Quantifying cognitive aging and performance with at-home gamified mobile EEG

Brian Murphy*†‡, Andrea Aleni§, Brahim Belaoucha§, John Dyer*, and Hugh Nolan§
* BrainWaveBank Ltd., 32-36 Dublin Road, Belfast, Northern Ireland
† Email: brian@brainwavebank.com
‡ EEECS, Queen's University Belfast, 18 Malone Road, Belfast, Northern Ireland
§ BrainWaveBank Ltd., One George's Quay, Dublin, Ireland

***Abstract*—Cognitive and clinical neuroscience often rely on small datasets, gathered from restricted populations in lab settings only. Dry-sensor EEG is a technology that can be made affordable and easy to use. When combined with gamification of experimental tasks, this could provide a new way to gather data on a large scale. Signal quality is a challenge, but the ease of gathering repeated samples can compensate for this. In this paper we present a new system that uses a novel wearable EEG headset, a tablet-based suite of cognitive games, and a cloud-based data analysis system. A large 12-week at-home feasibility study is described, yielding high adherence and user acceptability, even among older users. Data from that trial replicates well-established EEG and behavioural patterns from the cognitive neuroscience literature. Some initial machine learning analyses show that user age and cognitive performance level can be discriminated from single daily sessions, with much higher classification performance when aggregating over multiple days.**

***Index Terms*—wireless dry EEG, machine learning, at-home studies, cognitive health, gamification**

I. Introduction

Cognitive and clinical neurosciences often face practical limitations in how often and how widely brain activity can be sampled. In basic science this has led to questions about how far lab-derived results can be generalized: for instance, would similar effects be observed in normal daily environments, with ordinary daily tasks? [1] And how representative are university-recruited users of the broader population?

In clinical practice and research, many neurocognitive disorders emerge very gradually, or in sporadic episodes, making them difficult to detect in infrequent consultations – the 'snapshot' problem [2]. Both research communities have expressed concerns about the lack of statistical power inherent in small samples, and the publication bias that can result [3].

Conventional lab behavioural tasks are not always designed for repeated use, but gamification has been applied successfully to engage users longitudinally, and reveal changes in core cognitive functions with age [4]. EEG is a well-established, mobile and relatively simple neuroimaging technology that can be miniaturized, with the potential for large scale, low cost manufacture. Consumer devices have been used on a large scale to examine cognitive aging through band-power measures [5], though reports of mobile-recorded Event Related Potentials (ERPs) are more mixed [6, 7]. Research grade

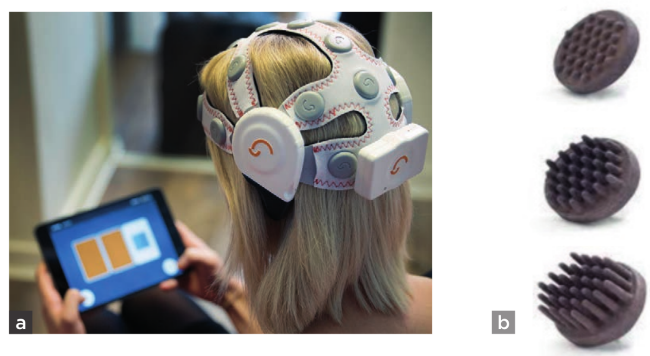

Fig. 1. 16-sensor headset (a), with soft polymer *Ag/AgCl* comb sensors (b)

mobile-EEG systems can yield excellent data, but are not designed for unsupervised use in the home, nor priced at a level that makes this viable [8].

In this paper, we present a new piece of EEG hardware, designed to be used by non-experts in the home and other unsupervised environments. We report on a large scale at-home field study of the technology, that sought 90 older adults to use the headset and a suite of gamified tasks on a quasi-daily basis over 3 months. Adherence, system usability and signal quality were assessed.

Of the large dataset collected (over 3 thousand sessions and 10 thousand individual pieces of task-driven EEG) we present a subset to demonstrate the feasibility of such an approach. We show that gamified versions of common tasks from experimental electrophysiology can yield expected behavioural and EEG effects, even when used by ordinary people alone in the home. We also report some preliminary machine learning analyses to predict individual characteristics: user age, and performance on a standard clinical screening test (MoCA [9]). Finally, we show the power of aggregation over multiple daily samples, using a bagging approach to substantially reduce error rates.

