Received July 13, 2019, accepted July 31, 2019, date of publication August 5, 2019, date of current version August 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933049

A New Paradigm for Region-Based P300 Speller in Brain Computer Interface

ZEKI ORALHAN
Department of Electrical Electronics Engineering, Nuh Naci Yazgan University, 38090 Kayseri, Turkey
(e-mail: zoralhan@nny.edu.tr)

ABSTRACT Electroencephalography-based brain computer interface systems could provide alternative communication methods for severely disabled people who cannot use their neuromuscular systems. The P300 signal is one of the event related potentials that are used for brain computer interface systems. The most important performance parameter of a P300 based brain computer interface system is information transfer rate that is calculated by using classification accuracy and P300 signal detection time. Moreover, P300 speller has a very critical role for classification accuracy and information transfer rate in a P300 based brain computer interface. Although most of studies are about row column based P300 speller in literature, region based P300 speller proved that has higher classification accuracy than row column based one. There are very few studies about region based P300 speller. This study aims to investigate methods for obtaining higher classification accuracy and information transfer rate with using region based P300 speller that constituted audio and visual stimulus. This is the first research that using audio and visual stimulus for a region based P300 speller in literature. Previous studies about region based P300 spellers focused on spellers with only visual stimulus types. Our new paradigm presents region based P300 spellers with only audio, only visual, and audio-visual stimuli. Audio-visual P300 speller structure is the newest model for region based spellers. The subject focused on the desired character stimulus. We used the stepwise linear discriminant analysis method for classification that either included the desired P300 signal or not. According to stepwise linear discriminant analysis, the mean classification accuracy value of the experiment was 90.31% with the audio-visual region based P300 speller. With this new paradigm, classification accuracy in the audio-visual P300 speller was improved 15.69% and 66,99% according to the visual only and audio only P300 speller that we used in the experiments, respectively.

INDEX TERMS Brain computer interface; P300; Human machine systems; P300 Speller.

I. INTRODUCTION

Brain computer interface (BCI) systems allow users to send commands or messages to an electronic system or computer without using their peripheral motor systems [1], [2]. BCIs are operated using cerebral activities. There are various BCIs in relation to the measurement methods of cerebral activities. Brain activity may be analyzed based on the functional magnetic resonance (FMRI), near infrared spectroscopy (NIRS), magnetoencephalography (MEG) and electroencephalography (EEG) measurement methods [3], [4]. The most commonly used method for BCIs is EEG, which is based on the measurement of electrical activity of the neurons in the brain with electrodes from the scalp. EEGs are preferred in BCI systems due to their high signal responses, which allow non-invasive measurements of real-time data to be processed [5]. Moreover, EEG devices present portability opportunities for users [7].

In EEG-based BCI systems steady state visual evoked potentials, P300 signals [6]–[18], motor neuron signals [8], slow cortical potential activities [9] are used. One of the most commonly used method is using P300 signals to give commands. P300-based BCI systems may be preferred because they are practical systems, have faster response and lack cognitive fatigue [10]. P300 is considered as an expectation signal. Sutton et al. discovered P300 waves in 1965 [11]. P300 is an event-related potential, so it may be considered as an endogenous signal. When the expected stimulus occurs,

The associate editor coordinating the review of this manuscript and approving it for publication was Samu Taulu.

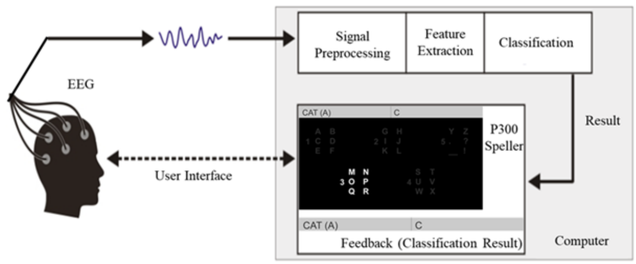

FIGURE 1. **Schematic diagram of P300 based BCI system. The system contains EEG signal acquisition system, signal preprocessing, feature extraction and signal classification, and a BCI is able to transfer the commands to external devices, providing feedback (classification result) to user.**

an internal response will fluctuate in the EEG signal positively and then usually at 250ms – 450ms after this event [12]. Since the peak point of the signal is around 300ms, it is called P300 or P3 that can be observed clearly in event-related potentials [37]. The value of these delay duration or latency depends on the rate of classification in the transition from one event to another [21]. Shortening latency results in better mental performance. Various paradigms are used to generate the P300 signal, the most common being the Oddball paradigm. All stimuli are presented equally frequently. However, because there are many stimuli to loop through, the attended stimulus is typically ''off'' or unlit. The rare instances it is on/lit make it an oddball (statistically speaking) that elicits a response by the subject. The response of the brain to the expected stimulus is the P300 signal. This expectation signal is more frequent in the parietal lobe [13]–[17].

The diagram in Figure 1 shows the structure of P300-based BCI systems. There are four significant steps: signal acquisition, preprocessing, feature extraction, and classification in a P300 based BCI system [38]. EEG signals are acquired on the scalp by electrodes. The acquired EEG data is firstly preprocessed for noise reduction and artifact removal. In the feature extraction step, important information that hidden in signal is extracted. Also, in this step dimensionality is reduced for classification [39]. Then desired signal is classified by classifier that was trained before. In a training part, the BCI user is asked to desire the specific targets. So, BCI system collects training dataset according to the individual BCI user. After signal processing, BCI sent the commands to the external devices and user checks classification result or feedback via user interface.

