BRAININFO 2021

The Sixth International Conference on Neuroscience and Cognitive Brain Information

ISBN: 978-1-61208-885-3

July 18 – 22, 2021

Nice, France

BRAININFO 2021 Editors

Ricardo Ron-Angevin, University of Málaga, Spain
Rory Lewis, University of Colorado, Colorado Springs, USA
Manuela Popescu, IARIA, EU/USA

BRAININFO 2021

Forward

The Sixth International Conference on Neuroscience and Cognitive Brain Information (BRAININFO 2021) continued a series of events dedicated to evaluating current achievements and identifying potential ways of making use of the acquired knowledge, covering the neuroscience, brain connectivity, brain intelligence paradigms, cognitive information, and specific applications.

The complexity of the human brain and its cognitive actions stimulated many researches for decades. Most of the findings were adapted in virtual/artificial systems in the idea of brain-like modeling them and used in human-centered medical cures, especially for neurotechnologies. Information representation, retrieval, and internal data connections still constitutes a domain where solutions are either missing or at a very early stage.

We take here the opportunity to warmly thank all the members of the BRAININFO 2021 technical program committee, as well as all the reviewers. We also kindly thank all the authors who dedicated much of their time and effort to contribute to BRAININFO 2021. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions. We also thank the members of the BRAININFO 2021 organizing committee for their help in handling the logistics of this event.

BRAININFO 2021 Chairs

BRAININFO 2021 Steering Committee
Erwin Lemche, Institute of Psychiatry, Psychology & Neuroscience | King's College School of Medicine and Dentistry, UK
Yu-Dong (Eugene) Zhang, University of Leicester, UK
Chih Lai, University of St. Thomas Minnesota, USA
William Speier, UCLA David Geffen School of Medicine, USA
Sule Yildirim-Yayilgan, Norwegian University of Science and Technology, Norway
Ricardo Ron-Angevin, University of Málaga, Spain
Rory Lewis, University of Colorado at Colorado Springs, USA

BRAININFO 2021 Industry/Research Advisory Committee
Konrad Ciecierski, Institute of Computer Science - Warsaw University of Technology / Clinic of Neurosurgery - Warsaw Institute of Psychiatry and Neurology, Poland
Nikola Kasabov, Auckland University of Technology, New Zealand
Domenico Ursino, University "Mediterranea" of Reggio Calabria, Italy
Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA

BRAININFO 2021 Publicity Chair
Mar Parra, Universitat Politecnica de Valencia, Spain
Alvaro Liebana, Universitat Politecnica de Valencia, Spain

BRAININFO 2021
Committee

BRAININFO 2021 Steering Committee

Erwin Lemche, Institute of Psychiatry, Psychology & Neuroscience | King's College School of Medicine and Dentistry, UK
Yu-Dong (Eugene) Zhang, University of Leicester, UK
Chih Lai, University of St. Thomas Minnesota, USA
William Speier, UCLA David Geffen School of Medicine, USA
Sule Yildirim-Yayilgan, Norwegian University of Science and Technology, Norway
Ricardo Ron-Angevin, University of Málaga, Spain
Rory Lewis, University of Colorado at Colorado Springs, USA

BRAININFO 2021 Industry/Research Advisory Committee

Konrad Ciecierski, Institute of Computer Science - Warsaw University of Technology / Clinic of Neurosurgery - Warsaw Institute of Psychiatry and Neurology, Poland
Nikola Kasabov, Auckland University of Technology, New Zealand
Domenico Ursino, University "Mediterranea" of Reggio Calabria, Italy
Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA

BRAININFO 2021 Publicity Chair

Mar Parra, Universitat Politecnica de Valencia, Spain
Alvaro Liebana, Universitat Politecnica de Valencia, Spain

BRAININFO 2021 Technical Program Committee

Mariano Alcañiz Raya, Institute of Research and Innovation in Bioengineering (I3B) - Universidad Politécnica Valencia, Spain
Saad Alqithami, Albaha University, Saudi Arabia
Gian Carlo Cardarilli, University of Rome Tor Vergata, Italy
German Castellanos-Dominguez, Universidad Nacional de Colombia, Sede Manizales, Colombia
Srijata Chakravorti, Vanderbilt University, USA
Shiaofen Fang, Indiana University Purdue University Indianapolis, USA
Monte Hancock, 4Digital Inc., USA
M. Shamim Kaiser, Jahangirnagar University, Bangladesh
Stamatis Karlos, University of Patras, Greece
Peter Kieseberg, St. Poelten University of Applied Sciences, Austria
Chih Lai, University of St. Thomas Minnesota, USA
Erwin Lemche, Institute of Psychiatry, Psychology & Neuroscience | King's College School of Medicine / Dentistry Department of Psychosis Studies | Section of Cognitive Neuropsychiatry, UK
Rory Lewis, University of Colorado at Colorado Springs, USA
Vasilis Megalooikonomou, University of Patras, Greece
Luis Alfredo Moctezuma, Norwegian University of Science and Technology, Norway
Marta Molinas, NTNU, Norway
Rafael Morales Herrera, Universidad de Castilla-La Mancha, Spain
Ricardo Nuno Vigário, Nova School of Science and Technology, Lisbon, Portugal
Izabela Rejer, West Pomeranian University of Szczecin, Poland

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Brain-Computer Interface Control of Smartphone Messaging Applications

Francisco Velasco-Álvarez, Álvaro Fernández-Rodríguez, Ricardo Ron-Angevin
Departamento de Tecnología Electrónica
Universidad de Málaga
Málaga, Spain
e-mail: {fvelasco, afernandezrguez, rron}@uma.es

***Abstract*— This work-in-progress paper presents an implementation of a Brain-Computer Interface (BCI) system focused on the control of the most common messaging applications of a smartphone: WhatsApp, Telegram, e-mail and Short Message Service (SMS). The control of these applications is achieved through the use of a virtual assistant running in the smartphone. The BCI system is based on the visual Row-Column Paradigm (RCP), which allows users to select several control commands and to spell messages that are converted to synthesized voice and received by the mentioned virtual assistant in the smartphone.**

Keywords- Brain-Computer Interface (BCI); P300; assistive technology; virtual assistant; messaging applications.

I. INTRODUCTION

Brain-Computer Interfaces (BCI) are a type of Assistive Technology (AT) that uses the brain signals of users to establish a communication and control channel between them and an external device (usually a computer) [1]. BCI systems may be a suitable tool to restore communication skills in severely motor-disabled patients, as BCI do not rely on muscular control. There are several diseases that cause severe impairment of motor skills in affected patients, such as Amyotrophic Lateral Sclerosis (ALS). AT can be used to control multiple devices, such as a wheelchair, a home automation system, or a verbal communication system [2]. AT should be able to be controlled through those output channels that the patient still has preserved, such as the voice, the eye gaze, movements of a finger, the head, the cheek or the tongue. However, in severe and progressive motor limitations (as is the case in ALS), most of these examples of AT may no longer be useful because they depend on some type of muscular channel that may be affected in the patient. In these cases, BCI may be a suitable option for those people who have completely lost the ability to move their muscles. The neuroimaging technique most used by BCIs is electroencephalography (EEG), possibly due to its relatively low cost and high temporal resolution [3]. One of the most extended EEG signal used in BCI system is the P300 evoked potential. P300 is a positive potential generally located in the parieto-occipital areas that appears about 250-500 ms (although the range can vary depending on numerous factors) after the presence of both an expected stimulus and a rare one [4]. Usually, the P300 is evoked through an oddball paradigm, in which the available items are highlighted pseudo-randomly while the user pays attention only to the desired item, thus resulting in a P300 potential after the stimulation of this desired item. After a predetermined number of iterations, the system averages the resulting EEG and determines which item the user wanted to select. This concept was adopted by [5] to propose a paradigm to control a text speller.

The loss of communication (mainly with family and caregivers) is considered by ALS patients as even more negative than the loss of physical aspects [6][7]. Therefore, this work will focus on the use of a BCI that could allow patient communication through some of the most common messaging applications on a smartphone: WhatsApp, Telegram, e-mail and Short Message Service (SMS). Other researchers have focused on the BCI control of daily use and domotic applications, for example the work in [8] uses visual event-related potentials to control a TV, an air conditioner and to make an emergency call. The authors in [9] propose a BCI control of Telegram and Twitter in a smartphone. A hybrid BCI (that uses as well electrooculography as input) presented in [10] allowed to control a web browser and an e-mail client. To our knowledge, the present work in progress is the first BCI proposal that allows to control some of the previously used applications by other BCI works (Telegram, e-mail) and WhatsApp and SMS as well.

There are some works that have already explored the idea of using voice commands sent to virtual assistants to facilitate integration between applications. Outside the field of BCI, the work in [11] can be highlighted. The authors used a proximity sensor on the fingers, feet or head (depending on the patient) to select commands in an application with a graphical interface which later allowed text to be converted to speech to verbalise the users' selections. The voice assistant used for this work was Google Assistant. This system allowed to control WhatsApp and YouTube. This work was based on a reduced set of possible actions: for WhatsApp, three contacts to choose from and three predetermined possible messages to send; for YouTube, three possible music/videos alternatives and four alternative-related events. In relation to the works that have used a BCI, the work in [12] has been the only one, to our knowledge, that has used a voice assistant (Amazon's Alexa) to control two devices (a light bulb and a fan).

The objective of this work is to present a communication bridge between the UMA-BCI Speller platform [13] (a BCI software developed by UMA-BCI group of the University of Malaga, UMA) and the messaging services of WhatsApp, Telegram, e-mail and SMS, through the use of Google Assistant on a smartphone. This paper presents a short

 ISBN: 978-1-61208-885-3

version of a full-length work that is currently under review [14].

This paper is organized as follows: section 2 and section 3 describe the system implementation and the control paradigm, respectively. The preliminary results are presented in section 3, followed by the conclusion and future works in section 4.

II. System Implementation

The aim of the BCI system was to generate voice commands that could be interpreted by a virtual assistant running in a smartphone. These voice commands were intended to read and send messages through various messaging services. In order to achieve that, a BCI system was implemented that could generate these commands in text form and convert them into voice. On the one hand, a laptop ran the software that presented the stimuli and registered and analysed the EEG. This software was the UMA-BCI Speller [13], a free tool that wraps BCI2000 [15] and simplifies its configuration and use. This software was used to spell and convert the control commands to voice. On the other hand, a virtual assistant was running on a smartphone. This virtual assistant was Google Assistant and it received and interpreted the voice commands, performing the corresponding action. The system implementation is shown in Figure 1. Through the EEG, the user can select commands from a computer control (in the example, "Read SMS") and spell messages to be sent. These commands and messages are converted to voice and received by the virtual assistant running in the smartphone. The assistant interprets them and performs the corresponding action (in the example, it informs the user that there is a SMS and it reads it).

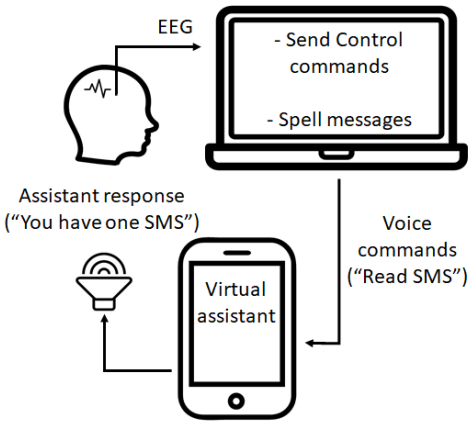


Figure 1. System implementation.

The UMA-BCI Speller includes a text prediction function that may help users when spelling words. As users choose the characters of a word (starting with the first character), the system proposes several predicted words based on the characters already written and the probability of occurrence based on a Spanish language specific corpus.

Once users had completed the spelling of the text command to send, they had to select a confirmation item in the interface so that the system could convert this command into speech. The Windows 10 Narrator (a text-to-speech feature) was used, particularly the voice named "Microsoft Helena" from the Spanish voice catalogue.

To avoid the influence of ambient noise on the understanding of the command by the virtual assistant, the voice commands were sent to the smartphone via a cable connection (using a mini-jack audio cable) connecting the laptop audio output with the smartphone microphone input. The output volume of the laptop was fixed throughout the whole experiment, so the assistant always received the same level of audio.

As the virtual assistant used in the experiment was Google Assistant, each command started with the words "*Ok Google…*", which is one of the wake-up keywords of the assistant. Two main types of voice commands were used: i) commands asking the assistant to read the received messages, e.g., "*Ok Google, read my WhatsApp messages*", and ii) commands asking the assistant to send a message using one of the messaging services installed in the smartphone, e.g., "*Ok Google, send a Telegram to* [contact], [message]". In addition to these, other voice commands were used that were needed to confirm or cancel actions, as we will explain below.

III. Control Paradigm

In order to send a command to the virtual assistant, users had to select items from different menus. The selection of an item followed the usual procedure in a P300 Row-Column Paradigm (RCP): users had to pay attention to the desired item (within a matrix of possible items) and mentally count the number of times it was highlighted. The timing of each selection for all the menus was the same, as all the interfaces consisted of a 7×7 matrix, even though in three of them there were dummy items (items that had no effect when selected). An item was selected after all the seven rows and columns were highlighted a certain number of times.

Four menus were implemented that allowed subjects to gradually form a sentence that would finally be converted to speech by the Windows 10 Narrator voice synthesiser (from now on, this conversion will be denoted as "speak"). The sentence to be spoken was present in the interface so that subjects decided when to indicate to the system to speak it. Some items added several predetermined words to this sentence, while other items were present to add individual letters to spell a message. The four menus are described next:

A. No Control (NC) menu.

This was a 7×7 matrix in which only one item was a valid command, and the other 48 items ("X") were dummy commands. The objective of this menu was to allow subjects to remain in a state where they could rest without generating control commands; the term "no control" is generally used in asynchronous systems to refer to such a state. The only valid command was a control command named "IC" whose selection changed the menu to an Intentional Control (IC) menu.

 ISBN: 978-1-61208-885-3

B. Intentional Control (IC) menu.

This was the main menu of the system, where subjects could choose what action they wanted to select. In a 7×7 matrix, ten valid options were available (the remaining 39 options were dummy non-visible items). This menu is shown in Figure 2. This menu presented ten options that can be grouped into three categories:

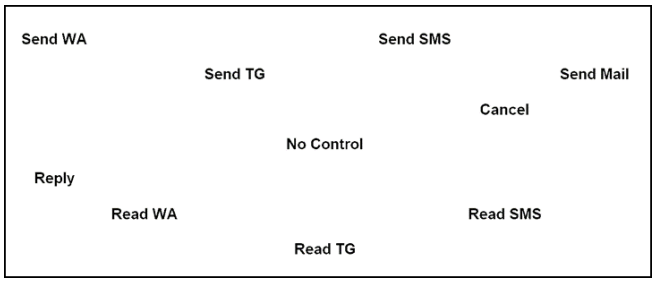


Figure 2. Intentinal Control (IC) menu.

