Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A MATHEMATICAL FRAMEWORK TO DECODE ATTENTION FROM CORTICAL RESPONSES FOR HEARING-IMPAIRED LISTENERS AND BCI APPLICATIONS

ANUDEEP PEDDI[1], DR. T. VENKATA RAMANA[2]

[1] Research Scholar Department of EECE,GITAM Uiversity, Visakhapatnam, Andhra Pradesh, India
Email:peddianudeep88@gmail.com

[2] Associate Professor,Department of EECE,GITAM Uiversity, Visakhapatnam, Andhra Pradesh, India.
Email:vteppala@gitam.edu

ABSTRACT

The science keeps advancing rapidly, new research topics keep emerging daily, and one such field is neuroscience. New computer models are being proposed that mimic human visual and auditory systems, with the former being the central area of focus. Humans are very good at focusing their attention on a required sound. This is not possible for people with hearing impairment because hearing aids amplify all the incoming signals. Our objective is to try and model the auditory system of humans, specifically on the topic of auditory attention. Our ears are always active and are fed with a large variety of sounds at each moment. We aim to model when our attention is grabbed by a particular sound amongst a large cacophony of sounds. If this is implemented on a hardware system, people with hearing issues can focus only on required sounds. This can be developed by using the concept of temporal response functions (TRFs), which show the linear relation between audio and EEG signals. We proposed a new mathematical framework to overcome the current challenges to predict the sound envelope. This obtained envelope is compared with the audio input given while the EEG data was recorded, using the concept of correlation. The correlation coefficients obtained for different values of regularization parameters are discussed. The proposed mathematical technique gave a better result compared to the existing state-of-the-art techniques.

Keywords: *Mathematical Modeling, Regression, Cocktail Party Problem, Auditory Scene Analysis, Auditory Attention Detection, Hearing Impairment solutions, Regression.*

1. INTRODUCTION

Humans have incredibly complicated auditory systems, and research in this field is gaining traction in recent times due to the advance in medical equipment. There is a big motivation to understand the human brain's response to audio stimuli as it can lead to progress in neuroscience, robotics, and brain-computer interfaces. Cognition is the capability to process information through stimuli that we get from the environment around us. There are different types of cognitive processes, and attention is the process that allows us to concentrate on certain activities or stimuli[1][2]. Attention is used in most daily tasks to be performed, and it controls and regulates the other cognitive processes like perception, thought, language, and learning [3]. The focus of this paper is on auditory attention, more specifically, selective auditory attention. It involves the auditory cortex of the human brain, and it is signified as the action which enables people to pay attention to specific sounds or speech stimuli. The scenario of a party where there are lots of sources of sound being heard simultaneously was popularized as the "cocktail party problem" in the 1950s [4]. It suggested that humans were adept at focusing on a specific sound source that they were interested in listening to and tuning out all the other disturbances [5]. Familiar characteristics of the speaker, like their tone of voice and distance from the speaker, helped filter out the other sounds. To completely understand the underlying processes in the brain, this area has been gaining more recent recognition from researchers.

2 LITERATURE

Every human ear can concentrate on one sound even though multiple sounds exist in the environment or surroundings. It happens for all the living beings in the environment. This is known as the "Cocktail Party Problem" [6]. Every human voice which is present in a noisy environment overlap with the frequency and time, which leads to

acoustic interference and can impair the clarity of speech [6]. It has been submitted that one's sensory memory subconsciously removes the entire unwanted event that evokes a specific functional reaction in an organ and identifies the required pieces, and transfers it to the human brain. This is the effect in which most people can listen to one voice instead in a group of noises [7]. This is a similar phenomenon that occurs when one suddenly detects a word with high importance rather than the unwanted event that evokes a specific functional reaction in an organ [8][9][10].
Most humans undergo cocktail party problems due to communicating groups and noisy surroundings. Similarly, the people with hearing impairment will receive the voice signal appropriately for communication, or due to the hearing aids, all the signals will be amplified and received [11][12]. We want to focus on the reception of the signals to make their lives more feasible.

