

Think to Speak - A Piezoelectric-EEG system for Augmentative and Alternative Communication (AAC) using Recurrent Neural Networks

Robert Sowah, Ryan Friedman+, Abdul R. Ofoli*, and Baffour Sarkodie-Mensah

Computer Engineering Department, University of Ghana
PMB 25, Legon, Accra-Ghana

+Electrical and Computer Engineering Department, Tufts University, Massachusetts, USA

*Electrical and Computer Engineering Department, University of Tennessee Chattanooga

rasowah@ug.edu.gh, ryan.friedman@tufts.edu, abdul-ofoli@utc.edu, and basrkodiemensah@gmail.com,

Abstract— The collection of individuals with severe speech and physical impairments (SSPI), is the target audience for the Think to Speak Augmentative and Alternative Communication (AAC) system. The slow communication rate of AACs accessible to the target audience renders them undesirable, exhausting to operate, and a barrier to social and economic inclusion. This research synergizes the use of Electroencephalography (EEG) and high-sensitivity piezoelectric sensor readings with a Long Short-Term Memory Recurrent Neural Network (LSTM RNN) to create a physically accessible AAC with performance comparable to 7.8 characters per minute communication rate. Since self-expression is inextricably linked with physical, mental, and emotional health, this research is of great significance to the estimated one percent of the global population with complex communication needs.

Keywords— Piezoelectric sensor, Electroencephalography (EEG), Long Short-Term Memory, Recurrent Neural Network, Complex Communication Needs.

I. INTRODUCTION

The field of study and collection of technological advancements explored to benefit individuals who cannot exclusively rely on spoken communication is referred to as Augmentative and Alternative Communication, or AAC [1]. One of the most prevalent speech conditions, dysarthria, is characterized by “difficulty in articulating words, caused by impairment of the muscles used in speech [2]” and affects 170 per 100,000 of the U.K. population [3, 13] and is often caused by “neurological disorders such as stroke, brain injury, brain tumors, and conditions that cause facial paralysis or tongue or throat muscle weakness [4].” If the individual is indeed capable of producing speech, it is usually unintelligible to unfamiliar communication partners therefore necessitating the use of an Augmentative and Alternative Communication (AAC) [5]. The American Speech-Language-Hearing Association’s (ASHA) compilation of research was one of the main resources used in this project. At their core, AACs can be reduced to three fundamental components: the *symbols* themselves (e.g. characters, images, brail), *symbol display* (e.g. keyboard, communication board, virtual environment), and *symbol selection* (via touch, eye-tracking, brain waves, et cetera).

Due to the physical limitations of the target audience, Electroencephalography (EEG) was chosen as the primary means of response selection due to the near universal accessibility of brain waves for any being capable of engaging in communication. EEG is the process of measuring the human brain’s electric field and the first documented recordings were

conducted by Hans Berger in the late 1920s [6]. EEG signals have low signal amplitude coupled with a low frequency bandwidth (1 Hz – 50 Hz) poses some challenges for the accurate measurement of brain waves, especially in the case of dry sensor signal acquisition [7]. For face-to-face communication, minimizing the disruption of natural eye contact between the user and the person being communicated to via the system is very important. Hence providing the user with both audio and visual feedback will make the *Think to Speak* system more physically accessible (i.e. for the visual or hearing paired) while also facilitating more natural communication. and hence the need for a piezoelectric sensor. Figure 1 shows a sample different EEG signals for brain activities.

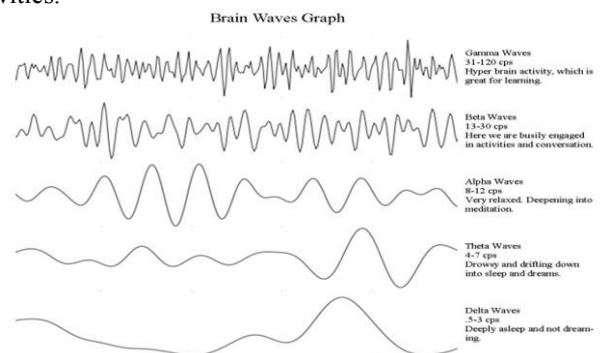


Fig. 1 Sample EEG Signal for brain activities

Artificial Neural Networks, or ANNs, are computational models derived from the intersection of primarily biology and neuroscience. ANNs consist of nodes (modeling neurons of the brain) connected by directed links (modeling synapses). Each link, or synapse, has a numeric weight associated with it corresponding to the “strength and sign of the connection [9].” A simplified neuron is depicted below in figure 2.

