion types, including additive Gaussian noise, color map noise, spatially correlated noise, quantization noise, denoising, Gaussian blur, and JPEG/J2K compression. Then, we cross-validate using the `LIVE` database Release 2 [3].

We introduce COHERENSI in Sec. III after we provide a short survey about related work in Sec. II. Then, we evaluate COHERENSI in Sec. IV, and conclude the paper in Sec. V.

II. RELATED WORK

Wang *et al.* [4] showed that human perception is more consistent with structural similarity as opposed to the classical Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). Structural similarity indices such as SSIM [4] and MS-SSIM [5] were shown to be more correlated with human error perception. Ponomarenko *et al.* developed a series of indices [6], [7], [8] that outperform MS-SSIM (in terms of consistency with subjective evaluation) for certain distortion types. Such indices are of the variety of PSNR but they take into account contrast sensitivity. Zhang *et al.* developed two feature-based indices, namely FSIM and its color version FSIMc [9] based on phase congruency and gradient magnitude. FSIMc was shown to collectively outperform all known indices

978-1-4799-7088-9/14/$31.00 ©2014 IEEE

for the various distortion types in the TID2013 database [2]. Other methods based on the frequency domain have also been used to analyze HVS properties, e.g., [10], [11].

In contrast to all the above methods, our proposed index runs a harmonic analysis of the error signal. While the index captures the luminance sensitivity, it also explores masking at the harmonic level. Tan and Ghanbari [12] developed a FR index based on harmonic analysis for the blockiness artifacts due to JPEG compression and a RR index based on harmonic gain and loss. The work presented in this paper differs from [12] in three aspects. First, it uses a more efficient harmonic analysis method. Second, it analyzes the phase spectrum of the error and one of the findings is that error phase carries information that cannot be detected by other indices. Third, the proposed index is designed to work with the general image distortion, without being limited to blockiness/blurriness.

III. The Proposed Similarity Index - Coherensi

Fig. 1 shows a block diagram of the single-scale Coherensi algorithm. It is composed of three major components: (*i*) harmonic chaos analysis, (*ii*) phase chaos analysis, and (*iii*) pooling. In the following, we explain each component. Then, we propose a multi-scale variation of Coherensi.

A. Harmonic Chaos

Biological visual systems, including the HVS, were shown to exhibit spatial frequency masking effects [13]. For example, it has been shown that striate cortex cells tuned to a certain spatial fundamental frequency ($1F$) respond to harmonics ($\sum nF$) if they are presented simultaneously with $1F$. The same cells do not respond to these harmonics in the absence of the fundamental frequency [13]. More spatial harmonic masking effects have also been reported [14]. Remarkably, harmonic masking effects are not limited to the *spatial* harmonics in the *visual* system. Similar masking effects have been observed for *temporal* harmonics in the *auditory* cortex [15].

Based on the above evidence, we reason that the quality perceived in the reference (error-free) image is based on a subtle spatial harmonic balance within the image. Upon observing the distorted image, subjects who evaluate the image quality will favor "well-behaved" (i.e., more uniform, or smooth) harmonic content in the error signal as it will interfere less with the harmonic balance achieved in the reference image. On the other hand, chaotic harmonic content in the error signal could highly interfere with the harmonic masking effect that is wired in the HVS, leading to poor subjective scores.

As shown in Fig. 1, Coherensi characterizes the behavior of harmonics by computing a harmonic behavior map $\mathbf{H}$, which we define as the 2D DFT magnitude of the 2D DFT magnitude of the 2D error gradient. Mathematically,

$$\mathbf{H}[\alpha,\beta] = \left|\mathscr{F}\left\{\left|\mathscr{F}\left\{\nabla\left(\boldsymbol{\Delta}[i,j]\right)\right\}\right|\right\}\right|, \qquad (1)$$

where $\mathscr{F}$ denotes the 2D DFT, $\boldsymbol{\Delta}[i,j]$ denotes the luminance error, which is the difference between the luminance of the distorted image and that of the reference image, and ∇ is the 2D gradient operator. In the following few paragraphs,