- Send messages. This group consisted of four commands: "Send WA", "Send TG", "Send SMS" and "Send Mail", that enabled users to send a message using WhatsApp, Telegram, SMS or e-mail, respectively. Once one of these commands was selected, the system wrote part of the sentence to be spoken, "*Ok Google, send a WhatsApp to*", "*Ok Google, send a Telegram to*", "*Ok Google, send a SMS to*" or "*Ok Google, send an e-mail to*" and then changed to a Spelling menu (Figure 3), so that the user could next spell out the receiver of the message and the message itself.
- Read messages. Three commands formed this group: "Read WA", "Read TG" and "Read SMS" used to read the messages received through WhatsApp, Telegram and SMS, respectively. The selection of one of these commands made the system speak the corresponding sentence: "*Ok Google, read my WhatsApp messages*", "*Ok Google, read my Telegram messages"* or "*Ok Google, read my SMS messages*". After the sentence was spoken, the system deleted it and automatically changed to the NC menu so that subjects could listen to the received messages, if any. After the virtual assistant read each received message, it asked the users if they wanted to reply to it or not. To do this, subjects first had to change to the IC menu where two commands were available related to replying to messages. These commands will be explained in the third group. After cancelling or replying to each message, the system continued to read the remaining messages, if any.
- Other commands. In this group, three commands were included: "Reply", "Cancel" and "No Control". The "Reply" command allowed users to reply to a WhatsApp, Telegram or SMS received message after the system read them. Once this command was selected, the system wrote the sentence "*Ok Google, reply*" and changed to the Spelling menu so that users could complete the sentence with the desired response (in a similar way to what was done with the "Send" commands); please note that in this case it is not necessary to specify the receiver of the message to be sent. The "Cancel" command was used to indicate to the system that the user did not want to reply to a received message. Once it was selected, the system wrote and spoke the sentence "*Ok Google, cancel*" and then deleted it and changed to the NC menu. Finally, the "No Control" command was presented in order to allow subjects to voluntarily change to the NC menu, in case they wanted to take a rest.

C. Spelling menu.

When users selected one of the "Send" or "Reply" options from the IC menu, the system changed to the Spelling menu (after adding some predetermined text). This menu is shown in Figure 3. Here, the users could spell out

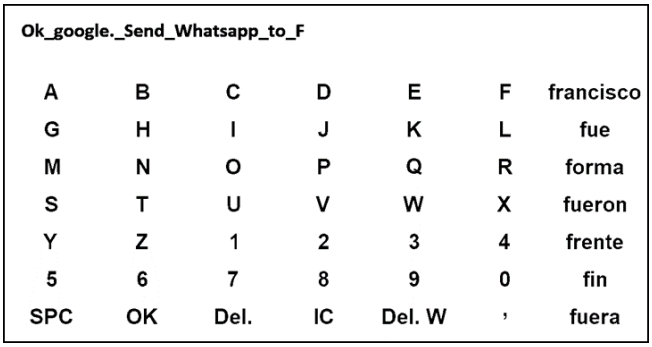


Figure 3. Spelling menu with 38 characters to spell, four control commands and seven available predicted words (last column, in Spanish)

the receiver and the message to send (or just the message in the case of the option "Reply"). It is worth mentioning that the message had to be spelled right after the receiver, only separated by a space. This menu consisted of a 7×7 matrix with spelling and control commands. The first six columns and rows corresponded to specific characters to be added (English alphabet letters and numbers). The last column was used to provide subjects with seven predicted words. The last row contained two characters ("SPC" (space) and ","), two delete commands ("Del." to delete a single character and "Del. W" to delete a complete word) and two control commands ("OK" and "IC"). The command "OK" was used to indicate to the system that the receiver (if needed) and the message to send were complete so that the written sentence could be spoken and interpreted by the virtual assistant. The "IC" command was used to return to the IC menu without generating any voice command (this was useful if a subject entered this menu unintentionally). The selection of "OK" or "IC" caused the current written sentence to be deleted (after speaking it in the case of "OK"), so a confirmation menu

 ISBN: 978-1-61208-885-3

was offered to subjects in order to avoid undesired selections of these two commands.

D. Confirmation menu.

Two valid commands were available (among other 47 non-visible dummy options) in a 7×7 matrix, "Confirm" and "Back". On the one hand, the "Confirm" command was used to corroborate the previous selection in the Spelling menu (that is, "OK" or "IC"). In the case of confirming an "OK" command, the system spoke (and deleted) the complete sentence so it could be interpreted by the virtual assistant, and it changed to the NC menu. In the case of confirming an "IC" command, the system deleted the written sentence and changed to the IC menu. On the other hand, the "Back" command was used to return to the Spelling menu in order to continue writing the sentence to be sent to the virtual assistant.

IV. RESULTS

Some preliminary tests have been carried out. In these, some healthy volunteers were asked to perform four tasks related to the four messaging applications: i) send a WhatsApp message with a predetermined message; ii) read the incoming SMS; iii) read an incoming Telegram message and reply to it with a free answer; and iv) send a free text e-mail to a contact chosen by them. The preliminary online results obtained, as well as the results of some questionnaires related to the subjective experience controlling the interface, support the viability of the proposed system.

V. CONCLUSION AND FUTURE WORK

The use of a virtual assistant to control the smartphone makes it possible to easily extend the functionality to other applications beyond messaging services, for example to control domotic devices through the smartphone. However, the use of a virtual assistant also presented some misinterpretation problems when the synthesized voice was not correctly understood by the assistant. Other issue related to the use of the virtual assistant arose when the assistant responded to a command in an unexpected way, as the system was based in a one-way communication (from the laptop to the smartphone), taking for granted that the virtual assistant would interpret this command correctly.

The future work based on this study is related to the extension of the functionality to domotic features, the improvement of the one-way communication from the BCI application to the virtual assistant, and the possibility of testing the system with motor-disabled patients.

ACKNOWLEDGMENT

This work was partially supported by the project SICCAU: RTI2018-100912-B-100 (MCIU/AEI/FEDER, UE) and by the University of Malaga ("Universidad de Málaga").

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, 2002, doi: 10.1016/S1388-2457(02)00057-3.

[2] R. Jamwal, H. K. Jarman, E. Roseingrave, J. Douglas, and D. Winkler, "Smart home and communication technology for people with disability: a scoping review," Disabil. Rehabil. Assist. Technol., vol. 0, no. 0, pp. 1–21, 2020, doi: 10.1080/17483107.2020.1818138.

[3] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," Sensors, vol. 12, no. 2, pp. 1211–1279, 2012, doi: 10.3390/s120201211.

[4] J. Polich, "Updating P300: An integrative theory of P3a and P3b," Clin. Neurophysiol., vol. 118, no. 10, pp. 2128–2148, Oct. 2007, doi: 10.1016/J.CLINPH.2007.04.019.

[5] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prothesis utilizing event-relatedpotencials," Electroencephalogr. Clin. Neurophysiol., vol. 70, no. 6, pp. 510–523, 1988.

[6] Z. Simmons, B. A. Bremer, R. A. Robbins, S. M. Walsh, and S. Fischer, "Quality of life in ALS depends on factors other than strength and physical function," Neurology, 2000, doi: 10.1212/WNL.55.3.388.

[7] S. H. Felgoise, J. L. Stewart, B. A. Bremer, S. M. Walsh, M. B. Bromberg, and Z. Simmons, "The SEIQoL-DW for assessing quality of life in ALS: Strengths and limitations," Amyotroph. Lateral Scler., vol. 10, no. 5–6, pp. 456–462, 2009, doi: 10.3109/17482960802444840.

[8] K. T. Sun, K. L. Hsieh, and S. R. Syu, "Towards an accessible use of a brain-computer interfaces-based home care system through a smartphone," Comput. Intell. Neurosci., vol. 2020, pp. 16–18, 2020, doi: 10.1155/2020/1843269.

[9] V. Martínez-Cagigal, E. Santamaría-Vázquez, J. Gomez-Pilar, and R. Hornero, "Towards an accessible use of smartphone-based social networks through brain-computer interfaces," Expert Syst. Appl., vol. 120, pp. 155–166, 2019, doi: 10.1016/j.eswa.2018.11.026.

[10] S. He et al., "EEG- And EOG-Based Asynchronous Hybrid BCI: A System Integrating a Speller, a Web Browser, an E-Mail Client, and a File Explorer," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 28, no. 2, pp. 519–530, 2020, doi: 10.1109/TNSRE.2019.2961309.

[11] G. E. Lancioni et al., "Mainstream technology to support basic communication and leisure in people with neurological disorders, motor impairment and lack of speech," Brain Inj., vol. 00, no. 00, pp. 1–7, 2020, doi: 10.1080/02699052.2020.1763462.

[12] V. K. K. Shivappa, B. Luu, M. Solis, and K. George, "Home automation system using brain computer interface paradigm based on auditory selection attention," I2MTC 2018 - 2018 IEEE Int. Instrum. Meas. Technol. Conf. Discov. New Horizons Instrum. Meas. Proc., pp. 1–6, 2018, doi: 10.1109/I2MTC.2018.8409863.

[13] F. Velasco-Álvarez, S. Sancha-Ros, E. García-Garaluz, Á. Fernández-Rodríguez, M. T. Medina-Juliá, and R. Ron-Angevin, "UMA-BCI Speller: An easily configurable P300 speller tool for end users," Comput. Methods Programs Biomed., vol. 172, pp. 127–138, Mar. 2019, doi: 10.1016/j.cmpb.2019.02.015.

[14] F. Velasco-Álvarez, A. Fernández-Rodríguez, F. J. Vizcaíno-Martín, A. Díaz-Estrella, and R. Ron-Angevin, "Brain-computer interface (BCI) control of a virtual assistant in a smartphone to manage messaging applications.", unpublished.

[15] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," IEEE Trans. Biomed. Eng., vol. 51, no. 6, pp. 1034–1043, 2004, doi: 10.1109/TBME.2004.827072.

 ISBN: 978-1-61208-885-3

Combining Self-Organizing Maps and Decision Tree to Explain Diagnostic Decision Making in Attention-Deficit/Hyperactivity Disorder

Anderson Martins Silva
Graduate Program
in Electrical and Computer Engineering
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: 71953371@mackenzista.com.br

Luiz Renato Rodrigues Carreiro
Graduate Program
in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: luizrenato.carreiro@mackenzie.br

Mayara Miyahara Moraes Silva
Graduate Program
in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: 72054298@mackenzista.com.br

Maria Cristina Triguero Veloz Teixeira
Graduate Program
in Developmental Disorders
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: mcris@mackenzie.br

Leandro Augusto da Silva
Graduate Program
in Electrical and Computer Engineering
Mackenzie Presbyterian University
São Paulo, Brazil
e-mail: leandroaugusto.silva@mackenzie.br

***Abstract*—Attention-Deficit/Hyperactivity Disorder (ADHD) presents in children and adolescents as a persistent pattern of inattention, hyperactivity, and impulsivity that interferes with their development. Computational studies on ADHD focus on measures of brain activity of the participants and a few use standardized cognitive tests or behavioral inventories to assess objective indicators for diagnosis. The paper presents a computational proposal in which the combination of two artificial intelligence methods is used to aid the identification of diagnostic indicators for ADHD. The proposal is to combine a neural network of self-organizing maps to group factors from standardized tests and inventories, and a decision tree to classify the most relevant factors. The study included 127 children and adolescents from 6 to 16 years old, 48 with ADHD diagnosis and 79 without ADHD (control group). The most relevant result of the study was the strong contribution of the scores of the Inventory of Behaviors for Children and Adolescents in the diagnosis of the disorder.**

***Keywords*—Self-Organizing Maps (SOM); Decision Tree; Attention Deficit/Hyperactivity Disorder (ADHD).**

I. INTRODUCTION

According to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition - DSM-5 [1] Attention-Deficit/Hyperactivity Disorder (ADHD) is a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development. The disorder is characterized by inattention involving, for example, difficulty sustaining attention in tasks or play activities, a state in which the mind seems elsewhere, even in the absence of any obvious distraction, difficulty to follow through with instructions and failing to finish schoolwork, often forgetful in daily activities, chores, or duties in the workplace, losing things, expressing excessive activity or restlessness, and inability to wait one's turn, always in ways that are excessive for one's age or developmental level. ADHD has its initial expressions in childhood and usually persists into adulthood, resulting in impairments in social, academic, and occupational functioning.

The diagnosis of ADHD is clinical, based on the individual's history and expression of symptoms. Because this diagnosis is often based on reports of symptom severity and because these symptoms are also part of other clinical conditions, the diagnostic difficulty is present in the daily lives of the interdisciplinary teams responsible for the evaluation process [2] [3]. Because of the complexity of the diagnostic evaluation, the American Association of Pediatrics recommends the use of an algorithm, both for evaluation and treatment of children and adolescents with ADHD [4]. To support clinical decision making, neuropsychological, behavioral, and adaptive functioning assessment procedures have often been used in conjunction with neurological assessments [5]. Considering the social importance involved in properly issuing a correct diagnosis of ADHD in both children and adolescents, studies must be proposed that discuss which are the best indicators of clinical-neurological, neuropsychological, and behavioral-adaptive diagnostic evaluation when children and adolescents present with complaints of inattention and hyperactivity. Furthermore, for appropriate assessments and interventions to be implemented, differential criteria are needed to correctly characterize and identify attention-deficit/hyperactivity among children and adolescents. Comprehensive assessments in this regard allow a better understanding of the complexity of each case for appropriate guidance, design of the therapeutic intervention, and evaluation of the need for educational and emotional support for patients and families [5].

Computational studies can help professionals in diagnostic assessments, especially using machine learning algorithms. Kam et al. [7] used an artificial intelligence algorithm called

 ISBN: 978-1-61208-885-3

decision tree for screening ADHD by monitoring the school activities of 153 children using 3-axial actigraph and obtained results consistent with previous studies. In turn, Lee et al. [8] analyzed the classification of ADHD in children through brain activity measurements. In their work, they used a neural network algorithm called self-organizing maps allowing categorizing characteristics of children with and without clinical indicators of ADHD.

Unlike previous proposals presented in the literature, this work aims to combine two artificial intelligence techniques. In the first step, standardized test results are grouped by means of Self-Organizing Maps (SOM) and, in a second step, the groups with a high level of overlap are analyzed using a decision tree algorithm, which helps discover which attribute is discriminative in the diagnosis of children and adolescents with suspected ADHD.

Besides Section I, that aims to contextualize the work and present the objective, the work is organized into six parts. Section II presents the theoretical framework and justification of the study. Section III presents the proposed use of two artificial intelligence algorithms to aid in the diagnosis. In Section IV, the procedures for developing the study are described, including the computational development with the application of two artificial intelligence techniques. In Section V, the contribution of standardized cognitive tests or behavioral inventories is described, as well as the proposal to solve the diagnostic doubt within the self-organizing maps and then the classification by the decision tree for understanding the characteristics of the diagnosis of the disorder. Finally, in Section VI, we present the conclusion and recommendations for further studies.

II. Related Work

A. Elements of Attention Deficit/Hyperactivity Disorder (ADHD)

ADHD is part of the group of neurodevelopmental disorders beginning in childhood, but a substantial proportion of children with ADHD remain relatively impaired into adulthood [9]. From a cognitive-behavioral point of view, it is characterized by deficits in several cognitive functions, such as attention, especially selective, sustained, alternating, and divided attention, deficits in inhibitory control, processing speed, organization, ability to inhibit distracting information, deficits in cognitive flexibility, hyperactivity behaviors, restlessness, and impulsivity. ADHD affects 5.29% of the world's child population. Of this population, 30% up to 70% maintain symptoms into adulthood [11] [12]. According to DSM-5 [1], ADHD can be classified according to the predominance of symptomatic axes as predominantly inattentive presentation, predominantly hyperactive-impulsive presentation, or combined presentation. Behavioral patterns are important in the diagnosis of ADHD. Here are some descriptions from parents regarding the children: difficulty listening, obeying, following routine rules, often postponing and forgetting daily activities, difficulty following direct instructions, difficulty regulating feelings of frustration, exacerbation of motor activity, maybe impulsive in changing activities before they are completed, having difficulty waiting their turn, may have impairments in social relationships. These behaviors may contribute to high-stress [13] family or school environments. Given the importance of collecting various pieces of information in cognitive neuropsychology and behavior analysis, the treatment and multivariate analysis of the data can help us obtain relevant information in understanding ADHD complaints, and the artificial intelligence techniques used become key elements in diagnostic discrimination.