The process in which every sound that we hear naturally is divided by the auditory system, and the sounds are overlapped and interleaved in time is auditory scene analysis (ASA) [13][14][15]. The components of these sounds are overlapped and interleaved in frequency. ASA is complicated because the human ear can access the single pressure wave that summates all the sound waves in the environment (human breath, sound of lighting, people walking etc.). A unique process to evaluate every incoming voice signal creates a mental description for each source [16]. These processes are based on the sound's incoming signals, which is the summation of all the other signals in the environment. ASA consists of the top-down and bottom-up methods. The bottom-Up method is operated only on basic cues which are present in the received signal. They are mandatory and automatic, which means they mostly don't depend on attention, which is very different from the top-down method, which focuses on the listener's expectations and experience, thus involving a high range of cognitive processes [13]. Most of the Bottom-up process for ASA is operated at low levels of the auditory system. Forward suppression and spectral filtering are crucial for the bottom-up process in ASA. On the other side, the top-down mechanism is used to work on the output based on the bottom-up mechanism, which occurs at the lower levels of sensory organs.

It is still unknown that how humans apparent cocktail party problem. The results are evidence that the amplitude of the speech envelope can be decoded out of EEG recordings. The stimulus reconstruction technique is used with EEG and MEG to analyze the continuous speech's neurophysiology aspects [8]. As EEG and MEG are non-invasive techniques, it is easy to implement the real-life scenario for any hearing-impaired person. EEG and MEG data is used to decode the attentional selection in any given environment where many speakers exist [17]. Temporal Response Function (TRF) is responsible for correlating the characteristics of the input speech signal, and its stimulus in the cortical response recorded respectively. Therefore, by existing a correlation for the data collected, we can conclude which TRF's can be widely used for the extraction of the required signal that is needed for the people with the impairment. The required TRF's can be used in the software connected to the hardware material, which can be developed for helping people with impairment. The methods like evoked response potential (ERP), when an audio stimulus was provided to the subject, only worked when the stimulus was short and had to be repetitive. Recently, newer studies were able to use continuous stimuli such as speech signals by employing the TRFs. However, these speech signals needed to have slowly varying envelopes to get accurate results. The TRF approach requires both audio stimulus and the corresponding EEG data to establish a linear relationship between the two. There are different methods to find TRFs, some of which have been mainly focused since they are relatively easy for implementation with good accuracies.

The concept of TRFs evolved due to the shortcomings of ERP as these required short stimuli in repeated intervals [17]. For continuous stimuli, a new approach was needed, and hence TRFs served the purpose. TRF is a linear stimulus-response model that provides a linear relationship between the provided input signal, which is speech, and the output signals, i.e., cortical response. TRF is used to predict the cortical response from the speech envelope, which is termed a forward model. Similarly, the equations can be altered so that the speech is predicted from the available cortical response, which is called the backward model. The backward model involves fewer complex calculations to find the TRF coefficients and is relatively easier to implement when compared to the forward model. Forward models are also called generative or encoding models as they define how the system generates or encodes information.

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

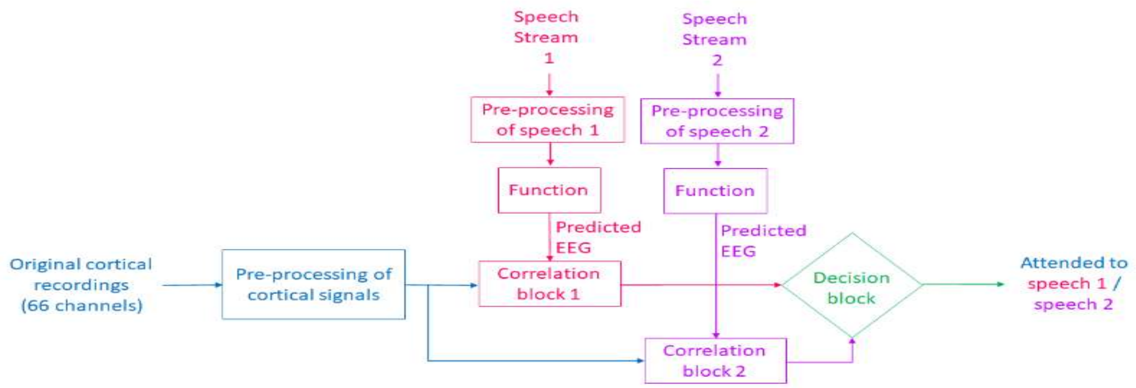

Fig. 1. Forward Model Block Diagram

Fig. 1 shows the forward model of TRF where EEG is predicted from the stimulus. The two speech streams are given to the pre-processing units, where the raw speech streams are converted to the required format. Pre-processing unit takes care of missing and noisy data in the input, if any. It converts the raw data into how the forward model of TRF needs the input data. After the pre-processed signals are passed, we may get the predicted EEG as output.