II. Methodology

A. Dry mobile EEG headset

BrainWaveBank have developed a wireless 16-channel dry sensor EEG headset (Fig. 1a), suitable for use in non-lab environments by non-experts. The analogue front-end is based on the ADS1299 chip-set from Texas Instruments [10], incorporating high input impedance of $1G\Omega$, a configurable driven

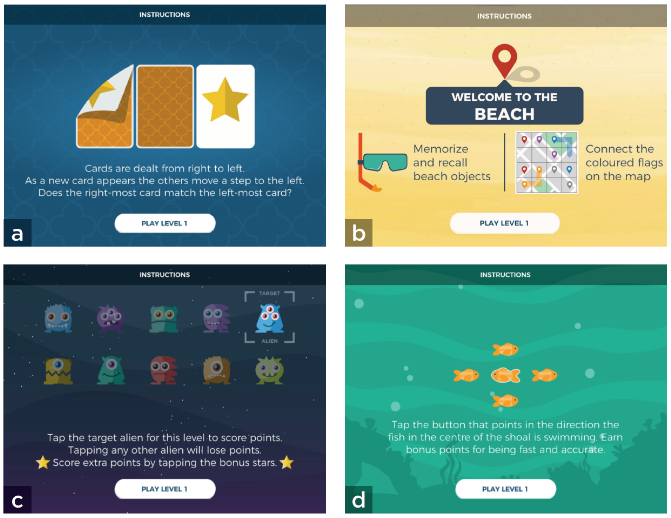

Fig. 2. Screen-shots of the four cognitive games used. In order, the N-Back (a), the visual DMTS (b), the Oddball (c) and the Flanker (d)

bias function for common-mode rejection, built-in impedance checking, and configurable gain and sampling rates. An on-board processor and Bluetooth module transmit 250Hz EEG data to another device (an Android tablet in this case), from where it is transferred to a cloud server for storage and processing. EEG recording and behavioural events are synchronised to $\pm 2ms$.

Flexible *Ag/AgCl* coated polymer sensors of a comb-design (Fig. 1b, ANT-Neuro/eemagine GmbH, [11]) are used to achieve a stable and dermatologically safe contact to the scalp, also through hair. The electronics and sensors are mounted on a flexible neoprene net for comfort and ease of montage. The incorporation of natural landmarks in the headset form-factor and the stretchable structure both enable consistent placement by non-experts in line with the 10-20 sensor system.

B. Tablet-based Gamified Cognitive Tasks

A series of tablet-based apps have been developed to implement well-known tasks from experimental electro-physiology in the form of engaging games, while covering a range of core cognitive functions. Currently, the suite of tasks (Fig. 2) covers: working memory (N-Back [12]); short-term memory (Delayed-Match-to-Sample [13]); attention and decision making (2-Stimulus Oddball [14]); inhibition and error awareness (Erikson Flanker [15]); and idling (resting state [16]). Individual task duration ranged from 2 to 7 minutes.

C. Users and Study Design

A total of 89 healthy older adults were recruited through a range of traditional and social media channels. Ages ranged from 40 to 79 (mean 58.8yrs, SD 8.8yrs) with a female/male mix of 69/20. We controlled for cognitive health using the MoCA cognitive screening test (threshold $\geq$24, [9]) and self-reports of current or historical neurological or psychiatric conditions. The study received ethical committee approval from Queen's University Belfast (QUB). Eleven users withdrew from the study over the 3-month participation.

Each user was asked to participate in a 12-week programme of regular at-home use of the technology, of 5 daily sessions of $\sim$20 minutes, each week. The first session of each user was collected at QUB premises and comprised all four games and the resting-state task. Users were then asked to complete a series of home recorded daily sessions, which each included a lifestyle/well-being self-report questionnaire, followed by a cycling selection of two cognitive games, and the resting-state.

D. EEG Preprocessing

Data was processed using our proprietary data processing suite, which is designed to quantify signal quality, and selectively aggregate individual trials to deal with the noise and signal variation seen in at-home dry-sensor recordings. Filtering was performed between 0.5 - 30Hz. Time-frequency analysis (short term Fourier transform with robust windowing) was used to extract a time-series of alpha band power across resting state in eyes-open and eyes-closed conditions. For ERPs, peak and area-based techniques were used to quantify the amplitude and timing of specific components [17].