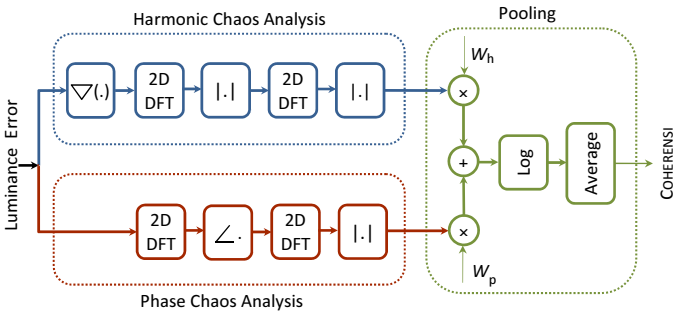
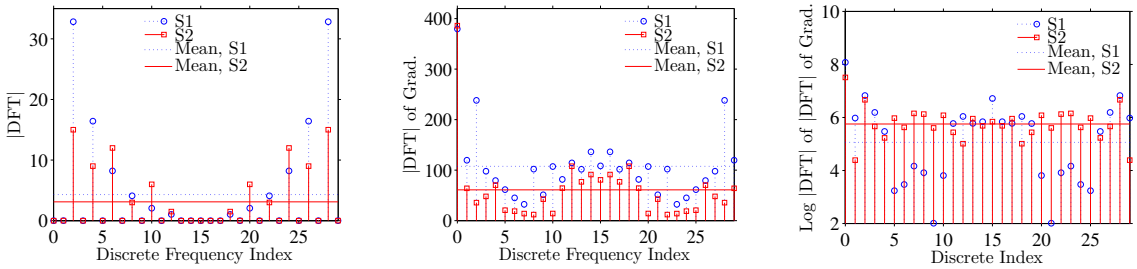


Fig. 1. Block diagram of the proposed similarity index, Coherensi. Thick lines represent array signals, while thin lines represent scalar signals.

we show how $\mathbf{H}$ relates to the HVS properties in terms of: (*i*) luminance sensitivity, which is captured by the energy in the error signal, (*ii*) frequency sensitivity and harmonic masking effects, which is captured by harmonic behavior, and (*iii*) local contrast sensitivity by using the gradient. Notably, this is somewhat similar to the cepstrum [16] of the error gradient without squaring or taking the logarithm for the inner DFT magnitude. Cepstrum variations are extensively used in sound/speech signal processing. We envision that the proposed Coherensi index will play an important role in image processing as cepstrum variations have played in the sound/speech processing . Thus, as we use the 2D DFT magnitude of 2D DFT magnitude as a transform to evaluate the *2D, spatial* harmonic content of the error, it is noteworthy that the 1D version of the proposed transform has been used to determine the *1D, temporal* harmonic content in sound signals for pitch recognition, which is another perceptual problem [17].

Fig. 2 illustrates how the proposed transform captures the harmonic behavior of simple 1D signals. Fig. 2a shows the two-sided DFT of two 1D discrete synthetic signals S1, S2 with a discrete fundamental frequency $f = 2$ and different harmonic behaviors. For illustration pusposes, the spectrum has ben synthesized as follows. While harmonic amplitudes in S1 decay smoothly in an exponential fashion, those of S2 decay chaotically in a "zigzag" fashion as shown in Fig. 2a. Further, the energy content of S1 is chosen to be higher than that of S2. Taking (a second) DFT of the signals in Fig. 2a should capture the chaos in the harmonic content because the zigzag behavior would constitute more and/or larger frequency components in the second DFT. However, we note that the second DFT coefficients will increase with the original signal energy as well as the harmonic chaos. Because of the typical decay of harmonics, we need to emphasize the high-frequency harmonics in order to emphasize the harmonic chaos. Without emphasizing high-frequency harmonics, the harmonic chaos contribution to the second DFT would be limited compared to the signal energy contribution. Thus, we emphasize the harmonic chaos contribution by computing the first DFT for the gradient instead of the signal itself. Fig. 2b shows the DFT magnitude of the gradient. Fig. 2c shows the second-DFT magnitude of the signal in Fig. 2b after taking the logarithm (which will be computed in the pooling step). Note that the



(a) Signal harmonics. (b) Gradient harmonics. (c) DFT Mag. of gradient harmonics.

Fig. 2. Illustration of the ability of the proposed transform to differentiate between two signals with different harmonic behaviors.

mean value that corresponds to S2 in Fig. 2c is higher than that for S1, indicating that the harmonic content in S2 is more chaotic, which is consistent with the zigzag decay of S2 harmonics. Therefore, Eq.(1) successfully characterizes the harmonic chaos in both signals. Furthermore, we observe that unlike the mean values in Fig. 2c, those in Fig. 2a and Fig. 2b are not consistent with the chaotic behavior of the signal, as they are higher for S1, whose harmonics are less chaotic. Finally, we note that although the energy of S1 is higher than that of S2, the mean values in Fig. 2c are still sensitive to chaos, and not just the energy content, which is the goal of the proposed index.

