32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

A Yarbus-Style Experiment to Determine Auditory Attention

P. Kidmose, M. L. Rank, M. Ungstrup, D. Looney, C. Park, and D. P. Mandic

***Abstract*— This paper presents an analysis of the merits of the original Yarbus experiment on eye movements with respect to judgments on differences in cognitive layer processes. The principles thus derived are applied to the development of an equivalent auditory experiment where, instead of eye movements, the response of the subject is observed by EEG measurements. Results from a preliminary trial are also included in which EEG analysis is used to ascertain the attended sound source in a multiple sound source environment. The investigation is part of ongoing research to improve the usefulness of hearing instruments and is also relevant in relation to other scientific investigations concerning the processing of sounds in complex acoustical environments by the human brain.**

I. INTRODUCTION

Modern hearing aids, despite advances in technology as well as hearing research in recent decades, still suffer from several deficiencies. The present investigation attempts to address amongst these a systemic shortcoming, namely that the intention of the person wearing the hearing aid is not easily, and by no means naturally, reflected in the advanced digital signal processing performed by the hearing aid.

Adjusting the hearing aid signal processing, based on user input, is limited to the fitting (including fine-tuning) of the hearing aid and program selection by a remote control during use (or other user input). Thus, no information is provided to the hearing aid about which of several sound streams from a mixture the user is currently attending. The processing in the hearing aid is in this sense blind; even were a perfect source separation possible, the hearing aid would have no way of figuring out which source or sources should be presented and which should be discarded as “noise”.

The hearing aid provides input to the perceptual system as illustrated in Fig. 1, however, there are no feedback connections from the perceptual system which would aid their operation. The aim of this research is to investigate whether electroencephalogram (EEG) recordings could provide such a feedback connection.

Previous work [1] has used EEG to model the degree of auditory attention and auditory brain control interface (BCI) studies have shown that, for a dual stimulus environment, it is possible to estimate the attended stimulus [2], [3]. The only cases considered used perceptually simple sound sources, such as frequency modulated tones, and contained convenient

P. Kidmose (prk@widex.com), M. L. Rank (mlr@widex.com), and M. Ungstrup (miu@widex.com) are employed by Widex A/S, 3540 Lynge, Denmark

D. Looney (david.looney06@imperial.ac.uk), C. Park (cheolsoo.park07@imperial.ac.uk), and D. P. Mandic (d.mandic@imperial.ac.uk) are with Department of Electrical and Electronic Engineering, Imperial College, London, SW7 2BT United Kingdom

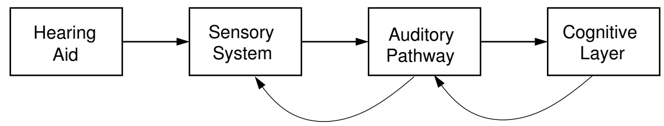

Fig. 1. Block diagram of the role of the hearing aid within the system of auditory perception which consists of: Sensory System (outer ear, middle ear, cochlear); Auditory Pathway (cochlear nucleus, auditory cortex); and Cognitive Layer (high level auditory layer, cognitive layer).

features which made it straightforward to identify related activity in the EEG. For our considered application in which the sound sources are considerably more complex, this is not sufficient. Instead, it is necessary to determine more robust features that are relevant to auditory attention in a real-world auditory environment.

A technical obstacle is that, in a complex auditory environment, an EEG recording can be influenced by several layers of brain functions: sensation, perception and cognition. However, only the perception and cognition layers can provide an insight into the relevant attention mechanisms. The challenge is therefore to design an experiment in which it is possible to isolate this cognitive information and the aim of this paper is to provide a framework for such investigations.

Following the seminal work by Yarbus [4] in which a fixed visual stimulus combined with different instructions was used to induce different cognitive responses in the investigation of visual scene analysis, we propose an corresponding experiment for auditory scene analysis. An experiment involving multiple trials was constructed where the auditory input and hence processes at the layer of sensation are identical for all trials, but the cognitive instruction is varied. This was achieved by instructing the subject to alternate their attention between the sound streams. Thus, systematic differences in observable measurements of the brain (EEG response) can be used to infer about changes at the cognitive layer.