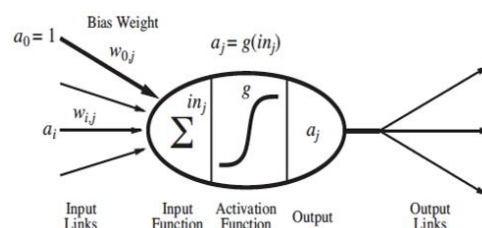


Fig. 2 McCulloch-Pitts model of Artificial Neuron

The arrows on the left of figure 2 illustrate a set of synapses (each having their own weights). They are fed into a weighted summation processing unit, effectively taking the linear combination of the synapse activations. The resulting summation then passes through a non-linear activation function to derive the neuron's output.

The type of network architecture is dictated by the manner in which the layers of the ANN are connected. A Recurrent Neural Network (RNN) has a feedback loop that provides the RNN cell with access to information in the present moment and in the recent past [11]. This feedback loop preserves information over multiple time steps such that correlations can be made between events far beyond just the current and most recent events.

The result is of supreme importance: sequential information is preserved in RNNs and can be used to form additional correlations and therefore improve model comprehension of inputted data. This is the primary reason why RNNs are applied to natural language processing (NLP) tasks. When the internal storage state of an RNN is unrolled and replaced by a long short-term memory (LSTM) cell, an LSTM Network is formed. The added complexity of an LSTM network results in the retention of information for even longer periods of time, as compared to traditional RNNs. This added memory, and therefore heightened level of comprehension, makes LSTM networks very attractive for NLP tasks. A traditional RNN architecture compared to that of an LSTM cell is depicted below [17] in figure 3.

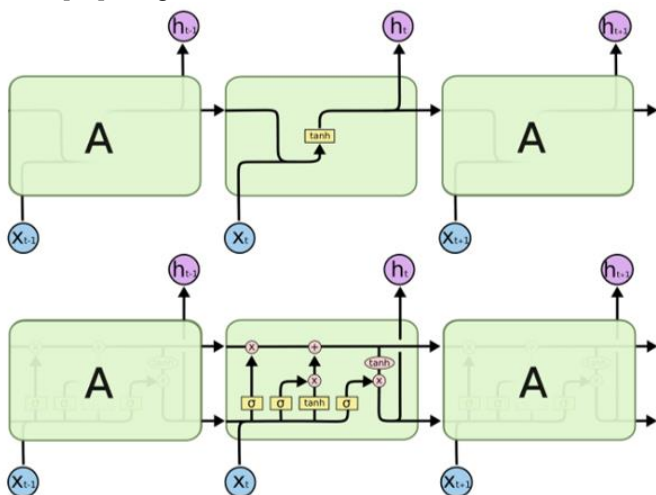


Fig. 3 traditional RNN architecture compared to that of an LSTM cell

The objective of this project is to create a more affordable and adaptable AAC system with LSTM RNN response generation and an EEG-piezoelectric sensor network to enable non-speaking users with severe motor to express themselves at a satisfaction rate of 80% and an accuracy rate of 95%.

In [5], the authors explore the use of a Voice-Input Voice-Output Communication Aid (VIVOCA) to sense disordered speech from a user with moderate to severe speech impairment (categorized by conversational intelligibility up to 50%), determine what the user is likely trying to communicate, and output a clear synthesized audio signal of the message. The system designers leveraged Hidden Markov Models (HMMs) to aid in the Automatic Speech Recognition (ASR) process. Testing of the system resulted in a mean recognition accuracy of 67% which “limit[s] the performance and usability of the device when applied in real usage situations.” This system will not be useful in the case of people with severe speech impairment.

In [18], the authors implemented selected EEG signal processing algorithms in asynchronous Brain-Computer Interface (BCI). The authors highlighted the necessity of removing power line frequency disturbance in EEG signal pre-processing.

The authors in [19] provided insight into one of the biggest design challenges, namely “the alignment of segments [needed] in order to extract event related waves hidden in the background EEG activity.” The authors also provided concise evaluations of methods to accurately identify the onset of stimuli some of which are already employed in mainstream EEG headsets used in this project.

