A WAVELET-BASED MODEL FOR FOVEAL DETECTION OF SPATIAL CONTRAST WITH FREQUENCY DEPENDENT APERTURE EFFECT

Hamed Hooshangnejad[1] , Shahriar Gharibzadeh[2], Farzad Towhidkhah[2]

[1] Department of Biomedical Engineering, Amirkabir University, No. 10, Behesht Alley, Nazak St., Fifth Golestan, Tehran, Iran, e-mail: h.hooshangnejad@gmail.com

[2] Department of Biomedical Engineering, Amirkabir University, No. 424, Hafez Ave., Tehran, Iran, e-mail: gharibzadeh@aut.ac.ir; Towhidkhah@aut.ac.ir

ABSTRACT

The main purpose of this study is to build a Computational model based on ModelFest dataset which is able to predict contrast sensitivity while it benefits from simplicity, efficiency and accuracy, which makes it suitable for hardware implementation, practical uses, online tests, real-time processes, an improved Standard Observer and retina prostheses. It encompasses several components, and in particular, frequency dependent aperture effect (FDAE) which is used for the first time on this dataset, which made the model more accurate and closer to reality. Shortcomings of previous models and the necessity of existence of FDAE for more accuracy led us to develop a new model based on Wavelet Transform that gives us the advantage of speed and the capability to process each frequency channels output. Considering our goal for building an efficient model, we introduce a new formula for modeling contrast sensitivity function, which generates lower RMS error and better timing performance. Eventually, this new model leads to having as yet lowest RMS error and solving the problem of long execution time of prior models and reduces them by almost a factor of twenty.

Keywords: visual system modeling, spatial vision, contrast detection, wavelet filter bank, multi-channel model, ModelFest, frequency dependent aperture effect.

INTRODUCTION

Spatial sensitivity modeling

Human visual system (HVS) is the most important and complicated sense of human body and it is also known as an optimal image processing system [3]. As a result, it has been the aim of numerous researches for so many years. For instance, the proposal of spatial-frequency selective mechanisms in vision is 50 years old. These researches were done in diverse fields, including spatial contrast detection, retina modeling, eye movements, etc.

Spatial contrast sensitivity as Watson & Ahumada [43] defined, is the ability to see alteration of light intensity in spatial patterns. Since spatial contrast sensitivity is the first stage of seeing, it has been the focus of many studies. One of the earliest studies was Ricco's law [15] which tried to describe the relation between stimulus area and contrast threshold and summation rule within an area. Later, by introducing contrast sensitivity function (CSF) [7], new models based on spatial filters were developed. The rule of CSF in modelling of HVS cannot be overlooked [28], and this great importance leads to both invention of several methods to measure human CSF [24] and divers investigations into different aspects of Contrast sensitivity function [22]. In 1969, Blakemore & Campbell [5] developed the idea of multiple spatial filters instead of single one. In addition, there were studies on variation of sensitivity with orientation [4] and eccentricity [27].

Although these studies tried to investigate HVS operation in one of these aspects, lack of a more general model which incorporates all or some of these properties was introduced by Watson [36], at the beginning of 21st century, to present his model which includes CSF, frequency channels and pooling. Later on, another improved model was introduced by Watson & Ahumada [43]. We will return to this model for more investigation in latter sections.

Designing and gathering of ModelFest data, helped scientists in the area of contrast sensitivity to design and validate their models using the same dataset; hence, we used this dataset to develop our model and compared our results with previous studies.

ModelFest dataset and previous analyses

ModelFest dataset consists of 43 grey scale stimuli (Figure 1) and Stimuli's properties are comprehensively described in [43]. It was gathered in two phases, first for nine observers and second for seven more and four trials for each observer. Since there is a detailed explanation of ModelFest dataset in [43], here we only point out their result of their calculation on dataset descriptive statistics which implies that, by assuming the homogeneity of variances, using the overall average over observers and trials for each stimulus for optimizing a model, the lowest achievable RMS error is 0.56 dB.

Some of the first studies on this dataset was conducted in [30], [34] and [8] but none of them used all dataset to develop a single model. In 2000, Watson used ModelFest phase one data to develop his model which consists of three parts, CSF filter followed by frequency channels and Minkowski pooling [36]. In that study, Watson tried different kinds of filter bank including Gabor channels, discrete cosine transform (DCT), and no-channel frequency channels part of model. This study suggested that DCT has poor performance and using Gabor channels will bring about a better result.

Watson [36] study was followed by Watson & Ahumada [43]; this investigation presented a more general model, including CSF, oblique effect, frequency independent aperture effect (FIAE), frequency channels and Minkowski polling. In [43], the effect of a number of CSF filters, and FIAE and oblique effect on RMS error was investigated while the Gabor channels were used. The result was that their model reached 0.79 dB RMS error, which is a good result.

More recently, another model was proposed by Bradley, Adams, & Geisler [6], and it reaches 1.09 dB RMS error on ModelFest dataset. Although this model of detectability is more physiological in inspiration and accounts for detectability across the entire visual field, which our model can easily be generalized to do so, as it will be discussed later, its advantages pale in comparison with its poor timing performance in practical uses.

Present analysis

In this study, we aim to develop a more efficient and accurate model based on Watson & Ahumada [43] model which is able to predict human contrast sensitivity. The uses of such a model is beyond count; for instance, this model can be applied to myriad problems of display measurement and inspection such as measurement of motion blur and mura [38, 41]. Further, this model can be used in machine vision applications, for example, Feature detection and letter identification [25, 42], and the more practical one which is predicting visibility of aircrafts [37] or even in future retina implants. Thus, it is obvious that a more efficient model that benefits from shorter execution time, lower complexity and fewer computations, which leads to smaller and lighter and less expensive hardware for implementing them, will be definitely desirable and beneficial. Furthermore, such an efficient model can be used in online tests and real-time processes. On the basis of Watson & Ahumada article [43], complex and time consuming channel type used in their model, termed as Gabor filters, made them omit the channels for practical uses and further investigations, but the fact remains that this omission not only increased the RMS error but also prevent us from benefiting from the existence of FDAE. After checking the effects of various types of channels on RMS error and execution time, the wavelet channels were chosen. The advantage of channels' existence helps us introduce FDAE which culminated in the as yet lowest RMS error of about 0.68 dB.
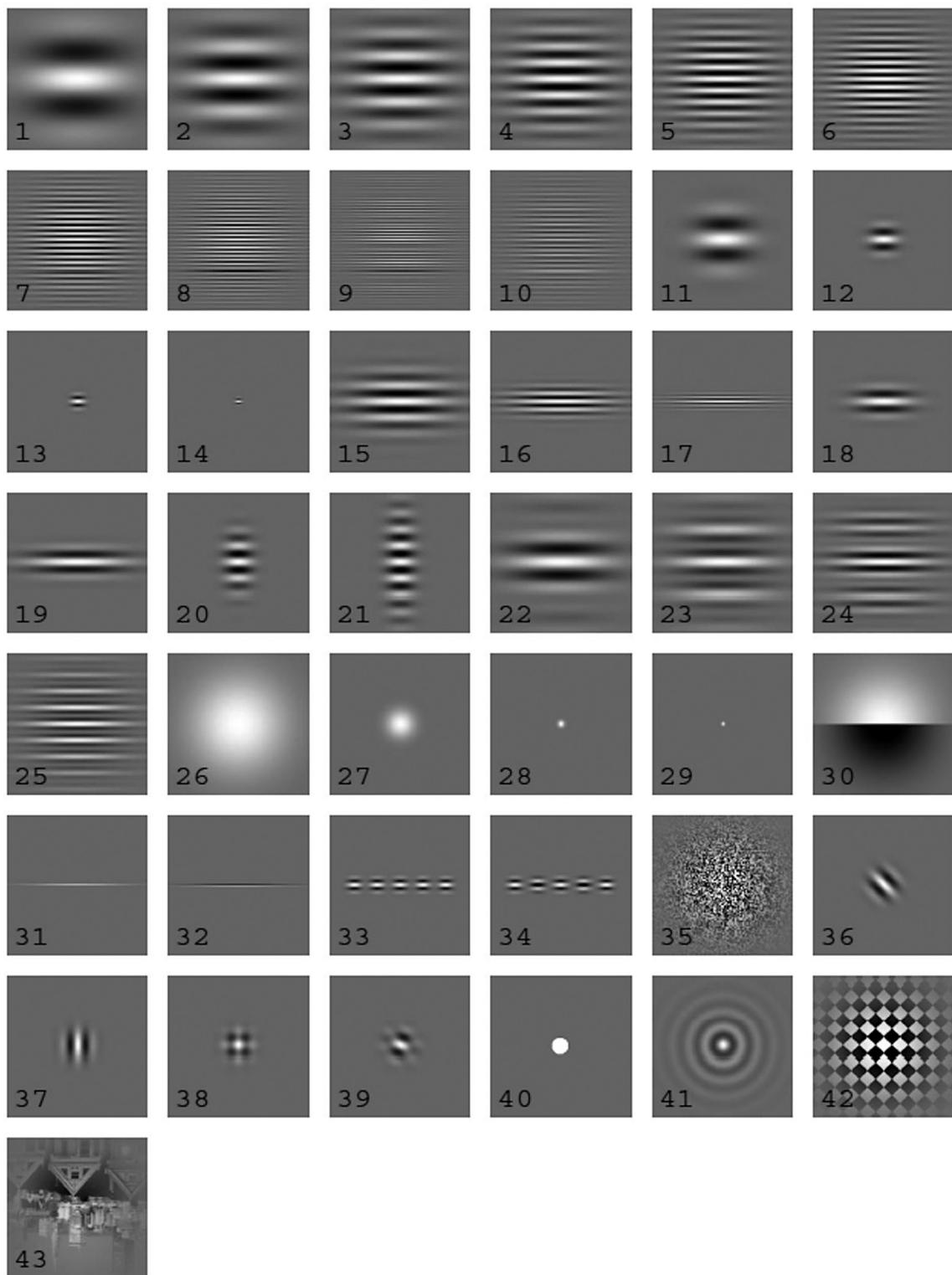
Fig. 1. 43 Stimuli of ModelFest dataset. These images were displayed in monochromic and the index number was not shown in display and their dimensions was 2.133×2.133 degrees [39]
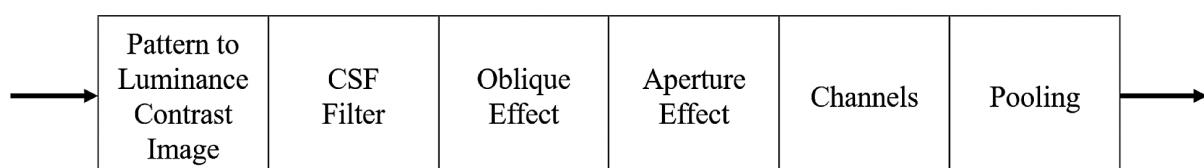