On the other hand, original cortical recordings of the experiment are applied to the pre-processing unit to process the noisy parts of the data. The output from the pre-processing of cortical recordings includes 66 EEG signals. The predicted EEG out of the TRF and pre-processed cortical signals are applied to the correlation blocks separately. For each speech stream, 66 correlation coefficients are produced out of the correlation block. These 66 correlation coefficients are the features to decide which speech stream listener has attended to.

The backward model of TRF is shown in fig. 2, where the envelope of the audio is approximated from the EEG. After pre-processing of cortical signals, the result is applied to TRF, which produces the predicted audio. From each speech stream, a correlation coefficient is produced from each correlation block. The decision block decides the attended speech stream by comparing the correlation coefficients, whichever the more significant is the attended.

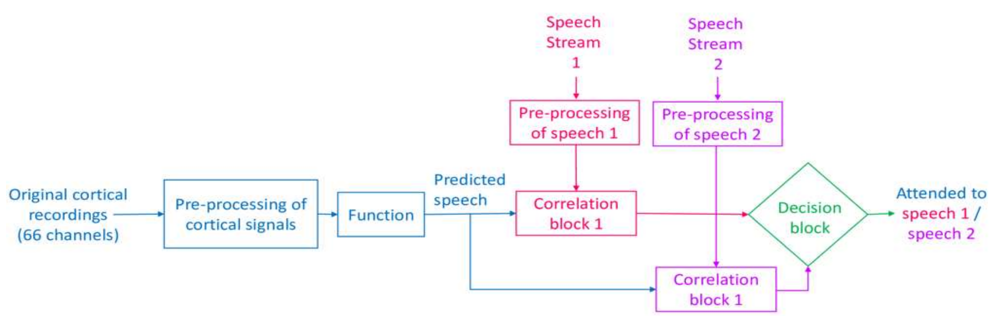

Fig. 2. Backward Model Block Diagram

3 EXPERIMENT AND DATA PROCESSING

The experiment consists of audio or speech stimuli recorded while a male and a female speaker read a fictional story. The sampling rate for the audio was 48KHz. Stories are divided into a total of 65 segments, each with 50-second-long samples. These stories (speech stimuli) are played via earphones, and the trials were randomized to represent a practical scenario best. The experiment was conducted in an electrically shielded room. When the subjects are listening to the speech stimulus, their cortical responses (EEG) are recorded. The Biosemi Active Two device with 64 channels and a 512 Hz sampling rate was used. Its high-resolution ADC meant that the sampling process gave precise values with minimal loss of information. 50 Hz line noise was eliminated from the EEG data and is passed through a 0.1Hz cut-off

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

frequency high pass filter. This filtered EEG data, and all the audio recordings were made available online publicly as a standardized dataset. By using further processing, the audio and EEG data were stored into a data structure which made it easy to access the information. For ease of processing and storage, both the speech and EEG signals were down sampled to 128 HZ.

The $\mathbf{50}\boldsymbol{Hz}$ line noise and harmonics we filtered are by applying convolution with $\frac{\mathbf{512}}{\mathbf{50}}$ sample square window of the EEG data. By using the FieldTrip EEG processing toolbox, the artifacts are removed

accordingly. From the EEG data, an artifact covariance matrix was computed at the samples $\boldsymbol{a\epsilon A}$ to solve the eigenvalues problem. So that the resultant eigenvalues, sorted by eigenvalues, explain the artifact and data covariance matrix maximum difference invariance. For further analysis, EEG is $\mathbf{1-9}\boldsymbol{Hz}$ band passed from a windowed synchronous linear finite impulse response filter. The window is shifted to produce a zero phase by its group delay. The estimation techniques to decode attention strive to draw a relationship between the features of the speech data and the cortical response. From the input speech data, we calculated the temporal envelope without reverberation. To glean the envelope rendition, gammatone filter bank is used and processes the attended and unattended speech streams. The audio envelope data was eventually down sampled using FFT based resampling technique, the sampling frequency of EEG.