B. Self-Organizing Maps (SOM)

According to Merényi et al. [14], a SOM network provides clustering and visual representation of data in low dimension. This technique preserves the topological structure of the data in a lattice of neurons. The grid can be defined as a rectangular or hexagonal grid, as in Figure 1, usually two-dimensional, in an ordered manner such that the most similar neurons are grouped with neurons that are close in the grid, and the opposite is true for less similar neurons that are far apart in the grid, providing a topological view of the data. All neurons in the grid must undergo exposure to different realizations of the input dataset to ensure that the self-organization process matures. The algorithm then proceeds to initially randomly choose synaptic weights with small values. Once the grid has been initialized, we have the presence of three essential processes used to construct the self-organizing map. They were summarized by Kohonen [15] and Kubat [16] as follows:

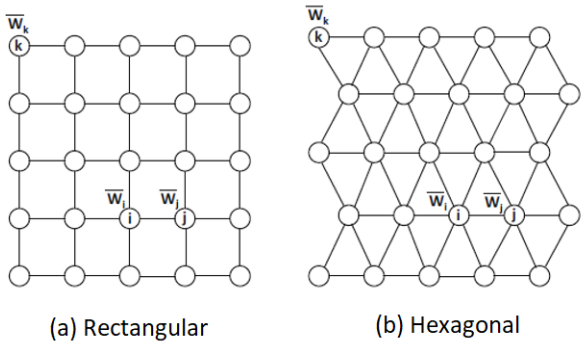


Figure 1. Topological Map - Rectangular and Hexagonal Grid ([17, p.451])

1) Competition: The synaptic weight vector is calculated for each j neuron of the grid with the same dimension as the input dataset through the inner product between the synaptic weight vector and the input data vector, this function being the basis for choosing the winning neuron. The maximization of this function has mathematical equivalence with the minimization of the Euclidean distance between the synaptic weight and input data vectors.

$$X = [x_1, x_2, ..., x_m]^T \quad (1)$$

X is the input vector of the space *m* transposed.

$$W_j = [w_{j1}, w_{j2}, ..., w_{jm}]^T, \ j = 1, 2, ..., l \quad (2)$$

W_j is the synaptic weight vector of each neuron in the grid.

$$i(X) = \arg\ \min_j \left\| X - W_j \right\|, \ j = 1, 2, ..., l \quad (3)$$

i(X) is the index that summarizes the competitive process between neurons.

2) Cooperation: The basis for cooperation between neighboring neurons is provided by the winner neuron that shows the spatial location of the topological neighborhood of neurons neighboring the winner $h_{j,i(X)}$;

$$h_{j,i(X)} = \exp\left(-\frac{d^2_{j,i}}{2\sigma^2}\right) \quad (4)$$

where $d_{j,i}$ is side distance and σ is the effective width of the topological neighborhood.

3) Adaptation: Neighboring neurons to the winner increase their discriminant function values based on the input dataset and as appropriate adjustments applied to their synaptic weights improve a subsequent input dataset.

$$W_j(n+1) = W_j(n) + \eta(n) h_{j,i(X)}(n)(X - W_j(n)) \quad (5)$$

where *n* equals epoch, $\eta(n)$ is the learning rate, and $h_{i,j(x)}(n)$ is the neighborhood function.

C. Decision Tree

A decision Trees is an Artificial Intelligence algorithm capable of organizing attributes from a dataset in priority, so that it can generate a path that leads to a decision for a classificatory attribute [18] [19].

The entropy (Shannon's) [20] measures the impurity of the dataset, being a measure of the heterogeneity of the input dataset (S) relative to its classification (c). The *Gain(S,A)* is given by the equation 6 and the entropy of *S* is given by the equation 7. Thus, the key factor is the use of a gain function that allows the attributes (A) to be compared to select the most relevant one. The attribute chosen is the one that maximizes the information gain which is calculated as being [20]:

$$Gain(S, A) = Entropy(S) + \Theta \quad (6)$$

S is the input dataset, *A* are the attributes Θ represents the probability of *A* multiplied by its entropy.

$$Entropy(S) = \sum_{k=1}^{c} -p_i log_2 p_i \quad (7)$$

$$\Theta = - \sum_{v \in Values(A)} p(A_v) Entropy(A_v) \quad (8)$$

The information gain is given by the equation 6 and represents the expected reduction in entropy when the value of the attribute *A* is known, since the process calculates the gain for each attribute, choosing the attribute with the highest gain to be tested in the set *S*. This process creates the division of objects to form the decision tree, giving rise to the node, labeling the attribute, and creating branches for each attribute value.

III. Proposed Method

The work presents a proposal for an unsupervised learning model as a method used in the identification of the neurons of the grid with greater diagnostic doubt of ADHD, that is, the diagnostic doubt in the neuron shows that it is difficult for both a machine learning algorithm and an expert to make a diagnosis. Thus, the paper brings a proposal to apply a decision tree on the neurons that show overlap to suggest which attributes are more discriminative. To understand this overlapping, the entropy (of Shannon) was calculated with the purpose of measuring the impurity of the neuron with its dataset, that is, the closer the entropy is to one, the greater the impurity of the neuron's dataset. Given this fact, a combination of SOM with the decision tree algorithm, which is a supervised model used in data classification to help identify one or more attributes from standardized assessment tools, such as cognitive tests and behavioral assessment inventories were sought. These tools were used to test the learning of ADHD characteristics. The objective of this decision tree algorithm was to verify the accuracy of the model for the confirmation of cases with ADHD diagnosis by identifying which assessment tools best contributed to this ADHD confirmation.

IV. Materials and Methods

The study sample consisted of 127 children and adolescents between 6 and 16 years old, 48 with a clinical diagnosis of ADHD and 79 from the control group, with no diagnosis of ADHD. The attributes that make up the neuropsychological tests and behavioral inventories applied in this study are Attention Cancellation Test (TAC), Trails (TT), Wechsler Intelligence Scale for Children (WISC-III), Wechsler Intelligence Scale for Children (WISC-IV), Wechsler Abbreviated Scale of Intelligence (WASI), Child Behavior Checklist for ages 6-18 (CBCL/6-18) and Teacher's Report Form for ages 6-18 (TRF/6-18). These attributes were normalized by the z-score method [21] to standardize the different scales of the attributes. The normalized data property is used to train the network SOM using the package available in R language [22]. In this library, the functions *somgrid* and *som* are used to parameterize and train the map, respectively. For the size of the map topology,

 ISBN: 978-1-61208-885-3

the dimension 4x4 was chosen. With this, the hypothesis of the study was to find neurons with a representative density of objects and with a significant class distribution.

With the trained map, the analyses made were the density of objects in each neuron, the distance between neurons, the quality of adjustment of the neurons, the contribution of the attributes in the formation of neurons, and the distribution of the label of each object in each neuron. In addition to the outputs analyzed, the representativeness of the number of objects contained in each neuron with the label attribute was sought in the table generated by the SOM. In this way, the neurons of greater relevance were identified, that is, with larger numbers of objects generated by the SOM algorithm.

From this point on, the entropy algorithm (Shannon's) was used on each neuron in the network to select the neuron with the highest class overlap along with the representativeness of objects that are difficult cases to diagnose.

By identifying neurons with overlapping classes, their objects are selected from the database generated by the SOM network for training and validation of the decision tree algorithm. The result of the decision tree brought a hierarchy of attributes in order of discrimination for cases of diagnostic doubt, and the validation of the algorithm shows the performance of the classification.

A. Rating Performance Evaluation

Table I shows the confusion matrix that was used to analyze the classification performance of the decision tree. The table indicates the prediction of the positive and negative scenarios, as well as current true and false scenarios [23]:

- TN is the correct number of negative predictions;
- FP is the number of false positive predictions;
- FN is the number of false negative predictions;
- TP is the correct number of positive predictions.

Table I. CONFUSION MATRIX

	Predicted Negative	**Predicted Positive**
Current False	TN	FP
Current True	FN	TP

From the confusion matrix, it is possible to measure the performance of the algorithm by calculating the accuracy, as follows:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (9)$$

$$Error = (FP + FN)/(TP + TN + FP + FN) \quad (10)$$

After this stage, it is possible to better understand the model's contribution to the understanding of Attention Deficit Hyperactivity Disorder, as well as to the diagnostic evaluation of patients. The next section presents the results obtained in this work.

V. RESULTS

The training result of the SOM network can be seen in two different visualizations, depicted in figure 2 and 3. figure 2 presents the attributes, common to the trials, graphically distributed in each neuron. The sizes indicate the contribution that each attribute has to the formation of the neuron. Note that neighboring neurons have similarities among the attributes. In figure 3, the diagnosis, an attribute that is not used in training the SOM, is projected on the map, allowing visualization of which neurons have the overlap of class 1 (group diagnosed with ADHD) and 2 (control group without ADHD). The network could not separate the diagnosed cases in neuron 4.

Table II presents for each neuron the percentage of objects of each class. Neuron 4 is the one with the highest concentration of objects (40%) and overlapping classes in the whole dataset.

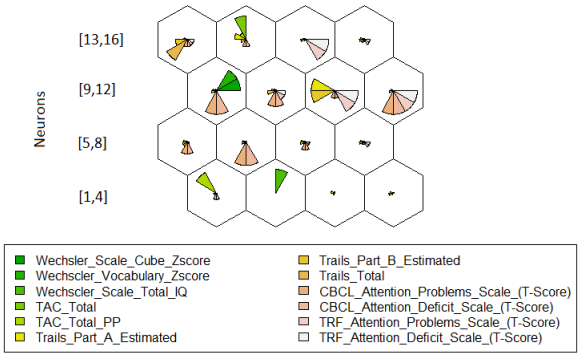


Figure 2. Contribution of the attributes in the formation of the neuron

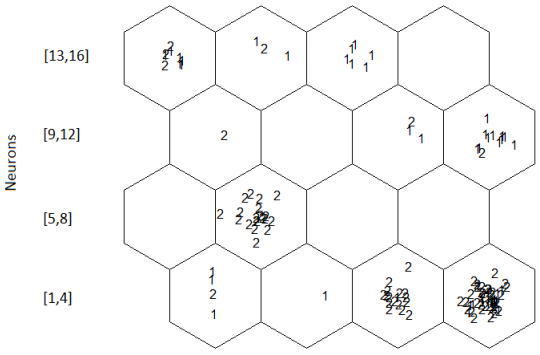


Figure 3. Scattering of objects diagnostic within neurons

The result of the decision tree with the data mapped onto neuron 4 can be seen in Figure 4. The result shows that the Child Behavior Checklist for ages 6-18 attribute, specifically the probability of attention problems scale (T-score) [24] [25] neurons had the highest discrimination.

Table II. COMPARATIVE DIAGNOSIS BY THE NEURON DIMENSION 4X4

	Diagnostic	1	2	Total
neuron	1	3 (6.2%)	1 (1.3%)	4 (3.1%)
	2	1 (2.1%)	0 (0.0%)	1 (0.8%)
	3	1 (2.1%)	16 (20.3%)	17 (13.4%)
	4	16 (33.3%)	35 (44.3%)	51 (40.2%)
	6	0 (0.0%)	20 (25.3%)	20 (15.7%)
	9	0 (0.0%)	1 (1.3%)	1 (0.8%)
	11	2 (4.2%)	1 (1.3%)	3 (2.4%)
	12	11 (22.9%)	1 (1.3%)	12 (9.4%)
	13	6 (12.5%)	3 (3.8%)	9 (7.1%)
	14	2 (4.2%)	1 (1.3%)	3 (2.4%)
	15	6 (12.5%)	0 (0.0%)	6 (4.7%)
Total		48 (100.0%)	79 (100.0%)	127 (100.0%)

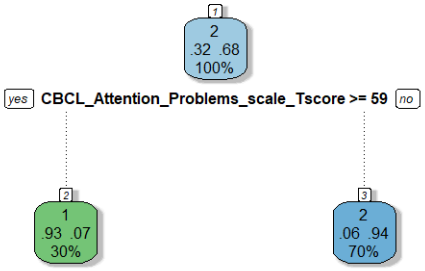


Figure 4. Decision Tree of Neuron 4

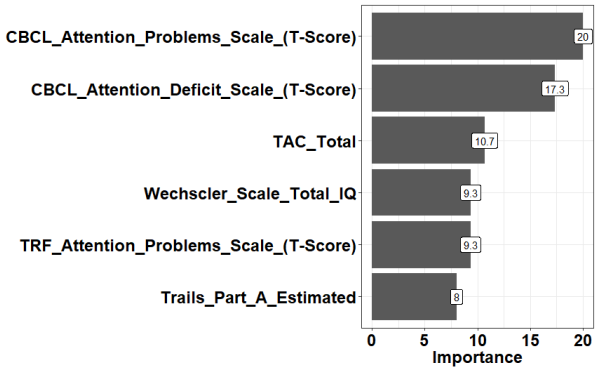


Figure 5. Importance of the attributes in neuron 4 by decision tree

Finally, Figure 5 allows you to visualize all six attributes with greater discrimination for cases with greater complexity relative to the integration of clinical evaluation and evaluation using tests for a confirmation of the diagnosis.

VI. Conclusion and Future Work

Data from the behavioral assessment inventory presented in [5] can generally be more susceptible to respondent bias because it is based on the answers of the subject. This bias is less so when using cognitive tests which are assessment measures applied directly to the person. Mathematical understanding and model generation is likely to become more difficult using only behavioral inventories. Since ADHD demands the use of both types of measures, in this study both tools were used to apply the decision tree. In the study, it was possible to group the children with and without ADHD by SOM, which made it possible to understand from the perspective of each grouping what was most important in their formation.

The self-organizing map contributed especially to the formation of groups and the understanding of clusters with class overlapping, which is the proposal of this work. In this case of overlapping to diagnose a disorder, the decision tree was used to classify the attributes that contributed to the formation of the ADHD group. With this, the predominance of characteristics that helped in the understanding of ADHD in children and adolescents in the study was observed.

The application of the decision tree identified six attributes, namely two of cognitive assessment and four of behavioral assessment, that showed relevant discrimination to make the diagnosis. The Child Behavior Checklist for ages 6-18 attribute the one that showed the highest discriminative power. However, the incidence of low T-scores on the attention problems scale and attention deficit scale does not necessarily imply that the child has ADHD. The results presented showed the difficulty and complexity of finding indicators that define ADHD, as already signaled by some authors [5] [6] [8] [26] [27]. Importantly, the diagnosis of ADHD is a clinical diagnosis that considers the measurement of behavioral correlates of attentional deficits and indicators of hyperactivity and impulsivity in more than one environment. With the Child Behavior Checklist for ages 6-18 attribute being a parent-reported measure, the validity of these two scales for identifying ADHD will likely be confirmed. However, when disregarding the scales, one should consider the evaluations made with the cognitive tests that directly make cognitive measurements and are essential to decide the diagnosis of ADHD. In this study, the tests that contributed the most to this decision tree were the Attention Cancellation Test (ACT) and the Trail Test (TT).

The study presented as a relevant factor the case of overlapping diagnoses of neurons when using the SOM and, in conjunction with the decision tree, was able to separate 88% of the cases. This way, future works can collaborate with the technique addressed in the study through supervised data procedures. These tools can help in making comparisons between results of standardized tests aiming to reduce possible biases of behavioral evaluations based on informants' reports. Future studies can test the same decision tree on larger samples to see if the attributes that showed higher accuracy are maintained. By doing so, the best indices of cognitive and behavioral assessment instruments that contribute to the increased accuracy of ADHD diagnosis may be identified. Since this study controlled for no comorbidities in the ADHD group, it is recommended for future studies to use sample groups with and without ADHD comorbidities from other psychiatric and neurodevelopmental conditions. This type of sample may allow the testing of new and more complex models due to the natural overlap of signs and symptoms between ADHD and some of these comorbidities.

