# **An Auditory Model For Brain Computer Interface**

Mrs.Anupama<sup>1</sup>, Rohith Reddy <sup>2</sup>, B.Manoj<sup>3</sup>

<sup>1, 2, 3</sup>Dept of Electronics and Communication Engineering <sup>1, 2, 3</sup> B.V Raju institute of technology, Narsapur, Telangana, INDIA

Abstract- Objective: The main objective of our project is to compare the accuracy within the subject comparison between auditory streaming beep stimuli and one that used word stimuli. We aimed to see whether this system is used for the people in the locked in state. Approach: We performed a counter balanced within the subject-comparison using auditory stimuli using beep sounds and the other using word stimuli. we took samples for two healthy subjects 25 trails each event resulting 100 trails on each subject. Main results. The EP1, EP2, EP3 and EP4 are 4 event-related potentials elicited by words varied more between subjects than those elicited by beeps. However, the difference between responses to attended and unattended stimuli was more consistent with words than beeps. Significance: Since performance using word stimuli was at least as good as performance using beeps, we recommend that auditory streaming BCI systems be built with word stimuli to make the system more pleasant and intuitive. Our preliminary data show that word-based streaming BCI is a promising tool for communication by people who are locked in.

*Keywords*- brain-computer interface, EEG, event-related potential, auditory attention, natural stimuli

# I. INTRODUCTION

People may become *locked in*—i.e. 'prevented from communicating by word or body movement—due to a variety of causes. For example, the locked-in state (LIS) may result from degenerative motor-neuron diseases such as amyotrophic lateral sclerosis (ALS), from acute polyneuropathies such as Guillan–Barre´ syndrome, or from stroke, tumours or trauma that affect the brainstem.

Various definitions and sub-categories of LIS allow for some residual voluntary movement. Many people in LIS communicate using eye movements or small muscle movements to give a yes-or-no answer in response to a verbal cue provided by a conversation partner. However, these movements may be very tiring, may weaken or disappear sporadically during the course of a day, and may decline over the course of months or years due to progressive motor-neuron degeneration. Conversation partners may find it difficult, and may vary widely in their ability, to recognize these movements unambiguously. The current article is the latest in a series in which we aim to optimize, and then translate into everyday use, a simple brain–computer interface (BCI) to address this problem. We aim to develop a communication device that can be cured by a conversation partner to elicit a single reliable yes-or-no answer at a time. We decode the user's intended answer from non-invasive EEG signals.

We have chosen to develop a BCI driven by nonvisual stimuli, due to the many problems that paralyzed people may experience regarding eye movement and eye health. Our previous works have shown that effective binary BCI systems can be driven purely by attention to auditory stimuli. Our design was inspired by the observation of Hillyard *et al* that selective attention modulates the early negative (80–100 msec) as well as later positive (300– 400 msec) event-related potential components in a dichotic listening task.

We have adopted an approach based on event-related potentials (ERPs) rather than steady-state evoked potentials (SSEPs), because we found ERPs to be more efficient. Auditory SSEPs seem to require longer trials to yield comparable accuracy.

We have adopted a streaming approach in which two simultaneous or interleaved trains of stimuli are presented, and the BCI is driven by the difference in responses between all attended and all unattended stimuli. For interfaces with two target classes, we find this to be more efficient than a sequential approach that relies on the difference in responses between standard stimuli and relatively rare targets. For more than two target classes, sequential approaches may allow a potentially larger rate of information transfer due to the fact that the non-target stimuli may be differentiated into multiple classes. This allows the efficient construction of speller systems, such as the so-called 'P300 speller' based on grids of visual stimuli or other spatial arrangements. It is possible to construct such speller systems using a purely auditory BCI approach but so far, such systems impose a high workingmemory load, as the subject must memorize, and mentally navigate through, multiple mappings between spatial locations and letters (or groups of letters).

Maximizing intuitiveness, and minimizing cognitive load, is one of the primary motivating factors behind the

current study. In our previous studies, our auditory BCIs were driven by abrupt beep stimuli typical of those used to elicit ERPs in psychophysiology research. Since these are devoid of intrinsic meaning, their association with the options 'yes' and 'no' is arbitrary, and is complicated by the fact that two types of stimuli (standard and target) are associated with each option. We found that experimental subjects and potential users consequently find the system difficult to understand. Many people also reported finding the beeps harsh and mildly unpleasant.