II. SYSTEM DESIGN AND DEVELOPMENT

The *Think to Speak* system uses dependent and (predominantly) independent clauses as the *symbol*, a Liquid Crystal Display (LCD) and speaker as the *symbol display*, and accomplishes *symbol selection* via EEG and piezoelectric signals. Due to the physical limitations of the target audience, Electroencephalography (EEG) is the primary means of response selection due to the near universal accessibility of brain waves for any being capable of engaging in communication. The high-sensitivity piezoelectric sensor allows for the subtlest of movements or muscle contractions to be registered by the system, effectively allowing for marginal improvements in the user's physical condition over time to map to even further increased AAC usability and satisfaction.

Put simply, it is a piezo-EEG self-learning Augmentative and Alternative Communication (AAC) System. The implementation design of a *Voice-Input Voice-Output Communication Aid* coupled with a long short-term memory (LSTM) Recurrent Neural Network (RNN) response generation results in increased levels of self-expression for non-speaking users with severe motor impairment. The LSTM network, originally created by [21], was implemented using Google's open-source software library, TensorFlow[®], due to its suitability for implementing neural networks. The Google Cloud Platform[®] was used to outsource the computational demand of training the model away from a singular local machine to Google's network of distributed computers.

A. System overview and functions

At its core, the *Think to Speak* AAC facilitates self-expression through five main system phases: environmental audio sensing/signal acquisition, signal processing, potential response generation, response selection, and selected response vocalization. The environmental sensing, signal processing, and selected response vocalization phases, via speech-to-text modules, sophisticated processing techniques (for both EEG and audio signals), and text-to-speech modules, respectively, have largely been fully optimized in a subset of communication systems known as VIVOCAs, or *Voice-Input Voice-Output Communication Aids* [20]. For this reason, this research primarily explores areas for growth within the remaining, underdeveloped phases of currently available AAC systems: **potential response generation and response selection**. These two components are intimately related—an communication system with state-of-the-art potential response generation is rendered useless if the user cannot physically select and convey their desired response (and vice versa), as is the case with many currently available systems [16].

Per every valid utterance registered via voice recognition, the *Think to Speak* system generates three unique and complete sentences (*symbols*), displays them sequentially on an LCD in order of decreasing confidence values, as determined by the loss function of the neural network (*symbol display*), and compares the averages of the user's Event Related Potentials (ERPs), the user's electrophysiological response to each displayed response, to determine the desired response (*symbol selection*). In essence, if a user's level of concentration surpasses the activation threshold and other Boolean logic conditions, it is vocalized via the final stage of the *Think to Speak* system, the speech generating device (SGD). The EEG data is transmitted to the microcontroller via Bluetooth and a UART.

An LSTM RNN simultaneously combats the issues of slow communication rates and undesirable potential responses by placing the majority of the decision-making burden on Artificial Intelligence, versus the user. Clear benefits include an increased communication rate and decreased cognitive burden, both key factors to *sustained* use of any Augmentative and Alternative Communication (AAC). The trade-off associated with this benefit, as with any reliance on computational intelligence, leave a margin for the occasional generation of undesirable, and sometimes even irrelevant potential responses. This approach lies in stark contrast to the more cumbersome individual character or symbol selection approach. The *Think to Speak* AAC effectively prioritizes communication speed at the expense of a degree of autonomy and customization in potential responses. Due to the fundamental nature of Neural Networks, the three potential response generated at each invocation will become increasingly tailored to the user.

B. Requirements, analysis and specifications

A Brainwave Starter Kit, the NeuroSky® MindWave® Mobile 2 shown in Fig. 4 was used to capture EEG waves for this research work. It consists of a single dry EEG electrode located at the Frontal Polar left hemisphere position and a reference and "ground" clip located on the ear (a relatively electrically neutral part of the body). According to the 10-20 International System, the electrode is located at the Frontal Polar left position [10]. According to [22], the hardware itself consists of an application-specific integrated chip (ASIC) known as the TGAM 1. Together with an MSP430, an affordable, low-power microcontroller from Texas Instruments designed precisely for embedded system applications, the TGAM 1 is responsible for raw EEG signal processing, feature extraction, and attention, meditation, and eye blink measurement (the three primary eSense® values). A single AAA battery is required to power the headset, providing for 8 hours of continuous device usage per battery.