Advanced data analysis was performed on the EEG recordings following evidence [5][6] which suggests that the degree of neuronal synchronisation within different cortical regions of the brain, specifically within the gamma band (30-80Hz), reflects the level of cognitive processing and can convey selective attention. Standard measures of synchronisation, such as coherence or crosscorrelation, are not appropriate in practice as they combine phase and amplitude information and are limited to the analysis of second order statistical signal properties only. To circumvent these issues we estimated the degree of synchronisation by combining phase synchrony and asymmetry using the approach detailed in [7] which facilitates highly localised analysis and is suitable for nonlinear and nonstationary data.

978-1-4244-4124-2/10/$25.00 ©2010 IEEE

II. YARBUS' EXPERIMENT ON EYE MOVEMENTS

The Russian psychologist Alfred L. Yarbus conducted an experiment in which he presented the subjects with a picture (The Unexpected Visitor), asked the subjects different questions about the picture, and recorded their eye movements while they investigated the picture to answer the question posed [4]. The subjects were each asked several question regarding the picture so that their eye movement patterns corresponding to different questions could be compared. A result from such a series of questions is shown in Fig. 2.

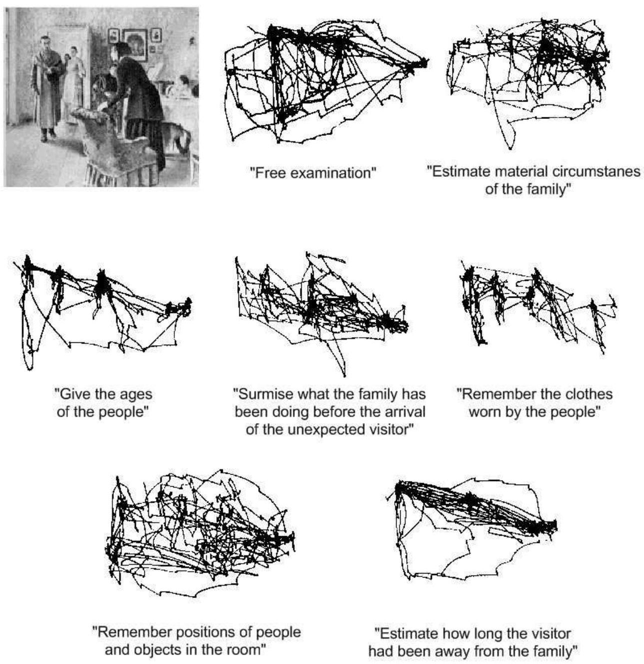

Fig. 2. The Unexpected Visitor by Yarbus [4]. Notice the different patterns of eye movements according to the different cognitive tasks.

The result is perhaps not surprising: the eye movement patterns differ according to the question posed. For some questions the faces of the people appearing in the picture are scanned, while for other questions the clothes, positions, or apparent gestures are more important. The conclusion drawn by Yarbus was that the patterns of eye movements are influenced by the cognitive layers of the brain. The input (the picture) was the same across all trials, but the instruction was different, causing different cognitive responses as indicated by the eye movement patterns.

To further analyse the result, we now propose a simplified model of the system of perception and movement. The system of perception consists of three layers (sensation, perception, and cognition) while movements are initiated by action units that activate relevant motor units. We can thus draw a simplified block diagram of the brain processes involved shown in Fig. 3. The "Unexpected Visitor" provides an input to the layer of *sensation*, which is further processed through the layers of *perception* and *cognition*. Processes at the two latter layers cause action units to initiate eye movements by activating relevant motor units, influencing the sensory input to the eyes.

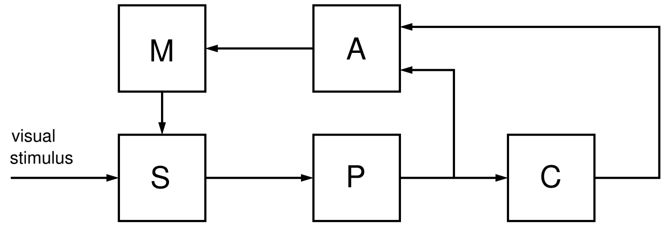

Fig. 3. Simplified block diagram of brain processes in the Yarbus experiment. Symbols M: motor, A: action, S: sensation, P: perception, and C: cognition.