E. Classification Analyses

In this paper we present analyses to establish that home-recorded data carries information that can make useful discriminations between individual users. As test-cases we look at arbitrary separated bands of age (40-55yrs vs 60-79yrs), and MoCA score (24-26 vs 28-30). Several linear and non-linear classifiers are used [18], and as input data we examine the utility of: (1) ERP waveforms from a single game task; (2) a broad set of EEG and behaviour derived metrics from several games on a single day; (3) aggregation of classifier estimates over multiple days.

The ERP waveform analysis uses an L1/L2-regularized (Elastic Net) classifier, with input data of 16 channels x 250Hz x 1.5s epoch length ($p = 6016$), over $n = 1196$ input session samples.

Combining derived metrics from brain activity and behaviour across multiple game tasks should be more informative, but poses a "large p, small n" learning challenge, exacerbated by the temporal and spatial autocorrelation properties and noise of multichannel EEG data. We selected $p = 82$ discrete metrics suggested by the literature including gender, time of day of recording, session number (for order and learning effects), aggregate behavioural measures (e.g. average reaction time), band power estimates, and summary ERP statistics (e.g. component peak and amplitude measures [17]). For this analysis results for both L1/L2-regularised Logistic Regression and Random Forests are reported.

For aggregation over multiple days, a straight-forward bagging approach was used that classified based on the cumulative probability up to that day, taken as the mean of daily classifier estimates. In all analyses, nested 5-fold cross-validation was used for a grid search over parameters; and sessions were partitioned block-wise so that each user's data was assigned to a single fold, never spanning testing and training partitions.

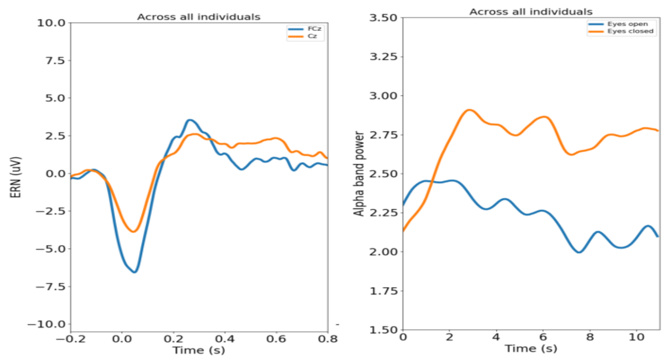

Fig. 3. Grand average Error-Related Negativity for two centre-line sensor locations (left) and first 10s of resting-state alpha power (right).

TABLE I
CLASSIFICATION USING ERP DATA FROM A SINGLE GAME SESSION

Oddball		Flanker	
Age	MoCa	Age	MoCA
70%	64%	67%	58%

III. RESULTS

A. Usability, adherence, data quantity

A total of 3,839 daily sessions were recorded, comprising 10,221 task-specific EEG recordings. This corresponds to an average adherence rate of 4 sessions per person per week. At the conclusion of the study, average judged system usability was 77.4 (corresponding to ‘good’/‘excellent’, on the System Usability Scale [19]). On an average session, 90% (i.e. 14 to 15) channels achieved a successful contact. All these metrics were similar for both younger and older (>65yrs) sub-cohorts.

B. Grand average EEG, band power and behaviour

Expected EEG and behavioural patterns were observed as predicted by literature from lab research. For instance, on the Flanker task, the ‘match’ condition had a faster reaction time (473ms grand average), compared to the ‘mismatch’ condition, which demands inhibition by the user (517ms). Its grand average ERP showed the expected Error Related Negativity, with a strong component at central and fronto-central locations, about 50ms after erroneous user responses (Fig. 3, left panel). On the resting state task, eyes-closed alpha power was reliably higher than eyes-open power (Fig. 3, right panel).

C. Classifying Individual Characteristics with ERP data, combined metrics, and multi-day aggregation

Table I demonstrates that estimates of cognitive performance and age can be made for individuals, based on a single game-driven recording of several minutes of EEG. Using this regularised linear classifier, Age determination appeared somewhat easier to estimate, and the Oddball task carried more information.