Hohne" et al found that natural stimuli (albeit still standardized, semantically empty syllables) can drive an auditory BCI better than artificial stimuli. Encouraged by this, we wished to know whether it was possible, without loss of performance, to adapt our auditory streaming BCI system to use spoken words instead of harsh, meaningless beeps. The potential advantages of word stimuli are three-fold. First, words can be selected that naturally reflect theoptions the user can choose (for example, 'yes' and 'no') making the system much easier to explain to beginning users. Second, the natural stimuli may be more pleasant, as Hohne" et al also reported. Third, voice stimuli are very well segregated by the auditory system into separate perceptual streams, whereas the streaming percept of isolated beeps may more easily break down. If we can improve the streaming quality of the stimuli, it may in future be possible to deliver both streams from the same audio speaker, and thereby remove the reliance on good directional hearing.

Spoken words have variable length and differing temporal distributions of stimulus energy. As a result, we expect that the resulting ERPs may look somewhat different from the textbook waveforms elicited by abrupt unnatural stimuli. And as we shall see, the responses to words also exhibit considerably more variability across subjects. Therefore, before we go any further, it is necessary to test whether we can change from beeps to words without loss of performance. The main laboratory study reported in this paper was a within-subject comparison of our previous stimulus configuration (*Beeps*) against a new stimulus condition involving the words 'yes' and 'no' (*Words*).

Our study also incorporated various additional aspects that aimed to further the progression of our system towards practical usability. This included a reduction in the number of EEG channels relative to previous studies: here, we use eight channels to minimize setup time and also so that we could use the eight-channel amplifiers that are part of our standard home BCI system. Furthermore, since EEG signals are highly non-stationary, particularly between sessions, we wished to evaluate the performance of our subjects using fixed classifier weights obtained on a previous session. Therefore, our laboratory design included two sessions for every subject. Finally, having established that the Words condition was at least as effective as the old Beeps condition, we took the first steps towards testing the system (Words condition only) with two potential users who are locked in.

# **II. MATERIALS AND METHODS**

#### 2.1. Subjects

Two healthy subjects, whom we will denote by the letters A&B, took part in the experiment Each subject attended for two sessions on separate days. None of them had any history of significant hearing problems or neurological defects. All subjects gave informed consent. Figure 1. Explains the electrodes placing on the scalp.



Fig.1 Electrode placement

#### 2.2. Hardware and software

EEG recordings were made using a custom eightchannel version of the g.USBamp amplifier (g.tec medical engineering GmbH, Austria) in conjunction with an eightchannel EEG cap (Electrocap Inc.). The cap used gelled 9 mm tin electrodes at positions F3, F4, T7, C3, Cz, C4, T8, and Pz of the extended international 10-20 system of Sharbrough et al, with the reference at TP10 (the right mastoid) and the ground electrode at TP9 (the left mastoid). The amplifier performed appropriate anti-alias filtering before digitizing at 24 bits and down sampling to 256 Hz. Data acquisition was performed using the BCI2000 software platform v.3.0; signalprocessing, classifier optimization and stimulus presentation were implemented in Python using the 'BCPy2000' add-on to BCI2000. Stimuli were delivered via the laptop's built-in soundcard to a pair of Sony MDR-V600 headphones worn by the subject. The software ran on a Lenovo ThinkPad T61p laptop with a 2.2 Ghz dual-core processor. Figure 2. Depicts the placing of EEG-cap on a person's scalp.



Fig.2 EEG cap placing on scalp

# 2.3. Stimuli and task design

A subject's first session began with a five-minute pre-recorded audio introduction explaining the experiment. It then consisted of 12 runs: 3 of one condition, 3 of the other condition, 3 more of the first condition, and 3 more of the second. In our case subjects are asked to hear the sounds and to imagine the same as per the instructions. The pre-recorded instructions were not played in the second session unless the subject asked to hear them.

Each run consisted of 25 trials. Each trial lasted around 5 s in total (including a few seconds' rest) and consisted of an attempt to listen to only the stimuli in the left earphone (toselect 'no') or only the stimuli in the right earphone (to select 'yes').