Fig. 2. Model structure composed of six parts

MODEL STRUCTURE

In this report, we used a model comprised of six parts including Luminance to Contrast Image, CSF Filter, Oblique Effect, Aperture Effect and Pooling (Figure 2). In the following sections, each component will be briefly discussed.

Model Input and Output

An input of model is one of the digital grayscale images of ModelFest dataset. Each image has 256 pixels width and length. Output of model is a real number that is the contrast threshold for that input image.

Contrast Image

The first stage is to converted each image to a contrast image that only contains the differences between pixels' intensity. This stage is similar to what happens in the retina when light is emitted on an on-center off-surround or off-center on-surround ganglion cells' receptive field (GCRF). In those GCRFs the response of the center and surround are subtracted from each other and ganglion cells only transmit differences, which is proved to be a more effective and optimized way than transmitting the exact intensity [18]. A contrast image will be then produced when the value of 128 , the nominal mean of intensities, is subtracted from each pixel intensity and then divide by 127 [36].

Contrast Sensitivity Filter (CSF)

After each stimulus is converted to contrast image, it is convolved by CSF like filter. This operation will be done by circular convolution's equivalence in frequency domain. Here we as-
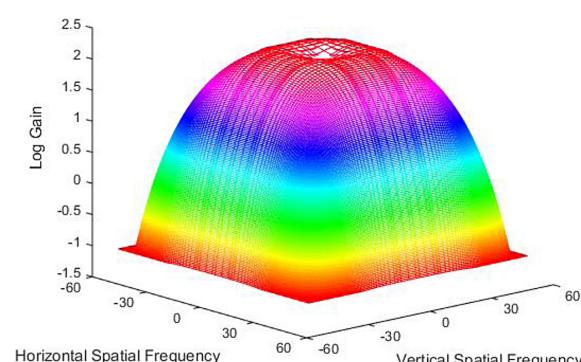
sume that Gaussian envelope of stimuli will decrease or even omit the border effect caused by circular convolution. Figure 3 shows an example of CSF filter, in which filter gain decreases in both high and low frequencies similar to human contrast sensitivity function.

Oblique Effect

Based on Berkley data [4], Watson & Ahumada modeled the oblique effect with the equation below, by the assumption of sinusoidal variation, and therefore, the following equation was obtained (Equation 1)[43].

$$O(f, \theta) = 1 - (1 - \exp(-\frac{f - \gamma}{\lambda}))\sin^2(2\theta) \quad if \quad f \succ \gamma \tag{1}$$
$$= 1 \qquad\qquad\qquad\qquad\qquad if \quad f \prec \gamma$$

In this equation $\gamma = 3.48$ cycles/degree is the frequency where decline in sensitivity starts and $\lambda = 13.57$ cycles/degree is the rate of decline per frequency. This effect is applied to the model as a filter in frequency domain (Figure 4).

Aperture Effect

This effect mainly introduced in 1981 [27]. it implies that sensitivity declines as a function of eccentricity. This phenomenon depends on two factors, first the distance from fixation point and second the frequency which the stimulus is emitted at. Traditionally, a Gaussian function is used to model this phenomenon, however, other functions such as witch's hat and samurai hat was proposed in recent years [2]. Given the quality of our model's fits we did not use these functions to avoid more complexity. The Gaussian function will apply to the model as a scaling function in spatial domain.
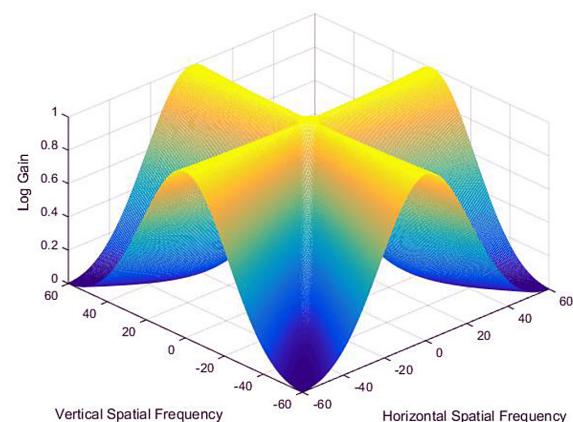

Fig. 3. Contrast Sensitivity Filter as a function of spatial frequency. Sensitivity gain declines at low and high frequencies

Fig. 4. Oblique Effect Filter implemented as a filter in frequency domain

$$G(r) = \exp(-\frac{r^2}{2\sigma^2}) \qquad (2)$$

where: σ is called the size of aperture and r is the distance from fixation point.

In this study, we attempted to model FDAE for the first time on ModelFest dataset, while Previous analyses have only modeled the FIAE using a solitary Gaussian function. Here, we used multiple Gaussians in different frequency channels with different sizes that are free to vary. The existence of FDAE will lead to fewer errors and for implementing its efficient frequency channels are needed. These Gaussians are applied with their centers aligned with the center of images. This is because of the assumption that observer's eye was fixated at the center of targets.

Frequency Channels

Based on physiological and psychophysical evidences, it is proven that HVS decomposes input stimuli to parallel frequency channels each has a distinct frequency band. Thus, the mainstay of the model is the channels. This part was included in previous analyses like [36] and [43].

In this article, we scrutinize the effect of six different types of channels five of which including wavelet filter bank, wavelet packets, real and complex dual tree complex wavelet transform (DT-CWT), and Berkeley Wavelet Transform (BWT) [44] were used for the first time. Another Channel's type termed as Gabor filters, which is known as the best model before our investigation, is simulated along with them, for comparison. Gabor filters' parameters are mentioned in Table 1 which are the same as what Watson & Ahumada used for their channels [43]. it is important to point out that while the Gabor function is very popular, other filter shapes have been considered too. A very recent paper introduced the Gaussian

Table 1. Gabor Channels' Parameters [43]

Number of frequencies	11
Number of orientations	4
Number of phases	2
Bandwidth (octaves)	1.4
Highest center frequency (cycles/degree)	30
Lowest center frequency (cycles/degree)	0.9375
Frequency spacing (octaves)	0.5
Orientation spacing (degrees)	45
Pyramid sampling	Yes

derivate filter as a suitable model [14], but since it was not based on ModelFest dataset, we did not include its result in this study.