A real time P300 based BCI system is designed by Farwell and Donchin with using P300 speller that was implemented for the first time in 1988 [6]. They proposed to 6 × 6 matrix of visual characters as a P300 speller. Each row and column flashes randomly. The BCI user gazes to one of the 36 visual characters and counts the number of times the gazed character is flashed. Flashing the 6 rows and 6 columns creates an oddball paradigm with the matrix that contain the gazed character rare instances [40]. Simply, the target visual character is chosen by according to the P300 potential is elicited in which row and column. Thulasidas et al. aimed to optimize BCI usability and BCI performance with P300 speller. They researched some variations of BCI system such as number of visual repetitions, and number of EEG channels. They showed that the classification accuracy saturates when 10 EEG channels were used. BCI P300 speller usability improved by using less repetition [41]. Sellers et al. compared matrix size of P300 speller. They compared 6 × 6 matrix P300 speller and 3 × 3 matrix P300 speller in their study. They observed higher P300 amplitude with using 6 × 6 matrix P300 speller. Also, they researched about inter stimulus interval time (175 ms and 350 ms) between P300 speller's character flashes. They reached higher accuracy using 3 × 3 matrix with 175 ms inter stimulus interval [19]. There are also studies about derivative row-column based P300 speller. Salvaris and Sepulveda changed some properties of Farwell and Donchin row-column based P300 speller. Color of visual character and screen background, character font and size are changed in their research for investigating higher accuracy. They observed that the highest classification accuracy was in the P300 speller with white visual character with black background. In the contrary of this the lowest classification accuracy was with small font size of visual characters [20]. Fazel Rezai and Abhari proposed new regional P300 speller for eliminating human perceptual errors in the row column based P300 speller. Moreover, they aimed to reach higher classification accuracy than row column based P300 speller. They used 7 visual regions in their proposed P300 speller. Each region contained 7 visual characters. This proposed P300 speller had 2 stages for detecting target character. In the first stage, BCI user gazed to region which contains target character. When the region was detected, second stage screen displayed on the user interface screen. In this stage, detected region's 7 characters were shown in different area on screen. According to oddball paradigm, character that was gazed by BCI user detected. Their research showed that region based P300 speller reached higher classification accuracy than row column based P300 speller, also, the error rate decreased 18.4% with the proposed P300 speller [22]. Ikegami et al. used 2 stages region based P300 speller and row column based P300 speller in their experiments with 7 amyotrophic lateral sclerosis (ALS) patients. They aimed to reach higher classification accuracy with region based P300 speller that they designed than row column based P300 speller. As a result of their study, they observed that experiments with region based P300 speller reached significantly higer classification accuracy than row column based P300 speller [23]. Pan et al. compared the performance of single character based P300 speller and region based P300 speller with 12 subjects. It was found that region based P300 speller had higher P300 amplitude and accuracy [24].

Additionally, ALS patients are not able to move their eyes horizontally in the locked in syndrome (LIS). Not only ALS patients LIS, also patients with various neurodegenerative syndromes has lack of eye movement ability [26], [27].

For this reason, BCI systems can design auditory stimuli based P300 speller [13], [28]. In 2009, Kübler et al. designed row column based auditory P300 speller in the form of 5 × 5 matrix for the LIS patients. Each row and column assigned with numbers from one to ten which were called randomly by a male voice. LIS patient focused to hear number that written on target's row and column. They found that classification accuracy was significantly higher with a similar visual P300 speller as compared with the proposed auditory P300 speller [29]. In 2011 Belitski et al. designed row column based audio-visual P300 speller. They used 6 × 6 matrix form also they carried out experiment with transposed matrix form too. Their row column based audio-visual P300 speller achieved highest performance than only audio and only visual based P300 speller [30]. In 2018, Zhao et al. contributed knowledge with their research row column based audio-visual based P300 speller. They also showed that better performance was reached with their hybrid P300 speller [42]. Qu et al. designed three dimensional audio based P300 speller. They combined audio-visual stimulus in their P300 speller. They reached higher accuracy with their proposed speller according to two dimensional P300 speller [43].

There are a lot researches about P300 speller from Farwell Donchin paradigm that was proposed until now. Although region based P300 speller was reached better performance than row column based P300 speller, most of studies were about row column based P300 speller and its derivatives [22]–[24]. Only audio, only visual, and audio-visual based row column P300 spellers were investigated with many studies by researchers from the 1990s to nowadays [42], [44], [45].

There is not adequate research about region based P300 speller as well there is lack of region based audio-visual P300 speller research. This situation motivated us to research about region based P300 speller with different stimuli types. Therefore, we presented a new region based P300 speller that combined auditory and visual stimuli and showed that the audio-visual region based P300 speller outperforms the audio only and visual only region based P300 speller. This is first study that combines auditory and visual stimuli for region based P300 speller.

II. MATERIALS AND METHODS

This study proposed a new region based P300 speller with audio-visual stimuli. Audio only, visual only, and audio-visual (hybrid) stimuli were presented to the participants for investigating better region based P300 speller performance.

A. PARTICIPANTS

Experiments were carried out with 7 healthy participants which were males. Participants' mean age was 27 ± 3. Moreover, the participants had normal hearing and vision based on self-report, and had not experienced with any psychiatric disorder in their life.

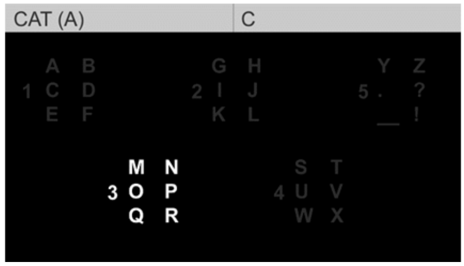

FIGURE 2. First stage of region based P300 speller with using audio-visual stimuli together.

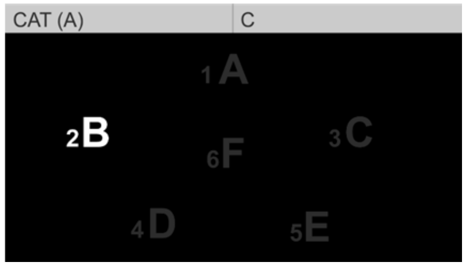

FIGURE 3. Second stage of region based P300 speller with using audio-visual stimuli together.

B. STIMULATION UNIT (P300 SPELLER STRUCTURE)

A region based P300 speller that was developed by Fazel-Rezai and Abhari, constituted of 49 characters in7 regions. Also, each region has 7 characters. They showed that their proposed region based P300 speller reached higher classification accuracy than Farwell and Donchin's proposed row column P300 speller. In this study, a new region based P300 speller was proposed according to research of regional based hypothesis. The main difference of our region based P300 speller that is consisting auditory and visual stimuli together for the first time in region based P300 spellers. Our P300 speller works in 3 modes. The first mode is based on visual stimulation, the second one is based on auditory stimulation, and third one is based on audio-visual stimulation as the hybrid P300 speller. Our P300 speller has 30 characters in 5 regions in the first stage as shown in Figure 2. Character selection is completed at 2 stages. At the first stage the region from where the BCI operator want to choose a character is detected. Than second stage starts. At the second stage, 6 characters in the detected region at first stage are displayed on the monitor as shown in Figure 3. Each region has a reference voice number in 2 modes which are visual only and audio-visual based P300 speller. This voice numbers are not visible in the mode of visual only based P300 speller. In the audio only based P300 speller, all region has white characters with black background that are displayed on the monitor. The reference numbers that are visible near regions were called out randomly by a male voice. Also, the screen is needed to see the content in each category in the auditory mode. In the visual only based P300 speller, each region intensifies from dark grey to white randomly without reference numbers.