 ISBN: 978-1-61208-885-3

ACKNOWLEDGMENT

We thank the Academic Excellence Program of the Coordination for the Improvement of Higher Education Personnel (CAPES-PROEX), Process number 1133/2019, (CAPES-PROSUC) in mode II, the National Council for Scientific and Technological Development (CNPq, Cases 307730/2017-4 and 307443/2019-1), the Mackenzie Research Fund (MACKPESQUISA) of Mackenzie Presbyterian University, and the Foundation for Research Support of the State of São Paulo (FAPESP, Cases 2018/01063-0 and 2019/20757-1) for financial support.

REFERENCES

[1] American Psychiatric Association, DSM-5: Diagnostic and Statistical Manual of Mental Disorders. Artmed publisher, 2014.

[2] E. M. Mahone and M. B. Denckla, Attention-deficit/hyperactivity disorder: a historical neuropsychological perspective. Journal of the International Neuropsychological Society: JINS, v. 23, no. 9-10, pp. 916-929, 2017, doi: 10.1017/S1355617717000807.

[3] E. Aretouli, How neuropsychology can inform our understanding of preschool ADHD: Clinical and research implications. Applied Neuropsychology: Child, vol. 8, no. 2, pp. 174-181, 2019, doi: 10.1080/21622965.2017.1421463.

[4] M. L. Wolraich et al., Clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents. Pediatrics, vol. 144, no. 4, 2019, doi: 10.1542/peds.2019-2528.

[5] L. R. R. Carreiro et al., Interdisciplinary neuropsychological, behavioral and clinical assessment protocol for children and adolescents with complaints of inattention and hyperactivity. Psychology: theory and practice, vol. 16, no. 3, pp. 155-171, 2014.

[6] S. Chandana and K. Vijayalakshmi, "An Approach to Measure and Improve the Cognitive Capability of ADHD Affected Children Through EEG Signals," 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), Mumbai, India, 2018, pp. 314-318, doi: 10.1109/ICALT.2018.00079.

[7] H. J. Kam et al. Development of a decision support model for screening attention-deficit hyperactivity disorder with actigraph-based measurements of classroom activity. Applied clinical informatics, vol. 1, no. 4, p. 377, 2010, doi: 10.4338/ACI-2010-05-RA-0033.

[8] S. H. Lee, B. Abibullaev, W. Kang, Y. Shin and, J. An, "Analysis of attention deficit hyperactivity disorder in EEG using wavelet transform and self organizing maps," ICCAS 2010, Gyeonggi-do, Korea (South), 2010, pp. 2439-2442, doi: 10.1109/ICCAS.2010.5670255.

[9] A. Koumoula, The course of attention deficit hyperactivity disorder (ADHD) over the life span. Psychiatrike= Psychiatriki, vol. 23, pp. 49-59, 2012.

[10] L. G. Leahy, Attention-deficit/hyperactivity disorder: A historical review (1775 to present). Journal of psychosocial nursing and mental health services, vol. 55, no. 9, pp. 10-16, 2017, doi: 10.3928/02793695-20170818-08.

[11] G. Polanczyk, The worldwide prevalence of adhd: a systematic review and metaregression analysis.American journal of psychiatry, Am Psychiatric Assoc, vol. 164, no.6, pp.942–948, 2007.

[12] V. Simon, Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis. The British Journal of Psychiatry, Cambridge University Press, vol. 194, no.3, pp.204–211, 2009.

[13] E. B. P. Benczik and E. B. Casella, Understanding the impact of ADHD on family dynamics and possibilities for intervention. Journal of Psychopedagogy, vol. 32, no. 97, pp. 93-103, 2015.

[14] E. Merényi, M. J. Mendenhall, and P. O'driscoll, Advances in self-organizing maps and learning vector quantization. Advances in Intelligent Systems & Computing, vol. 295, 2016.

[15] T. Kohonen, Self-organized formation of topologically correct feature maps. Biologicalcybernetics, Springer, vol. 43, no. 1, pp. 59–69, 1982.

[16] M. Kubat, Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7. The Knowledge Engineering Review, vol. 13, no. 4, pp. 409-412, 1999.

[17] C. C. Aggarwal, Deep Reinforcement Learning. In: Neural Networks and Deep Learning. Springer, Cham, 2018, pp. 373-417.

[18] M. J. Berry and G. S. Linoff, Data mining techniques: for marketing, sales, and customer relationship management, John Wiley & Sons, 2004.

[19] I. H. Witten, E. Frank, M. A. Hall, and C.J. Pal, Data Mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2016.

[20] T. Mitchel, Machine learning. mcgraw hill education (ise editions), 1997.

[21] A. E. Curtis, et al., The mystery of the Z-score. AORTA Journal, vol. 4, no. 4, p. 124, 2016.

[22] R. Wehrens and J. Kruisselbrink, Flexible self-organizing maps in kohonen 3.0. Journal of Statistical Software, vol. 87, no. 1, pp. 1-18, 2018.

[23] S. Ruuska et al., Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. Behavioural processes, vol. 148, pp. 56-62, 2018, https://doi.org/10.1016/j.beproc.2018.01.004.

[24] H. Budak, S. E. Tasabat, A modified t-score for feature selection. Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi A-Uygulamalı Bilimler ve Mühendislik, vol. 17, no. 5, pp. 845-852, 2016, doi:10.18038/aubtda.279853.

[25] T. M. Nolan et al., Child Behaviour Checklist classification of behaviour disorder. Journal of paediatrics and child health, vol. 32, no. 5, pp. 405-411, 1996.

[26] R. A. Barkley and K. R. Murphy, Attention-deficit hyperactivity disorder: A clinical workbook. Guilford Press, 2006.

[27] F. L. Cibrian, G. R. Hayes, and K. D. Lakes, Research Advances in ADHD and Technology. Synthesis Lectures on Assistive, Rehabilitative. and Health-Preserving Technologies, vol. 9, no. 3, pp. i-156, 2020, doi: 10.2200/S01061ED1V01Y202011ARH015.

 ISBN: 978-1-61208-885-3

Synthesis of Neonate Connectomes for Artificial Sentience and Common Sense

Michael Bihn
Dept. of Computer Science
University of Colorado at Colorado Springs
Colorado Springs, USA
email: mbihn@uccs.edu

Rory Lewis
Dept. of Computer Science
University of Colorado at Colorado Springs
Colorado Springs, USA
email: rlewis5@uccs.edu

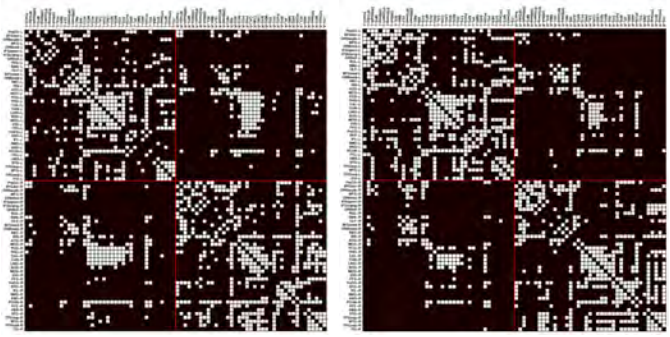
***Abstract*—In the ongoing research effort of synthesizing sentience into artificial intelligence, we propose a modular network that emulates neurological synaptic evolution in the neonate brain. Our hypothesis is that if one were to successfully develop a synthesized emulation of human's six-month hippocampus as it initializes adult-like glucose usage and synaptic density which is generally accepted in the domain of neuroscience as being the foundation of human sentience, then so can human sentience be injected into the synthesized replication of said six-month hippocampus. Accordingly, we present a theoretical proposition that facilitates a significant step towards overcoming the commonsense challenge that state-of-the-art artificial intelligence systems are still grappling with today; where even the most powerful artificial intelligence systems are void of the common sense of a three year old: That lemons are sour, that things fall towards the ground and that they, as children, can pretend to be somebody else. Herein, we present a methodology to efficiently promulgate the research goal of integrating sentience and common sense reasoning into artificial intelligence, taking a neurological rather than a psychological approach.**

***Keywords*—Sentience, Common Sense, AI.**

I. INTRODUCTION

In 1766, Immanuel Kant theorized that human knowledge is a combination of priori knowledge where knowledge is acquired independently of any particular experience, and posteriori knowledge, which is derived from experience that we reason from our senses being affected by our surrounds [1]. Nowadays, it is accepted that to achieve common sense processing, an entirely new method will need to be invented [2] [3] and, that this new method will require priori and posteriori knowledge [4]- [5]. The design of a posteriori knowledge component shall seamlessly communicate with the artificial intelligence system. In 2020, Shanahan *et al.* [6] examined the common sense of animals and concluded that there must exist, in each animal, some innate knowledge that allows them to learn without words.

We postulate that, because common sense in humans and animals require priori and posteriori knowledge, so should we design sentient machines. Our premise is that, to emulate human and animal common sense, one needs to mathematically emulate developments in neuroscience which will include discovering that the directionality of brain waves in the cortical regions of the brain form different frequency bands [7], that functions emerge from the flow of information linking distant cortical regions [8] that a modular network topology is present in the brain from the first days of life. [9] [10] and that the complexity of our multi-connected connectome network [11] has been decoded [12].



(a) (b)

Fig. 1. Connectivity Matrice: Mapping connectome evolution in 78 cortical regions, excluding subcortical and cerebellar regions, in the brains of a) two-week, and (b) one-year old infant humans [13]. The white squares represent connectivity between the lobes of the horizontal and vertical axis.

A. Neuronal Pathway

We focus on the neuronal network between regions of the brain, as defined by Automated Anatomical Labeling (AAL) for length and local efficiency [13] where Yap *et al.* used a connectivity matrix to group neuronal regions into three distinct communities, as shown in Fig. 1. When comparing the synaptic evolution on the matrices from a cohort of two week old and one year old children, one observes a complex neuronal mesh comprised of multiple additions and pruning of the network. Fig. 2 illustrates how integrating the Kamada-Kawai layout with the Pajek software package [14] shows three distinct neuronal communities [13].

Recently, Fornito *et al.* showed that connectomes have an inheritable complex topology that suggests a genome-wide association that can be either excitatory or inhibitive [16]. Rosenburg *et al.* showed that at ~6 months, the hippocampus has adult like glucose use and synaptic density [17] and Szalkai *et al.* built four consensus brain graphs from a cohort of 106 individual brain graphs and set directions by popular

 ISBN: 978-1-61208-885-3

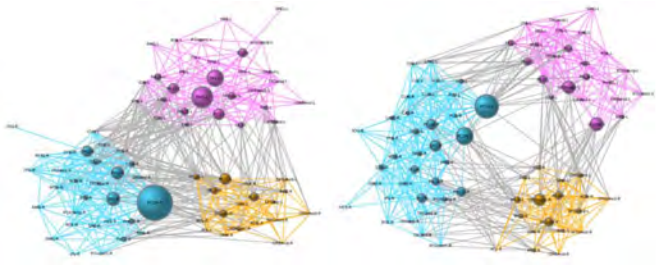


Fig. 2. Spring Embedding Visualization: Where the sizes of the vertices are weighted by Yap *et al.* using Freeman's logarithmically scaled node betweenness algorithm, [13] [15] .

TABLE I
THE PATTERNS OF CHANGE

Pattern	Number		Pattern	Number
000	2120		100	161
001	64		101	28
010	53		110	39
011	136		111	402

vote and in so doing, proved that i) axons have directionality from the soma to the end of the axon, and that, ii) 82% of the directions were the same for all four groups [18].

B. Neuronal Pathway Insights

At the biological level, the authors first accept that neuronal pathways are constantly changing over time with rapid myelination occurring in the first year of life [13]. Secondly, the authors acknowledge that brain modules (highly connected groupings of brain lobes) at two weeks old and one year old are not the same. Thirdly, the authors accept that using connectomes is the state-of-the-art for mapping these pathways yet, unfortunately, it is evident that, in essence, using a two dimensional array, that has no geometric similarity with the brain, at all, to track these neurological events is a shortcoming. It is here that the authors found their motivation to procure what is in this paper; a more robust and precise means to emulate, represent and measure in a computer the neuronal pathway evolution that occurs at the biological level. It should also be noted that Yap *et al's* 3-D representations of the neuronal pathways suggest that i) there are hubs inside the modules that are most likely connected to many lobes in that module, that ii) there are bridges with edges which connect the hubs of different modules, and that iii) a non hub leaf node in one module may never connect to a non hub leaf node in another module.

C. Neuronal Instantiation

The authors note that because lobe-to-lobe connections are initialed as early as two weeks, only to be turned off at one year, and then turned back on again at two years, while others are turned off at two weeks, turned on at one year, and then off again at two years [13], the authors have focused on studying whether this seemingly random and chaotic process has patterns that when found will enlighten researchers in this domain as to how human sentience is formed, and have said formation replicated synthetically in a machine. The numbers inherent in each of these patterns are represented in Table I. The first bit of the pattern is dependent on the two week connectome, the second bit represents the one year connectome and the third represents the two year connectome. We note that there are 2120 possible connections that never occur and 402 connections that never break. Additionally, it is of interest that the connectomes have 630 connections and that the number of connections added from two weeks to one year is 189, which is also the number disconnections from two weeks to one year. The number of connections added from one year to two years is 92, which is also the number of disconnections from one year to two years. This reflects the slowdown of myelination over time. Yap *et al.* have found the growing efficiency of the brain in [13].

The rest of the paper is structured as follows. In Section 2, we present our research objectives. In Section 3, we present our experiments. In Section 4, we present how we plan to move from a prototype to the real model. We make our conclusion in Section 5.

II. RESEARCH OBJECTIVES

The aforementioned research has lead the authors to answer four questions. 1) What is the purpose of the changing lobe to lobe connections? 2) Is there a development phase to the neuronal pathways? 3) Is there an initialization phase? 4) At which connectome do humans start independent thinking and if so, what pathways are crucial to this independent thinking?

A. Transversal Definition

We start by defining the Transversal propagation of connectivity and pruning by focusing on forming a means to measure the path distance from one lobe to the next and present (1) and (2):

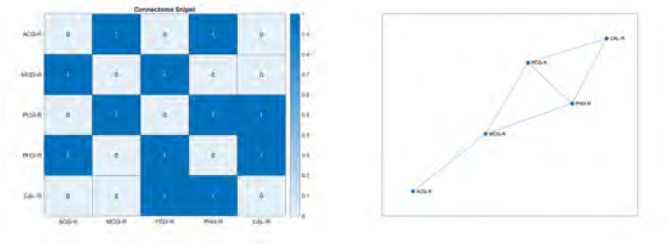
$$SSP = SingleShortestPath(lobe_i, lobe_j) \quad (1)$$

$$T_i = \sum_{j=1, j\neq i}^{78} SSP(i,j) \quad (2)$$

where $lobe(i)$ and $lobe(j)$ are the brain lobes as indexed by i and j from the automatic anatomical labels provided by Montreal Neurological Institute [13]. The Single Shortest Path is the length of the single shortest path defined in the connectome graph preliminaries. Additionally, we define Ti, transversal of the ith node, as the sum of all the shortest paths from the ith lobe to the jth lobe excluding the path to from the ith lobe to itself. 78 is the number of brain lobes in our connectome. Each of the 78 brain lobes is fully connected to all the other 77 lobes in the brain. Accordingly, we reference Fig.3 (b) and determine the single shortest path from the right Anterior Cingulate Gyrus lobe (ACG-R) to the right Median Cingulate Gyrus lobe (MCG-R) is one.