Pooling

The last part of the model is responsible for converting 2D patterns into a real number, indicating the contrast threshold. Because of its operation, it is named as the pooling stage. This stage is related to image quality assessment field [35] and although there are divers methods to use for this purpose [11, 23, 46], here we used Minkowski pooling method which has been used in numerous previous studies [6, 36, 38, 41, 43]. This method is inspired by summation property of V1 cells [16, 27, 40]. Equation 4 shows Minkowski Pooling formula.

$$R = [\sum_{x=1}^{Nx} \sum_{y=2}^{Ny} p_x p_y \mid c_T r_{x,y} \mid^{\beta}]^{1/\beta} \qquad (3)$$

Here c_T is the contrast threshold and $r_{x,y}$ is the processed pixel. β is the summation exponent and p_x, p_y are pixel length and width in degrees [43]. Determining $R = 1$ as the threshold of detection, CT can be calculated from equation 5.

$$CT = [\sum_{x=1}^{Nx} \sum_{y=2}^{Ny} p_x p_y \mid r_{x,y} \mid^{\beta}]^{-1/\beta} \qquad (4)$$

When frequency channels exist, we use the subsequent equation for pooling the channels contrast thresholds.

$$CT = [\sum_{q=1}^{Q} CT_n^{-\beta}]^{-1/\beta} \qquad (5)$$

Here CT_n is contrast threshold of each of the channels and N is number of the channels. The advantage of using Minkowski pooling is that it incorporates peak detection model ($\beta = \infty$) [40], probability summation ($\beta = 3$) and energy summation ($\beta = 2$). It should also be noted that in the case of the discredited high threshold model of contrast detection, the exponent depends on the slope of the psychometric function, and beta between 2 and 4 is probably reasonable in many cases.

It should also be noted that the psychometric function's shape might be diagnostic for some of the contrast detection models. Steeping of this function can be associated with uncertainty, but

the fact remains that good estimation can be acquired by several threshold runs, and single run will not provide enough data for estimating this function. Thus, we simulated our model for 10 times and 5-fold cross validation, as it will be explained later.

FREQUENCY CHANNELS

In this section we will describe various types of channels used in this study. As it was mentioned above, there are physiological evidences that HVS uses several frequency channels for stimuli processing. Thus, this kind of processing can be implemented using a component that decomposes its input to multi frequency bands. Filter banks are used for such a process and each filter bank can be designed by one of the mathematical transforms, such as Discrete Fourier transform filters array, known as Gabor filters [36], short time discrete Fourier transform (ST-DFT), discrete cosine transform (DCT) and discrete wavelet transform (DWT).

The results of Watson investigations on Gabor filters and DCT implied that DCT filter bank has frustrating performance, compared to Gabor filters with the error more than twice as high[36]. Here we used wavelet transform as the main mathematical tool for creating filter banks. Wavelet packets, real and complex dual tree complex wavelet transform (DT-CWT), and Berkeley Wavelet Transform are different derivations of a wavelet transform.

Wavelet Filter bank

Wavelet filter bank or discrete wavelet transform is a more recent concept compared to Fourier transform and it was introduced by Morlet due to time-frequency resolution deficiencies of short time Fourier transform [26]. As we mentioned earlier, we used different types of wavelet filter bank and since wavelet transform is a well-known mathematical transform, here we only briefly explain the distinctions between these types.

The first type is the conventional wavelet filter bank or, in other words, discrete wavelet transform that divide frequency bands by first halving the whole frequency band and then continue dividing it by just halving the lower band and leave the higher band intact [20]. As a result, focus of DWT is on low frequencies. The second type of

filter bank that we used is wavelet packet transform, which divides frequency bands in the way that not only does it halve the lower band, but it also halves higher bands during decomposition. Wavelet packets transform divides high and low frequencies in the same way that DWT divides low frequencies hence using wavelet packets leads to a better control over partitioning time frequency plane [12].

The third and fourth filter banks which are called Real DT-CWT (RDT-CWT) and Complex DT-CWT (CDT-CWT) are much newer than previous ones. We can describe these transforms as forms of DWT that uses dual tree of wavelet filters to obtain their real and imaginary parts for their complex output coefficients. Unlike previous ones, these two focus on orientation rather than frequencies. Real DT-CWT has six directional wavelets in six distinct directions. Complex DT-CWT has the same six distinct directional wavelets, except in this kind of DT-CWT there are two phases for each direction by using twice as many wavelets as the Real type.

The last derivation of wavelet transform we used was BWT. BWT is a two-dimensional triadic wavelet transform, and it is composed of four pairs of mother wavelets, at four orientations. In each pair one wavelet has even symmetry and the other has odd symmetry. This wavelet transform shares some of the features of V1 neurons' receptive filed, and by scaling and transition of the whole set, the wavelet constitutes an orthogonal basis [44].

Gabor Channels

The sixth type of frequency channels that we simulated is Gabor channels. Since we utilized these frequency channels with the same parameters (Table 1) as [36] and [43], we will not continue describing it in this article, and we suggest reading reference articles for more information.

CONTRAST SENSITIVITY FILTERS (CSF)

In this part, we will introduce the CSFs which we used in this research. As it was pointed out previously here, we mainly used three types of CSFs namely, Log-Sensitivity Interpolation (LSI, [36]), HPmH ([43]) and the last one, which we introduced, termed as DoS (Difference of Sigmoids). CSFs are somehow band pass filters, and

therefore they can be built by subtracting two low pass filters, as it was done in HPmH and DoS. Filtering operations were done by using frequency equivalence of circular convolution. Border effect can be ignored by considering the Gaussian envelopes of stimuli and influence aperture effect on processed image.

Log-sensitivity Interpolation

This CSF, introduced by [36], has 11 free parameters. 10 of these parameters are for 10 gains at stimuli 1-10 frequencies. Additionally, an extra parameter was assigned to frequency 0. For the purpose of limiting the interpolation, the gain at the last frequency (256 Cycles/degree) was fixed to –50 dB.

HPmH

This CSF was introduced by Watson & Ahumada [43] based on HPmG filter that was suggested by [29] and it uses the difference of two secants, each of which serves as a low pass filter. This CSF has 5 parameters to be assigned.

$$G_{HPmH}(f; f_0, f_1, a, p) = \sec h[f/f_0]^p \\ - a \times \sec h[f/f_1] \quad (6)$$

In this equation f_0 and f_1 scale frequencies in high and low frequency lobes respectively. indicates the effect of low frequency lobe on the high one and p gives us more flexibility.

DoS

We introduced this function as the difference of two sigmoid functions which each of them implements a low pass filter, and by subtracting them from each other a band pass filter can be generated. The chief reason why we introduced this CSF is because it has 2 less exponent function than HPmH, which leads to less computations and contributes to better execution time. This filter has 7 free parameters.

$$G_{DoS}(f; b_1, b_2, a_1, a_2, c_1, c_2, g) = \\ g(b_1 \times sigm[f, a_1, c_1] \quad (7) \\ - b_2 \times sigm[f; a_2, c_2])$$

In this formula g is the gain of the whole filter, b_1 & b_2, a_1 & a_2 and c_1 & c_2 are the weights, scales and shifts of sigmoid functions at high and low frequency lobes, respectively.

Other CSFs

As we wanted to compare our CSF with conventional filters we have simulated 7 more CSFs including DoG, MS, HmH, HmG, LP, EmG, YQM, HPmG in two conditions. First, when the channels exist, and second when they do not. All these CSFs are introduced in [43].

RESULTS

Model implementation and Optimization

We used MATLAB® programming language for implementing our model. Parameters of each configuration of model were estimated by using five MATLAB® built-in methods including Fminsearch, Fminunc, pattern search algorithm, genetic algorithm and Particle Swarm. Although no optimization algorithm can guarantee that it can find the global minimum, we utilized the hybrid method of Fminunc plus Fminsearch for finding the lowest minimum. These results must be considered as the upper bound of attainable minimum. Each optimization was done for at least ten times, to check the consistency of answers.

Considering the fact that Gaussian envelopes of stimuli can minimize the border effect, we used circular convolution for implementing all filtering operations.

We validated our model using the RMS error (RMSE) in dBs. Since this method of validation (RMSE) is completely prone to overfitting, here we used 5-fold cross validation for training our model. Although here we assumed that the results of prior studies were obtained through the same process, since nothing is mentioned in their articles, there is no evidence to give credence to this assumption [6, 43]. Thus, their results are not without reservations about the performance of their model when cross validation is used. One of the most challenging tasks regarding cross validation was selecting the members of folds. Due to the scarcity of stimuli and similarities of some of them (for instance, Stimuli number 1-14), we separate the data set into 4 distinct group (first group: 1-21, 36-37, second group: 22-25, 32-33, 38-39, third group: 26-29,30 and forth group: rest of stimuli) base on each stimulus' trait. Then, the members of each fold for each irritation (here we chose 10 irritations) will be randomly selected from these 4 groups in a way that each fold have at least one stimulus of each group. After train-

ing the model on 4 of these folds, the model was tested on the hold-out one. Using equation 8 the error for training and testing stage was calculated. Therefore, all the following reported results are the results of the model that its parameters are equal to the average of obtained parameters of 5-fold cross validation with 10 irritations of randomly selecting fold members.

$$RMS \quad error = \sqrt{\frac{1}{J}\sum_{j=1}^{j=J}(C_j - P_j)^2} \qquad (8)$$

In this equation C_j is the threshold from dataset and P_j is the model predicted value. Note that both of these values must be used in dBs. J is the number of stimuli of hold-out fold in testing stage and all other folds in training.