In the Words condition, the left stream consisted of a synthesized male voice saying 'no' twenty five times. Randomly on each trial, 1, 2 or 3 out of the last 5 'no' stimuli were instead target stimuli in which the voice said 'no'. Here all the conditions are given to the subjects randomly i.e., In our context we were running 4 different runs and each run of 25 times so that overall 100 trails were given to the subjects randomly each of 25 times and the samples were collected, As we are comparing the best out of word stimuli and beep stimuli we collect the samples regarding to it. In the process of giving it to the subjects a synthesized male voice saying 'yes' heard 25times in the trails only in the right stream and the same voice saying 'no' heard 25 times in the trails only in the trails only in the left stream.

The Beeps condition was conceptually identical to the Words condition, but the standard stimuli were 150 msec beeps at 512 Hz (left) or 768 Hz (right), and the target stimuli were amplitude-modulated versions of the standard beeps. The synthesized vocal cue on each trial was, for example, '*listento* <*LATERALIZED BEEP*> to say 'yes'.' There was no need for subjects to look at a screen. We asked them to keep their eyes still during stimulus presentation by fixating a spot marked on the wall.

Power spectral density- PSD is calculated by Fourier transforming the estimated autocorrelation sequence which is found by nonparametric methods. One of these methods is Welch's method. The data sequence is applied to data windowing, producing modified period grams. The information sequence xi (n) is expressed as

$$x_i(n) = x(n+iD), \quad n = 0, 1, 2, \dots, M-1$$
  
while  $i = 0, 1, 2, \dots, L-1;$ 

Take iD to be the point of start of the ith sequence. Then L of length 2M represents data segments that are formed. The resulting output periodgrams give

$$\overset{\approx}{P}_{xx}^{(i)}\left(f
ight)=rac{1}{MU}{\left|\sum_{n=0}^{M-1}x_{i}(n)w(n)e^{-j2\pi fn}
ight|^{2}}.$$

Here, in the window function, U gives normalization factor of the power and is chosen such that

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n)$$

Where w (n) is the window function. The average of these modified periodgrams gives Welch's power spectrum as follows

$$P_{xx}^{W} = \frac{1}{L} \sum_{i=0}^{L-1} \tilde{P}_{xx}^{(i)}(f).$$

#### 2.4. Signal processing and classification

The signal-processing methods were categorized into different stages. First the signal was filtered using equi-ripple band pass filter. Features were computed by taking the right–left difference of the within-trial averages of 600msec epochs. We applied wavelet packet decomposition. For n levels of decomposition, the WPD produces  $2^n$  different sets of coefficients (or nodes) as opposed to (3n + 1) sets for the DWT. However, due to the down sampling process the overall number of coefficients is still the same and there is no redundancy.

From the point of view of compression, the standard wavelet transform may not produce the best result, since it is

limited to wavelet bases that increase by a power of two towards the low frequencies. It could be that another combination of bases produce a more desirable representation for a particular signal. The best basis algorithm by Coifman and Wickerhause finds a set of bases that provide the most desirable representation of the data relative to a particular cost function (e.g. entropy).

Separate subject-specific classifiers were maintained for the two conditions: one for each subject's Beeps data and one for each subject's Words data. In the first session, classifiers were re-trained after every new run of 25 trials. In the second session, the final classifier weights from the first session were used and kept fixed.

# **III. RESULTS**

## 3.1. Performance

The main question of our laboratory study was whether there was any significant difference in performance between the Words and the Beeps stimulus conditions. Addresses this question by showing the results of the withinsubject comparison of the two conditions. Each data-point represents the data from one subject: points that lie above the diagonal indicate better performance in the Words condition than in the Beeps, and points below the diagonal indicate better performance on Beeps than on Words.

On average, Words narrowly beat Beeps. The mean and standard deviation of % correct scores across subjects was  $76.9\% \pm 11.1$  for Words, and  $73.0\% \pm 10.6$  for Beeps. Neither of these gains was significant at the 5% level in a Wilcoxon signed rank test.

Whether a subject performs better in Words or Beeps depends very much on the individual. We performed 100 trials on each subject collecting the samples, and on each condition we performed 25 trials i.e randomly performing 4 conditions of listening to left beep, right beep, a synthesized male voice of 'yes' to right ear and the final condition of listening to synthesized male voice of 'no' to the left ear.