The single shortest path from ACG-R to the right Posterior Cingulate gyrus lobe (PCG-R) is two. For ACG-R to the right

 ISBN: 978-1-61208-885-3



(a) (b)

Fig. 3. Connectivity: This shows a snippet of the connectome and the graph produced. (a) connectome, and (b) graph

Parahippocampal gyrus lobe (PHG-R) it is also two. However, for the ACG-R to the right Calcarine cortex (CAL-R) there are multiple paths, the shortest being a length of three. T_ACG-R is the transversal of node ACG-R and is defined by 2 as the sum of all the SSP of A, therefore T_ACG-R=1+2+2+3=8.

$$MaxTransversal = \max(T_i, i = 1..78) \quad (3)$$

$$MinTransversal = \min(T_i, i = 1..78) \quad (4)$$

where the $MaxTransversal$ is defined as the maximum of traversals, see (2), of all the nodes transversal, i stepping through the lobes from 1 to 78 lobes, as shown in(3). $MinTransversal$ is defined as the minimum of all the nodes transversals, defined in (2), as shown in (4). Once again this calculation looks at the lobes i from lobe 1 through lobe 78, similarly to the $MaxTransversal$.

B. Defining Maximum Traversal Length

We have determined that a common denominator to correctly replicating human sentient neurological evolution in a machine, lasers in on how accurately one can define and measure the maximum traversal length. The maximum traversal length on the two week connectome is 257; see 3. The maximum traversal for the one year connectome is 219, and 212 for the two year connectome. We examine the maximal transversal of 257 at two weeks, 219 at one year and 212 at two years. Preliminary findings show that, while the connectome is reducing the maximum length path over time, *at the same time*, the minimal transversal path is increasing from 107, at two weeks, to 127 at one year, and 128 at two years. It is also interesting and not understood why the number of connections remains stable at 630 during this period.

C. Optimizing Maximum Traversal Length

The optimization of maximum transition length with the constraints of pairs of connections (connection, disconnection) over time and stable 630 connections over time could lead to a model that predicts the connectome development. Of course, more data and analysis is necessary to order the transitions of the connectome. This gives us an optimization problem, namely, minimize the maximum transition length, with the constraint of stable number of 630 connections. This optimization is how we propose to predict which transitions will occur. Left unbounded, the model would keep pruning and connecting, thus we shall just look at the best 189 from two weeks old to one year old and compare our predictions to the actual changes that occur. We will then start with the one year old and execute the same optimization, this time only taking the best 92 pairs of disconnect and connect.

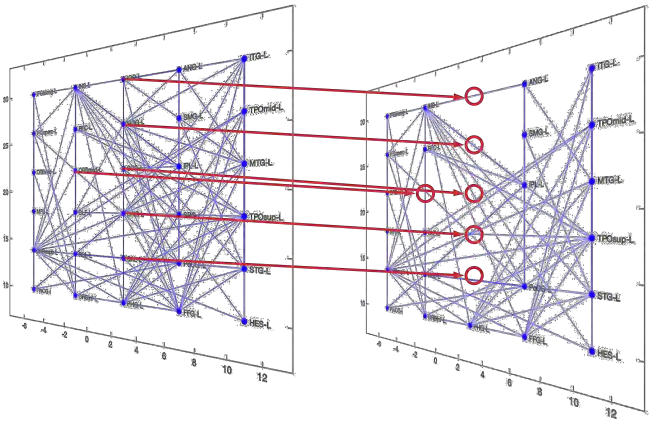


Fig. 4. The red lines emphasize a few lobes moving out of Module one

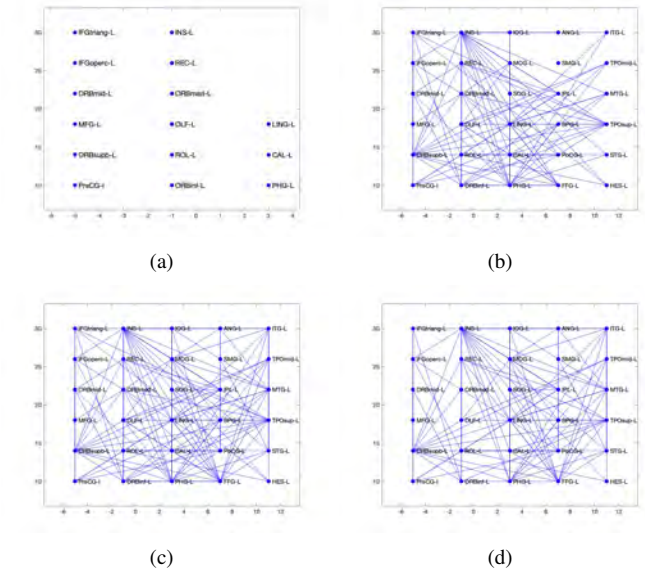
A further challenge is matching the modules at different ages. Fig. 4 shows how certain brain lobes will move from one module of the brain to another. This occurs due to the changing connections those lobes have with other brain lobes. This figure provides the reader with insight to view the changes in the figures of the next section. In the next section, we examine the development of the connectome from conception to one year old. Noting the changes that occur just inside the first lobe for simplicity of explanation. The changes that occur across all 78 lobes, modifying from 3 brain modules to four brain modules at two years old are too complex to put on one sheet of paper. The actual lobes involved in particular changes are details to be reconciled by our predictions.

III. Experiments

The authors hypothesize that it is possible to measure the path from any node to another node in terms of the number of edges traversed through the connectome. Our justification for this proposition is that because we know that as the brain develops from three to four modules, it is logical to expect that the total number of edges traversed is decreasing. To formulate our methodology we base our experiments off of the following determinate. Fig. 4 shows the detail that should be noted as stepping through the determinate.

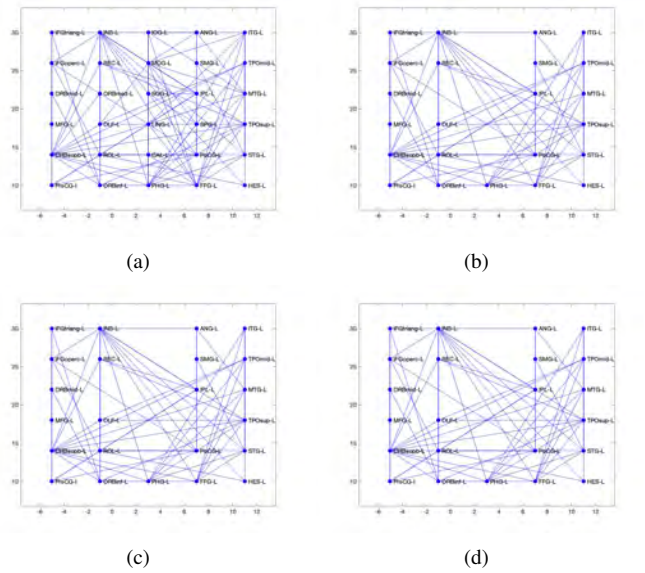
1) At conception no brain lobes exist.
2) Over time, the brain lobes form and connections are established.
3) Fig. 5(a) shows the first brain lobes appearing in module one of the brain.
4) Fig. 5(b) shows the neuronal pathways being formed between some of the existing lobes in module one of the brain. The same growth is occurring in the second and third modules of the brain.

 ISBN: 978-1-61208-885-3



(a) (b) (c) (d)

Fig. 5. Connectome Proliferation: Isolating the first highly connected lobes of the module 1 of the brain, we see the following stages (a) Module one shows initial lobes appearing. (b) all lobes in module one appear with first synapses, (c) first brain module complete at two weeks, and (d) after two weeks, initial edge pruning occurs.



(a) (b) (c) (d)

Fig. 6. Connectome Pruning (a) After two weeks, edge pruning completed. (b) After two weeks, some nodes break away from module 1. (c) After two weeks, all nodes to break away from module 1, have done so., and (d) At one year, the first new node moves into module 1.

5) Fig. 5(c) shows the the connectome lobe 1 completed at two weeks of age.
6) Once the two week connectome is formed, we know that connections will be broken and others formed leading to the one year connectome. Fig. 5(d) shows synaptic connections being broken. While the prototype shows multiple edges dropping, we note i) that this is only module one of the brain and ii) it is highly likely, that while a synapse/edge is disconnected, another synapse is formed, maintaining the full connectedness of the connectome and the 630 connections.
7) Fig. 6(a) shows all the edges dropped that are not in the one year connectome. While this gives the reader an overview of how the connectomes are changing over time. One should note that even though it may appear that all the lost edges are dropped before adding nodes and edges to achieve the one year connectome, evidence shows that this is *not* what happens in the human brain. The lobes *do not go away* but rather connect to another module of the brain. The edge modifications are most likely intermingled. We need to know if a neuronal pathway is dropped if the myeline becomes available to form another synapse/edge.
8) Fig. 6(b) shows nodes moving out of module 1 between two weeks and one year. A node switches modules when it becomes more connected to the nodes/lobes of the new module then its current module.
9) Fig. 6(c) shows the rest of the nodes moved out of module 1 and into another module. This is the final node/lobe state that is expressed in the one year connectome.
10) Fig. 6(d) shows a node from another module moving into module 1.
11) Fig. 7 shows how at one year lobes from other modules have become more connected to the module 1.

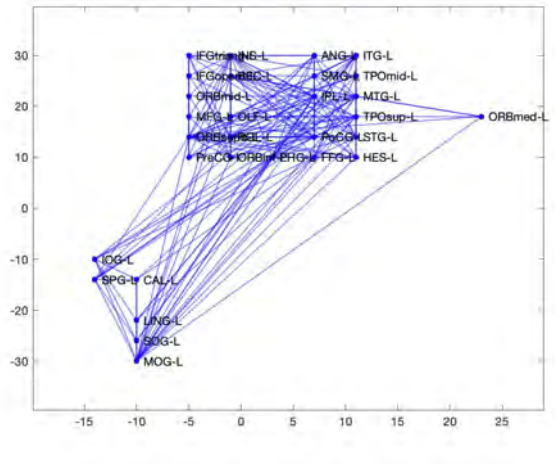


Fig. 7. Nodes and edges are added to complete module 1 for the one year connectome

The ordering of these modifications is most definitely not occurring en masse, not all edge deletions, then all node deletions, then all node additions, then all edge additions. As already stated the lobes do not go away, but rather become connected to one of the other modules. Therefore, the nodes should not go away, but rather be connected to lobes in another module of the brain. Noting that from one year old to two years old a forth module appears. The graph with all 78 lobes/nodes and 630 synapsis/edges is incomprehensible. The connections are created when the synapse strength had crossed a certain

threshold. The connection is broken when the synapse strength drops below that threshold.

The rapid myelination in the first year eventually slows down. We need some measure of this myelination availability and usage in forming a synapse to be able to predict when the next neuronal pathway can be formed. If the myeline is not available, the synapse and neuronal pathway will not be formed. We know what connections exist at two weeks old. With enough data we should be able to measure the total length of these pathways and determine the myeline needed to form them. This could lead to a side study on what mechanism provides the myeline and what factors such as nutrition, O2, and time influence said mechanism. Simultaneous growth in the three lobes could provide competition for resources. We need to measure the myeline production and usage over time. Only then can we use regression analysis (Time Series) to plot the brains resources availability and use over time. Once these curves can be predicted, then they can be used to spot deviation in individuals brain growth. These deviations can then be catalogued, accumulated and used for finding disease or giftedness.

Another possibility is to find an arbitrary variable for the synaptic length growth potential over time. Rather then measure the myeline availability measure the total synapse length over time. [13] provides average synapsis length over time, an average implies a sum divided by a count. If this data is available, we should be able to regress the growth in number and growth of length of the synapsis over time. These growth rates along with a scripted ordering of synapsis connectivity will then predict when connections are built.

Fornito *et al.* [16] suggest that the DNA explicitly controls the brains growth. This infers that if there was a resource problem the brain may not have formed properly, potentially altering the expectations of a DNA study. In which case, it may be better to compare the connectome development to disease expression, rather than the DNA to disease expressions. This leads the authors to conclude that it raises the importance of researching *how* the brain is formed rather than studying the DNA process at the neurological level.

IV. From Prototype to the Real Model

o model the real connectome development we dynamically mimic four constraints, 1) the maximal traversal minimization over time, 3) that there are always 630 connections, 3) that the average synapse length is always increasing, and 4) that the connectome remains fully connected.

In order to accomplish this, we shall implement connectome changes in pairs, one disconnect paired with a connect. These pairs shall be created by ordering all the connects in increasing synapse length and the disconnects in increasing synapse length. Then, pairing the smallest connect with the smallest disconnect, to give the list of pairs to be implemented over time. We then shall ensure that the connectome remains fully connected over time. The pairs of actions(connect/disconnect) will need to be examined for orphan creation, where a node/lobe has a degree of 1 and that link is being deleted.

TABLE II
The Action Table

Action	the action to be taken connect or disconnect
LobeFrom and To	the lobes being connected or disconnected
Day and Time	the projected day and time the action will occur
Length	the length of the connection
TimeToComplete	the time necessary to complete the action

Since we know that the connectome stays fully connected, we know there is another connect in the list that reconnects that node/lobe. The two actions shall then swap partners, such that the disconnect and connect will maintain that particular node/lobe inclusion in the connectome, and maintain the fully connectedness of the connectome. To maintain the full connectedness, the connection will be created before the disconnect.

Once the actions list is refined, the actions can then be implemented over time. The rate of increase of myelin availability applied to project when the next action will occur. That is, when there exist enough myeline available to satisfy the net gain of pathway creation. We shall build the action table, see Table II, to provide the time ordered list of connectivity growth.

The length of synapsis should be known. [13] shows the average synapse length is increasing over time. The average is the sum of all the lengths divided by the number of synapsis. In future experiments we may find the synapse growth rate could replace the myelin availability function, or provide the myeline available function with another variable. Considering a biological process is not instantaneous, we shall set a length of time to complete the action.

The rate of myeline availability or rate of synapsis length growth shall be determined by regression analysis. We have the data points that from two weeks to one year there are 189 pairs of actions taken. While we do not know what the pairs are, we do know what the 189 disconnects are, and we do know what the 189 connects are. From one year to two years there are 92 pairs of action, likewise we know the 92 connects, and the 92 disconnects. With only two points we can only have a linear function. Research into the ConnectomeDB will provide more defined pairs of actions over time, thus more data points for the regression. Several regression attempts (linear, quadratic, polynomial, exponential, logarithmic) shall be executed to find the best correlation coefficient, r.

Assessment of the Accuracy of the algorithms developed will be possible when more connectome data becomes available. The National Institute of Health (NIH) is sponsoring a Baby Connectome Project that began in 2016 where the data will be available to NIH sponsored researchers [19].

V. Conclusion and Future Work

In considering where neuroscience research has lead we reiterate our original concept which was to produce a minimal ontology, recognizing that a toddler has minimal common sense and must learn it from experiences. In pursuing the neuroscience, we find the connectome, a representation of the

 ISBN: 978-1-61208-885-3

brains fully connected network between 78 cortical regions. But rather than having a stable connectome as a child, the connectome is more stable as an adult. What we have run into is the innate process of building the networked brain. The neuroscience shows us the result of the innate process of building the brain, but does not, as of yet, show us how it is built.