The headset is both Bluetooth® (BT) and Bluetooth® Low Energy (BLE) compatible meaning that information is passed at a radio frequency of 2.4 GHz. The operational frequency of the device was set to 50 Hz, the electrical grid frequency of Ghana (the location of *Think to Speak* AAC development). Since the headset is equipped with a Class 2 BT module, it has a maximum output power of 2.5 mW and a max effective operation range of 10 meters. The headset outputs 12-bit raw EEG signals/brainwaves in the frequency range of 3.0 – 100.0 Hz at a sampling rate of 512 Hz. The outputted brainwaves, in order of increasing frequency are Theta, Alpha, Beta, and Gamma brainwaves. A Fast Fourier Transform (FFT) is performed on

the raw EEG signal by the TGAM 1, providing a clear representation of the band power of the brainwaves.



Fig.4 NeuroSky MindWave Mobile 2 Headset

The Piezoelectric Sensor, LDT0-028K, shown in fig. 5, is made of a 28 μm thick flexible piezoelectric Polyvinylidene Fluoride (PVDF) film that generates a small alternating current and large voltage (on the order of $\pm 90\text{V}$) in response to applied mechanical stress that causes any displacement from the equilibrium or neutral axis. A 1 M Ω resistor is placed in parallel across the leads of the piezoelectric sensor to drop the generated voltages down to safe levels for processing by the Arduino. Applications and some characteristics of piezoelectric sensor is documented in [8].

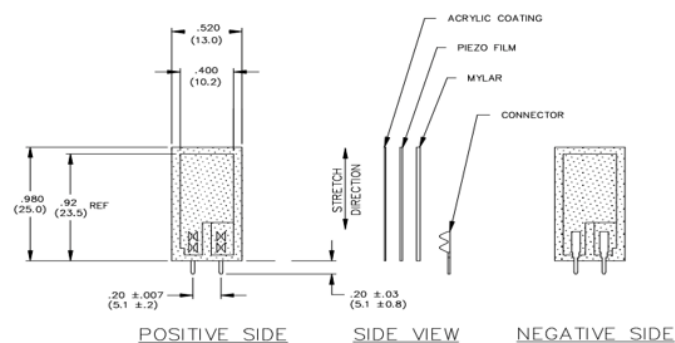


Fig. 5: TE LDT0-028K PVDF Piezoelectric Sensor

C. System design and development process

Per every valid utterance registered via voice recognition, the *Think to Speak* AAC generates three unique and complete sentences, displays them sequentially on an LCD in order of decreasing confidence values, as determined by the loss function of the LSTM RNN, and compares the averages of the user's Event Related Potentials (ERPs), the user's electrophysiological response to each displayed response, to determine the desired response. In essence, if a user's level of concentration surpasses the threshold, as determined by EEG signal processing via NeuroSky Inc.'s proprietary eSense® algorithms, it is selected and vocalized via the final stage of the *Think to Speak* system, the speech generating device (SGD). The block diagram in fig.6 below illustrates the fundamental processes involved in carrying out the aforementioned procedure.

D. System modelling and simulation

Proteus®, is a software design suite used for electronic design automation, was the primary means of system modelling with a snapshot shown in fig. 7. Due to limitations on available components in Proteus® libraries, a few of the components used in the simulation are not the actual components physically

implemented in the *Think to Speak* AAC. The simulation revealed the desired range of 10 K Ω potentiometer values to ensure a proper voltage on the V_{DD} pin of the LCD, and thus optimal brightness for the display.