We also wished to know whether, having trained a classifier on one session, we could expect good performance on future sessions without the necessity to perform additional supervised trials for re-calibration.Finally, we got the accuracy of 60% for beep sounds and 70% for word stimuli for subject1.For the subject2we got an accuracy of 65% for beep sounds and 80% for word stimuli. By this we can directly say that for the people who are in locked-in syndrome feel very

easy for communicating using word stimuli rather than word stimuli.

Having established that we can adopt a design based on Words stimuli without loss of performance relative to our older Beeps design, we then wished to verify that the Words approach could work for potential users who were locked in shown in figure 3.



Fig.3 waveforms of response based on auditory stimuli

Figure3. above shows the wave forms, first waveform represents the beep sound on the left ear, second wave form represents the beep sound on the right ear, third wave form represents the word Yes and the last waveform represents the word No.

## 3.2. Features

Figure 4shows the patterns that we can see in the responses to Beeps stimuli. Each panel is an eight-channel  $\times$  600 msec image of the epoch following stimulus onset. Column(i) is simply the averaged response to all stimuli: left stimulus onset is at time 0 and right stimulus onset is at 250 msec. The colour scale from -4 to +4 $\mu$ V is common to all plots in column (i). The primary sensory response, the N1 at a latency of 100 msec, appears as a dark stripe at 100 msec (elicited by left stimuli) and at 350 msec (elicited by right stimuli). This feature is very consistent across subjects—even subject M, who was not consistently awake throughout the experiment. For some subjects (F and G) it is possible that the response in the two ears is different due to an difference between the two ears' sensitivity to sound.

Columns (ii)–(iv) are all contrast plots showing *d*prime measures of signal-to-noise ratio, i.e. the mean voltagedifferences divided by standard-deviations of the differences, across stimulus epochs. They share a common colour scale of -0.6 to  $+0.6\sigma$ . Column (ii) shows the difference between standards and targets in the unattended stream, and column(iii) shows the same for the attended stream. In both, the N2a component (known as the mismatch negativity, or MMN) is visible as a dark stripe at around 200–250 msec. To some extent in both, although to a greater extent in the attended stream (column iii) the P3 response is visible as a light feature immediately following the MMN.

Column (iv) contrasts all attended versus all unattended stimuli (onset at time 0) and is therefore most closely related to BCI performance. Note that this set of features is very variable from subject to subject. Although some subjects show similar patterns to each other (for example, subjects B and D), there are some subjects whose patterns appear to have opposite polarity to each other (for example, D and G). Such individual differences underscore the need for subject-specific classifier training.

The picture looks very different for the Words condition. The format and colour scaling are identical. The subjects inare now reordered according to their performance in the Words condition on day1.Note that the average response in column(i) is now much more variable from subject to subject, as we mentioned in the introduction. The MMN-P3 pattern is no longer consistently visible in the unattended stream (column ii) and, although it is visible in the attended stream results. (column iii), it is also more variable from subject to subject. Note, however, that despite the increased variability of the primary sensory responses, the attended-versusunattended contrast shows less inter-individual variation than it did for Beeps. Many of the subjects, particularly the betterperforming ones, show the same negative feature 300 msec after stimulus onset (most visible in subjects B, J, K, G and C) although there are still subjects (e.g. D and I) who appear to show the opposite polarity while still performing above chance. The negative feature at 300 msec is visible during H3's free-selection trials (although not during the cuedselection trials, during which performance was lower).

## **IV. DISCUSSION**

Our results show that an online binary auditory streaming BCI can be built with as few as eight EEG channels, and use single trials in which the critical EEG segment is less than 4.5 s long, and still achieve 77% correct on average (93% for the best subject). We also find that there is no significant disadvantage, but rather some non-significant tendency towards improved performance, in switching to more-intuitive natural stimuli (voices saying 'yes' and 'no') instead of the abstract, unpleasant beeps used in previous studies.

The finding that natural, intuitive stimuli are at least as effective as abstract stimuli spurred us to conduct a preliminary assessment of the effectiveness of this system when used by people in the locked-in state. We presented simple yes-or-no questions to two people who could only make small, limited muscle movements because of advanced amyotrophic lateral sclerosis (ALS). Both subjects were able to use the system to answer questions correctly at roughly the same level of accuracy that we observed among healthy volunteer subjects. This is an encouraging sign that the system will translate well to the target user group.