B. Phase Spectrum Chaos

It is widely believed that the phase spectrum highly contributes to perception. Further, it believed that local phase distortion is more perceived than global phase distortion [18]. Therefore, we measure the degree of local chaos in the phase spectrum of the error as a measure of perceived distortion. We employ an analogous approach to that of harmonic chaos analysis. We take the magnitude of 2D DFT of the phase spectrum, which on average captures the chaos in the phase spectrum. Mathematically, the phase chaos map at a given scale is computed as follows:

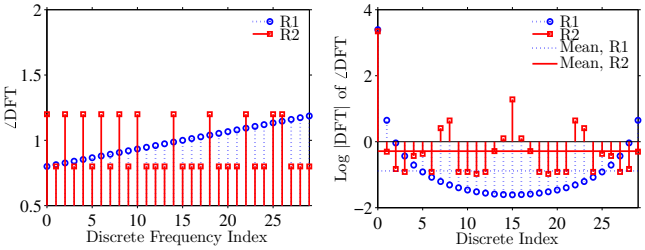
$$\mathbf{P}[\alpha,\beta] = \left|\mathscr{F}\left\{\angle\mathscr{F}\left\{\mathbf{\Delta}[i,j]\right\}\right\}\right|, \tag{2}$$

where $\angle$ denotes the phase operator.

Again, as a simple illustration, Fig. 3a shows synthesized phase spectra of two 1D signals R1, R2. While the phase spectrum of R1 is smooth, R2 exhibits an oscillating (more chaotic) phase spectrum. Fig. 3b shows the DFT magnitude of the phase spectrum with higher mean value for R2 (the more chaotic) than for R1 (the less chaotic). Therefore, the average DFT magnitude of the phase spectrum can capture the phase chaos. Unlike the harmonic chaos analysis, the gradient here is irrelevant since phase does not show the same rapid decay observed in harmonic amplitudes.

C. Pooling

The harmonic chaos analysis and the phase chaos analysis components yield two 2D maps, which we first combine (Fig. 1) using a weighted sum operator. Then, we logarithmically scale the combined map, and take the mean value.



(a) Different phase spectra. (b) Log |DFT| of phase spectrum.

Fig. 3. Ability of the proposed transform to capture phase disorder.

Formally, COHERENSI is computed as:

$$\text{COHERENSI} = E\left\{\ \log(w_h\mathbf{H} + w_p\mathbf{P} + \varepsilon)\right\}, \tag{3}$$

where E is the expectation operator, and w_h and w_p are weight parameters, and ε is a stability parameter to avoid taking the logarithm of zero.

D. Multi-scale variation

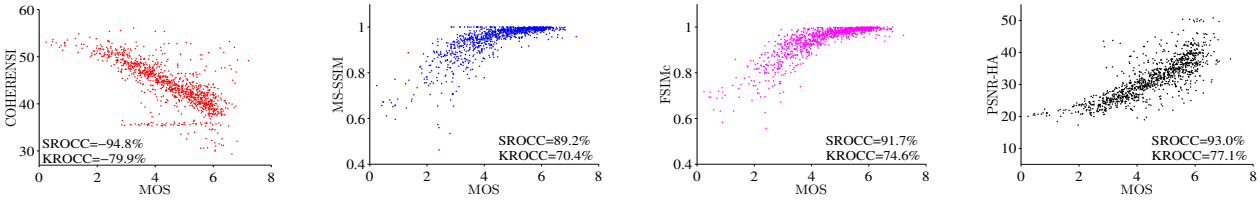
Multi-scale COHERENSI repeats the single-scale computation for several downsampled (by powers of 2) versions of the error map. The multi-scale index is computed as a weighted average of the single-scale ones. The more downsampled versions receive higher weights. The rationale is that errors perceived at more abstract levels are indeed larger than those perceived at less abstract levels. Thus, the error map version downsampled by 2^i will be assigned a weight of $(1+\delta i)$, where δ is a design parameter. Consequently, for the full-scale error map, $i=0$, the weight is 1. After that, the weights increase linearly to exceed 1 as we start to downsample.

The computational complexity of multi-scale COHERENSI is dominated by the DFT and number of scales. Thus, it can be shown that the worst-case complexity is $O(NS\log(S))$, where N is the number of scales and S is the image size.