A systematic approach to define the requirements of synthesizing the brain needs to be taken. Mapping known capabilities to cortical regions and corresponding tool that has been built or needs to be built. Evaluating all existing tools for input, processing and output. All inputs shall be received from the network that simulates the connectome. All outputs shall be delivered to other cortical consumers through the synthesized connectome. The collective process shall mimic our notion of common sense. What started as an attempt to minimize the scope of the common sense problem has led us to the extensiveness of the brains innate development in the first years of life. What we set upon to build is not stable, but rather time dependent, adding a fourth dimension as we observe the connectome evolve to its adult stability.

This research has revealed that the max traversal shortens with time. In essence, the shorter max traversal provides a shorter path from one brain lobe to all the other brain lobes. This quicker transmission of brain signals provides humans with an increasing brain speed as we age. This in turn provides us a mathematical means to differentiate between "*quick wittedness*" versus "*not the sharpest tool in the shed*". Both of these phrases are common judgments of a person's level of common sense. We have determined that a common denominator to correctly replicating human sentient neurological evolution in a machine, lasers in on how accurately one can define and measure the maximum traversal length.

For our future research we will be studying how Prescott *et al.* [20] have built their humanoid robot, named ICub, based on psychological division. Prescott proposes using the human cognitive architecture, yet they take a psychological self approach to their brain inspired control architecture. However, they make no mention of connectomes, or the latest neural networks of the brain. Conversely, the authors have full faith that studying and replicating the neurological approach is more realistic and shall prove more fruitful in the long term.

REFERENCES

[1] I. Kant, "The possibility of experience," *Journey into Philosophy: An Introduction with Classic and Contemporary Readings*, p. 42, 2016.

[2] E. Davis and G. Marcus, "Commonsense reasoning and commonsense knowledge in artificial intelligence.," *Commun. ACM*, vol. 58, no. 9, pp. 92–103, 2015.

[3] R. Zellers, Y. Bisk, R. Schwartz, and Y. Choi, "Swag: A large-scale adversarial dataset for grounded commonsense inference," *arXiv preprint arXiv:1808.05326*, 2018.

[4] S. A. Bini, "Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care?," *The Journal of arthroplasty*, vol. 33, no. 8, pp. 2358–2361, 2018.

[5] A. Talmor, J. Herzig, N. Lourie, and J. Berant, "Commonsenseqa: A question answering challenge targeting commonsense knowledge," *arXiv preprint arXiv:1811.00937*, 2018.

[6] M. Shanahan, M. Crosby, B. Beyret, and L. Cheke, "Artificial intelligence and the common sense of animals," *Trends in Cognitive Sciences*, 2020.

[7] F. Babiloni *et al.*, "Estimation of the cortical functional connectivity with the multimodal integration of high-resolution eeg and fmri data by directed transfer function," *Neuroimage*, vol. 24, no. 1, pp. 118–131, 2005.

[8] P. Bartolomeo, "The quest for the critical lesion site in cognitive deficits: problems and perspectives," *Cortex*, vol. 47, no. 8, pp. 1010–1012, 2011.

[9] P. Hagmann, M. Kurant, X. Gigandet, P. Thiran, V. J. Wedeen, R. Meuli, and J.-P. Thiran, "Mapping human whole-brain structural networks with diffusion mri," *PloS one*, vol. 2, no. 7, p. e597, 2007.

[10] O. Sporns, *Networks of the Brain*. MIT press, 2010.

[11] A. Sokolov Arseny, G. Cristina, F. Elda, P. Giulia, R. Philippe, V. Dimitri, and K. J. Friston, "Brain network analyses in clinical neuroscience," *Swiss Archives of Neurology, Psychiatry and Psychotherapy*, vol. 170, no. 6, 2019.

[12] E. Fischi-Gomez, E. Muñoz-Moreno, L. Vasung, A. Griffa, C. Borradori-Tolsa, M. Monnier, F. Lazeyras, J.-P. Thiran, and P. S. Hüppi, "Brain network characterization of high-risk preterm-born school-age children," *NeuroImage: Clinical*, vol. 11, pp. 195–209, 2016.

[13] P.-T. Yap, Y. Fan, Y. Chen, J. H. Gilmore, W. Lin, and D. Shen, "Development trends of white matter connectivity in the first years of life," *PloS one*, vol. 6, no. 9, p. e24678, 2011.

[14] V. Batagelj and A. Mrvar, "Pajek-program for large network analysis," *Connections*, vol. 21, no. 2, pp. 47–57, 1998.

[15] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[16] A. Fornito, A. Arnatkevičiūtė, and B. D. Fulcher, "Bridging the gap between connectome and transcriptome," *Trends in Cognitive Sciences*, vol. 23, no. 1, pp. 34–50, 2019.

[17] A. Rosenberg, *How history gets things wrong: The neuroscience of our addiction to stories*. MIT Press, 2019.

[18] B. Szalkai, C. Kerepesi, B. Varga, and V. Grolmusz, "High-resolution directed human connectomes and the consensus connectome dynamics," *PloS one*, vol. 14, no. 4, p. e0215473, 2019.

[19] B. R. Howell, M. A. Styner, W. Gao, P.-T. Yap, L. Wang, K. Baluyot, E. Yacoub, G. Chen, T. Potts, A. Salzwedel, *et al.*, "The unc/umn baby connectome project (bcp): An overview of the study design and protocol development," *NeuroImage*, vol. 185, pp. 891–905, 2019.

[20] T. J. Prescott and D. Camilleri, "The synthetic psychology of the self," in *Cognitive architectures*, pp. 85–104, Springer, 2019.

 ISBN: 978-1-61208-885-3

Decoding Imagined Auditory Pitch Phenomena with an Autoencoder Based Temporal Convolutional Architecture

Sean Paulsen
Department of Computer Science
Dartmouth College
Hanover, USA
e-mail: paulsen.sean@gmail.com

Lloyd May
Computer Research in Music and Acoustics
Stanford University
Stanford, USA
e-mail: lloydmaydart@gmail.com

Michael Casey
Department of Computer Science
Dartmouth College
Hanover, USA
e-mail: michael.a.casey@dartmouth.edu

***Abstract*—Stimulus decoding of functional Magnetic Resonance Imaging (fMRI) data with machine learning models has provided new insights about neural representational spaces and task-related dynamics. However, the scarcity of labelled (task-related) fMRI data is a persistent obstacle, resulting in model-underfitting and poor generalization. In this work, we mitigated data poverty by extending a recent pattern-encoding strategy from the visual memory domain to our own domain of auditory pitch tasks, which to our knowledge had not been done. Specifically, extracting preliminary information about participants' neural activation dynamics from the *unlabelled* fMRI data resulted in improved downstream classifier performance when decoding heard and imagined pitch. Our results demonstrate the benefits of leveraging unlabelled fMRI data against data poverty for decoding pitch based tasks, and yields novel significant evidence for both separate and overlapping pathways of heard and imagined pitch processing, deepening our understanding of auditory cognitive neuroscience.**

***Keywords*—neuroimaging; neuroscience; auditory cognition; deep learning.**

I. Introduction

A. Motivation

Brain decoding is the problem of classifying the stimulus that evoked given brain activity. Music's well-defined structure and the wealth of previous results about the neural representation of that structure are thus an appealing foundation upon which to approach this problem. Our primary goal was to train a machine learning classification model to predict the pitch-class of a note (the relative position of the note within the key) given an input of brain activity evoked by that note. We hypothesized that such a classifier would achieve significant results for three tasks: trained and tested on neural activity when the note is actually heard (hereafter referred to as the "heard task"), the same when the note is only *imagined* ("imagined task"), and most importantly, trained on neural activity when the notes are heard but evaluated on data when the notes are imagined ("cross-decoding task") to test for overlap between heard and imagined pathways. To our knowledge, the cross-decoding task had not been done before. Toward these ends, we obtained functional Magnetic Resonance Imaging (fMRI) data from musically trained participants while they both heard and imagined particular pitches. We further detail our scanning protocol in the Methods and Materials section.

Training machine learning models on such voxel data is challenging, though, primarily due to the scarcity of relevant and labelled data to be used for training, and our experiments were no exception. However, Firat et al. [5]'s work on visual memory brain decoding addressed this challenge of fMRI data poverty in a novel and effective way. More specifically, Firat et al. hypothesized that unlabelled fMRI data, which are normally deemed irrelevant and discarded, contain information about overall patterns of brain activity and can therefore be exploited in brain decoding classification tasks. Their architecture began with a sparse autoencoder [10] to perform unsupervised learning of neural activation patterns latent in unlabelled fMRI data. These patterns then served as filters in a temporal Convolutional Neural Network [12] to encode the labelled fMRI data into a non-linear, more expressive feature space. We refer to the inputs of this pipeline as "unencoded datasets" and the outputs as "encoded datasets" throughout this work. Thus, the encoded dataset is the result of filtering the task-dependent fMRI data by the patterns latent in task-independent data. Firat et al. then demonstrated improved performance of Multi-Voxel Pattern Analysis (MVPA) classifiers trained and tested on encoded datasets compared to unencoded datasets.

B. Our Approach

In Section 2 of this paper, we expand on the architecture of Firat et al. by adapting their autoencoder-tCNN pipeline from the visual domain to our novel auditory domain task of decoding imagined pitch. Section 3 presents our results, in which our encoded datasets are *essential* for successful decoding of the imagined task, as well as first-of-their-kind significant results on the cross-decoding task. Section 4 discusses these results in the greater context of our goals and motivations. In particular, that this work demonstrates for the first time, to the best of our knowledge, that temporal filtering of fMRI data for an auditory task not only improves the performance of MVPA classifiers, but can also reveal fundamental, learnable attributes of auditory imagery that would go undetected by machine

 ISBN: 978-1-61208-885-3

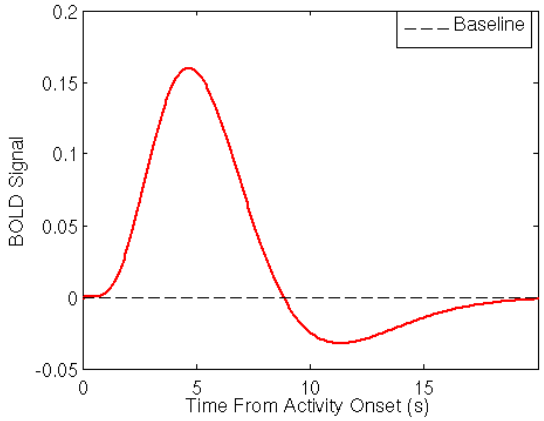


Fig. 1. Hemodynamic Response Function (HRF) plotted as a 6-TR timeseries. [14].

learning models trained on unencoded datasets. Section 5 details our methods and materials: participant selection, fMRI scanning protocol, hardware for training models, and statistical methods for evaluating our final classifiers. Section 6 concludes this paper and explores future work.

II. Architecture Design

A. Neural Activation Pattern Training Data

Each fMRI scan yielded a timeseries of 3-dimensional voxel data, where the value of each voxel represented the intensity of neural activity at that geographic location in the brain. We used the Python Multi-Variate Pattern Analysis (PyMVPA) [8] library to store and transform fMRI data throughout the experiment. When we imported a participant's fMRI data, PyMVPA flattened the 3D voxel data into a single spatial dimension by concatenating along two axes (during which all voxels are preserved), restricted to one of twenty selected Regions Of Interest (ROIs) at a time, and provided a mapping back to 3D space for that ROI. Thus, we began with a matrix VT of V-many voxels, which depended on each ROI, by T timesteps, which was 1864 for all participants and ROIs.

The Hemodynamic Response Function (HRF) in Figure 1 depicts the rise and fall of the intensity value of a voxel in response to a stimulus across 12 seconds. The time between images in our fMRI scans (TR) was 2 seconds, therefore the HRF would be observed across 6 timesteps in a given voxel. We thus expected any other latent activation patterns to occur across 6 timesteps as well. We therefore compiled our training data by sampling 1x6 windows of data from the matrix VT. Collecting every possible such window would provide the largest set of training data, but we believed the extreme overlap in that case could cause unpredictable bias during training. Spacing the samples out by exactly 6 timesteps would remove overlap, but could induce a different bias with every sample beginning and ending where another sample begins and ends, possibly limiting the kinds of patterns we expose to the model during training. Sampling with a stride greater than 6, however, might unnecessarily reduce the total size of our training set. Therefore our method considered each possible 6-TR window, then added it to the training data with probability 1/6. This allowed us to sample windows of training data that can begin at any timestep across the entire scan, while balancing our desire to both reduce overlap and minimize reduction of the training set. We further discarded any sample overlapping with labelled timesteps to avoid any possibility of downstream circularity. In summary, we collected 6-TR windows of unlabelled fMRI data, for each participant, for each ROI, to learn neural activation patterns latent in that participant in that ROI.

B. Learning the Patterns

We implemented a sparse autoencoder model to perform unsupervised learning of the latent temporal neural activation patterns among each region's voxels without the need for hand-crafted features. The sparse autoencoder was implemented with the Keras [3] library in Python. The model input was encoded by a dense layer with sparsity enforced by an "activity regularizer" parameter $\rho = .001$, hereafter referred to as the "sparsity constraint," and then rectified linear unit (ReLU) activation functions were applied to obtain the encoded version of the input. We refer to the preceding steps as the "encoding layer" throughout this paper. Each encoding layer had fourteen neurons in its dense layer, obtained via grid search on $\{8, 10, 12, 14, 16, 18\}$. Each neuron's set of trained weights would then serve as a filter for obtaining the encoded dataset. The decoding layer was also dense, with six neurons (recall that this layer attempts to reproduce the six-dimensional input) and ReLU activations. The model was optimized via backpropagation to minimize the mean squared error between the output of the decoding layer and the input using the "adamax" optimizer [11].

C. Filtering with Temporal Convolution

For each combination of participant and ROI, we extracted the set of learned neural activation patterns from the corresponding trained encoding layer and used them as filters in a tCNN to obtain the corresponding encoded dataset. Our tCNN pipeline is depicted in Figure 2. More specifically, we performed a 1D full convolution on the VT matrix along its time axis with each of the fourteen trained neurons as the temporal filter. This resulted in fourteen response matrices for each combination of participant and ROI. Note that a full convolution means each response matrix had the same dimensions as VT.

We expected the voxels to exhibit locally correlated activations [13], so we employed max pooling to extract spatial information from the filtered data in our response matrices. Recall, though, that VT is the result of flattening the 3D voxel space to 1D, and therefore voxels next to each other in VT are not necessarily next to each other geographically in the brain. Firat et al. [5] did not detail their solution to this problem of 3D max-pooling with 1D data, so we devised our own method. Recall that PyMVPA provided a mapping back to the 3D voxel space of unencoded voxel values for each ROI, so we directly we backfilled the original 3D space with the values of each response matrix.

 ISBN: 978-1-61208-885-3

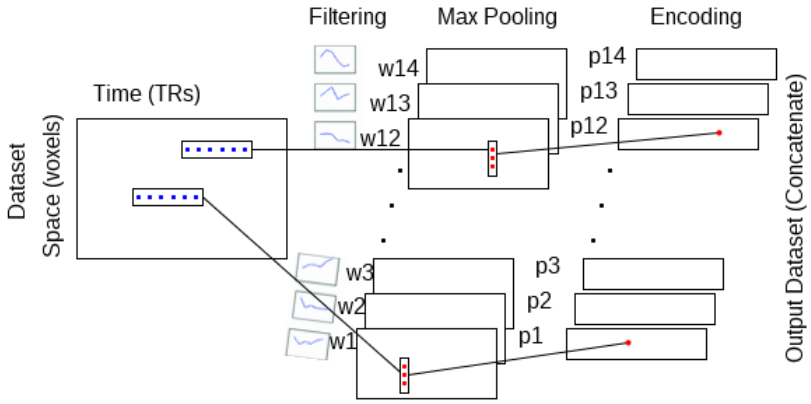


Fig. 2. Our tCNN pipeline from voxel space to the encoded dataset. The filters are the neurons extracted from each trained autoencoder and represent neural activation patterns.