The paradigm was suitable for eliciting and examining a number of well-known ERP components in a relatively short time, and may therefore be valuable for developing a profile of a new user. Such a profile will yield useful information about the user's ability to hear the stimuli, hear equally well with both ears, discriminate differences between stimuli, and follow instructions for attending to one side or another. It may therefore allow preliminary inferences to be made about the subject's state of consciousness and perceptual and cognitive state.

While abrupt beep stimuli produced N1, MMN and P3 event-related potential (ERP) components that were very consistent across subjects, the difference between responses to attended and unattended beeps was very variable, even among subjects who performed well. Since this difference drives the BCI, it is clearly very important to tailor classifier weights individually to each subject. By contrast, word stimuli produced less-consistent N1, MMN and P3 responses, but the difference between attended and unattended stimuli was actually more consistent across subjects. The most consistent feature of the difference wave was a negative peak around 300 msec after stimulus onset. The energy of each word stimulus was spread out over about 350 msec with a relatively slow attack, rather than concentrated within the first 150 msec with a very sudden attack, as with the beep stimuli. As a result, the effective latency of the crucial negative component should perhaps be considered to be smaller than 300 msec-but, for the same reason, it is difficult to be certain which of the wellknown ERP components, if any, most closely corresponds to it.

Finally, we show that there is a significant loss of performance when classifier weights are transferred from session to session, which remains as a challenge for the development of signal-processing algorithms. We had hoped to be able to use these data to develop and examine adaptive semi-supervised methods that might help us to transition from session to session with minimal loss of performance. However, it seems that the current data were unsuitable for an offline examination of this issue. The steady decline in performance over the course of the second session can be eliminated by updating the classifier after each run, but even with supervised re-training, average performance on day 2 does not increase. So, it seems that our day 2 data are inherently noisier than the day 1 data. This is not purely a result of using fixed classifier weights, since neither the 6-7 percentage-point drop in performance, nor the decline as a function of time, occurs when evaluating day 1 data using fixed weights transferred from day 2 instead of incrementally re-trained weights. We examined trends in classifier bias over time in the incrementally re-trained data, and also the variability of the classifier weights over time, but found no obvious difference between day 1 classifier solutions and day 2 classifier solutions that might explain the differing character of the two days' results. Rather, we suspect that the drop in performance from day to day, and the decline as a function of time, must result from an online interaction between the use of fixed weights and other factors intrinsic to the subject's state (for example, motivation). It seems that even an early feasibility test of an adaptive semi-supervised algorithm designed to overcome this session-to-session transfer problem may require collection of BCI data performed with the candidate algorithm running online.

While the current results are in the appropriate performance range for comparing accuracy across conditions and avoiding ceiling effects, for future practical use we will want to increase the absolute level of accuracy above 90% for the majority of users. This may potentially be accomplished by lengthening the trials, or ideally by using trials whose length is determined by a dynamic stopping mechanism. Further performance improvements might be expected if future studies investigate the effect of changing the speed of stimulus delivery, and the effect of user training—the results of Hill and Scholkopf suggest that the latter in particular may be possible because BCI performance is predicted by subjects' accuracy in giving an overt behavioural response to the targetcounting task.

We conclude that our paradigm is a promising candidate for a simple, practical BCI system that could be used by a locked-in person to communicate 'yes' or 'no' via an intuitive non-visual interface during a conversation with a human partner. The value of such non-visual approaches may go beyond application to people who cannot see. One of our subjects with ALS, who can see a computer screen well and is able to use an on-screen keyboard via eyebrow movement, welcomed this novel non-visual access method, telling us, 'my eyes get tired, but never my ears.' This suggests that the approach may provide an attractive ergonomic alternative for some tasks, even for users who can see well.

# V. ACKNOWLEDGMENTS

This work was supported by BVRIT. We thank our guide and lab instructor for their helpful suggestions.We also thank our department for their encouragement.