Working process of audio only and visual only based P300 speller was combined for the mode of audio-visual

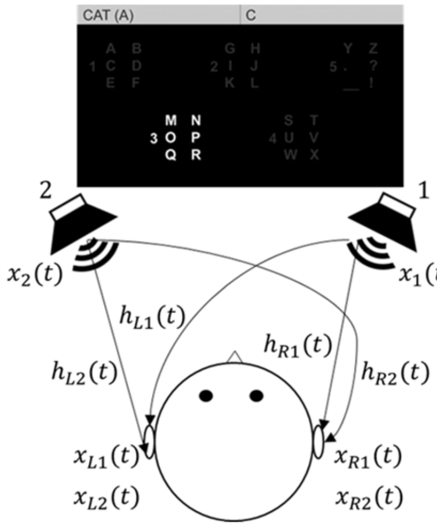

FIGURE 4. Second stage of region based P300 speller with using audio-visual stimuli together.

based P300 speller. It is simply, while a region intensifies, the region voice reference number is called randomly by a male voice. In the Audio only and audio-visual mode, 2 speakers were used for hearing the reference numbers. The speakers were arranged right and left side the monitor. Odd and even reference numbers were spoken through speaker 1 and speaker 2, respectively as shown in figure 4.

Figure 4 shows that the propagation of sound wave that from the speaker to BCI user ears. Head related transfer function (HRTF) is a transformation of this propagation. In the time domain of HRTF is head related impulse response. As shown in figure 4, $h_{R1,R2}(t)$ and $h_{L1,L2}(t)$ denotes the impulse responses of right and left ears, respectively. $x_{1,2}(t)$ denotes the pressure of the sound source and $x_{R1,R2}(t)$ and $x_{L1,L2}(t)$ denotes the pressure at the ears as shown in figure 4.

In time domain, pressure at the ears are convolutional of the sound source and the head related impulse response as written in equation 1.

$$x_{L,R}(t) = h_{L,R}(t)^{*}x(t) = \int_{-\infty}^{\infty} h_{L,R}(t-\tau)x(\tau)d\tau \qquad (1)$$

In the frequency domain of equation 1 can be written as in equation 2.

$$X_{L,R}(\omega) = F(h_{L,R}(t)^{*}x(t)) = H_{L,R}(\omega)X_{L,R}(\omega) \qquad (2)$$

HRTF can be obtained from equation 2. HRTF is give in equation 3.

$$H_{L,R}(\omega) = \frac{X_{L,R}(\omega)}{X(\omega)} \qquad (3)$$

Measurement of $H_{L,R}(\omega)$ vary depending on angle. Moreover, HTRF is dependent on distance that if center of head and sound source less than 1 meter. This situation is called near field HTRF that has difficult measurement since needs very sensitive equipment. In the experiments we have not measure this value. Our experiment set, head distance from sound source was less than 1 meter and as shown in figure 4, head distance and angle is same between the both speakers. Speaker 1 is more effective to right ear than left ear as speaker 2 is to left one. If the auditory stimuli are spatialized to match the spatial configuration on the screen, three dimensional sounds would be created from stereo speakers and could adjusted effective distance or angle for each BCI user. Thus, audio based P300 speller performance can be improved via using HTRF [46], [47].

In the experiments, tasks are seen on the upper left side of the proposed P300 speller screen. Selection characters are seen on the upper left side of the screen. Our proposed P300 speller was developed with the C# coding.

C. DATA ACQUISITION

In the experiments, CleveMed BioRadio device and BioCapture were used for EEG signals recording non-invasively. The device manufacturer is Great Lakes Neuro Technologies which is an American firm. Gold cup electrodes were placed on the scalp according to the international 10/20 system. The reference electrode was placed on the left earlobe. Experiments were carried out in a soundproof room. We used sampling rate as 500 Hz for EEG signals recording.

D. EXPERIMENTAL PROCEDURE

In the experiments we used our proposed region based P300 speller with 3 modes. We used a liquid crystal display (LCD) for showing user interface of the P300 speller. LCD had a resolution of 1366×768 pixels. Experiment subjects' head was 60 cm away from the LCD, horizontally. All subjects participated the experiments with using 3 modes of P300 spellers which audio only, visual only, and audio-visual region based P300 spellers. Each mode of P300 spellers was used in different days by an experiment subject.

While the experiment was operating by audio P300 speller, all regions were in white color, the auditory stimulus was heard for 275 ms, randomly for each region. After hearing the voice, there was the a 125 ms of silence time which was the time to wait between the two stimuli called the inter stimulus interval time. After silence time, the stimulus in the other region was active. We have chosen interval time according to the research of Sellers et al. that showed the difference effect of stimulus on and off time as we mentioned in introduction section [19]. A sequence was completed when each auditory stimulus was heard once randomly. The first stage of P300 speller consists 5 regions as we mentioned in figure 2. The first stage completion period of a sequence was 2 seconds as shown in Figure 5. A trial consists repeated 8 sequences for. We aimed to increase true P300/false P300 ratio with a trial. When the trial of the first stage was completed, screen was darkened for 3.8 seconds than second stage of P300 speller screen was seen as shown in figure 3. The second stage of P300 speller consisted 6 regions were characters that detected in the first stage after the trial was completed. The second stage completion period of a sequence was 2.4 seconds.

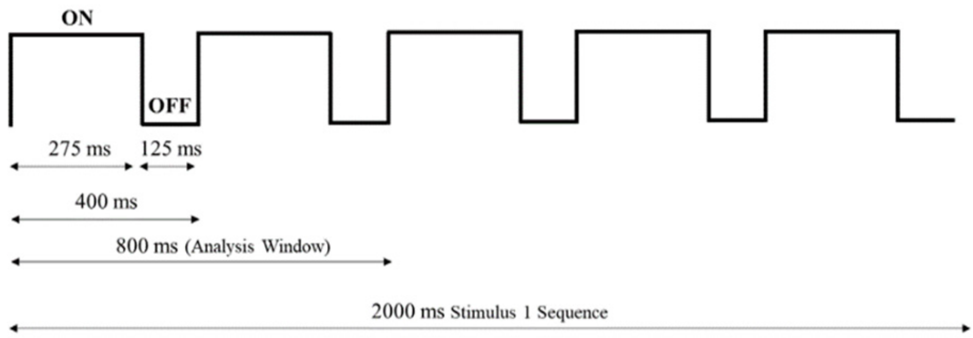

FIGURE 5. A timeline of one stimulus-presentation sequence. Audio only, visual only, and audio-visual stimuli lasted 400ms each; 275ms ON continued by a 125ms inter stimulus interval (OFF) time. Totally, stimuli were presented randomly, every 2500ms. For analysis, the response window used began with the stimulus ON and lasted for 800 ms.