Moreover, we used autocorrelation test for further validation of our model predictions, by assuming that our model is a zero order moving average (MA (0)) model. Based on this method, a suitable model with unbiased parameters estimation is obtained if the prediction errors are uncorrelated. Figure 5 shows the prediction errors can approximately be deemed as uncorrelated with 95% confidence bound, and therefore, our estimation is nearly unbiased.

Optimized model Predictions

Here we illustrate the performance of model by showing the predicted values along with dataset's thresholds. The black line indicates the error for each stimulus. Figure 6A shows the prediction and errors when no channels exist, and Figure 6B displays the same variables when
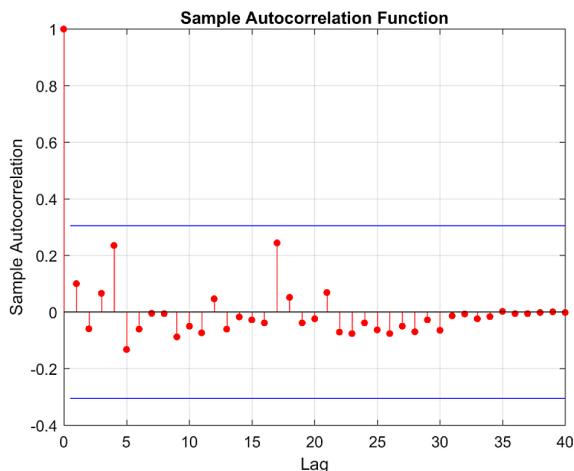
Wavelet channels are used. As it can be seen in Figure 6, there is a substantial change in error for two stimuli of 35 and 43 compared to the others when channels are added to the model, we will talk about this later on.

Contrast Sensitivity Functions

In this section we talk about the effect of different types of CSFs on RMS error. For this purpose, we compared our introduced CSF with 10 conventional CSFs. As it is shown in Figure 7, the first three CSFs from left, which are used for further investigations, are the best CSFs amongst all the others. This investigation was done in two configurations, no channels (Fig. 7A) and wavelet channels (Fig. 7B). It is worth noting that here we used FIAE and later we will add FDAE for more examinations.

It is also notable that using the channels causes the DoS to have lower error, compared to HPmH which performs better in no channel configuration. Figure 8 shows three types of CSFs we mostly used in this article. It is evident that all these three CSFs are similar to a typical human contrast sensitivity function [31, 45], and it proves that our new model
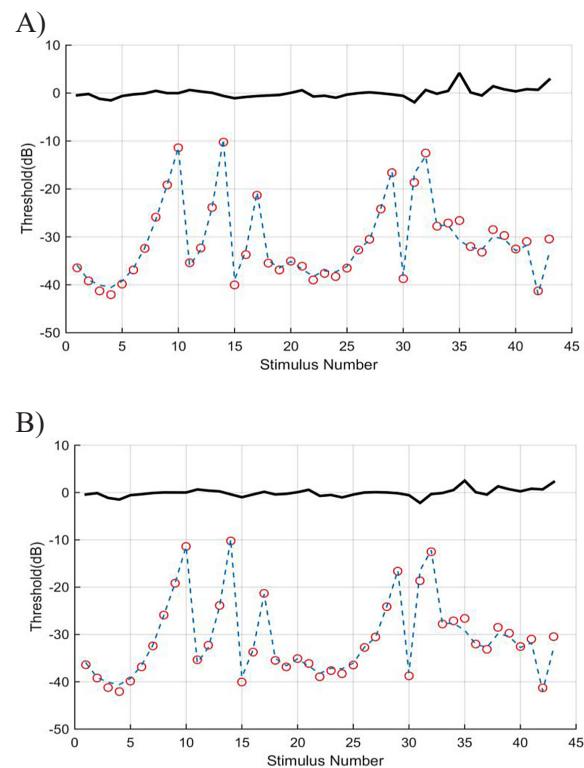

Fig. 6. Optimized model predictions (red dots) alongside thresholds (Blue dashed line) and errors (black solid line) for each stimulus. (A) is the predicted values when no frequency channels exist, (B) shows the predicted values when wavelet channels are used for frequency channels
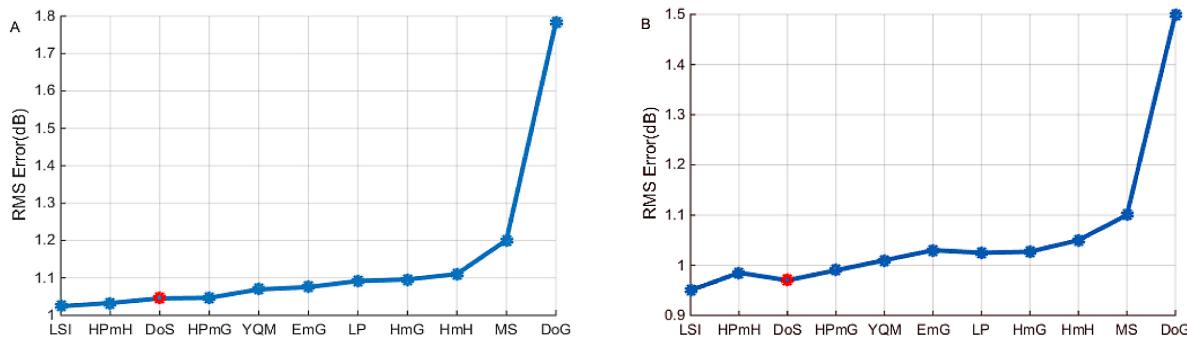

Fig. 5. Autocorrelation test of prediction errors

Fig. 7. Comparison of 11 CSFs including Dos which we introduced. (A) displays RMS error for NO channel model. (B) depicts the RMS errors when wavelet channels and FIAP were used

is close to human physiology. As we intended to have both advantages of speed and low RMS error for the rest of the article we have used DoS CSF along with wavelet channels. Although using LSI will lead to a little decrease in model error, as it will be shown later, it causes a tangible increase in execution time of our model (Figure 11).

Investigation into Channels

As we mentioned earlier in this study we used six types of frequency channels listed below. along with these six channels we compared Retina-V1 model, the latest model on this dataset, as well [6].
- Gabor Filters (used by [36, 43])
- Wavelet filter bank (DWT) (used for the first time)
- Wavelet packets (used for the first time)

- Real DT-CWT (used for the first time)
- Complex DT-CWT (used for the first time)
- BWT (used for the first time)

In using wavelet two significant and essential points should be considered. The first point is decomposition level which displays the level of model's complexity. Another point is the mother wavelet selected that specifies the fundamental kernel for filter bank [32, 33]. By obtaining RMS error for 16 diverse mother wavelets and 4 decomposition levels we realized that the least error is acquired by using "Bior 6.8" as the mother wavelet and three levels of decomposition. Since other types of wavelet based channels have nearly the same principle as the wavelet filter bank [19], we used the same configuration for other types.
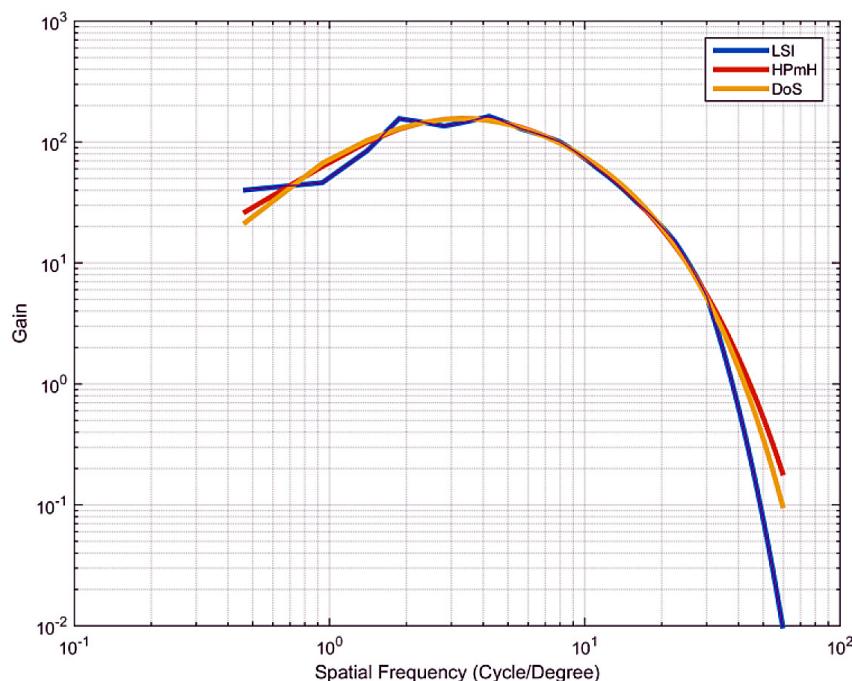