#### REFERENCES

- [1] Plum F and Posner J B 1966 *The Diagnosis of Stupor andComa* 1st edn (Philadelphia, PA: FA Davis)
- [2] Bauer G, Gerstenbrand F and Rumpl E 1979 Varieties of the locked-in syndrome *J. Neurol.***221**77–91
- [3] Hochberg L R and Anderson K D 2012 BCI users and their needs *Brain–Computer Interfaces: Principles and Practice*1st edn ed J R Wolpaw and E W Wolpaw (Oxford: Oxford University Press) pp 317–23 chapter 19
- [4] Sharma R, Hicks S, Berna C M, Kennard C, Talbot K and Turner M R 2011 Oculomotor dysfunction in amyotrophic lateral sclerosis Arch. Neurol.68857–61
- [5] Hill N J, Lal T N, Bierig K, Birbaumer N and Scholkopf<sup>\*</sup> B 2005 An Auditory Paradigm for Brain– ComputerInterfaces (Advances in Neural Information Processing Systems vol 17) ed L K Saul, Y Weiss and L Bottoupp 569–76 http://books.nips.cc/papers/files/nips17/NIPS2004\_0503.p df
- [6] Hill N J and Scholkopf B 2012 An online braincomputer interface based on shifting attention to concurrent streams of auditory stimuli J. Neural Eng.9026011
- [7] Hill N J, Moinuddin A, Hauser" A-K, Kienzle S and Schalk G 2012 Communication and control by listening: toward optimal design of a two-class auditory streaming brain–computer interface *Front. Neurosci.***6**181
- [8] Hillyard S A, Hink R F, Schwent V L and Picton T W 1973 Electrical signs of selective attention in the human brainScience 182177–80
- [9] Kim D-W, Hwang H-J, Lim J-H, Lee Y-H, Jung K-Y and Im C-H 2011 Classification of selective attention to auditory stimuli: toward vision-free brain–computer interfacing *J. Neurosci. Methods*197 180–5
- [10] Lopez M-A, Pomares H, Pelayo F, Urquiza J and Perez J 200009 Evidences of cognitive effects over auditory steady-state 00r0esponses by means of artificial neural networks andits use in braincomputer interfaces *Neurocomputing***72** 3617–23
- [11] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing eventrelated brain potentials *Electroencephalogr. Clin. Neurophysiol.***70** 510–23
- [12] Williamson J, Murray-Smith R, Blankertz B, Krauledat M and Muller K-R 2009 Designing for uncertain,

asymmetric control: interaction design for braincomputer interfaces*Int. J. Hum.-Comput. Stud.* **67**827-41

- [13] Furdea A, Halder S, Krusienski D, Bross D, Nijboer F, Birbaumer N and Kubler<sup>..</sup> A 2009 An auditory oddball (P300) spelling system for brain-computer interfaces*Psychophysiology* **46**617–25
- [14] Klobassa D S, Vaughan T M, Brunner P, Schwartz N E, Wolpaw J R, Neuper C and Sellers E W 2009 Toward a high-throughput auditory P300-based brain–computer interface *Clin. Neurophysiol.***120**1252–61
- [15] Schreuder M, Rost T and Tangermann M 2011 Listen, you are writing! speeding up online spelling with a dynamic auditory BCI *Front. Neurosci.***5**112
- [16] Hohne" J, Schreuder M, Blankertz B and Tangermann M 2011 A novel 9-class auditory ERP paradigm driving a predictive text entry system *Front. Neuroscience*599
- [17] Hohne" J, Krenzlin K, Dahne" S and Tangermann M 2012 Natural stimuli improve auditory BCIs with respect to ergonomics and performance J. Neural Eng.9045003
- [18] Bregman A S 1994 Auditory Scene Analysis (Cambridge, MA: MIT Press)
- [19] Sharbrough F, Chatrian G E, Lesser P R, Luders" H, Nuwer M and Picton T W 1991 American electroencephalographic society guidelines for standard electrode position nomenclature J. Clin. Neurophysiol.8200–2
- [20] Schalk G, McFarland D, Hinterberger T, Birbaumer N and Wolpaw J 2004 BCI2000: a general-purpose braincomputer interface (BCI) system *IEEE Trans.Biomed. Eng.* 511034–43
- [21] Schalk G and Mellinger J 2010 A Practical Guide toBrain–Computer Interfacing with BCI2000 (London:Springer)Hill N J, Schreiner T, Puzicha C and Farquhar J 2007 BCPy2000 http://bci2000.org/downloads/BCPy2000.