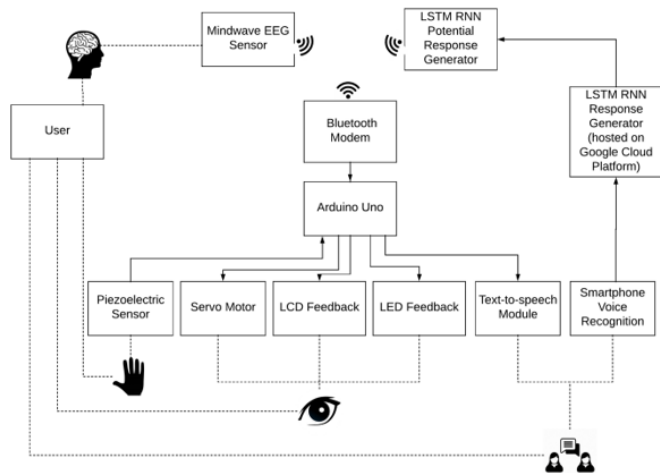


Fig. 6 Overall system block diagram

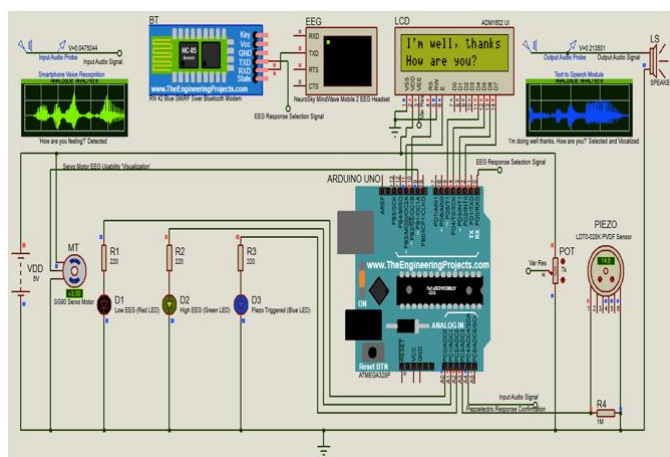


Fig. 7 Proteus Simulation Model

III. IMPLEMENTATION AND TESTING

The primary goal of the testing phase is to capture both quantitative and qualitative data pertaining to the communication rate and level of user satisfaction. The communication rate, defined as the number of desired symbols uttered per minute, is measured from the time the speaking partner speaks into the smartphone voice recognition until the point that the primary AAC user's desired response is vocalized via the output speaker. Since communication rates of commercially available (speller) systems is often quantified in terms of letters per minute, if, say, the *Think to Speak* user is able to vocalize the response "I am well, thank you" in 30 seconds, the communication rate would be the number of characters in the selected response (15) divided by the number of minutes required to produce the response (.5), resulting in an average communication rate of 30 characters/minute.

The proposed testbed consists of 15 pairs of speaking partners in three environments varying from low to medium to high background noise. For each of the three testing environments, two pairs participate in 10 trials, two pairs participate in 20 trials, and one pair participates in 30 trials. Each speaking pairs

are split halfway into two sub-testbeds, one limited exclusively to EEG signal acquisition and one with piezoelectric signal acquisition. Communication rates are calculated per each utterance exchange and a communication rate average was calculated for each pair of speakers. This testing methodology encapsulates the key elements outlined under ASHA's critical performance measures [23].

The Google Cloud Platform[®] was used to outsource the computational demand of training the model away from a singular local machine to Google's network of distributed computers in Europe.

A. Results and Discussions

Due to the predictable high performance of the *Think to Speak* AAC with piezo activation, the predictable low performance nature of the system when solely reliant on EEG signals (due to the inadequate nature of the single dry electrode EEG reader), and time constraints, the following testing methodology was created but not carried out. Preliminary trials of the *Think to Speak* AAC with piezoelectric sensor response selection yielded a communication rate of 133.49 characters/minute and a communication accuracy of 98.0%. Preliminary trials of the *Think to Speak* AAC with solely EEG response selection yielded a communication rate of 26.77 characters/minute and an unacceptable communication accuracy of 20.0%.

While training the LSTM model on a local machine, reducing the training data size to 2,000 resulted in nearly gridlocking the entire laptop (MacBook Air) due to the minimal 4 GB of Random-Access Memory (RAM). This first training attempt was terminated after 20 minutes and the output flag indicated that after the 20 minutes, only 8 iterations of training had been conducted. As a result of these unforeseen challenges, the project was migrated to the Google Cloud Platform[®] (GCP). Adjusting the training data set to 80 entries (near the minimum data size while still avoiding a 'sample larger than population' value error due to the pool of size of the random batch generation of the training data), GCP was able to train and save the model at iteration 1000 in one hour and fifteen minutes.