Fig. 8. DoS along with LSI and HPmH CSF after optimization. As it can be seen all these three CSFs resemble Human contrast sensitivity function
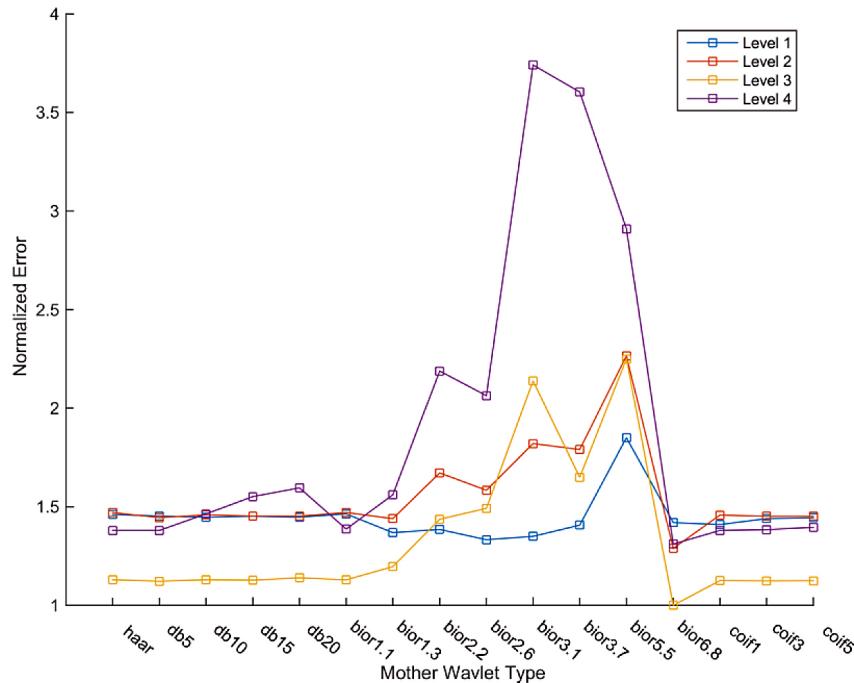
Fig. 9. Obtaining model RMS error by simulating 16 diverse mother wallet and four decomposition levels

By simulating the six aforementioned frequency channels types using FIAE and the mentioned configuration for wavelet based channels, we got the results shown in Figure 10. Based on this observation, Gabor filters have the least error among other types, wavelet channels are the second and wavelet packets are the third one. It should also be noted that the Retina-V1 model has the highest error among the others.

According to this result (Figure 10) one may pose the question that why we need wavelet channels when Gabor filters has the least RMS error. To answer this question, first we have to show timing performance of the model in various configurations. Figure 11 depicts runtime performance of the 15 model configurations. In this investigation, we first allowed the models (model configurations) to be optimized by trying to find minimum error using the previously mentioned algorithms. Then, we ran the models for all 43 stimuli, and at the same time, we measured the execution time using the built in MATLAB® function "tic, toc". According to Figure 11B runtime performance of Gabor filters is so poor in comparison with other types of channels especially wavelet channels, which has the least runtime after the no channel model.

For a better illustration of this matter, we display timing performance alongside error performance (Figure 12). Despite the fact that LSI has the best error performance between other CSFs, DoS has simultaneously good er-

ror ratio and timing performance (Figure 11), and 11 free parameters of LSI that should be determined, which increased model complexity considerably, made us to use DoS as CSF for further investigations, such as Figure 12. According to Figure 12, although Gabor channels with 0.81 dB (SD = 0.005) error has the least RMS error amongst the others, 0.15 dB lower than wavelet channels, wavelet channels has a significant timing advantage over Gabor filters. Wavelet filters' runtime is 1.01 seconds, which is about 20 times faster than Gabor filters. This disadvantage of Gabor channels, according to Watson & Ahumada [43], give rise
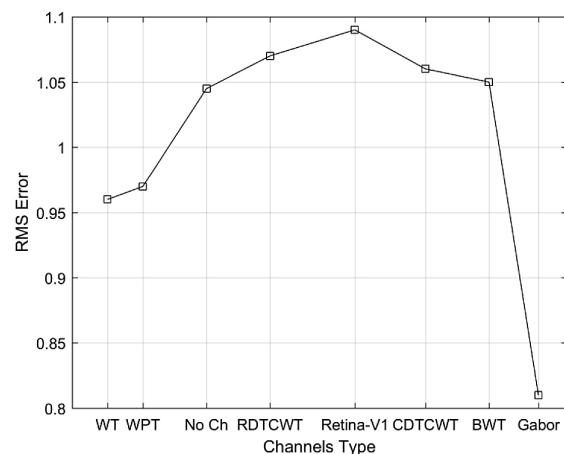

Fig. 10. Error performance of six wavelet configurations alongside no channels and Retina-V1 model
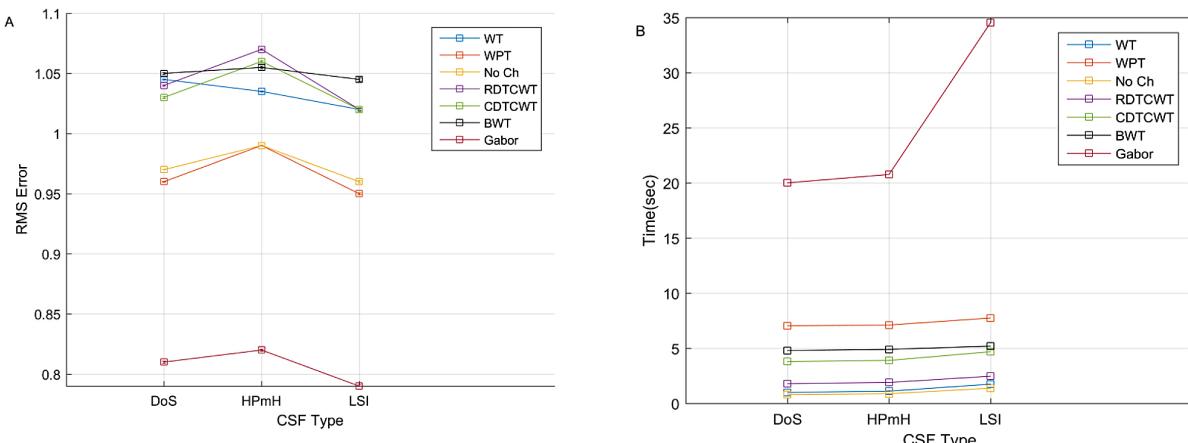
Fig. 11. (A) Model's RMS error for five types of frequency channels alongside three types of CSF. (B) Runtime of model for 15 different configurations

to the omission of channels in practical models. This means that because of the complexity and slowness of Gabor filters, the mainstay of model which is necessary for implementing FDAE, and therefore, further decrease in RMS error has to be omitted. In contrast, using wavelet channels, which has the vital advantage of speed along with having the least error after Gabor filters, will allow us to add FDAE to the model, and eventually, it leads to a much more realistic model with as yet lowest RMS error. Thus, owing to the advantages and good performance of wavelet channels, it was chosen for modelling the channels part of model.

Another notable matter is that as Figure 12 depicts, Retina-V1 Model, in spite of its nearly similar structure to V1, has both poor error and especially timing performance compare to the other model configurations.

Aperture Effect and Introducing FDAE to the Model

In this part, first, we investigated the relation between aperture size (σ) and summation exponent (β). Using FIAE with no channels and allowing the aperture size (σ) to vary when summation exponent is fixed in specific range
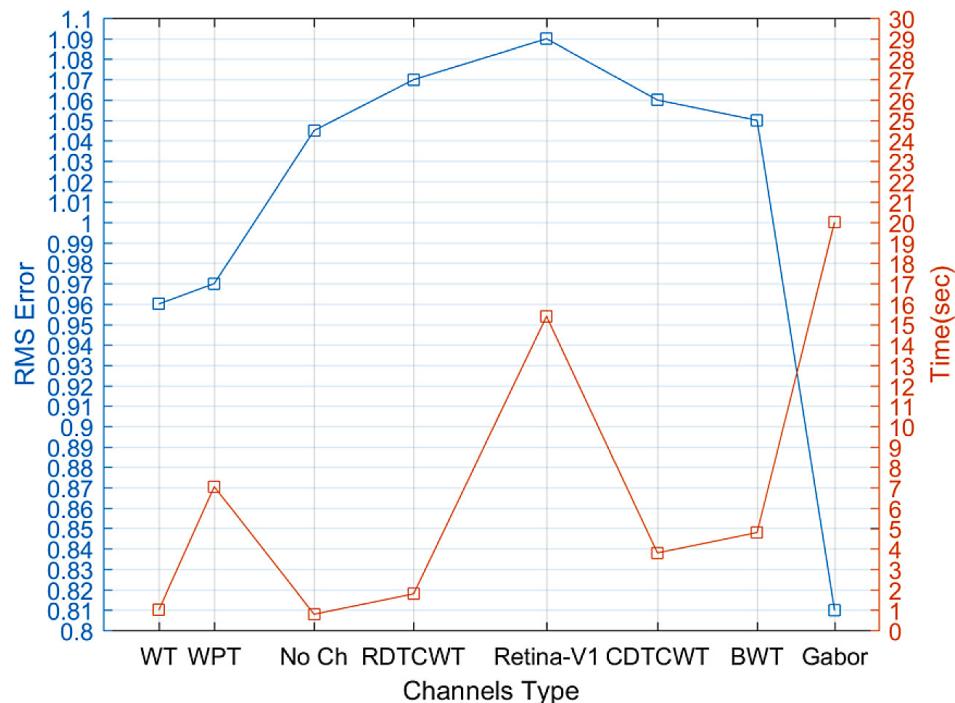

Fig. 12. Timing performance of channels alongside error performance

between 2 and 3, we measured the RMS error for each fixed summation exponent value. Aperture size will be then determined, when the model reaches the lowest achievable minimum. According to [27] study on effect of eccentricity on contrast sensitivity, aperture effect will cause sensitivity to decrease by 0.5 dB/cycle, but more recent studies change these classic results. The investigations in [1, 2] maintain that the sensitivity decrease by about 1 dB/cycle.