For 3-dimensional spatial max-pooling, we proposed a pooling cube of tunable dimensions $[c_1,c_2,c_3]$ moving exhaustively throughout each 3D space with no overlap, storing the maximum value within the cube at each step in a list. The jagged 3D voxel structure of each ROI was padded on all sides with zeroes due to the way PyMVPA maps back from 1D to 3D, so these zeroes needed to be accounted for. We certainly did not want to record a zero as a max-pooled value when the pooling cube is full of these padding zeroes, and more subtly we did not want to record voxel values on the jagged fringes as max-pooled values when they were being compared almost entirely to padding zeroes. Our solution was a tunable parameter z_0 which we called "zero threshold". The maximum value within the cube was only recorded as a max-pooled value when the proportion of non-zero values within the pooling cube exceeded z_0. Our [2, 2, 2] pooling cube and zero threshold of 0.6 were obtained via grid search.

We performed our method of 3D max-pooling on each timestep for each of the response matrices, applied hyperbolic tangent to each list of max-pooled values, and finally concatenated the lists for each timestep. The result of the concatenation was the encoded dataset for that participant and ROI. A repository of our code is available upon request.

D. Pitch Decoding Classifiers

For each participant and ROI, we partitioned the labelled fMRI data by whether the corresponding pitch was heard or imagined. The heard samples were split further in half, with each half serving in turn as training data and testing data for an MVPA classifier. We stored the trained classifiers' predictions on the respective test sets with their corresponding pitch-class labels. Our analysis of classifier performance on the heard task was performed on the union of the two halves of test set predictions for each participant and ROI. The imagined task was evaluated similarly. For the cross-decoding task, we trained the classifier on all heard data, then predicted the labels of all imagined data. We calculated group level significance for each task and ROI using a t-test between per-participant prediction mean accuracies and null decoding model mean accuracies, detailed further in the Methods and Materials section.

III. Results

A. Temporal Filter Results

Figure 3 shows twenty learned temporal filters (i.e, trained neurons) uniformly at random across the encoding layers of all participants and ROIs. Six weights connect each such neuron to the input layer, one for each timestep in the input, so we plotted the raw values of each sampled neuron's weights as a timeseries. This allows us to visually evaluate the learned filters as a pattern of neural activity. Observe that several of these patterns are good approximations of the HRF, which we expected most of the autoencoders to learn. Note further that none of the patterns are dominated by a single weight, which is to say that the models were not biased toward any particular timestep in the input data. This was the intent of our careful creation of each autoencoder's training data.

B. Brain Decoding Results

Table I contains the results of our pitch decoding experiments. We evaluated the group-level statistical significance of the multivariate classifiers' ability to outperform chance in each of our regions of interest. The region of interest is given in the first column. The second column indicates the task, as explained above. The next two columns give the accuracy and False Discovery Rate (FDR)-corrected p-values when the classifiers were trained and evaluated with their respective encoded dataset, and the last two columns give the same information on the unencoded dataset. Observe one of our critical results, that thirteen of the fifteen successful regions *required* the encoded dataset to obtain statistical significance. Eleven of the fifteen significant results were for the imagined task, and indeed *all* of these regions required the encoded dataset for significance.

IV. Discussion

A. Architecture Discussion

Our first goal was to learn auditory neural activation patterns latent in 6-TR windows of unlabelled fMRI data with sparse

 ISBN: 978-1-61208-885-3

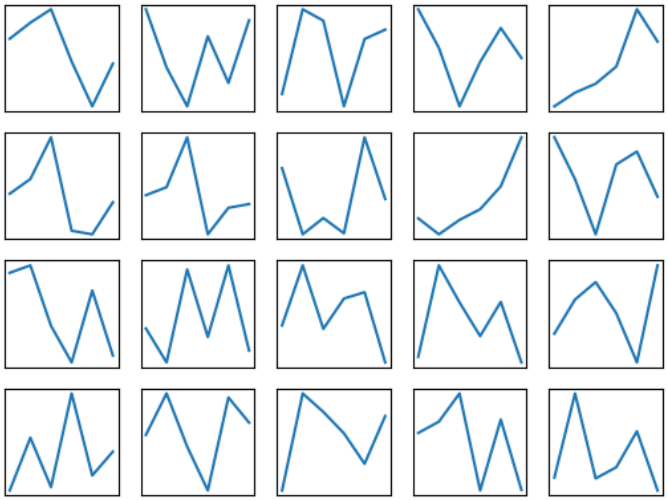


Fig. 3. Learned temporal filters, sampled uniformly at random across all sparse autoencoders. Each consists of six weight values, one for each timestep. The HRF appears to have been learned by several of the selected neurons.

TABLE I
WITHIN-SUBJECT CLASSIFIER RESULTS. FDR-CORRECTED P-VALUES FOR ALL ROIS WITH SIGNIFICANT RESULTS. THE ENCODED DATASETS ENABLED THE CLASSIFIERS TO OBTAIN SIGNIFICANT RESULTS IN THIRTEEN OF THE FIFTEEN SIGNIFICANT REGIONS.

		Encoded Dataset		Unencoded Dataset	
Region of Interest	**Task**	**WPC Accuracy Mean (Min, Max) baseline = 0.1429**	**FDR-corrected pvals (threshold = 0.05) (20 ROIs)**	**WPC Accuracy Mean (Min, Max) baseline = 0.1429**	**FDR-corrected pvals (threshold = 0.05) (20 ROIs)**
Left Heschl's Gyrus	H	0.1642 (0.1250, 0.1964)	**0.0039**	0.1523 (0.0833, 0.2262)	0.5808
Right Superior Temporal Sulcus	H	0.1394 (0.0833, 0.2143)	0.8554	0.1754 (0.1190, 0.2560)	**0.0071**
Left Inferior Frontal Gyrus (Orbitalis)	I	0.1625 (0.0893, 0.2381)	**0.0368**	0.1485 (0.0952, 0.2024)	0.6475
Left Precentral Gyrus	I	0.1607 (0.1190, 0.2262)	**0.0368**	0.1530 (0.0952, 0.2202)	0.5228
Left Superior Temporal Gyrus	I	0.1684 (0.1190, 0.2202)	**0.0087**	0.1586 (0.1131, 0.2143)	0.2326
Left Supramarginal Gyrus	I	0.1642 (0.0952, 0.2440)	**0.0355**	0.1502 (0.0893, 0.2083)	0.6291
Left Insula	I	0.1649 (0.1310, 0.2440)	**0.0163**	0.1478 (0.0833, 0.2024)	0.6475
Right Superior Temporal Sulcus	I	0.1604 (0.1131, 0.2083)	**0.0180**	0.1499 (0.0893, 0.2083)	0.6291
Right Inferior Frontal Gyrus (Triangularis)	I	0.1688 (0.0952, 0.2500)	**0.0368**	0.1642 (0.1131, 0.2202)	0.0996
Right Precentral Gyrus	I	0.1719 (0.1071, 0.2321)	**0.0103**	0.1569 (0.0952, 0.2560)	0.5228
Right Superior Temporal Gyrus	I	0.1726 (0.1190, 0.2560)	**0.0124**	0.1453 (0.1012, 0.1845)	0.8149
Right Supramarginal Gyrus	I	0.1656 (0.1250, 0.2440)	**0.0251**	0.1586 (0.0774, 0.2440)	0.5228
Right Insula	I	0.1649 (0.1190, 0.2381)	**0.0124**	0.1506 (0.0893, 0.2024)	0.6291
Right Superior Temporal Gyrus	X	0.1628 (0.1310, 0.1964)	**0.0157**	0.1492 (0.1071, 0.2083)	0.6445
Right Rostral-Middle Frontal Gyrus	X	0.1509 (0.1250, 0.1905)	0.3619	0.1642 (0.1131, 0.2143)	**0.0202**

autoencoders. We took care to avoid subtle biases when we collected our training data for the autoencoders by minimizing the overlap of the samples while allowing for the possibility of a sample to begin at any timestep in the scan. We plotted the weights of twenty uniformly randomly sampled encoder-layer neurons as timeseries to visualize the neural activation patterns that those neurons represented. These visualizations reassured our efforts in two ways. First, several of them are good approximations of the HRF, which we expected to be learned by one of the neurons in most of the autoencoders. Second, none of the patterns are dominated by a single timestep, and the peaks of activity are fairly well distributed across the timesteps, which was the intent of our training data collection method. These considerations, along with the success of our brain decoding classifiers, provide evidence that each neuron learned a latent auditory neural activation pattern, accomplishing our first goal.

Our second goal was to generate a collection of encoded datasets by transforming the unencoded voxel data VT in terms of the neural activation patterns learned by each autoencoder's encoding layer. Thus, the final step of our architecture was a modified tCNN. We used each of the learned activation patterns as temporal filters by convolving them with their respective VT along the time axis and applying our own method of 3D max pooling. Concatenating the pooled matrices for each participant and ROI yielded our encoded datasets, thus achieving our second goal. Note that the final dimension after concatenating was dependent on the size of the 3D pooling cube and the number of filters. In our experiments the encoded datasets' dimensions ended up being roughly equal to the dimension of their respective unencoded dataset. However, one could increase the size of the pooling cube or learn fewer temporal filters if

 ISBN: 978-1-61208-885-3

the dimensionality were a burden on computing. That would, of course, be a tradeoff with performance, but it is nevertheless valuable to have a mechanism for dimension reduction available in this pipeline.

B. Brain Decoding Discussion

Our third goal was to train a machine learning classifier to predict the pitch-class labels of heard and imagined pitches, trained and tested on fMRI data of twenty selected regions of interest. We hypothesized that such classifiers would outperform chance with statistical significance, and that the classifiers would achieve higher accuracy when trained on encoded datasets versus the unencoded datasets. We used the PyMVPA library to train multi-class Support Vector Machines (SVMs) with linear kernels on each of the encoded datasets and each of the unencoded datasets. Each classifier's accuracy was calculated on a held out test set, and the accuracies were averaged across participants for each ROI. Finally we calculated the group-level significance of the accuracies and controlled the FDR by correcting our p-values for multiple comparisons. Further details are in the Methods and Materials section.

As shown in Table I, the statistical significance of outperforming chance relied almost entirely on the encoded datasets. For the imagined task the classifiers did not obtain significant results in any ROIs using *unencoded* data. Indeed, training on the encoded datasets did not merely nudge almost-significant p-values past the threshold, but quite the opposite. Our encoded datasets enabled the classifiers to reduce their p-values by more than an order of magnitude in most regions in Table I, and *two* orders in some, indicating that the encoded dataset reveals fundamental, learnable attributes of auditory imagery that would otherwise remain undetected by machine learning models trained on unencoded data. Thus, we achieved our third goal and obtained statistically significant evidence of our hypothesis in the case of the imagined task. Moreover, the significant results on the cross-decoding task provide a critical novel result- statistically significant evidence of geographical overlap between heard and imagined sound.

Eleven of the fifteen significant results were achieved on the imagined pitch decoding task. This is explained by the greater cognitive involvement in imagining versus hearing sound. That is, imagining sound is a more involved activity than listening, evoking stronger, wider signals that are easier for the autoencoder to detect and learn.

The heard and cross-decoding tasks both achieved two significant results, one each on the encoded and unencoded datasets. In both cases of significant unencoded datasets, the p-value for the respective encoded dataset was at least an order of magnitude worse. For the heard task, the two regions are near each other- Heschl's Gyrus and the Superior Temporal Sulcus are both auditory cortex areas in the superior temporal lobe. Therefore, while the inconsistency of the encoded dataset on the heard task requires further study, the results on the heard task are geographically consistent. On the other hand, the significant regions on the cross-decoding task are in separate lobes and non-adjacent. The Right Rostral-Middle Frontal Gyrus is interesting because significant results were achieved on the cross-decoding task with the unencoded dataset with a p-value at least an order of magnitude better than any other region for that task and dataset. Further, for the heard and imagined tasks, the encoded dataset improved the p-values in this region. Thus, the significant result in the Right Rostral-Middle Frontal Gyrus is curious, piquing further study.

V. Methods and Materials

A. Participant Selection

Participants possessed at least 8 years of formal music training or professional performance experience in Western tonal music, and they completed the Bucknell Auditory Imagery Scale (BAIS) [7] and the Bregman Musical Ability Rating Survey [9]. Twenty-three such participants passed the screening process and provided their written informed consent in accordance with the Institutional Review Board at Dartmouth College. Each subject was compensated $20 US upon completion of the scan.

All scanning used a 3.0 T Siemens MAGNETOM Prisma MRI scanner with a 32-channel head coil and Lumina button box with four colored push buttons. Each scan performed a T2* weighted single shot echoplanar (EPI) scanning sequence with a repetition time (TR) of 2 sec and 240mm field of view with 3mm voxels, yielding 80 voxel by 80 voxel images with 35 axial slices for a total of 224,000 voxels per volume. We used the fMRIPrep software [4] to perform motion correction, field unwarping, normalization, and bias field correction preprocessing, as well as brain extraction and ROI parcellation, on the raw T2* BOLD data.

B. fMRI Protocol

Each participant's fMRI scan consisted of 8 runs of 21 musical trials. Each scan was randomly assigned either the key of E Major or F Major, which was not known by the participant. We designed each run to collect data for either the heard task or the imagined task, alternating from run to run. Each trial began with an arpeggio in the assigned key for the participant to internally establish a tonal context, followed by a cue-sequence of ascending notes in their assigned major scale. After a randomized time interval, the participant either heard the next ascending note in the scale, or was instructed to imagine the next ascending note, depending on the run. The following four seconds (2 TRs) of scanning collected from all trials constituted the labelled data for the heard and imagined tasks. Next, a probe tone was played, and the participant rated the probe tone's goodness of fit in the tonal context from 1 to 4. We excluded the data of any participant with at least 20% of their ratings missing, or whose ratings did not reflect internalization of the tonal hierarchy. Thus, we excluded the data of six of the twenty-three participants.

Previous literature on imagined and heard tonal pitch-classes directed us to twenty regions of interest in the frontal, temporal, and parietal lobes according to the Desikan-Killiany (D-K) atlas in Freesurfer [6]. The D-K ROIs are large cortical regions, reducing the burden of correcting for multiple comparisons

 ISBN: 978-1-61208-885-3

compared to a larger quantity of smaller regions. Further, the D-K ROIs are consistent with the scales of relevant previous literature. The full table of the ROI atlas indices, cortical labels, and corresponding Brodmann areas is available on request.

C. Autoencoder Training

The autoencoders were trained on Intel Xeon E5 processors, either 2.3, 2.6, or 3.2 GHz for 30 epochs on Dartmouth's Discovery High Performance Cluster with an average training time of approximately 3 hours. 10% of the training data were held out as a validation set during training to prevent overfitting via early stopping. For each combination of participant and ROI, we trained ten autoencoders and kept the model with the lowest validation accuracy after 30 epochs. This was to avoid the rare but observed case where an autoencoder failed to find any minima during training.