B. Performance evaluation and limitations of system

The suboptimal communication accuracy for the trials that did not accept piezoelectric sensor input can be largely attributed to the inadequate sensor readings from the EEG headset. The single-electrode nature of the reader is inadequate for reliable potential response selection. Well before the training and evaluation phase, the work of [17, 24] hinted towards the challenges posed by using the more affordable but less cohesive Neurosky[®] MindWave[®] headset (\$99.99 USD) as compared to the near-medical grade 14 dry electrode Emotiv[®] EPOC headset (\$719.00 USD). While designing the best system possible would have resulted in the use of the EPOC, that would simultaneously quadruple the overall system cost of the AAC, effectively making it inaccessible to a large number of individuals with limited income.

The main challenge faced throughout system development was multi-channel serial communication. Both hardware TTL serial communication via UART logic and software serial communication were implemented and the Uno was chosen over the Mega due to its affordability (roughly half the price).

As [14] noted, the fact that “people who use AAC rarely produce linguistic demands; they typically allow the speaking partner to initiate topics and control the conversation. Since desired responses are generated in response to the speaking partner’s speech, the AAC user does lose some level of control over the conversation since it is effectively guided by the speaking partner. This trial user-reported sub-optimal self-expression, and therefore unrealized autonomy, needs to be addressed in future extensions to the *Think to Speak* system.

IV. CONCLUSION AND FUTURE WORK

A. Conclusion

Through an extensive review of existing works, employment of a multiple input-output machine learning protocol, and a reconfigurable sensor-actuator system, the body of work that is the *Think to Speak* AAC provides a unique contribution to the field of Augmentative and Alternative Communication. Despite unsatisfactory EEG desired response selection due to a poor choice in EEG readers and the inability to deploy the trained LSTM RNN model on the Google Cloud Platform[®], the *Think to Speak* AAC has potential to be both commercialized and beneficial to the near 1% of the global population living with complex communication needs. The path to the development of a *Think to Speak* AAC that is effective, affordable, and desirable to operate, necessitates implementation of one or more of the following future work recommendations.

B. Future Work

From the work in [18], applying the BCI2000 software suite to replace the closed-source feature extraction and signal processing aspect of the Mindwave[®] Mobile at an affordable price provides a natural project extension for the *Think to Speak* AAC system. On the neural network architecture, training the RNN with a larger and more suitable language corpora and constructing an LSTM network with more hidden units, layers, and finely tuned bias parameters and synapse weights would improve the quality of responses generated [12]. Along the same lines, leveraging the Augmentative Communication Quantitative Analysis program (ACQUA) to quantify system metrics (specifically communication rate and accuracy) provides a simple means of stratifying the *Think to Speak* AAC among other currently available AACs. The authors in [15] spoke to the “unnecessary redundancy of language,” such as the presence of vowels that the brain (or computer) can fill in based on the selection of pairs or collections of ordered consonants. Certain spoken languages around the world, such as Hebrew, demonstrate this phenomenon in that vowels are rarely used in printed text. Should the *Think to Speak* AAC extend into a typing aid, applying conditional probability, such as Bayes Theorem, to infer redacted information would be a beneficial system extension.