The lowest RMS error will be attained when β is 2.4 (Fig. 13A). With this β, the aperture size is 0.51. This leads to 6.02 dB decline in 0.58 degree, and it conforms to Baldwin & et al. [1, 2] result in 10.38 cycle/degree. The obtained result seems to be plausible, since here we used FIAE, which consist of only one Gaussian as the aperture effect, and the frequency range is 0 to 60 cycle/degree, so the 10.38 cycle/degree falls within the frequency range.

Contrary to Watson & Ahumada model [43] which includes slow and complex Gabor filters that prevented them from investigating the relation between aperture size (σ) and summation exponent (β) when channels exist, here by using wavelet channels we are able to perform the same study on relation between β and σ, when channels exist. On the basis of Figure 13B, it can be seen that error will be minimum when β is around 2.4 and the corresponding σ will be 0.63 degree (SD = 0.003). Using the same calculation as above and based on [1, 2] result, the equivalent frequency will be 8.1 cycle/degree. This result seems perfectly reasonable because level-3 wavelet channels mostly focus on first quarter of frequency band (0 to 15 cycle/degree) and based on wavelet coefficient, most of the information is in the first quarter. Hence, it can be anticipated that aperture size will try to reach the middle point (7.5 cycle/degree) of this part of frequency band for highest effectiveness.

Ultimately, we aimed to add FDAE to the model. Since we chose level three wavelet decomposition, there would be four distinct circular frequency bands. As a result, we used
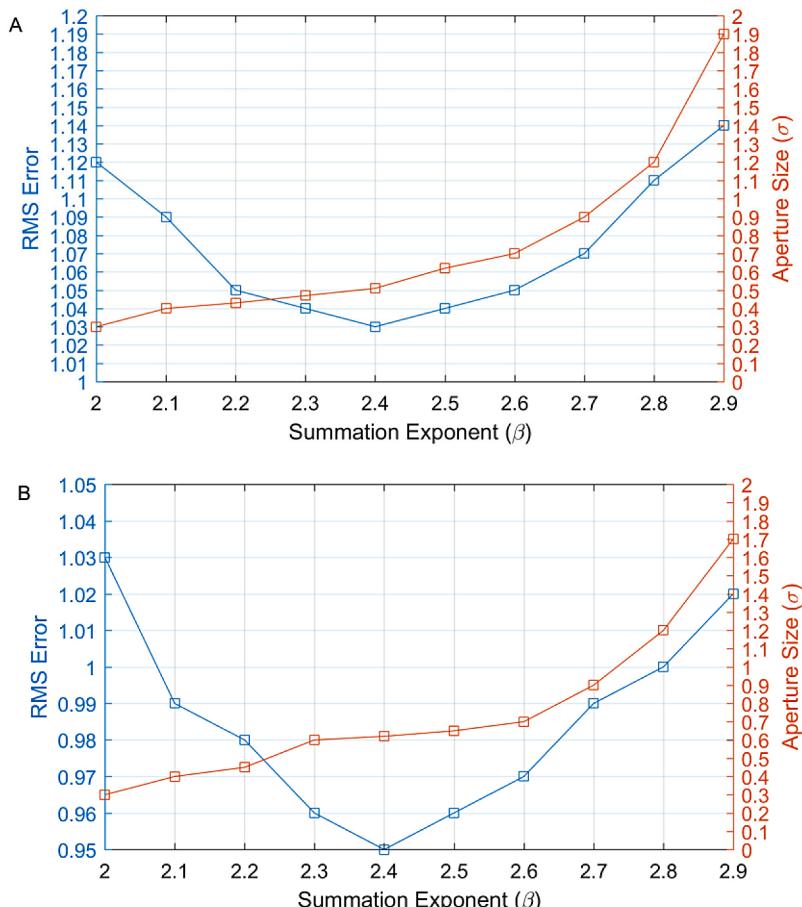

Fig. 13. This curve depicts the relation between aperture size and summation exponent. Other conditions: no channels and FIAE(A) and wavelet channels and FIAE(B)

four Gaussians, each of which is applied to one of the frequency bands. Each Gaussian has an aperture size σ separated from the other ones, and they are free to vary during optimization phase. Then we permitted the model to be optimized using the dataset to reach the least RMS error, and at last we obtained a result that lends credence to the accuracy of our model. Starting from a same value, the aperture sizes reach 0.14 (SD= 0.003), 0.23 (SD= 0.006), 0.505 (SD= 0.002), and 1.47 (SD= 0.004) degrees at the end of the optimization phase. The corresponding frequencies to these aperture sizes ,based on [1, 2] result, are respectively 38, 22.3, 10.1, and 3.5 cycle/degrees. This result completely conforms to both physiology and models assumption. First, based on discrete wavelet transform concept, the frequency bands from low to high band are 0-7.5, 7.5-15, 15-30, and 30-60. It is conspicuous that all of the aforementioned frequencies are within the corresponding wavelet frequency bands, and they are almost at the middle of them. Second, the aperture size of frequency bands of higher frequencies is lower than frequency bands of lower frequencies. This shows that in high frequencies bands the rate of decline in sensitivity by increase in frequency is much higher than low frequencies bands. From the physiological point of view, human eye has lower sensitivity in high frequencies and our result is completely consistent with this phenomenon.

The last but not least result is that using FDAE has a significant impact on error. As Figure 14 clearly shows, using FDAE with wavelet channels decreases RMS error by 0.17 dB, so the model RMS error would be 0.68 (SD= 0.004) which is the hitherto lowest error that a model has achieved in literature. Another important matter is that as it can be seen there is no point regarding Gabor channels and FDAE. This is because as we pointed out earlier due to slowness and complexity of Gabor channels, it is almost practically impossible to use a set of apertures, whose σ is free to vary, along with Gabor channels. In contrast, using wavelet channels which have the advantage of speed and simplicity as well as accuracy allowed us to use FDAE, which eventually leads to a model with the as yet least RMS error.

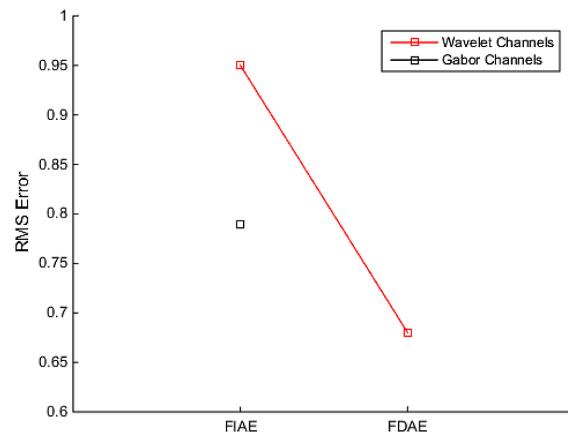Figure 15 shows the model predicted values and contrast threshold simultaneously,

Fig. 14. The effect of adding FDAE on error when wavelet channels are used. Using Gabor channels prevent us from adding FDAE to model due to the complexity and slowness of model

alongside errors for each stimuli. These predicted values are for final model with a complete set of components including, contrast image converter, DoS CSF filter, oblique effect, wavelet channels, FDAE, and pooling.

DISCUSSION

Channels

One of the noteworthy points here is why we used wavelet channels in the first place. First, unlike conventional filter banks (for instance Gabor filters) wavelet filter bank only needs to design two filters. This advantage would be better understood, by considering that the Gabor filters that was used in [36, 43] needs 88 filters to be designed. Although neurons in the primary visual cortex are widely believed to form the substrate of psychophysical spatial frequency channels, and that the spatial receptive field structure of these neurons is often described with a Gabor function [10, 17], what will be the benefit of this channels if it should be omitted for practical uses due to slowness and complexity? On the other hand, using wavelet channels will help us not only by lowering the number of filters to be designed, but also by preserving the mainstay of the model (channels). This gives us a chance for further improving and adding new features such as FDAE, which we introduced here.