Table II shows corresponding results using the combination of ERP-derived metrics, and aggregate measures of behaviour from a range of daily tasks (total <15min of task/EEG time), and compares a linear and non-linear classifier. Compared to the previous model (based on a single game ERP at a time), session-wise classification of Age could be improved substantially, by more than 10 percentage points, reducing error rates by about a third, while maintaining a good balance between sensitivity and specificity. For cognitive performance (MoCA) there was no clear improvement in this case.

TABLE II
CLASSIFICATION USING COMBINED METRICS FROM A DAILY SESSION

Metric	Random Forest		Logistic Regression	
	Age	MoCa	Age	MoCA
N	1006	771	1006	771
Accuracy	82%	59%	78%	60%
AUC	0.82	0.57	0.78	0.60
Sensitivity	78%	40%	79%	60%
Specificity	85%	74%	78%	58%

TABLE III
CLASSIFICATION USING CUMULATIVE EVIDENCE FROM MULTIPLE DAYS

Cum. Weeks	Random Forest		Logistic Regression	
	Age	MoCA	Age	MoCA
1	77%	59%	83%	57%
2	85%	65%	86%	59%
3	90%	58%	87%	62%
4	90%	66%	83%	66%

Table III illustrates the effect of aggregating information from multiple daily sessions, with a simple bagging approach. Again, Age proves easier to classify. Taking all days cumulatively up to the end of week 2 (10 daily sessions) yields a modest improvement, but by the end of week 3 the error rate has fallen to about 10% for the Random Forest classifier. Again MoCA showed less clear improvement.

Fig. 4 illustrates how information builds up over days to improve prediction for individuals. In many, but not all cases, as evidence accumulates, users who were misclassified (younger users in the top left quadrant, older users in the bottom right quadrant), trend towards their correct classification.

IV. DISCUSSION AND CONCLUSION

In this paper we present a new system and approach to regularly gather research-quality EEG and behavioural data from ordinary people in their homes, using dry-sensor mobile technology. We report a suite of gamified tasks we have developed to engage users, while probing specific aspects of their cognition. In a field study adherence remained high through-out the 12-week study (averaging 4 sessions per week), and we gathered a large database of task-specific EEG data. Initial analyses of ERP, band power, and behavioural data confirmed that a range of effects from the literature (event related negativity, incongruency effect on latency, and alpha reactivity) were replicated in home-gathered data.

We present some preliminary classification analyses as proof of principle, that home-collected dry-EEG can be used to make determinations about individual users – in this case their age, and cognitive performance on a commonly used clinical screening tool. ERP waveforms from single games carry valuable information, and can be supplemented by game behaviour

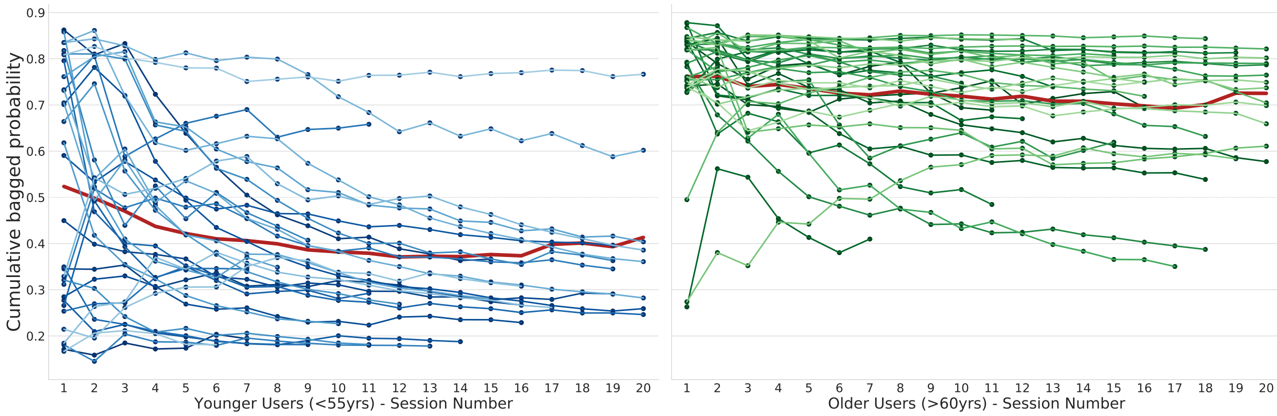

Fig. 4. Cumulative bagging of daily data: blue and green lines represent individual younger and older users respectively. Red traces show mean bagged probability for each true class.

metrics. Crucially, aggregating evidence from multiple daily samples substantially increased classification performance – for instance prediction of age scale rose to 90% using 3 weeks worth of sessions.