Character was selected after a trial was completed in the second stage. So, a character selection period was 39 seconds.

In the experiments, we used three different modes (visual only, audio only, audio-visual) of P300 speller with the same stimulus duration, inter stimulus interval, and screen transition time. In the visual P300 speller mode, the auditory stimulus was not heard, and regional reference numbers were not visible on the screen, and each region flashed once randomly in a sequence, while the other four regions were dark gray. Again, 8 sequences were repeated. So, 1 trial was completed. In the audio-visual P300 speller mode, auditory stimuli and visual stimuli were presented synchronously with stimulation time and inter stimulus interval time.

Training and test sessions were held in three different modes of P300 speller in the experiments. For each mode of P300 speller, training sessions were carried out before the test session. The experiment subjects were desired to select some words in the test sessions. These words were "DOG", "FUN", "CAT", "BIP" and "SEX" which consists 15 different letters (characters). Each training session carried out 2 times. Thus, we had training dataset which contains 30 characters totally. After the training session, test session was carried out. The experiment subjects were desired to select 8 different words that are "DIP", "SUN", GOOD", "BUS", "NODE", "FOG", "NUT" and "EXACT". So, in the test session, we had dataset that constituted 28 characters.

E. DATA ANALYSIS

We used CleveMed BioRadio device with 7 EEG channels for data acquisition. These EEG channels were P7, P8, Pz, C3, C4, Cz, and Fz that has P300 signal dominantly [40]. EEG electrodes were placed on scalp according to 10/20 international system as shown in Figure 6.

As a first step, the acquired EEG data were pre-processed. In pre-processing step, we used the fifth order Butterworth band pass filter that has 0.1 and 35 Hz cut off frequencies. This filter was used for eliminating electrical network noise and passing EEG signals that has the dominant frequencies in brain. EEG signals are very sensitive for eye blinking, eye vertical and horizontal movements, and muscle movements

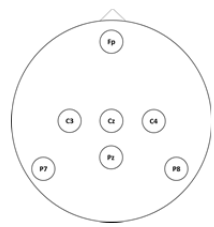

FIGURE 6. EEG channel layout that using 7 channels (C3, C4, Cz, Pz, P7, P8, and Fz) for analysis.

in any part of body. These artifacts have higher amplitude than desired EEG signals. We used winsorization process for removing these artifacts [48]. The 10% most extreme values in the EEG samples from each electrode were replaced value by the most extreme value from the remaining samples of that channel. Thus, effects of artifacts were reduced. We have calculated signal to noise ratio (SNR) value in dB unit before and after preprocessing according to the equation 4.

$$SNR = 10 \log \left(\frac{P_{signal}}{P_{noise}}\right) \tag{4}$$

We have observed to EEG data averagely +17 dB SNR value after the preprocessing step. Before the filtering SNR was at minus values. After the preprocessing, 800 ms that began at stimulus onset status as shown in Figure 5 [28]. For the analysis, data from each EEG channel were isolated by window. 400 signal points were obtained by isolated data with 800 ms window and using 500 Hz sampling frequency. We reduced signal points to 40-point dataset from 400-point dataset with using sub sampling. Each EEG channel data that in the same time period were added to endpoint of each other. Thus, concatenating of data segments process was created a single feature vector before classification step. We constituted training dataset for the classification process with using 280-point dataset that obtained with addition. We showed an example in Figure 7 for the 280-point dataset according to subject-4 response in training session with using audio-visual P300 speller. As shown in Figure 7, the averaged each training dataset from 7 different channel were added endpoint of each other.

TABLE 1. A trial record matrix of first stage of region based P300 speller.

1st Stage	1st Sequence	2nd Sequence	3rd Sequence	4th Sequence	5th Sequence	6th Sequence	7th Sequence	8th Sequence	Σ
1st Region	0	0	1	1	1	0	0	1	4
2nd Region	0	0	1	0	0	0	1	0	2
3rd Region	0	0	0	1	1	0	0	0	2
4th Region	0	0	0	0	0	0	1	0	1
5th Region	0	1	0	1	0	0	0	0	2

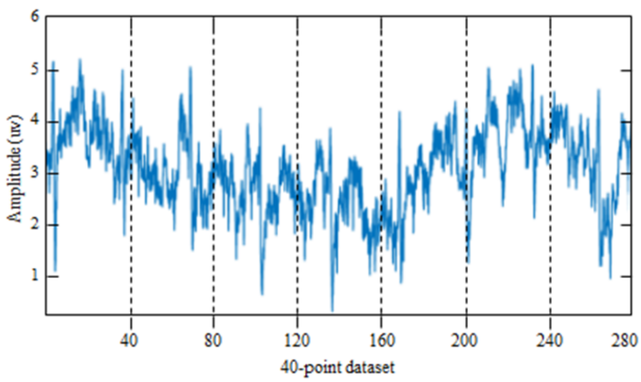

FIGURE 7. The averaged each training dataset from 7 different channel were added endpoint of each other.

We used the stepwise linear discriminant analysis (SWLDA) for the signal classification and weight generation. SWLDA is the extension form of the Fisher's linear discriminant analysis (FLDA) [49], [50]. SWLDA provides the higher signal classification accuracy in P300 speller based BCI applications [6], [50], [51].