4 MATHEMATICAL MODELING OF THE CORRELATION FUNCTION

Mathematically, there exist several ways that output is dependent on its input to any system. The coefficient of correlation which an important parameter in connecting the input and output of the experiment can be written as

$$\boldsymbol{r} = \frac{\sum \mathbf{PQ}}{\sqrt{\sum \mathbf{P^2Q^2}}}$$

where $\boldsymbol{P}$- deviation of $\boldsymbol{p}$, Q - deviation of $\boldsymbol{q}$ and

$$\boldsymbol{P} = (\boldsymbol{p}-\overline{\boldsymbol{P}}), \boldsymbol{Q} = (\boldsymbol{q}-\overline{\boldsymbol{Q}})$$

where $\overline{\boldsymbol{P}}$, $\overline{\boldsymbol{Q}}$ are means of the series p, q

When deviations are taken from an assumed mean

$$\boldsymbol{r} = \frac{\sum \boldsymbol{PQ} - \frac{\sum \boldsymbol{P} \sum \boldsymbol{Q}}{\boldsymbol{N}}}{\sqrt{\left[\sum \boldsymbol{P}^2 - \frac{(\sum \boldsymbol{P})^2}{\boldsymbol{N}}\right]\left[\sum \boldsymbol{Q}^2 - \frac{(\sum \boldsymbol{Q})^2}{\boldsymbol{N}}\right]}}$$

where,

N = no. of terms

$\sum \boldsymbol{PQ}$- product of derivations of p and q from assumed mean

$\sum \boldsymbol{P^2}$ - total of the squares of the deviations of 'p' from assumed mean

$\sum \boldsymbol{Q^2}$- total of the squares of the deviations of 'q' from assumed mean

$\sum \boldsymbol{P}$ - total of the deviations of 'p' from assumed mean

$\sum \boldsymbol{Q}$ - total of the deviations of 'q' from assumed mean

When the number of observations is very large the data is classified into a two-way frequency distribution. Then correlation of coefficient is

$$\boldsymbol{r} = \frac{\sum \boldsymbol{fPQ} - \frac{\sum \boldsymbol{fP} \sum \boldsymbol{fQ}}{\boldsymbol{N}}}{\sqrt{\left[\sum \boldsymbol{fP}^2 - \frac{(\sum \boldsymbol{fP})^2}{\boldsymbol{N}}\right]\left[\sum \boldsymbol{fQ}^2 - \frac{(\sum \boldsymbol{fQ})^2}{\boldsymbol{N}}\right]}}$$

where f - frequency, P, Q - deviated values

The correlation coefficient is a measure of the degree of covariability between 2 variables. At the same time, the regression establishes a functional relation between dependent and independent variables so that the former can be predicted for a given value of the latter. In correlation, both p and q are random variables, while in regression, p is a random variable, and q is a fixed variable.

The general format of the regression equation of Q on P can be represented as

$$\sum \boldsymbol{Q} = \boldsymbol{Na} + \boldsymbol{b} \sum \boldsymbol{P}$$

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\sum PQ = a\sum P + b\sum P^2$$

The derivations of P and Q from their means can be represented as

$$Q - \bar{Q} = r\frac{\sigma q}{\sigma p}(P - \bar{P})$$

where $\bar{P}$ - mean of p, $\bar{Q}$ - mean of q

The regression coefficient of Q on P is

$$b_{QP} = \frac{\sum PQ}{\sum P^2} = r\frac{\sigma q}{\sigma p}$$

$$r^2 = b_{PQ}.b_{QP}$$

If we take deviations from the assumed mean, the relation can be represented as,

$$Q - \bar{Q} = r\frac{\sigma q}{\sigma p}(P - \bar{P})$$

$$r\frac{\sigma q}{\sigma x} = \frac{\sum dpdq - \frac{\sum dp \sum dq}{N}}{\sum dq^2 - \frac{(\sum dq)^2}{N}}$$

Considering the above correlation and regression relations in forward models, input, p, is speech and output, q, is EEG and for backward model, it is the receiver.

It can be represented in the form of an equation as

$$\hat{Q} = PW$$

where $\hat{Q}$ – model prediction of time dimension 't' vector

P – model input matrix with the channel dimension 'c' and dimension 't'

W – linear TRF model parameter

$$P = [p_{(t,f),c}]$$

and $Q = [q_t]$

As discussed, backward models outperform forward models, and our interest here is working on backward model.

$$W = (P^T P)^{-1} P^T Q$$

Hence the input applied to the model is 66 channel EEG data, and output is predicted speech.

Filter coefficients can be estimated by the existing regression techniques [18]

$$W = (P^T P)^{-1} P^T Q \quad (1)$$

where $P^T P$ – estimated covariance matrix

$P^T Q$ – estimated cross-covariance matrix

By following this technique, no additional hyperparameters need optimization.