IV. EXPERIMENTAL EVALUATION

A. Experimental Setup

We evaluate COHERENSI using the `TID2013` database [2] and the `LIVE` database Release 2 [3]. We focus on the `Simple`, `Actual`, and `Noise` categories of `TID2013`, as they comprise 11 of the most real-life encountered distortion types, including additive Gaussian noise, color map noise,



(a) COHERENSI (b) MS-SSIM (c) FSIMc (d) PSNR-HA

Fig. 4. Scatter plots of COHERENSI and other metrics vs. MOS, plotted collectively for the `Noise`, `Actual`, and `Simple` categories of `TID2013`.

spatially correlated noise, quantization noise, Gaussian blur, JPEG/J2K compression, and denoising. The limited space does not allow to introduce the COHERENSI variations that fully address the other 3 distortion categories of the `TID2013` database [2], which are kept for a future publication that combines all distortion types with a detailed analysis. We compare the performance of COHERENSI against the best available indices that performed best for the considered distortion categories: MS-SSIM [5], PSNR-HA [8], FSIMc [9]. The comparison is based on the absolute value of Spearman and Kendall rank order correlation coefficients (SROCC, KROCC) with the mean opinion score (MOS).

Parameter design is based on a 20% random sample of the `TID2013` 1,375 images used in evaluation. We have observed very low sensitivity around values given below. At the end of this section, we evaluate COHERENSI with the `TID2013`-based parameters using the `LIVE` database Release 2 [3]. The parameters and implementation considerations are as follows:

- Number of scales: $N = 4$ (full, 1/2, 1/4, and1/8 scales).
- Scaling fraction: $\delta = 0.18$.
- Logarithm stability parameter: $\varepsilon = 0.1$.
- Gradient: *two consecutive* applications of the 2D Sobel operator to the *absolute* values of the error map.
- Luminance values are normalized between 0 and 1.
- For downsampling, bicubic interpolation is used where the output pixel value is a weighted average of pixel values in the nearest 4-by-4 neighborhood.

B. Results

Table I shows the absolute values of SROCC and KROCC for COHERENSI versus the best-known similarity indices for `TID2013` as given in [2]. SROCC for COHERENSI improves over the best index by 1.9%, 1.2%, and 1.8% for the `Simple`, `Actual`, and `Noise` distortion categories, respectively . Based on KROCC, COHERENSI improves over the best index by 3%, 1.5%, and 3.2%, respectively for the same categories. The overall improvement over the 3 categories is **1.8%** and **2.9%** for SROCC and KROCC, respectively.

Table I includes an entry for "COHERENSI-No Phase", which is COHERENSI without phase chaos analysis (i.e., with $w_p = 0$). Note that even without phase chaos analysis, COHERENSI still outperforms the best-known indices by an overall of 1.3% and 2% for SROCC and KROCC, respectively. In addition, the phase chaos analysis component of COHERENSI yielded more consistency with MOS by contributing extra 0.5% and 0.9% for SROCC and KROCC, respectively.

TABLE I
PERFORMANCE OF COHERENSI ON THE `TID2013` DATABASE.

	Noise	Actual	Simple	Overall
	|SROCC| (%)			
COHERENSI	**94.2**	**95.0**	**97.1**	**94.8**
COHERENSI-No Phase	**93.5**	**94.4**	**96.8**	**94.3**
MS-SSIM	87.3	88.7	90.5	89.2
FSIMc	90.2	91.5	94.7	91.7
PSNR-HA	92.3	93.8	95.3	93.0
	|KROCC| (%)			
COHERENSI	**79.0**	**80.2**	**85.0**	**80.0**
COHERENSI-No Phase	**77.9**	**79.3**	**84.3**	**79.1**
MS-SSIM	67.9	69.7	72.0	70.4
FSIMc	72.2	74.2	79.2	74.6
PSNR-HA	76.0	78.7	81.8	77.1

Fig. 4 shows scatter plots of COHERENSI, MS-SSIM, FSIMc, PSNR-HA versus MOS for the studied categories of `TID2013`. We note that Fig 4a shows an overall linear trend, which indicates that, in its current state, COHERENSI is by nature linearly consistent with MOS without much post-processing, as it is the case with other indices.

For further cross-validation, we fix the parameters above (tuned for a sample of `TID2013`) and we show the performance using the `LIVE` database Release 2 [3]. Table II compares SROCC of COHERENSI and MS-SSIM as reported in [19], and FSIMc as reported in [20]. We skip KROCC for space limitation. COHERENSI outperforms MS-SSIM and FSIMc across more distortion types and the overall SROCC. It is noteworthy, that COHERENSI which is based on luminance only outperforms FSIMc, which takes color into consideration.