Whether P300 signal is existed or not in the EEG signal can be defined as a binary classification problem that has a decision hyper plane defined in the equation 5.

$$w \cdot x - b = 0 \tag{5}$$

According to the equation 4, x denotes the feature vector that we mentioned before, w denotes the vector of feature weights, b denotes the bias term. Since we used region based P300 speller, P300 is elicited for one of the region intensification or voicing. According to the equation 6 and 7 as a classification result is taken as feature vector that maximum of the sum scored.

$$predicted\ 1st\ stage = \max_{1st\_stage} \left[\sum_{i_{1st\_stage}} w \cdot x_{i_{1st\_stage}} \right] \tag{6}$$

$$predicted\ 2nd\ stage = \max_{2nd\_stage} \left[\sum_{i_{2nd\_stage}} w \cdot x_{i_{2nd\_stage}} \right] \tag{7}$$

According to this design the predicted character is in sub-set of the predicted 1^{st} stage. Than the character is selected in the 2^{nd} stage as shown in equation 7. SWLDA is a method that selects the suitable discriminant function by adding specific channel and time domain amplitude information to the linear equation. Coefficients of a discriminant function is determined by using training dataset. In present research, the points were chosen from the 400-point dataset for classifying process. We used p-value for estimation point selection. This value was added to the discriminant function.

In initial, the discriminant function did not consist any features of signal. At each step, the feature which has the value of $p < 0.1$ was added to the discriminant function. Thus, it was evaluated as an input function. The feature which has the value of $p > 0.15$ was eliminated via backward stepwise analysis. This stepwise analysis was continued until no more features were existed to neither to be eliminated nor to be used as an input. After this process, the signal was classified whether consisting P300 signal or not.

By the way information transfer rate (ITR) shows that a BCI overall performance. As shown in equation 8, ITR can be calculated with using classification accuracy that is denoted by "P", total numbers of stimulus that is denoted by "N", and selection time that is denoted by "T".

$$ITR = \frac{60}{T} \left[log_2 N + P log_2 P + (1-P)\, log_2 \left[\frac{1-P}{N-1} \right] \right] \tag{8}$$

III. RESULTS

EEG data that acquired from all subjects were analyzed. In experiments, according to equation 6 and 7, classification result of each trial with 8 sequences were recorded to the matrix for detection of a character. The matrix recording was based on whether P300 signal was existed or not in the EEG signal. As shown in Table 1 and 2, we wrote the "1" to cell that intersection of sequence and region if the EEG signal included P300 wave. After a trial finished, we counted value of "1" row based in the matrix. A region that has a highest counted value was selected as a classification result. We showed a trial result of subject-4 as an example in Table 1 and 2 for first stage and second stage, respectively. According to the Table 1 that is related to 1st stage of P300 speller, as a result of a trial, the 1st region was detected with the highest counted value. The 1st region

TABLE 2. A trial record matrix of second stage of region based P300 speller.

2nd Stage	1st Sequence	2nd Sequence	3rd Sequence	4th Sequence	5th Sequence	6th Sequence	7th Sequence	8th Sequence	Σ
1st Region	0	0	0	0	0	1	0	0	1
2nd Region	0	0	0	0	0	0	0	1	1
3rd Region	0	1	1	0	1	1	1	0	5
4th Region	0	0	1	0	0	1	0	1	3
5th Region	0	0	0	0	0	1	0	0	1
6th Region	0	0	0	0	0	0	0	0	0

TABLE 3. Outcome of the experiment.

Subject	Visual P300 Speller			Audio P300 Speller			Audio – Visual P300 Speller		
	Classification Accuracy (%)	Peak Amplitude (μV)	Peak Latency (ms)	Classification Accuracy (%)	Peak Amplitude (μV)	Peak Latency (ms)	Classification Accuracy (%)	Peak Amplitude (μV)	Peak Latency (ms)
1	82.14	4.88	392.37	64.29	3.53	400.43	89.29	5.76	347.38
2	78.57	4.43	451.22	57.14	3.17	417.61	85.71	4.94	393.52
3	78.57	5.02	388.64	60.71	4.01	398.42	92.86	4.73	358.49
4	71.43	4.32	309.44	46.43	3.82	493.21	85.71	5.09	400.64
5	78.57	4.66	304.75	60.71	4.31	419.50	96.43	5.35	302.40
6	75.00	4.72	415.86	46.43	3.84	596.36	92.86	5.83	284.64
7	82.14	5.14	332.74	42.86	3.72	498.53	89.29	5.02	318.85
Average	78.06			54.08			90.31		

that in the 1st stage of P300 speller contains the letters "A", "B", "C", "D", "E" and "F". Letters of the detected region were displayed in 6 regions on the screen in the 2nd stage of P300 speller. According to the Table 2 that is related to 2nd stage of P300 speller, as a result of a trial, the 3rd region was detected with the highest counted value. Thus, we could define subject-4's selected letter as "C" that is placed in the 3rd region of the P300 speller.

According to the SWLDA method, we calculated classification accuracy of our proposed region based P300 speller as a percentage. This calculation was carried out for all subjects according to the experiments with 3 modes P300 speller. The result of classification accuracy, peak latency, and peak amplitude values of the subjects were shown in Table 3 for each mode of P300 speller. The average classification accuracies were 54.08%, 78.06%, and 90.31% in the audio only, visual only, and audio-visual P300 speller modes, respectively. It is understood that the average classification accuracy in the audio-visual P300 speller was significantly higher than the other modes of P300 speller. Classification accuracy in the audio-visual P300 speller was improved 15.69% and 66,99% according to the visual only and audio only P300 speller that we used in the experiments, respectively.

In the experiments we observed that negative ERP by subject-5. As shown in Table 3, classification accuracy reached over 80% in the visual only P300 speller by 3 subjects. Moreover, 3 subjects exceeded the classification accuracy 90% in the audio-visual P300 speller. But in the experiments with audio only P300 speller, classification accuracy did not reach over %70 with any subject.

The mean latency in the audio-visual P300 speller reached its peak amplitude in 344ms which was shorter than other modes of speller that we used in the experiments. The longest mean latency time was 461ms in the audio only P300 speller. The average peak amplitude values were 4.74μV and 5.25μV in visual only and audio-visual P300 speller modes, respectively. However, the lowest average peak amplitude value was 3.77μV in audio only P300 speller.

Figure 8 shows that classification accuracy of all sequences in the experiments with different spelling modes. All subjects in experiments with 3 modes P300 spellers achieved higher classification accuracy in the last sequences than initial sequences.