Therefore, (1) can be expressed as

$$\sum_{j=1}^{a} p_{(t,f)} W_j = Q_i \quad (2)$$

where $i = 1, 2, 3, \ldots\ldots n$

(2) can be written in the form of a matrix as

$$Q = PW \quad (3)$$

where $P = \begin{bmatrix} p_{11} & p_{12}\ldots\ldots & p_{1p} \\ p_{21} & p_{22}\ldots\ldots & p_{2p} \\ \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & p_{np} \end{bmatrix}$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

$$Q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix}$$

The (2) is applicable for an overdetermined system. For such a system, no exact solution can be determined.

Method 1[19]:

In linear regression, an $nx\,1$ column vector q is projected onto the column space of the nxa design matrix P, whose columns are highly correlated. The estimator coefficient in the earlier model $W \in R$ by which the columns are multiplied to get the orthogonal projection PW is

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$W = (P^T P + \lambda I)^{-1} P^T Q \tag{4}$$

where I is an identity matrix of size a x a, λ is regularization parameter.

Method 2:

(4) can be reframed as

$$W = \left((1-\lambda)P^T P + \lambda b I\right)^{-1} P^T Q \tag{5}$$

where b −average eigenvalue trace of the covariance matrix.

when regularization parameter, λ becomes zero, (5) becomes (1). Similarly, when $\lambda = 1$, the covariance estimator becomes diagonal. Therefore, this method penalizes extreme eigenvalues more smoothly.

Proposed method:

In continuation to the earlier relations between P, Q and W

Assume

$$\frac{\partial w_i}{\partial i} \approx (w_{i+1} - w_i) \tag{6}$$

where $w_i$and w_{i+1} are neighboring filter pairs.

The new function of the filter can be approximated as

$$W = [(P^T P + \lambda C)^{-1} P^T Q] \tag{7}$$

$$C = \begin{bmatrix} 1 & -1 & 0 & .... & 0 \\ -1 & 2 & -1 & .... & 0 \\ 0 & -1 & 2 & .... & 0 \\ . & . & . & ...... & . \\ . & . & . & ...... & . \\ 0 & 0 & 0 & ...... & -1 \\ 0 & 0 & 0 & ...... & 1 \end{bmatrix} \tag{8}$$

In this model, the adjacent columns of P have a strong correlation when P includes the shifts. The filter endpoints may be affected due to the channels of their neighbours.

5 RESULTS

We considered the correlation coefficient to substantially differentiate in a more analogizing approach to measure the classification performance. The behavior of different relations taken from the existing methods are compared with the proposed model. This lets us understand a behavioral, experimental, and quantitative perspective how it can impact future research and the state-of-the-art models. All the models, including the proposed (except method 1), have a regularization parameter, λ, that plays a formidable and consequential role in affecting the overall accuracy in correlating and predicting the produced audio attended with the original speech streams. The regularization parameter, λ, is varied with five different values for all the methods i.e., $\mathbf{0.1, 0.25, 0.5, 0.75, 1.0}$. Different methods discussed above are applied with the regularization parameter, λ, with the said values for comparison.

Table 1. Comparison Of Regularization Parameter And Obtained Correlation Coefficients For The Existing Methods And Proposed Method

Regularization parameter, λ	**Obtained correlation coefficients**		
	Method 1	**Method 2**	**Proposed**
0.1	0.5911	0.5589	0.5589
0.25	0.5917	0.5594	0.5915
0.5	0.5915	0.5608	0.5912
0.75	0.5914	0.5673	0.6121
1	0.5913	0.5869	0.5909

In table. 1, the obtained correlation coefficient values for different regularization parameters are applied to the methods discussed. The variation between the obtained correlation coefficients fluctuates asymmetrically when the regularization parameter is changed. From the observed values, we cannot conclude how speech can correlate to the audio envelope predicted from the change in regularized parameters. But the proposed model gave a better correlation value at λ=0.75, whereas method 1 gave a better correlation value at 0.25

Fig. 3 shows the area under the curve of the correlation coefficient for the transition of regularized parameters from 0.1 to 1.0. The area under the curve in fig. 3 gives the average value of sensitivity (or specificity) for all possible values of specificity (or sensitivity)[20], i.e., the three-dimensional variation of change of correlation coefficient with the change of regularization parameter can observe.