The second advantage of using wavelet channels is that its output size has the same size as its input. This will become much more
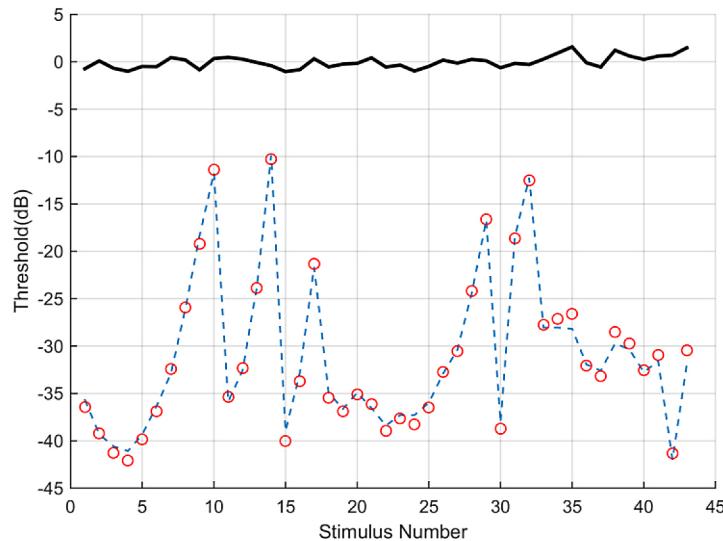
Fig. 15. shows the predicted values (red dots) alongside threshold values (blue dashed line) and error (black solid line) when FDAE is added to model with wavelet channels

important when timing performance is a significant aspect of the model. To shed light on this matter, consider that we have a Gabor filter array with only ten frequency bands, which will lead to 256×256×10=655360 pixels for output, if we assume input size is 256×256 pixels and circular convolution is used for filtering operation (by using conventional convolution more output pixels will be produced.). In contrast, in the same situation, wavelet filter bank by using a particular algorithm which decreases the output size, and at the same time it avoids aliasing. Consequently, it has only 256×256 pixels for output, which is ten times as low as Gabor filters output. Using down sampling for reducing output pixels of Gabor channels, not only arises the aliasing problem, but it is also not as efficient as wavelet channels.

One more notable matter is that using channels causes a reduction in error, especially in stimuli number 35 and 43 (Figure 5). On the basis of Figure 16, both of these stimuli are wide bands. This can be well justified by the fact that these stimuli have a large coefficient in all frequency bands. This considerable error reduction compared to other stimuli seems to be because of a more efficient pooling. It means that using wavelet channels causes the pooling stage to act much more efficiently for wide band stimuli than when we do not use it.

Another important point is that despite the fact that BWT is more physiological based Wavelet, it is not as accurate as Gabor filters. Moreover, contrary to wavelet channels, this type of channels

has a poor timing performance. Thus, a channels type with the efficiency of wavelet channels, accuracy of Gabor channels and physiology basis of BWT is desirable. One of the things that can inspire a new channels type is the way that GCRFs work. In this channels type an array of GCRFs with multitude of sizes and positions can be utilized. So much like the mosaic samp(x) function represented by Retina-V1 model [6], although it needs much better timing performance to be able to be compared with Wavelet channels in practical uses.

Contrast sensitivity filter

Here we will mainly discuss why DoS has better timing and error performance than HPmH and LSI. First, HPmH consists of subtracting two sech functions from each other, and each sech function is comprised of two exponent function as it is shown in equation 9.

$$\sec h(x) = \frac{2}{e^x - e^{-x}} \tag{9}$$

Conversely, the DoS CSF, we introduced here, consists of two sigmoid functions and each sigmoid function only has one exponent function, so in the long run it would be rational that DoS will have better timing performance than HPmH.

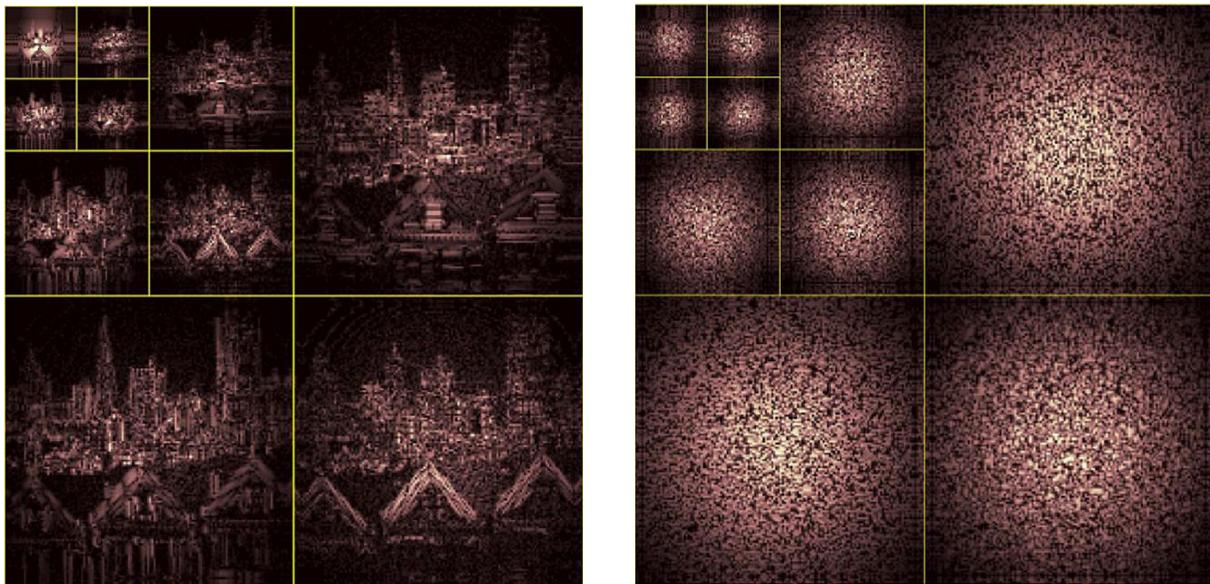$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

Fig. 16. Level-three discrete wavelet decomposition of two stimuli is shown. (A) shows the wavelet coefficient for stimulus number 35 and (B) shows coefficients for stimulus number 43. Both pictures justify the fact that both of these stimuli are wide band, since they have many large coefficient (intense and bright lights) in all frequency bands

Another important fact is that having two more free parameters to vary during the optimization phase than HPmH, makes DoS more flexible, so it will have slightly better performance when channels are added to the model. Thus, it causes a more reduction in error. In contrast, although LSI CSF leads to a little more decline in error, its 11 free parameters which are to be determined impose a superfluous complexity on model. Thus, use of DoS lead to a much more efficient model.

Aperture effect

One of the main impetuses of conducting this research was to introduce FDAE into previous models. Therefore, we tried to replace previous models' inefficient and complex Channels with a more expeditious one, so it can encompass frequency channels for practical models and more complex investigations. After using wavelet channels as a more preferable choice for frequency channels, on the basis of its advantages, we add FDAE to the model which results in a great reduction in RMS error. There are two main reasons for such decrease in error.

First, adding four Gaussians with free σ (aperture size) to vary instead of one (as it was in FIAE) will make our model much more flexible. More flexibility will provide our optimization algorithm a better chance of finding lower

minimum which is closer to global minimum, and therefore it decreases RMS error. Second, as Robson & Graham [27] claimed in their article, high frequency components will pass through smaller aperture and low frequency component pass through larger aperture. Therefore, using one aperture (as FIAE) oblige the aperture size parameter to somehow stay in the middle of the range, and this nearly compulsory value will cause error to increase in stimuli with high frequency components which need smaller aperture, and stimuli with low frequency components which need larger aperture. On the other hand, using several apertures with free sizes to vary, will prevent error from increasing due to mentioned process. Thus, based on two aforementioned reasons, it is plausible that not only does adding FDAE make our model closer to reality, but it also reduces error and makes our model more accurate.

Pooling Stage

In this study, for the purpose of comparison we used the same pooling method used by almost all of the previous analyses, termed as Minkowski Pooling. Despite its advantages such as simplicity and implementing several models including peak detection, energy summation, and probability summation with only one formula, it has a serious disadvantage which should be noted. By replac-

ing equation 5 in equation 6 the resulting equation would be equation 12. This equation clearly shows that contrast threshold will be obtained by summing all processed pixels after channels raising to the power β and then the result of this summation to the power of . Thus, most of channels' effect would be canceled due to similar summation of all output pixels. Consequently, channels will have much less effect on error reduction as it should have.

$$CT = [\sum_{n=1}^{N}\sum_{y=1}^{N_y}\sum_{x=1}^{N_x} p_{x,n}p_{y,n}|r_{x,y,n}|^{\beta}]^{-\frac{1}{\beta}} \qquad (11)$$

Alternative decoding rules that decodes the responses of visual mechanisms in other ways have been considered in [9], analytically in [13] and implicitly in [21], but of course, not in a practical model of contrast detection and on ModelFest dataset. Be that as it may, considering the quality of our cross validation fit, popularity among prior models and our interest and attempt to avoid unnecessary complexity encouraged us to use Minkowski pooling. Testing new pooling methods in this model, requires a comprehensive and distinct investigation, since the effect of this new pooling stage on different parts of the model including aperture size, mother wavelet and decomposition level, and the efficiency of model should be determined. Thus, it is out of the scope of this study.