Information transfer rate values are given in Figure 9 that shows the values in bit per second (bpm). With audio-visual P300 speller, ITR was improved 29.11% according to the

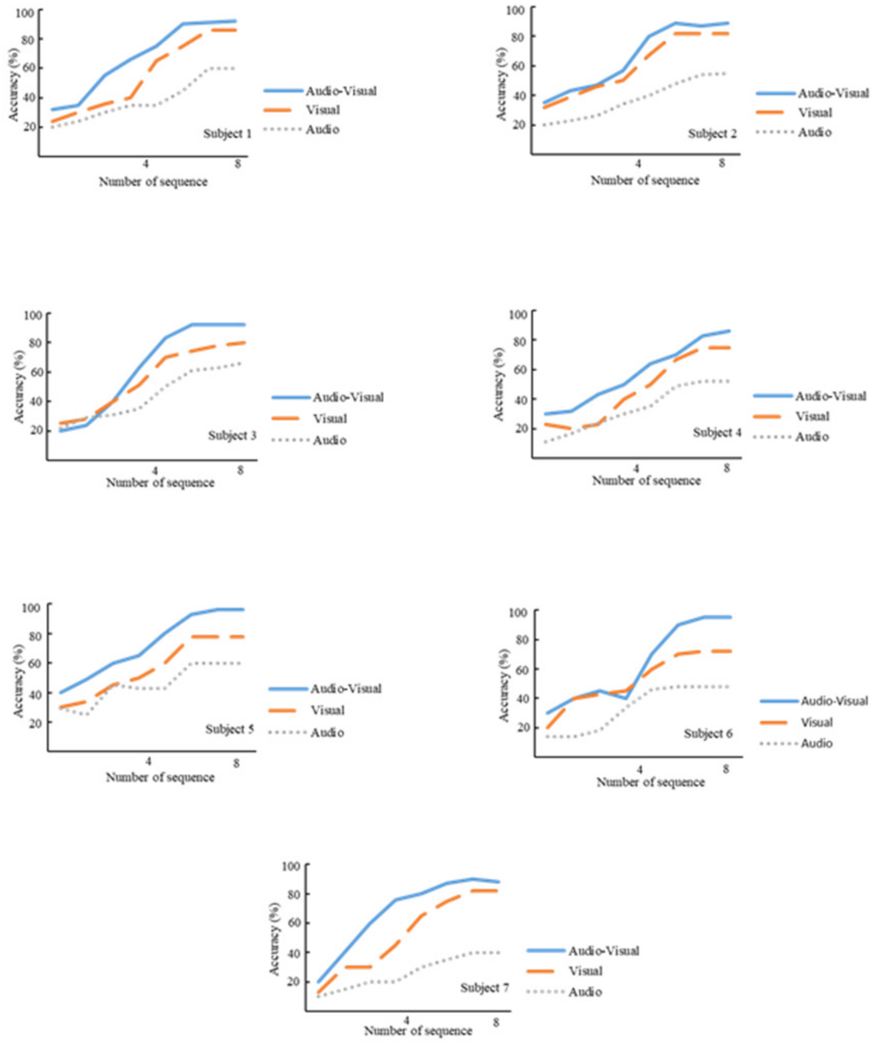

FIGURE 8. Spelling accuracy rate as a function of number of sequences for each user. The solid line represents spelling accuracy for the audio-visual, the pointed line for the audio, and the dashed line for the visual region based P300 speller.

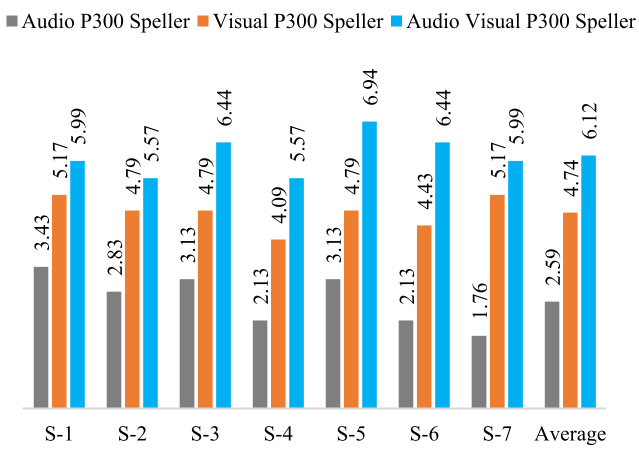

FIGURE 9. Information transfer rate values in bit per second (bpm) with all subjects.

visual P300 speller that we used in the experiments. This parameter is a key role for BCI overall performance.

Consequently, in this study the proposed region based audio-visual P300 speller has higher classification accuracy and ITR than region based only visual and only audio P300 spellers that we used in the experiments. We also performed a one-way ANOVA on classification accuracy, peak latency and amplitude of each subject's P300 response according to the average value of electrodes. Classification accuracy values at different hypothesis has statistically significant difference ($p < 0.05$). The amplitude and latency ANOVA did not detect a significant difference across trials ($p > 0.05$). According to the equation 8, classification accuracy, number of targets, and signal detection time is very important for ITR that shows the BCI overall performance. According to the results, our new paradigm has proved its higher performance.

IV. DISCUSSION

This study investigated whether or not it is possible to use a region based audio-visual P300 speller for subjects with high classification accuracy and ITR. Although previous studies suggested region based visual P300 spellers until our study, there has been no research about region based audio-visual (hybrid) P300 spellers. Fazel-Rezai and Abhari and Pan et al. showed that their approach had higher accuracy rates than the Farwell-Donchin paradigm that had a row column P300 speller [6], [22], [24].

Although region based P300 speller with visual stimulus has higher performance than row column based P300 speller with visual stimulus, there are very few studies about region based research. Moreover, there are studies row column based P300 speller with using audio and visual stimulus together as a hybrid. For instance, Klobassa et al. showed that a row-column P300 speller with audio-visual stimulus had higher classification accuracy than the visual one. Moreover, Klobassa et al. reached the 5.6 bpm as a maximum ITR [28]. In our study with using region based audio-visual P300 speller, we reached the 6.9 bpm and 6.12 bpm as maximum and mean ITR values, respectively. There is no study about region based P300 speller with using audio-visual stimulus together in the literature. We proposed the P300 speller with audio-visual stimulus as a new paradigm for higher performance than others.

In this study, we improved the classification accuracy and ITR in the audio-visual P300 speller that was improved 15.69% and 29.11% according to the visual only P300 speller that we used in the experiments. These two parameters are very significant for BCI overall performance. Our new paradigm showed that its higher performance.