Journal of Theoretical and Applied Information Technology
31st March 2024. Vol.102. No 6
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

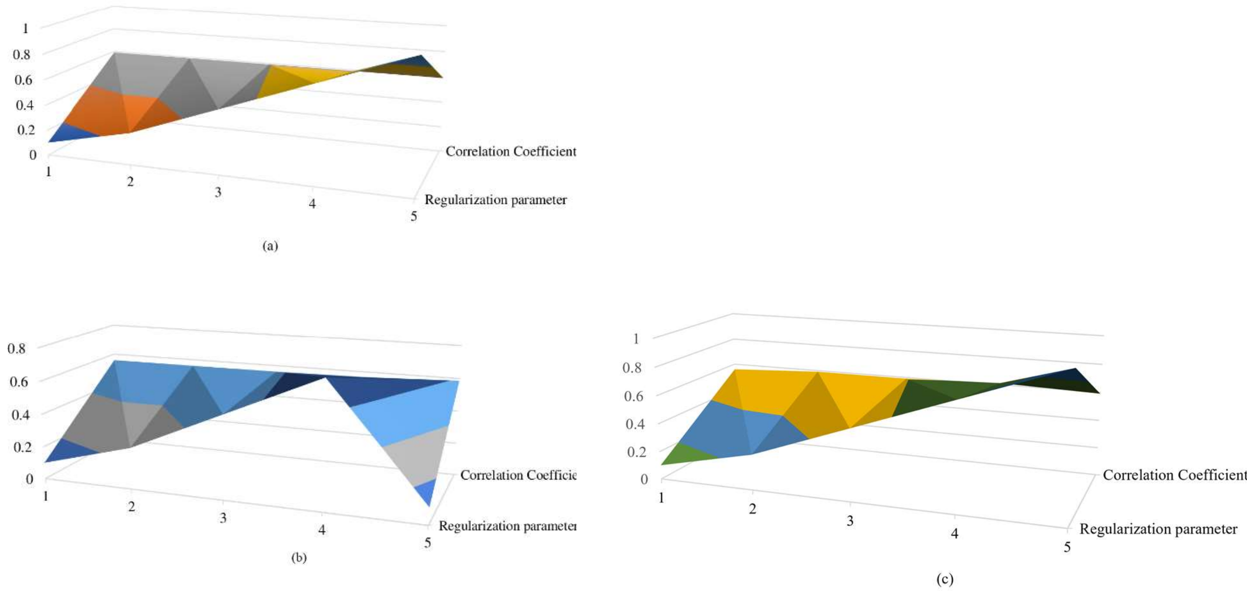

Fig. 3. Area Under Curve Of Correlation Coefficient For Transition Of Regularized Parameters From 0.1 To 1.0

In fig. 4, the plot shows the variance of aggregate values of correlation coefficient with regularization parameter for $\boldsymbol{\lambda} = \mathbf{0.1, 0.25, 0.5, 0.75, 1.0}$. The variation of the proposed model is also functioning similarly because the dependencies on the regularization parameter are multiplied with the identity matrix of size $\boldsymbol{aXa}$ to fit the model to correlate the audio envelope from TRF with the speech stream. A sharper transition can be observed in fig. 3 (c), which can give the possible correlation between predicted and attended signals.

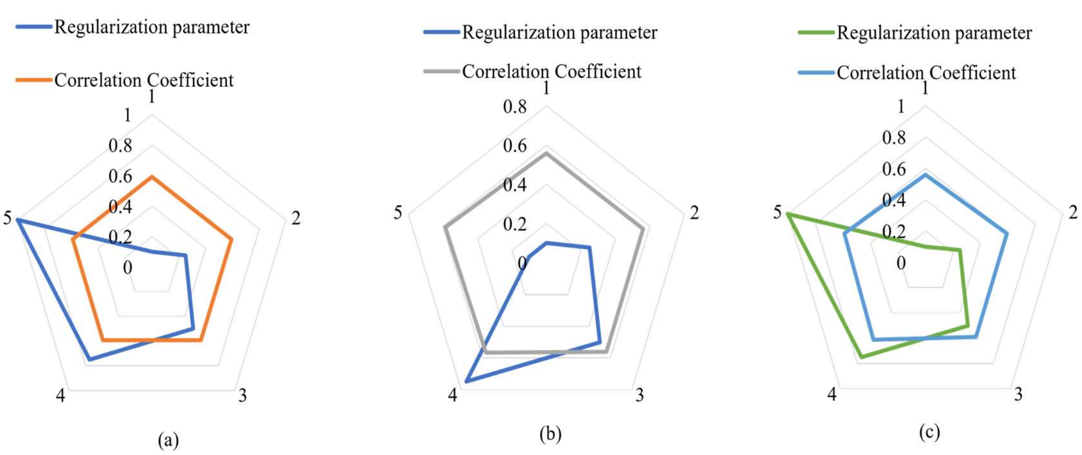

Fig. 4. Variance of the aggregate values of correlation coefficient with regularization parameter for $\lambda = 0.1, 0.25, 0.5, 0.75, 1.0$