REFERENCES

- [1] D. Beukelman and P. Mirenda, *Augmentative and Alternative Communication: Supporting Children and Adults with Complex Communication Needs*, Third. Baltimore: Paul H. Brookes Publishing Company, 2005.
- [2] The Editors of the American Heritage Dictionaries, Ed., *The American Heritage Dictionary of Medicine*, 2nd ed. Boston: Houghton Mifflin Publishing Company, 2015.
- [3] P. Enderby and J. Emerson, “Speech and language therapy: does it work? (Education & Debate),” *Br. Med. J.*, vol. 312, no. 7047, p. 1655, 1996.
- [4] “Dysarthria - Symptoms and causes - Mayo Clinic.” [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/dysarthria/symptoms-causes/syc-20371994>. [Accessed: 04-Oct-2018].
- [5] M. S. Hawley *et al.*, “A Voice-Input Voice-Output Communication Aid for People With Severe Speech Impairment,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 23–31, Jan. 2013.
- [6] M. Tudor, L. Tudor, and K. I. Tudor, “[Hans Berger (1873-1941)--the history of electroencephalography].,” *Acta Med. Croatica*, vol. 59, no. 4, pp. 307–13, 2005.
- [7] J. Malmivuo, V. Suihko, and H. Eskola, “Sensitivity distributions of EEG and MEG measurements,” *IEEE Trans. Biomed. Eng.*, vol. 44, no. 3, pp. 196–208, Mar. 1997.
- [8] Li Tianze, Zhang Xia, Jiang Chuan, and Hou Luan, “Analysis of the characteristics of piezoelectric sensor and research of its application,” in *2009 18th IEEE International Symposium on the Applications of Ferroelectrics*, 2009, pp. 1–4.
- [9] S. Russell and P. Norvig, *Artificial Intelligence - A Modern Approach*, Third. Upper Saddle River: Prentice Hall Series In Artificial Intelligence, 2010.
- [10] M. Hassib, A. Schmidt, A. Sahami, and N. Henze, “Mental Task Classification Using Single-Electrode Brain Computer Interfaces CR-Klassifikation,” Stuttgart, 2012.
- [11] C. Nicholson and A. Gibson, “A Beginner’s Guide to LSTMs and Recurrent Neural Networks | SkyMind,” *SkyMind*, 2018. [Online]. Available: <https://skymind.ai/wiki/lstm>. [Accessed: 18-Nov-2018].
- [12] G. Hoffman, “Introduction to LSTMs with TensorFlow - O’Reilly Media,” 2018. [Online]. Available: <https://www.oreilly.com/ideas/introduction-to-lstms-with-tensorflow>. [Accessed: 17-Nov-2018].
- [13] S. Creer, P. Enderby, S. Judge, and A. John, “Prevalence of people who could benefit from augmentative and alternative communication (AAC) in the UK: determining the need,” *Int. J. Lang. Commun. Disord.*, vol. 51, no. 6, pp. 639–653, Nov. 2016.
- [14] B. Wisenburn and D. Jeffery Higginbotham, “An AAC Application Using Speaking Partner Speech Recognition to Automatically Produce Contextually Relevant Utterances: Objective Results,” *Augment. Altern. Commun.*, vol. 24, no. 2, pp. 100–109, 2008.
- [15] A. Kübler *et al.*, “The thought translation device: a neurophysiological approach to communication in total motor paralysis,” 1998.
- [16] B. O’Keefe, N. B. Kozak, and R. Schuller, “Research priorities in augmentative and alternative communication as identified by people who use AAC and their facilitators,” *AAC Augment. Altern. Commun.*, vol. 23, no. 1, pp. 89–96, 2007.
- [17] M. Yeo, L. Jiang, E. Tham, and W. Xiong, “Evaluation of a low-cost alternative communication device with brain control,” *Ind. Electron. Appl. (ICIEA), 2015 IEEE 10th Conf.*, pp. 229–232, 2015.

- [18] A. Majkowski, M. Kolodziej, and R. J. Rak, "Implementation of selected EEG signal processing algorithms in asynchronous BCI," in *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, 2012, pp. 1–3.
- [19] S. Petrichella, L. Vollero, F. Ferreri, V. Di Lazzaro, and G. Iannello, "TRS-TMS: An EEGLAB plugin for the reconstruction of onsets in EEG-TMS datasets," in *13th IEEE International Conference on Bioinformatics and BioEngineering*, 2013, pp. 1–4.
- [20] M. S. Hawley *et al.*, "A voice-input voice-output communication aid for people with severe speech impairment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 23–31, 2013.
- [21] S. Ram, "Practical seq2seq," 2016. [Online]. Available: <http://complx.me/2016-12-31-practical-seq2seq/>. [Accessed: 17-Nov-2018].
- [22] I. NeuroSky, "MindWave," 2018. [Online]. Available: <https://store.neurosky.com/pages/mindwave>. [Accessed: 17-Nov-2018].
- [23] "Proloquo2Go - AssistiveWare." [Online]. Available: <https://www.assistiveware.com/products/proloquo2go>. [Accessed: 03-Oct-2018].
- [24] A. Al-Abdullatif, H. Al-Negheimish, L. Al-Mofeez, N. Al-Khalifa, L. Al-Andas, and A. Al-Wabil, "Mind-controlled augmentative and alternative communication for people with severe motor disabilities," *Innov. Inf. Technol. (IIT)*, 2013 9th Int. Conf., pp. 107–112, 2013.