On-going and future work will include more granular data in the input (despite the "large p, small n" problem that presents), and this becomes more feasible as our database grows. In diagnostic and screening contexts classifiers are often appropriate to give clinical professionals simple guidance for decision making, but regression models will allow us to look at these individual characteristics as continuous variables. Including participant information into models, in a way that does not artificially bias results, is another challenge we are working on.

We have a particular interest in understanding how age affects cognitive performance, and how its measurement can aid in the tracking of cognitive decline and early detection of neurodegeneration. More generally, the platform could be applied to any scientific question that would benefit from the sampling of recordings of brain activity from larger numbers of people, on multiple occasions, in more ecological settings

ACKNOWLEDGMENTS

This study was funded by the Innovate UK Biomedical Catalyst programme. A. Buick contributed to study design, and data collection was carried out by A. Buick, L. Watson and A. Sweeney. B. McGuinness and P. Passmore provided clinical oversight.

REFERENCES

[1] Donna M. Spooner and Nancy A. Pachana. "Ecological validity in neuropsychological assessment". In: *Archives of Clinical Neuropsychology* 21.4 (2006), pp. 327–337.

[2] Saul Shiffman et al. "Ecological Momentary Assessment". In: *Annual Review of Clinical Psychology* 4.1 (2008), pp. 1–32.

[3] Alexander A. Aarts et al. "Estimating the reproducibility of psychological science". In: *Science* 349.6251 (2015). eprint: arXiv:1011.1669v3.

[4] Fiona McNab et al. "Age-related changes in working memory and the ability to ignore distraction". In: *Proceedings of the National Academy of Sciences* 112.20 (2015), pp. 6515–6518.

[5] Ali Hashemi et al. "Characterizing Population EEG Dynamics throughout Adulthood". In: *eNeuro* 3.6 (2016).

[6] Stefan Debener et al. "How about taking a low-cost, small, and wireless EEG for a walk?" In: *Psychophysiology* 49.11 (2012), pp. 1617–1621.

[7] Anthony J. Ries et al. "A Comparison of Electroencephalography Signals Acquired from Conventional and Mobile Systems". In: *Journal of Neuroscience and Neuroengineering* 3.1 (2014), pp. 10–20.

[8] Vojkan Mihajlovi et al. "Wearable, Wireless EEG Solutions in Daily Life Applications: What are we Missing?" In: *Ieee Journal of Biomedical and Health Informatics* 19.1 (2015), pp. 6–21.

[9] Ziad S. Nasreddine et al. "The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment". In: *Journal of the American Geriatrics Society* 53.4 (), pp. 695–699.

[10] Texas Instruments. *Low-noise, 8-channel, 24-bit analog front-end for biopotential measurements ADS1299*. 2012.

[11] P Fiedler et al. "Modular multipin electrodes for comfortable dry EEG". In: August (2016), pp. 5705–5708.

[12] Shanshan Dong et al. "Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty". In: *Brain Research* 1616 (2015), pp. 146 –156.

[13] Axel Mecklinger. "On the modularity of recognition memory for object form and spatial location: a topographic ERP analysis". In: *Neuropsychologia* 36.5 (1998), pp. 441 –460.

[14] Christoph S Herrmann and Robert T Knight. "Mechanisms of human attention: event-related potentials and oscillations". In: *Neuroscience & Biobehavioral Rev.* 25.6 (2001), pp. 465–476.

[15] Nick Yeung et al. "The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity". In: 111 (Nov. 2004), pp. 931–59.

[16] Christi E. D. Alloway et al. "The Alpha Attenuation Test: Assessing Excessive Daytime Sleepiness in Narcolepsy-Cataplexy". In: 20 (May 1997), pp. 258–66.

[17] Steven Luck. *An Introduction to The Event-Related Potential Technique*. Jan. 2005.

[18] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12 (2011), pp. 2825–2830.

[19] John Brooke. "SUS: A quick and dirty usability scale". In: 189 (Nov. 1995).