III. AUDITORY EQUIVALENT OF THE YARBUS EXPERIMENT

For our purpose we wish to conduct an experiment to determine whether the auditory attention of a subject in a complex acoustical environment can be gauged by means of EEG measurements. We conjecture a model for the brain processes involved in focusing auditory attention, shown in simplified schematic form in Fig. 4. The sound stream provides an input to the layer of sensation, which is further processed through the layers of perception and cognition, whereas feedback from the cognitive layer to the perceptual layer is involved in focusing the attention of the subject. Such a model is supported by evidence that behavioral context, including attention and intention, affect even basic perceptual processes [8].

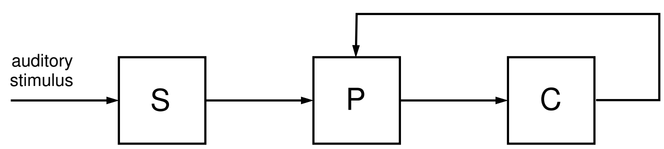

Fig. 4. Simplified block diagram of brain processes involved in auditory attention. Symbols S: sensation, P: perception, and C: cognition.

We consider the following point to be of fundamental importance for investigations concerning the cognitive response to different stimuli, whether the response is observed by eye movements, EEG, Magnetic Resonance Imaging (MRI), or similar: *the low-level input must be controlled in order to be able to draw conclusions about differences in processing at the perceptual or cognitive level*. This is the important lesson to be learned by the brilliant experiment constructed by Yarbus. In [9] a similar observation for the case of Auditory Evoked Potentials (AEP) was made.

Applying this principle to the auditory case, an experiment was constructed involving multiple trials where the auditory input and hence processes at the layer of sensation are identical for all trials, but the instruction is varied in analogy to the original Yarbus experiment. We record the EEG during each trial and subsequently develop feature extraction and classification in order to distinguish between instructions with a probability which is significant. In the sequel, this experiment construction will be referred to as an *auditory Yarbus experiment* due to the conceptual resemblance with the original, visual Yarbus experiment. Table I summarizes a number of analogies between the two experiments.

TABLE I

ANALOGY BETWEEN THE ORIGINAL VISUAL YARBUS EXPERIMENT AND THE AUDITORY YARBUS EXPERIMENT.

Visual Experiments	Eye movement determines the stimulus to the sensory organ and hence the low-level response in the brain differs with the instruction.	The search for information in the picture is physical (eye movement).	The eye movement pattern depends on the input (picture) as well as the instruction.	Interpreted locally the the eye movement pattern somehow reveals the local information rate or entropy of the input.	Interpreted globally the eye movements seem connected with the subjects interpretation of the input and the instruction.	The correlation between input and eye movement pattern and the interpretation thereof indicate that the response is governed also on a cognitive level.	Repetitions across different test subjects result in 'similar' patterns. (Averaging over different test subjects is possible).	Conjecture: Repetitions of the experiment with the same test subject could cause a weaker response.
Auditory Experiments	The low-level stimulus to the sensory organs is the same and hence the low-level response of the brain does not differ with the instruction.	The search for information in the audio stream is psychological or cognitive.	Conjecture: A similar dependency exists for the feedback in the auditory experiment.	No conjecture.	No conjecture.	There is no conjecture of a correlation between input and measurement or interpretation thereof.	No expectation of cross-subject correlation.	Conjecture: Repetitions of the experiment with the same test subject will not to the same degree result in a weakening of the response.

To justify the auditory Yarbus paradigm, the following experimental setup was used for eight volunteers with healthy hearing (mean age 30 years, median age 25 years, five male, three female). An auditory stimulus, a mixture of music and speech, was played to the subject through two loudspeakers placed at approximately +/- 60 degrees from the front of the subject. Speech was presented from the right speaker (relative to the subject) and music from the left. The volumes of speech and music were loosely calibrated so that both sources were loud and clear (60-80 dB SPL range) and so as to facilitate selective attention to each.