V. CONCLUSION

We presented COHERENSI, a new FR IQA index based on error spectrum chaos analysis. Using harmonic chaos and phase chaos analyses, COHERENSI captures the 3 important HVS properties and improves over the state of the art by nearly 3% in rank order correlation for about a dozen of the most encountered distortions. Further, the phase chaos analysis contributed nearly one-third of the reported performance improvement. The performance is cross-database validated.

TABLE II
ABSOLUTE VALUE OF SROCC% FOR `LIVE` DATABASE RELEASE 2.

	Jpeg 2k	Jpeg	WN	G.Blur	FF	Overall
COHERENSI	**97.1**	**95.5**	**98.9**	94.9	92.7	**94.5**
MS-SSIM	94.7	93.0	97.7	96.7	**95.4**	94.2
FSIMc	95.7	91.3	97.1	**97.1**	95.1	92.0

REFERENCES

[1] S. Westen, R. Lagendijk, and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," in *International Conference on Acoustics, Speech, and Signal Processing*, Detroit, MI, USA, May 1995.

[2] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. of the 4th European Workshop on Visual Information Processing EUVIP2013*, Paris, France, June 2013.

[3] H. Sheikh, L. C. Z.Wang, and A. Bovik, "LIVE Image Quality Assessment Database Release 2." [Online]. Available: http://live.ece.utexas.edu/research/quality

[4] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004.

[5] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. of the 37th IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 2, Nov 2003.

[6] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," in *CD-ROM Proc. of the 2nd International Workshop on Video Processing and Quality Metrics*, Scottsdale, AZ, USA, 2006.

[7] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *CD-ROM Proc. of the 3rd International Workshop on Video Processing and Quality Metrics*, Scottdale, AZ, USA, 2007.

[8] N. Ponomarenko, O. Eremeev, L. V., K. Egiazarian, and M. Carli, "Modified image visual quality metrics for contrast change and mean shift accounting," in *Proc. of 11th International Conference on Experience of Designing and Application of CAD Systems in Microelectronics*, Polyana-Svalyava, Ukraine, 2011.

[9] L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, no. 5, 2011.

[10] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "On the performance of human visual system based image quality assessment metric using wavelet domain," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 6806, Mar 2008.

[11] M. Narwaria, W. Lin, I. V. McLoughlin, S. Emmanuel, and L. Chia, "Fourier transform-based scalable image quality measure," *IEEE Transactions on Image Processing*, vol. 21, no. 8, 2012.

[12] M. Qadri, K. Tan, and M. Ghanbari, "The impact of spatial masking in image quality meters," *Global Journal of Computer Science and Technology*, vol. 11, no. 3, 2011.

[13] D. Albrecht and R. De Valois, "Striate cortex responses to periodic patterns with and without the fundamental harmonics," *Journal of Physiology*, vol. 319, 1981.

[14] A. Norcia, S. McKee, Y. Bonneh, and M. Pettet, "Suppression of monocular visual direction under fused binocular stimulation: evoked potential measurements," *Journal of Vision*, vol. 5, no. 1, 2005.

[15] H. Alphei, D. Pschel, and A. Kohlrausch, "Temporal and spectral masking effects of harmonic complex tones," in *Audio Engineering Society Convention 82*, Mar 1987.

[16] B. Bogert, M. Healy, and J. Tukey, "The quefrency alanysis of time series for echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum and Saphe Cracking," in *Proc. of Symposium on Time Series Analysis*, Providence, RI, USA, 1963.

[17] S. Marchand, "An efficient pitch-tracking algorithm using a combination of fourier transforms," in *Proc. of the COST G-6 Conference on Digital Audio Effects*, Limerick, Ireland, Dec 2001.

[18] K. Vilankar, L. Vasu, and D. Chandler, "On the perception of band-limited phase distortion in natural scenes," in *Proc. of Human Vision and Electronic Imaging*, vol. 7865, no. 1, San Francisco, CA, Jan 2011.

[19] A. Moorthy and A. Bovik, "Perceptually significant spatial pooling strategies for image quality assessment," in *SPIE Conference on Human Vision and Electronic Imaging*, San Jose, California, Jan 2009.

[20] H. W. C. Cheng, "Quality assessment for color images with tucker decomposition," in *Proc. of ICIP'12*, Orlando, FL, Sep 2012.