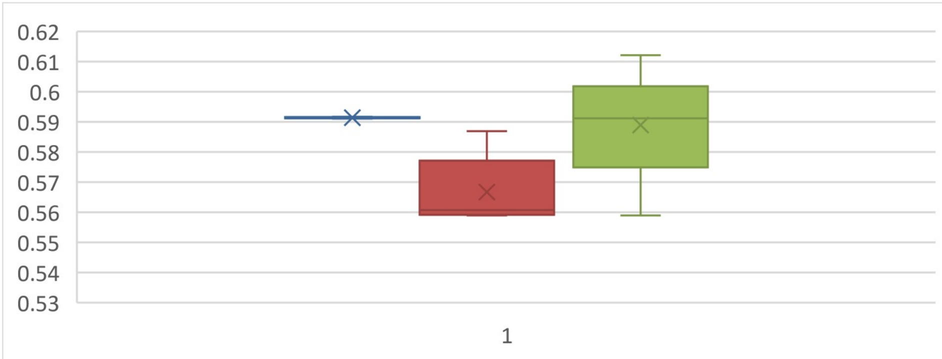

Fig. 5. Population Observation From Table. 1

between methods 1,2, and proposed

We can observe from fig. 5, that, in all the methods including proposed, more commonly the correlation coefficient is ranging between 0.55 to 0.59 according to the regularization parameter taken from 0.1 to 1.0. But the proposed model shows that there is a huge rise in prediction accuracy of the model to correlate the envelope of the audio to speech stream. Fig. 6 shows the regularization parameter wise comparison of different models for the obtained correlation coefficients.

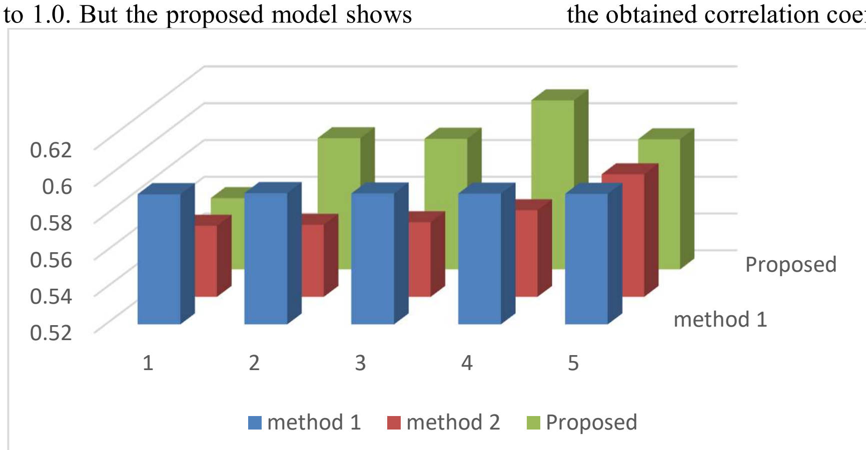

Fig. 6 Regularization Parameter Wise Comparison Of Different Models For The Obtained Correlation Coefficients

6 CONCLUSION

The central ground to study and explore the possibilities of auditory attention detection is to build up a better hearing aid technology that allows hearing-impaired people to retrieve the normal hearing, at least to some part. In a multi-speaker scenario, the current methods' performance is down because hearing aids indistinctly amplify all the speakers. To overcome this hindrance, there is a strong need to inform hearing aids, i.e., the hearing aid should auto detect the user attention and attenuate all other sounds than what they are attending. So that, the signal processing or machine learning techniques can help to improve the enhancement of the speakers attending. The proposed model can achieve a correlation accuracy of about 62%. Overall, if real-time hints of an individual's attentional state are provided, there is a chance for better hearing aid technology to assist older people or any hearing-impaired listener. The current pace of research in this field may deliver a more productive system to perform better in real-time scenarios by overcoming the limitations in decoding, which can serve other domain applications like education, health, BCI games.