- As a proof of principle, a series of sub-experiments were performed on two subjects only[1]. Three recording sessions of 20 trials were conducted[2] in which the subject was instructed to attend the speech or music only (10 music trials, 10 speech trials). In between each trial, an additional recording was conducted to record baseline EEG activity to be used in later analysis. These recordings were made with the gUSBamp biosignal acquisition device at a sampling frequency of 4.8kHz.
- The framework was extended to a larger number of subjects. For each subject, a single recording session was conducted (16-24 trials) without baseline recordings. As before, the subject was instructed to attend either the music or the speech[3] for a given trial. These recordings were made with the gMOBIlab+ portable biosignal acquisition system at a sampling frequency of 256Hz.

EEG was recorded from electrode positions FC3, FC4, FC5, FC6, C3, C4, T7 and T8 with reference to the right ear lobe according to the 10-20 system.

IV. ANALYSIS

Analysis of the sub-experiments was achieved by calculating the percentage change in spectrum power, induced by attending the relevant sound stream, relative to spectrum activity in the absence of stimulus (the baseline recording) within the frequency range 35-45Hz. A percentage increase is defined as event related synchronisation (ERS) and a decrease as event related desynchronisation (ERD) [10]. Results that exceeded +/- 150% were disregarded as outliers.

Analysis for the larger study was achieved by estimating the degree of neuronal synchronisation, within the frequency range 30-80Hz, for pairs of electrodes using the features:

- phase synchrony – the temporal locking of phase information between the electrodes, and
- asymmetry – the lateralization of spectral power between the electrodes.

The features were estimated within a unified bivariate empirical mode decomposition (BEMD) [11] framework which facilitates a highly localised comparison, in time and frequency, between two signals. Furthermore, the adaptive nature of BEMD makes it suitable for the analysis of nonstationary data such as EEG. For more information regarding the BEMD synchronisation framework see [7]. The recorded data for each trial was divided into a set of 6 subsegments each of length 4s. For each subject, features obtained for 70% of the subsegments were used to train a support vector machine (SVM)[4], and classification was performed on the remaining subsegments with the goal of correctly estimating whether the subject attended speech or music. The selection of subsegments used for training was performed in a random fashion. In this way, the SVM was retrained and used to perform classification a total of 50 times. The classification performance was taken to be the average of these 50 outcomes. This same analysis was repeated, including a full retraining of the SVM, for other electrode pairs and subjects.

[1]For the sub-experiments only, the language of the speech stimulus was that of the first language of the subject.

[2]Each recording session was performed on different days.

[3]The speech stimulus used was identical for each subject.

[4]The toolbox for the SVM code is available from [12].

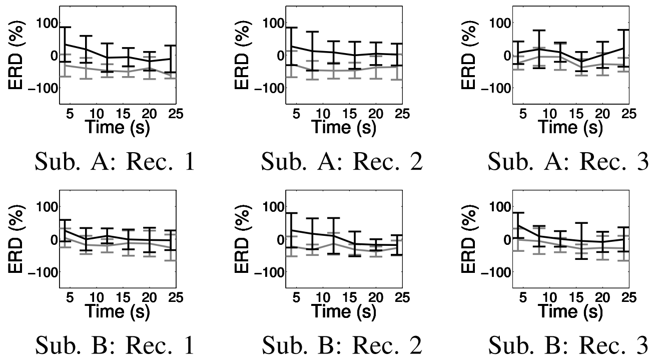

Fig. 5. ERD/ERS results for subjects 'A' and 'B' for three recording sessions (denoted by Rec. #). For each session, the average results for the 10 speech trials are shown in gray and the average for the 10 music trials in black. The distance between the error bars denote two standard deviations.

V. RESULTS

ERD/ERS results for the proof of principle experiments are shown in Fig. 5. Each subplot shows the average ERD/ERS values across time for 10 music trials and 10 speech trials (disregarding some trials as outliers[5]). Each subplot shows that different degrees of event related synchronisation exists in the speech and music trials for the two subjects.

In the case of the larger experimental study, classification performances for the eight subjects using the BEMD synchronisation features combined with normalised power spectrum features are given for various electrode pairs in Table II. The highest classification performance was obtained with the electrode pair FC5/FC4 which gave a median classification performance of 71%.

TABLE II
CLASSIFICATION RATES FOR 8 SUBJECTS USING BEMD SYNCRONISATION.