Moreover, in this study, the subjects reported that the audio-based P300 speller needed subjects to pay more attention than the visual and audio-visual P300 spellers. Previous studies mentioned that mental and visual fatigue may decrease the P300 signal's amplitude [31]. Hill et al. also observed high error rates in an audio-based P300 speller [32], and Nijboer et al. observed high error rates in their audio-based paradigms [33].

Studies shows that visual stimuli evoke stronger ERP responses than auditory stimuli [30], [34]. According to our experiment we observed that the same result. Furthermore, in our study we also showed that multi modal stimulation as auditory and visual stimuli together has stronger ERP responses than mono modal stimulation. Multi modal stimulation might help the BCI user attention, and support to the loss of a single sensory modality [30]. Moreover, P300 speller users of our experiments reported that tasks are more understandable in multi modal stimulation. These findings might be the reason of ERP response as a stronger while using audio-visual P300 speller.

Some studies showed that complex stimuli are reasons for longer latency. Moreover, task difficulty affects peak latency [34]–[36]. In this study, we observed a longer latency value with the audio P300 speller.

It may be concluded that in this study most important things are that region based audio-visual P300 speller was proposed as a new paradigm for the first time. As a result, this proposed P300 speller had higher performance than visual only and audio only P300 spellers that we used. Also previous studies proved that region based visual only P300 speller has higher performance than row column based P300 speller. Also the reason of better performance with region based P300 speller could be less visual stimuli crowding effect [25]. According to the ITR equation (shown in equation 8), BCI overall performance depends on classification accuracy, number of targets, and P300 signal detection time. We showed that region based audio-visual P300 speller had higher performance. Our performance can be improved by variation of number of target, classification methods, and P300 signal detection time with further studies. In P300 based BCI systems, Linear Discriminant Analysis (LDA) and its extension methods are using for signal classification. In this study, we used SWLDA method and tested whether this method is applicable for the audio, visual and audio-visual P300 speller structures. We showed that P300 signal was elicited with using SWLDA method. In the future, we can investigate whether neural networks are well established for our proposed speller. It could affect the classification accuracy rates.

ACKNOWLEDGMENT

The authors would like to thank each subject for participation in the experiments for creation database.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002.

[2] J. Wolpaw, N. Birbaumer, W. Heetderks, D. McFarland, P. Peckham, G. Schalk, E. Donchin, L. Quatrano, C. Robinson, and T. Vaughan, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Feb. 2000.

[3] B. He and Z. Liu, "Multimodal functional neuroimaging: Integrating functional MRI and EEG/MEG," *IEEE Rev. Biomed. Eng.*, vol. 1, pp. 23–40, 2008.

[4] J. del R. Milan and J. M. Carmena, "Invasive or noninvasive: Understanding brain-machine interface technology," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 1, pp. 16–22, Jan./Feb. 2010.

[5] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalogr. Clin. Neurophysiol.*, vol. 78, pp. 252–259, Mar. 1991.

[6] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, 1988.

[7] M. De Vos, M. Kroesen, R. Emkes, and S. Debener, "P300 speller BCI with a mobile EEG system: Comparison to a traditional amplifier," *J. Neural Eng.*, vol. 11, no. 3, 2014, Art. no. 036008.

[8] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.

[9] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor, "The thought translation device (TTD) for completely paralyzed patients," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 190–193, Jun. 2000.

[10] G. Townsend, J. Shanahan, D. B. Ryan, and E. W. Sellers, "A general P300 brain–computer interface presentation paradigm based on performance guided constraints," *Neurosci. Lett.*, vol. 531, no. 2, pp. 63–68, 2012.

[11] S. Sutton, M. Braren, J. Zubin, and E. R. John, "Evoked-potential correlates of stimulus uncertainty," *Science*, vol. 150, no. 3700, pp. 1187–1188, 1965.

[12] R. C. Panicker, S. Puthusserypady, and Y. Sun, "An asynchronous P300 BCI with SSVEP-based control state detection," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1781–1788, Jun. 2011.