REFERENCES

[1] D. E. Broadbent, *Perception and communication*. London: Pergamon Press, 1958.

[2] . A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cogn Psychol*, vol. 12, no. 1, pp. 97–136, Jan. 1980, doi: 10.1016/0010-0285(80)90005-5.

[3] J. A. Deutsch and D. Deutsch, "Attention: Some Theoretical Considerations," *Psychol Rev*, vol. 70, no. 1, pp. 80–90, Jan. 1963, doi: 10.1037/H0039515.

[4] C. EC, "Some experiments on the recognition of speech, with one and with two ears," *J Acoust Soc Am*, 1953.

[5] B. Rakerd, P. F. Seitz, and M. Whearty, "Assessing the cognitive demands of speech listening for people with hearing losses.," *Ear Hear*, vol. 17, no. 2, pp. 97–106, Apr. 1996, doi: 10.1097/00003446-199604000-00002.

[6] attention shifts to one's name in an irrelevant auditory channel?," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 21, no. 1. American Psychological Association, US, pp. 255–260, 1995. doi: 10.1037/0278-7393.21.1.255.

[7] M. K. Pichora-Fuller and P. E. Souza, "Effects of aging on auditory processing of speech.," *Int J Audiol*, vol. 42 Suppl 2, pp. 2S11-6, Jul. 2003.

[8] J. R. Kerlin, A. J. Shahin, and L. M. Miller, "Attentional gain control of ongoing cortical speech representations in a 'cocktail party,'" *J Neurosci*, vol. 30, no. 2, pp. 620–628, Jan. 2010, doi: 10.1523/JNEUROSCI.3631-09.2010.

[9] R. M. Warren, J. A. Bashford, E. W. Healy, and B. S. Brubaker, "Auditory induction: reciprocal changes in alternating sounds," *Percept Psychophys*, vol. 55, no. 3, pp. 313–322, May 1994, doi: 10.3758/BF03207602.

[10] . A. G. Samuel, "Knowing a word affects the fundamental perception of the sounds within it," *Psychol Sci*, vol. 12, no. 4, pp. 348–351, Nov. 2001, doi: 10.1111/1467-9280.00364.

[11] T. L. Arbogast, C. R. Mason, and G. Kidd, "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," *J Acoust Soc Am*, vol. 117, no. 4 Pt 1, pp. 2169–2180, Apr. 2005, doi: 10.1121/1.1861598.

[12] T. Baer, B. C. Moore, and S. Gatehouse, "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times.," *J Rehabil Res Dev*, vol. 30, no. 1, pp. 49–72, 1993.

[13] A. S. Bregman, "Auditory Scene Analysis: The Perceptual Organization of Sound," 1990, doi: 10.7551/MITPRESS/1486.001.0001.

[14] F. B. R. Parmentier, "The cognitive determinants of behavioral distraction by deviant auditory stimuli: a review," *Psychological Research 2013 78:3*, vol. 78, no. 3, pp. 321–338, Dec. 2013, doi: 10.1007/S00426-013-0534-4.

[15] F. Pulvermüller and Y. Shtyrov, "Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes," *Prog Neurobiol*, vol. 79, no. 1, pp. 49–71, May 2006, doi: 10.1016/J.PNEUROBIO.2006.04.004.

[16] C. Alain and K. Tremblay, "The role of event-related brain potentials in assessing central auditory processing," *J Am Acad Audiol*, vol. 18, no. 7, pp. 573–589, Jul. 2007, doi: 10.3766/JAAA.18.7.5.

[17] M. R. Harter, S. L. Miller, N. J. Price, M. E. LaLonde, and A. L. Keyes, "Neural processes involved in directing attention," *J Cogn Neurosci*, vol. 1, no. 3, pp. 223–237, 1989, doi: 10.1162/JOCN.1989.1.3.223.

[18] B. Zdaniuk, "Ordinary Least-Squares (OLS) Model," *Encyclopedia of Quality of Life and Well-Being Research*, pp. 4515–4517, 2014, doi: 10.1007/978-94-007-0753-5_2008.

[19] B. Zdaniuk, "Ordinary Least-Squares (OLS) Model," *Encyclopedia of Quality of Life and Well-Being Research*, pp. 4515–4517, 2014, doi: 10.1007/978-94-007-0753-5_2008.

[20] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.," *Biometrics*, vol. 44, no. 3, pp. 837–845, Sep. 1988.