Electrode pair \ Subject	A	B	C	D	E	F	G	H
FC5/FC6	77	65	66	71	62	77	82	67
FC5/FC4	77	61	69	69	82	73	76	54
FC5/C4	79	71	67	69	66	72	77	59
FC5/T8	79	65	66	73	64	71	75	60
T7/T8	68	57	54	68	67	54	77	88

We can draw several conclusions from the results.

- The input (the sound mixture) was the same across all trials, but the instruction was different, causing different cognitive responses.
- The attended stimulus can be estimated from the EEG with a probability which is significant.
- Mirroring the original Yarbus experiment, it can be stated that the recorded EEG signals are influenced by the cognitive layers of the brain.

[5] In the case of subject 'A', 1 trial was disregarded from rec. session 1. In the case of subject 'B', 2 trials were disregarded from rec. session 1, 1 trial from rec. session 2 and 1 trial from rec. session 3.

However, this has been an initial study which needs to be elaborated on with further experiments. Several sources of ambiguity are listed below.

- Speech was dominant in the right loudspeaker and music in the left in all trials. The significant difference observed in the data could be a spatial difference between attending a source on the left or right.
- The significant difference observed in the data may be due to a difference in attentional load between the two instructions, meaning that the focus of attention is more difficult and straining in one case than in the other.
- It is possible that other subject activity during the act of attending the stimuli (tapping feet, humming) has influenced the recordings. This source of error is also mentioned in [2].

VI. CONCLUSION

The auditory Yarbus-style experiment provides a formal methodology to assess if auditory attention in a complex acoustical environment is observable. The measurement data obtained within this framework show that selective auditory attention to different sources cause systematic changes in brain wave measurements, a conclusion that can only be drawn because of the rigorous design of the experiment. Insofar as cognitive processes remain unobservable by brain wave measurements, the results of the experiment support the conjectured feedback mechanism from the cognitive layer to the layers of auditory sensation and/or perception.

REFERENCES

[1] W. Nager, O. Rosenthal, I. Bohrer, W. A. Teder-Salejarvi, and T. F. Munte, "Human event-related potentials and distraction during selective listening," *Neuroscience Letters*, vol. 297, pp. 1–4, 2001.
[2] N. J. Hill, T. N Lal, K. Bierig, N. Birbaumer, and B. Schölkopf, "An auditory paradigm for brain-computer interfaces," *Advances in Neural Information Processing Systems 17*, pp. 569–576, 2005.
[3] M. G. J. Kallenberg, "Auditory selective attention as a method for a brain computer interface," M.S. thesis, Master's thesis at the Department Music, Mind, Machine at the NICI, Radboud University, Nijmegen, 2006.
[4] A. L. Yarbus, *Eye Movements and Vision*, New York:Plenum, 1967.
[5] T. Womelsdorf and P. Fries, "The role of neuronal synchronization in selective attention," *Current Opinion in Neurobiology*, vol. 17, pp. 154–160, 2007.
[6] J. Fell, G. Fernandez, P. Klaver, C. E. Elger, and P. Fries, "Is synchronized neuronal gamma activity relevant for selective attention?," *Brain Research Reviews*, vol. 42, pp. 265–272, 2003.
[7] D. Looney, C. Park, X. Yili, P. Kidmose, M. Ungstrup, and D. P. Mandic, "Towards estimating selective auditory attention from EEG using a novel time-frequency-synchronisation framework," in *Proceedings of the 2010 International Joint Conference on Neural Networks (accepted)*, 2010.
[8] M. K. Tanenhaus and S. Brown-Schmidt, "Language processing in the natural world," *Philosophical Transactions of the Royal Society B*, vol. 363, no. 1493, pp. 11051122, 2008.
[9] C. C. Wood, W. R. Goff, and R. S. Day, "Auditory evoked potentials during speech perception," *Science*, vol. 173, pp. 1248–1251, 1971.
[10] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, pp. 1842–1857, 1999.
[11] G. Rilling, P. Flandrin, P. Goncalves, and J. M. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, pp. 936–939, 2007.
[12] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "SVM and kernel methods Matlab toolbox," Perception Systemes et Information, INSA de Rouen, Rouen, France, 2005.