[13] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, "An auditory oddball (P300) spelling system for brain-computer interfaces," *Psychophysiology*, vol. 46, no. 3, pp. 617–625, 2009.
[14] S. G. Horovitz, P. Skudlarski, and J. C. Gore, "Correlations and dissociations between BOLD signal and P300 amplitude in an auditory oddball task: A parametric approach to combining fMRI and ERP," *Magn. Reson. Imag.*, vol. 20, no. 4, pp. 319–325, 2002.
[15] S. J. Radlo, C. M. Janelle, D. A. Barba, and S. G. Frehlich, "Perceptual decision making for baseball pitch recognition: Using P300 latency and amplitude to index attentional processing," *Res. Quart. Exerc. Sport*, vol. 72, no. 1, pp. 22–31, 2001.
[16] F. Nijboer, E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, N. Birbaumer, J. R. Wolpaw, and A. Kübler, "A P300-based brain–computer interface for people with amyotrophic lateral sclerosis," *Clin. Neurophysiol.*, vol. 119, no. 8, pp. 1909–1916, 2008.
[17] V. Kolev, T. Demiralp, J. Yordanova, A. Ademoglu, and Ü. Isoglu-Alkaç, "Time–frequency analysis reveals multiple functional components during oddball P300," *NeuroReport*, vol. 8, no. 8, pp. 2061–2065, 1997.
[18] Z. Oralhan and M. Tokmakçi, "The effect of duty cycle and brightness variation of visual stimuli on SSVEP in brain computer interface systems," *IETE J. Res.*, vol. 62, no. 6, pp. 795–803, Apr. 2016.
[19] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. M. Wolpaw, "A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance," *Biol. Psychol.*, vol. 73, no. 3, pp. 242–252, 2006.
[20] M. Salvaris and F. Sepulveda, "Visual modifications on the P300 speller BCI paradigm," *J. Neural Eng.*, vol. 6, no. 4, 2009, Art. no. 046011.
[21] A. Magliero, T. R. Bashore, M. G. H. Coles, and E. Donchin, "On the dependence of P300 latency on stimulus evaluation processes," *Psychophysiology*, vol. 21, no. 2, pp. 171–186, 1984.
[22] R. Fazel-Rezai and K. Abhari, "A region-based P300 speller for brain-computer interface," *Can. J. Elect. Comput. Eng.*, vol. 34, no. 3, pp. 81–85, 2009.
[23] S. Ikegami, K. Takano, K. Kondo, N. Saeki, and K. Kansaku, "A region-based two-step P300-based brain–computer interface for patients with amyotrophic lateral sclerosis," *Clin. Neurophysiol.*, vol. 125, no. 11, pp. 2305–2312, 2013.
[24] J. Pan, Y. Li, Z. Gu, and Z. Yu, "A comparison study of two P300 speller paradigms for brain–computer interface," *Cogn. Neurodyn.*, vol. 7, no. 6, pp. 523–529, 2013.
[25] H. Strasburger, "Unfocussed spatial attention underlies the crowding effect in indirect form vision," *J. Vis.*, vol. 5, no. 11, pp. 1024–1037, 2005.
[26] L. Averbuch-Heller, C. Helmchen, A. K. E. Horn, R. J. Leigh, and J. A. Büttner-Ennever, "Slow vertical saccades in motor neuron disease: Correlation of structure and function," *Ann. Neurol., Off. J. Amer. Neurol. Assoc. Child Neurol. Soc.*, vol. 44, no. 4, pp. 641–648, 1998.
[27] A. J. Suminski, D. C. Tkach, and N. G. Hatsopoulos, "Exploiting multiple sensory modalities in brain-machine interfaces," *Neural Netw.*, vol. 22, no. 9, pp. 1224–1234, 2009.
[28] D. S. Klobassa, T. M. Vaughan, P. Brunner, N. E. Schwartz, J. R. Wolpaw, C. Neuper, and E. W. Sellers, "Toward a high-throughput auditory P300-based brain–computer interface," *Clin. Neurophysiol.*, vol. 120, no. 7, pp. 1252–1261, 2009.
[29] A. Kübler, A. Furdea, S. Halder, E. M. Hammer, F. Nijboer, and B. Kotchoubey, "A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients," *Ann. New York Acad. Sci.*, vol. 1157, no. 1, pp. 90–100, 2009.
[30] A. Belitski, J. Farquhar, and P. Desain, "P300 audio-visual speller," *J. Neural Eng.*, vol. 8, no. 2, 2011, Art. no. 025022.
[31] K. E. Misulis and T. Fakhoury, *Spehlmann's Evoked Potential Primer*, 3rd ed. Oxford, U.K.: Butterworth-Heinemann, 2001, pp. 26–37.
[32] N. J. Hill, T. N. Lal, K. Bierig, N. Birbaumer, and B. Schölkopf, "An auditory paradigm for brain-computer interfaces," in *Advances in Neural Information Processing Systems*, 1st ed. Cambridge, MA, USA: MIT Press, 2005, pp. 569–576.
[33] F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N. Birbaumer, and A. Kübler, "An auditory brain–computer interface (BCI)," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 43–50, 2008.
[34] J. Katayama and J. Polich, "Auditory and visual P300 topography from a 3 stimulus paradigm," *Clin. Neurophysiol.*, vol. 110, no. 3, pp. 463–468, 1999.
[35] L. A. Combs and J. Polich, "P3a from auditory white noise stimuli," *Clin. Neurophysiol.*, vol. 117, no. 5, pp. 1106–1112, 2006.
[36] G. F. Hagen, J. R. Gatherwright, B. A. Lopez, and J. Polich, "P3a from visual stimuli: Task difficulty effects," *Int. J. Psychophysiol.*, vol. 59, no. 1, pp. 8–14, 2006.
[37] S. H. Patel and P. N. Azzam, "Characterization of N200 and P300: Selected studies of the event-related potential," *Int. J. Med. Sci.*, vol. 2, no. 4, pp. 147–154, 2005.
[38] J. Thomas, T. Maszczyk, N. Sinha, T. Kluge, and J. Dauwels, "Deep learning-based classification for brain-computer interfaces," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 234–239.
[39] S. Krishnan and Y. Athavale, "Trends in biomedical signal feature extraction," *Biomed. Signal Process. Control*, vol. 43, pp. 41–63, May 2018.
[40] M. Fabiani, G. Gratton, D. Karis, and E. Donchin, "Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential," *Adv. Psychophysiol.*, vol. 2, no. 3, pp. 1–78, 1987.
[41] M. Thulasidas and C. Guan, "Optimization of BCI speller based on P300 potential," in *Proc. 27th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Jan. 2006, pp. 5396–5399.
[42] X. Zhao, G. Cui, L. Yuan, T. Tanaka, Q. Zhao, and J. Cao, "A hybrid brain computer interface based on audiovisual stimuli P300," in *Proc. IEEE Int. Conf. Consum. Electron.-Asia (ICCE-Asia)*, Jun. 2018, pp. 206–212.
[43] J. Qu, F. Wang, Z. Xia, T. Yu, J. Xiao, Z. Yu, Z. Gu, and Y. Li, "A novel three-dimensional P300 speller based on stereo visual stimuli," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 4, pp. 392–399, Aug. 2018.
[44] T. Raij, K. Uutela, and R. Hari, "Audiovisual integration of letters in the human brain," *Neuron*, vol. 28, no. 2, pp. 617–625, 2000.
[45] F. Wang, "A novel audiovisual brain-computer interface and its application in awareness detection," *Sci. Rep.*, vol. 5, Jun. 2015, Art. no. 9962.
[46] G. D. Romigh, D. S. Brungart, R. M. Stern, and B. D. Simpson, "Efficient real spherical harmonic representation of head-related transfer functions," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 921–930, Aug. 2015.
[47] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha, "P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep. 2014, pp. 53–61.
[48] T. Castermans, M. Duvinage, M. Petieau, T. Hoellinger, C. De Saedeleer, K. Seetharaman, A. Bengoetxea, G. Cheron, and T. Dutoit, "Optimizing the performances of a P300-based brain–computer interface in ambulatory conditions," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 1, no. 4, pp. 566–577, Dec. 2011.
[49] U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 115–125, Jan. 2008.
[50] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 speller," *J. Neural Eng.*, vol. 3, no. 4, pp. 299–305, 2006.
[51] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15–21, Jan. 2008.

ZEKI ORALHAN was born in Turkey. He received the B.Eng. and Ph.D. degrees in electrical electronics engineering from Erciyes University, Kayseri, Turkey, in 2008 and 2016, respectively.

From 2008 to 2014, he was a senior engineer in a high tech firm. From 2014 to 2017, he was a Research and Development Advisor of Türk Telekom. Since 2018, he has been an Assistant Professor with the Electrical Electronics Engineering Department, Nuh Naci Yazgan University. He has also a technology firm in a techno park center of university. His research interests include brain computer interface, artificial learning, data mining, and signal processing.