CONCLUSION

In this study, we mainly tried to present a new, more accurate and efficient model of contrast detection, based on previous models and especially the model presented in [43]. We showed that on the basis of the ModelFest dataset, this model can reach RMS error as low as 0.68 (SD= 0.004) as yet lowest RMS error. Although this model may seem to be more complex than the previous ones, because of its more complex mathematical basis, by using the fast wavelet transform algorithm it would be much simpler to be implemented. Thus, the model's better performance is not due to complexity.

We developed a new and unprecedented channel type based on wavelet transform and filter bank with a number of advantages. The first advantage is its simplicity, only two filters have to be designed for a multi band filter bank that culminate in much more simplicity than conventional filter banks. Second, its output is computed

much faster than Gabor filters, and thus, it is a better choice than Gabor filters. Third, the number of output pixels is almost the same as input, without any further need for separate down sampling, while this down sampling is needed for channels like Gabor filters that eventually will lead to aliasing problem if it is used without precautions. Last but not least, contrary to Gabor filters which forced the omission of the channels from model in practical uses, due to the above mentioned advantages, wavelet channels help us to preserve channels as the mainstay of the model for adding FDAE and further decrease in RMS error.

We introduced new frequency dependent aperture effect termed as FDAE into the model. Not only does the model become much closer to reality, but also a considerable decline in error by 0.17dB can be achieved. In addition to making our model more accurate, using this along with using wavelet channels that provide us speed privilege, give us a chance to use this model for physiological purposes that need accuracy and real-time processing such as retina implants and prostheses.

Lastly, we introduced DoS as a new CSF which performs better than all other CSFs, in terms of error and timing performance. This means that this CSF can be used in wider range of future practical models due to its better performance.

Acknowledgements

REFERENCES

1. Baldwin, A.S. and T.S. Meese, Fourth-root summation of contrast over area: No end in sight when spatially inhomogeneous sensitivity is compensated by a witch's hat. Journal of vision, 15 (15), 2015, 1–4.

2. Baldwin, A.S., T.S. Meese, and D.H. Baker, The attenuation surface for contrast sensitivity has the form of a witch's hat within the central visual field. Journal of vision, 12 (11), 2012, 23–23.

3. Batchelor, B.G., Machine Vision Handbook. 2012: Springer.

4. Berkley, M.A., F. Kitterle, and D.W. Watkins, Grating visibility as a function of orientation and retinal eccentricity. Vision Research, 15 (2), 1975, 239–244.

5. Blakemore, C.t. and F. Campbell, On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. The Journal of physiology, 203 (1), 1969, 237–260.

6. Bradley, C., J. Abrams, and W.S. Geisler, Retina-V1 model of detectability across the visual field. Journal of Vision, 14 (12), 2014, 12–22.

7. Campbell, F.W. and J. Robson, Application of Fourier analysis to the visibility of gratings. The Journal of physiology, 197 (3), 1968, 551–566.

8. Carney, T., et al. Modelfest: Year one results and plans for future years. in Electronic Imaging. 2000. International Society for Optics and Photonics.

9. Clatworthy, P., et al., Coding of the contrasts in natural images by populations of neurons in primary visual cortex (V1). Vision research, 43 (18), 2003, 1983–2001.

10. Daugman, J.G., Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. JOSA A, 2 (7), 1985, 1160–1169.

11. Gong, M. and M. Pedersen, Spatial pooling for measuring color printing quality attributes. Journal of Visual Communication and Image Representation, 23 (5), 2012, 685–696.

12. Gonzalez, R.C., Digital image processing. 2009: Pearson Education India.

13. Goris, R.L., et al., A neural population model for visual pattern detection. Psychological review, 120 (3), 2013, 472.

14. Goris, R.L., E.P. Simoncelli, and J.A. Movshon, Origin and Function of Tuning Diversity in Macaque Visual Cortex. Neuron, 88 (4), 2015, 819-831.

15. Graham, C. and R. Margaria, Area and the intensity-time relation in the peripheral retina. American Journal of Physiology--Legacy Content, 113 (2), 1935, 299–305.

16. Graham, N., Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. Vision research, 17 (5), 1977, 637–652.

17. Jones, J.P. and L.A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of neurophysiology, 58 (6), 1987, 1233–1258.

18. Levin, L.A., et al., Adler's Physiology of the Eye. 2011: Elsevier Health Sciences.

19. Mallat, S., A wavelet tour of signal processing. 1999: Academic press.

20. Mallat, S., A wavelet tour of signal processing: the sparse way. 2008: Academic press.

21. May, K.A. and J.A. Solomon, Connecting psychophysical performance to neuronal response properties I: Discrimination of suprathreshold stimuli.

22. McAnany, J.J. and K.R. Alexander, Spatial contrast sensitivity in dynamic and static additive luminance noise. Vision Research, 50 (19), 2010, 1957–1965.

23. Meese, T.S. and R.J. Summers, Theory and data for area summation of contrast with and without uncertainty: Evidence for a noisy energy model. Journal of vision, 12 (11), 2012, 1–9.

24. Pelli, D.G. and P. Bex, Measuring contrast sensitivity. Vision Research, 90, 2013, 10–14.

25. Pelli, D.G., et al., Feature detection and letter identification. Vision research, 46 (28), 2006, 4646–4674.

26. Polikar, R., The story of wavelets. Physics and modern topics in mechanical and electrical engineering, 1999, 192–197.

27. Robson, J. and N. Graham, Probability summation and regional variation in contrast sensitivity across the visual field. Vision research, 21 (3), 1981, 409–418.

28. Rovamo, J., R. Franssila, and R. Näsänen, Contrast sensitivity as a function of spatial frequency, viewing distance and eccentricity with and without spatial noise. Vision Research, 32 (4), 1992, 631–637.

29. Tyler, C.W., Theory of texture discrimination of based on higher-order perturbations in individual texture samples. Vision Research, 44 (18), 2004, 2179–2186.

30. Tyler, C.W. and C.-C. Chen, Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. Vision research, 40 (22), 2000, 3121–3144.

31. Valberg, A., Light vision color. 2007: John Wiley & Sons.

32. Vetterli, M. and C. Herley, Wavelets and filter banks: Theory and design. Signal Processing, IEEE Transactions on, 40 (9), 1992, 2207–2232.

33. Vetterli, M. and J. Kovacevic, Wavelets and subband coding. 1995: Prentice-hall.

34. Walker, L., S. Klein, and T. Carney. Modeling the Modelfest data: decoupling probability summation. in Optical Society ofAmerica Annual Meeting, Digest ofTechnical Papers, pp. SuC5. 1999.

35. Wang, Z. and X. Shang. Spatial Pooling Strategies for Perceptual Image Quality Assessment. in Image Processing, 2006 IEEE International Conference on. 2006.

36. Watson, A., Visual detection of spatial contrast patterns: Evaluation of five simple models. Optics Express, 6 (1), 2000, 12–33.

37. Watson, A., C.V. Ramirez, and E. Salud, Predicting visibility of aircraft. PloS One, 4 (5), 2009, 5594.

Journal of vision, 15 (6), 2015, 1–8.

38. Watson, A.B., 31.1: Invited Paper: The Spatial Standard Observer: A Human Vision Model for Display Inspection. SID Symposium Digest of Technical Papers, 37 (1), 2006, 1312–1315.

39. Watson, A.B. ModelFest. 1999; Available from: http://vision.arc.nasa.gov/modelfest/.

40. Watson, A.B., Probability summation over time. Vision research, 19 (5), 1979, 515–522.

41. Watson, A.B., Spatial standard observer. 2012, US Patent 8,139,892.

42. Watson, A.B. and A.J. Ahumada, Letter identification and the Neural Image Classifier. Journal of vision, 15 (2), 2015, 10–15.

43. Watson, A.B. and J.A.J. Ahumada, A standard model for foveal detection of spatial contrast. Journal of Vision, 5 (9), 2005, 1–6.

44. Willmore, B., et al., The berkeley wavelet transform: a biologically inspired orthogonal wavelet transform. Neural Computation, 20(6), 2008, 1537–1564.

45. Yantis, S. and H. Pashler, Stevens' Handbook of Experimental Psychology: Vol. 1. Sensation and Perception. 2002, New York: Wiley.

46. Zewdie, C.G., M. Pedersen, and W. Zhaohui. A new pooling strategy for image quality metrics: Five number summary. in Visual Information Processing (EUVIP), 5th European Workshop on. 2014.