D. MVPA Classifiers

For each ROI, we partitioned the labelled fMRI data of each participant into two halves according to whether the pitches were heard or imagined. We then split the heard data in half, with each half serving in turn as training data and testing data for a multi-class SVM with linear kernels. We implemented the SVMs with the libSVM support vector machine library [2]. We then pooled the classifier's predictions on each of the two rounds of test data into a single set, along with their corresponding pitch-class labels. Our analysis of the heard task was performed on this collection of predictions and labels for each participant and region of interest. The imagined task was evaluated similarly. For the cross-decoding task, the classifier used all heard data for training, then predicted the labels of all imagined data. We calculated group level significance for each task using a t-test between per-participant prediction mean accuracies and null decoding model mean accuracies. We used Monte Carlo simulation to calculate the null models, repeating each classifier's training and testing 10,000 times with randomly permuted target labels and storing the mean overall accuracy. We corrected the group-level p-values for multiple comparisons using the method in Benjamini and Hochberg [1], which strictly controls the FDR of a family of hypothesis tests.

VI. Conclusion and Future Work

In this work, we adapted the architecture and pipeline of Firat et al. [5] from the visual domain to the auditory domain. Latent neural activation patterns were learned from unlabelled fMRI data, which are normally discarded, in order to generate our encoded datasets, which improved the performance of downstream MVPA classifiers. On the task of decoding the pitch class of imagined sound from fMRI data, the encoded datasets enabled the classifiers to outperform chance with group-level statistical significance in eleven ROIs. This demonstrated for the first time, to the best of our knowledge, that exploiting unlabelled fMRI data to perform temporal filtering for an auditory task not only improves the performance of MVPA classifiers, but can also reveal fundamental, learnable attributes of auditory imagery that would go undetected by machine learning models trained on unencoded datasets. Further, the group-level classifier performance on the cross-decoding task in two ROIs provided our novel statistically significant evidence of geographical overlap between heard and imagined sound.

There are several immediate directions for future work. First is toward an end-to-end architecture for this task, rather than a disconnected training session to obtain the encoded datasets. Second is toward decoding/cross-decoding the other information in our fMRI protocol, such as the timbre (clarinet or trumpet) of the heard or imagined sound. Third is toward the generalization of our pipeline to other fMRI datasets with auditory tasks. Fourth is a deeper dive on the ROIs with significant cross-decoding results, as these results did not quite match our expectations.

References

[1] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," Journal of the Royal Statistical Society, Series B (Methodological), vol. 57(1), pp. 289–300, 1995.

[2] C. C. Chang and C. J. Lin, "Libsvm: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2(3), pp. 1–27, 2011.

[3] F. Chollet, *Keras*, [Online], https://github.com/keras-team/keras 2021.06.16

[4] O. Esteban et al., "fMRIPrep: a robust preprocessing pipeline for functional mri," Nature Methods, vol. 16(1), pp. 111–116, 2019.

[5] O. Firat, E. Aksan, I. Oztekin, and F. T. Yarman Vural, "Learning deep temporal representations for fmri brain decoding," in proceedings of the Machine Learning Meets Medical Imaging workshop in conjunction with ICML, K. Bhatia and H. Lombaert, Eds. Springer International Publishing, pp. 25-34, 2015.

[6] B. Fischl, "Freesurfer," Neuroimage, vol. 62(2), pp. 774–781, 2012.

[7] A. Halpern, "Differences in auditory imagery self-report predict neural and behavioral outcomes," Psychomusicology: Music, Mind and Brain, vol. 25, pp. 37–47, March 2015.

[8] M. Hanke et al., "Pymvpa: a python toolbox for multivariate pattern analysis of fmri data," Neuroinformatics, vol. 7(1), pp. 37–53, 2009.

[9] M. Hanke et al., "A high-resolution 7-tesla fmri dataset from complex natural stimulation with an audio movie," Scientific Data, vol. 1, May 2014.

[10] K. Kavukcuoglu, M. Ranzato, R. Fergus, and Y. LeCun, "Learning invariant features through topographic filter maps," in proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1605–1612, 2009.

[11] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[12] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 156–165, 2017.

[13] F. Pereira and M. Botvinick, "Information mapping with pattern classifiers: a comparative study," Neuroimage, vol. 56(2), pp. 476–496, 2011.

[14] D. Stansbury, *fMRI in neuroscience: the basics*, [Online], https://theclevermachine.wordpress.com/tag/finite-impulse-response-model 2021.06.16

 ISBN: 978-1-61208-885-3

Mechanical Dermal Stimulation to Modulate the Interoceptive Network in Sleep-disordered Populations

Gina Sensale, Sahithi Garikapati, Jean Toher, Hanna Villa, Sean Hagberg
Feelmore Labs, Inc.
Brooklyn, NY, USA
e-mail: gina@feelmorelabs.com, sahithi@feelmorelabs.com, jmtoher@gmail.com, hannakvr@gmail.com, sean@feelmorelabs.com

***Abstract*— More than 80 million adults are affected by insufficient sleep in the United States, yet effective, drug-free therapies with few or no side effects are lacking. A 30-day open-label study was conducted, in which participants (n=25) reporting poor sleep (assessed via Pittsburgh Sleep Quality Index) were recruited to test a novel wearable mechanical neuromodulation device. The device was designed to modulate the interoceptive system via the affective touch network. 86% of participants showed an overall improvement in sleep after 30 days of device use.**

Keywords-sleep; neurostimulation; interoception.

I. INTRODUCTION

There is growing evidence that humans are hard-wired for receiving and processing affective (slow, light) touch via unmyelinated C-Tactile Afferent neurons (CTAs). Neuroimaging studies have found information from CTAs is primarily processed in the insula cortex, not the somatosensory cortex where myelinated inputs are processed [1]. This distinction is important to note when considering the growing evidence in support of the fundamental role of the insula in interoceptive processing. Affective touch, in turn, is associated with feelings of calm, relaxation and social connectedness, all related to improved interoceptive regulation.

Interoception is recognized as the processing of internal bodily signals such as changes in physiological states, via the Central Nervous System (CNS) [2][3]. Information is continuously exchanged through ascending and descending pathways between the CNS and the periphery, allowing for interoceptive information such as a change in temperature, to influence various physiological systems and processes, including sleep [3][4]. Furthermore, research has shown that interoceptive sensitivity and awareness may disrupt several processes, during sleep initiation and while asleep [2][3][4]. Dysregulated interoception, a mismatch between perceived and actual bodily states, is a hallmark of affective disorders and may be an underlying mechanism for sleep-related disorders. In a 2016 report, an estimated 83 million adults in the United States suffered from insufficient sleep; however, few drug-free therapeutic interventions exist that have few side effects and are effective in treating sleep-related disorders [7]. We developed a novel mechanical stimulation device, targeting the affective touch pathway, and thereby modulating the interoceptive system to improve sleep.

The objective of this research is to develop a device designed to stimulate the interoceptive system via the affective touch pathway, and then test the technology in the real world, targeting symptoms associated with dysregulated interoception. Here, we targeted poor sleep, measured by the Pittsburgh Sleep Quality Index (PSQI) [6].

The rest of this paper is organized as follows. Section II outlines study methods, including a description of the participant sample and the device technology used, an ethics statement, participant responsibilities, and measures used to assess sleep quality. Section III contains the detailed results of the present study and Section IV closes out the paper with the conclusion.

II. METHODS

A. Participants

25 adults (14 females, 11 males), out of a total of 245 screened, were enrolled in the study after meeting the criteria for poor sleep, measured by the PSQI (Global PSQI $\geq$ 10). Ages ranged from 24 to 60 years old (mean=35).

B. Mechanical Stimulation Device

A simple headband with small piezoelectric actuators at the distal ends, seen in Figure 1, was developed to deliver short bursts of very low intensity, low frequency mechanical stimulation, targeting CTAs associated with the affective touch pathway. The actuators were positioned behind the ears, on the mastoid, for convenience.

Figure 1. Mechanical Stimulation Device Prototype.

 ISBN: 978-1-61208-885-3

The specific wave form was derived from a combination of empirical study (changes in alpha power pre/post stimulation in in-lab studies over 2 years) and known response characteristics of the CTA mechanoreceptors (low intensity, low frequency ~10 Hz).

C. Study Procedure

All study procedures were reviewed and approved by an ethical board (Solutions IRB, #: FWA00021831) [8]. The 25 participants who consented to the study were familiarized with the device, instructed in its use, and the device was donned by participants. A researcher, with feedback from the participant, reduced the intensity to the lowest perceivable level for that individual. Participants could increase/decrease the intensity throughout the 30-day study. The device logged usage time and intensity. The participant then had their first 20-minute session in our lab to assess any side effects and ensure competence in using the device. They were instructed to use the device for 20 minutes every day, within an hour of their normal bedtime. In addition, participants wore a wrist device to track sleep (Garmin VivoSmart 4).

D. Measures

Changes in sleep quality were assessed via PSQI, a 1-Item Sleep Quality Rating Scale, and the Garmin VivoSmart 4 wrist device. The PSQI is a self-report measure assessing sleep quality and disturbances over a 1-month timeframe. The 1-Item Sleep Quality Rating Scale is a self-report measure utilizing a scale of 1 to 5, where 1 represents little to no sleep at all, and 5 represents great sleep (no problems falling or staying asleep).

III. RESULTS

In the sample of 25 participants, 3 participants were excluded from analysis due to lack of compliance with the study protocol (i.e., device usage, completing the study). The following results will be reported as Mean ± standard deviation. 86% of compliant participants (n=22) reported an overall improvement in sleep, measured by Global PSQI scores shown in Figure 2, where lower scores indicate improvement in symptoms. Global PSQI scores improved by 43% on average (Pre: 9.8 ± 3.0, Post: 5.2 ± 2.6).

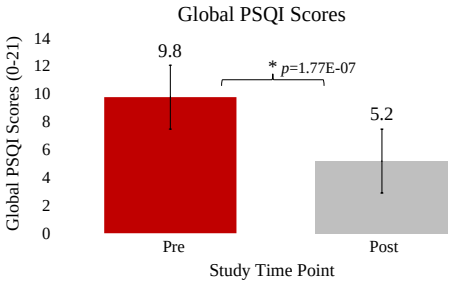


Figure 2. Global PSQI scores decreased on average, after 30 days of device use, representing an improvement in sleep.

More specifically, 91% of participants reported improvement in sleep quality, 77% reported falling asleep faster, and 68% reported a reduction in daytime drowsiness, shown in Figure 5, displaying average PSQI component scores pre and post 30 days of device use. Similarly, sleep hours increased 65 minutes on average (Pre: 6.60 ± 0.29, Post: 7.45 ± 0.45), shown in Figure 3, and self-reported sleep quality, shown in

Figure 4, improved significantly (Pre: 3.71 ± 0.31, Post: 4.25 ± 0.10). Sleep hours were assessed via a commercial wrist Photoplethysmography (PPG) device (Garmin VivoSmart 4, chosen as it was most reliable for sleep time in earlier studies). No devices appeared, in our assessment, to accurately measure sleep stages, nor is PPG a suitable substitute for a hypnogram.

Subjects rated the device as simple and easy to use. There were no known adverse effects and some minor transient side-effects (headache, skin irritation) that subsided with use.

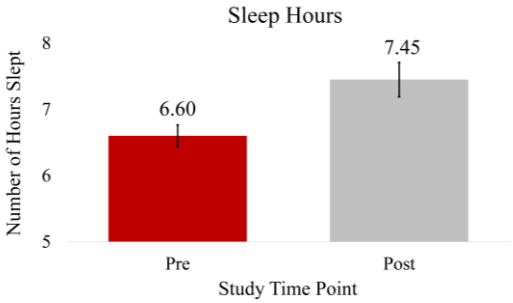


Figure 3. Sleep Hours increased on average, after 30 days of device use (measured by Garmin VivoSmart 4).

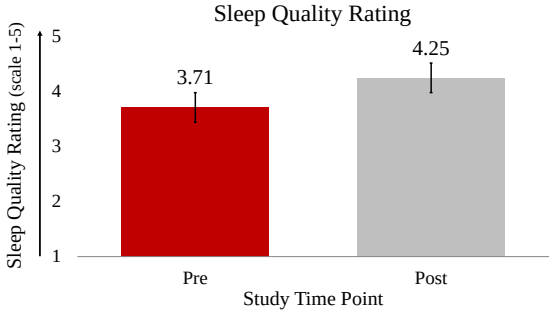


Figure 4. Sleep Quality Rating increased on average, after 30 days of device use.

IV. CONCLUSION AND FUTURE WORK

This is the first human study to evaluate mechanical stimulation of the affective touch pathway in a sleep-disordered population. Although the trial is small, open-label, and used early prototypes, the results were significant and participants clearly thought they benefited from the use of the device. A confirmatory Randomized

 ISBN: 978-1-61208-885-3

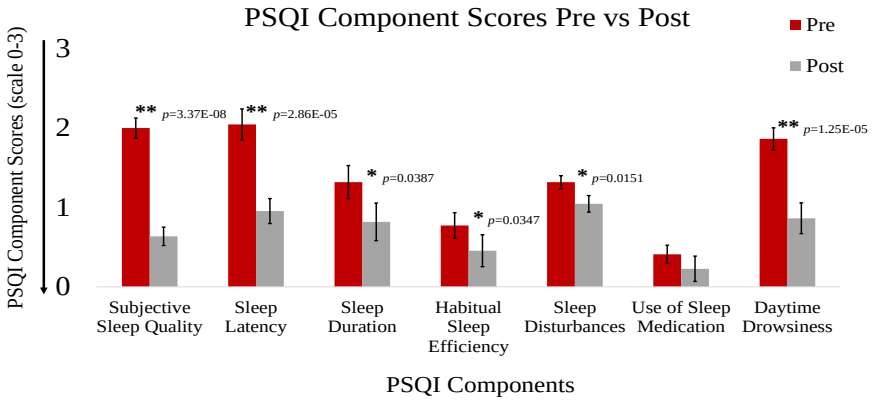


Figure 5. PSQI Component Scores before and after 30 days of device use, representing an overall improvement across 7 dimensions of sleep on average, most notably in Subjective Sleep Quality, Sleep Latency, and Daytime Drowsiness.

Control Trial (RCT) is underway to address limitations such as sample size and inclusion of a control group, which will be completed in late 2021.

REFERENCES

[1] R. Ackerley, I. Carlsson, H. Wester, H. Olausson, and H. B. Wasling, "Touch perceptions across skin sites: differences between sensitivity, direction discrimination and pleasantness," Frontiers in Behavioral Neuroscience, vol. 8, pp. 54, Feb. 2014, doi:10.3389/fnbeh.2014.00054.

[2] S. S. Khalsa et al., "Interoception and mental health: a roadmap," Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, vol. 3, pp. 501-513, June 2018, doi:10.1016/j.bpsc.2017.12.004.

[3] Y. Wei and E. J. W. Van Someren, "Interoception relates to sleep and sleep disorders," in Current Opinion in Behavioral Sciences, vol. 33, pp. 1-7, June 2020, doi:10.1016/j.cobeha.2019.11.008.

[4] E. F. Pace-Schott et al., "Physiological feelings," Neuroscience & Biobehavioral Reviews, vol. 103, pp. 267-304, Aug. 2019, doi:10.1016/j.neubiorev.2019.05.002.

[5] A. Schulz and C. Vögele, "Interoception and stress," Frontiers in Psychology, vol. 6, Jul. 2015, doi:10.3389/fpsyg.2015.00993, URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4507149/ [access date: 5/10/2021].

[6] D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer, "The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research," Psychiatry Research, vol. 28, pp. 193-213, May 1989, doi: 10.1016/0165-1781(89)90047-4.

[7] Y. Liu et al., "Prevalence of healthy sleep duration among adults — United States, 2014," Morb Mortal Wkly Rep, vol. 65, pp. 137-141, Feb. 2016, doi: doi:10.15585/mmwr.mm6506a1

[8] Solutions IRB https://www.solutionsirb.com/

 ISBN: 978-1-61208-885-3