060315

Youth Science Canada
Sciences jeunesse Canada

CWSF ESPC
Canada-Wide Science Fair
Expo-sciences pancanadienne

Jacky Zhao

Senior / Innovation

Augmenting Hearing Capabilities of the Deaf with Natural Language Processing

My project involved developing a portable device capable of transcribing speech into braille to augment the hearing capabilities of the deaf. The device is able to recognize speech via a speech recognition software that I built using a category of AI called machine learning. This solution solves a lot of problems with current solutions, being easy to use and completely independent of visual cues.

Project Forms

4.1A Humans Low Risk (edited_signed_4.1A_Humans_Low_Risk (2) (1).pdf)
4.1A Informed Consent – Your Letter of Information (informed_consent_letter_of_information_en (2).pdf)
4.1A Informed Consent – Your Permission Form (informed_consent_blank_en (1).pdf)

Augmenting the hearing capabilities of the deaf using Natural Language Processing

Jacky Zhao, St. George's School

1. Background

Over 460 million people in the world have disabling hearing loss which can be a danger to their physical and mental health. In children, having hearing loss can greatly impact an individual's ability to learn and understand language. [1] Later in life, the deaf are not able to respond to auditory information like warnings and alarms and could suffer from mental health issues and social stigma. In fact, the global economy loses as much as $750 billion dollars due to loss of productivity. Current solutions that address these problems have pitfalls that prevent them from being accessible. Lip Reading and American Sign Language are incredibly dependent on visual cues and take a long time to learn how to interpret them accurately. [2] Other, more technical solutions, such as hearing aids and cochlear implants, are prevented from being accessible because of their high purchase and maintenance costs and risky surgical procedures. [3][4]

2. Objective

To design and create an embedded device that allows those who are hearing-impaired to have access to auditory information without visual cues.

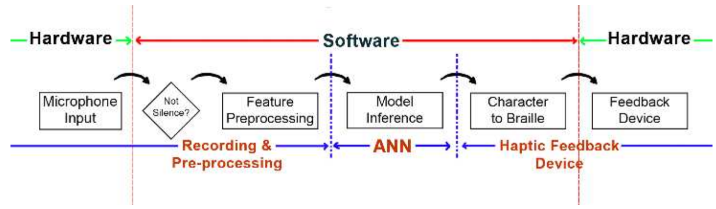
3. Proposition

We propose a novel device that could be developed to transcribe speech to tactile messages to be recognized. To allow for ease of use and portability, the whole solution is designed to be compact enough to fit in an embedded device. Through further research and development, it became clear that several technologies (speech recognition, haptic feedback device) needed to be integrated together to create a functioning solution.

4. Implementation.

The top-level functions that need to be implemented are:

1) Microphone input to audio file
2) Audio file to character sequence
3) Sequence to haptic feedback



4.1 Sound recording/Pre-processing

In a real environment, the target speech is always mixed with background noise. A noise removal technique called spectral masking was implemented to get large enough signal to noise ratio [5]. We can use an operation to find the mel-frequency cepstral coefficients (MFCCs). MFCCs take into account the shape of the vocal tract (e.g. the tongue, teeth) by analyzing the envelope of the short-time Fourier Transform. This allows the computer to represent complex speech with a simple array. The first coefficient is discarded and replace with RMSE, and first derivatives are added [6].

4.2 ANN-Based Speech recognition

For computers, recognizing speech is much more complex than simply identifying patterns and sounds [7]. Speech is extremely variable--different people speak in different ways. Artificial Neural Networks, or ANNs, would allow us to tackle a lot of these problems by mimicking a biological neural network, allowing it to 'learn' how to deal with issues like variation in accents, pauses, tone, and volume [8].

4.2.1 Key Elements of implementing the ANN

- **Code Framework**: Tensorflow
- **Platform**: Laptop with i7-6700HQ processor, 16G of RAM, and a NVIDIA 970M GPU
- **Dataset**: LibriSpeech Corpus [9].

4.2.2 Developed software utilities and programs

Name	**Description**
tf_model.py	Network architecture construction and training loop.
denoise.sh	Bash script that removes background noise from audio via sound profiling
braille_util.py	Converts character sequences to braille sequences and interfaces with GPIO
gpiostates.py	Manages the pilot light and state locking
init.sh	Serves as a daemon that spawns necessary child processes to process data and gets an output
load_meta.py	Loads ANN from saved model '.meta' and runs preprocessed data through the loaded model

4.2.3 ANN Model Architecture

Feature	**Reason**
LSTM - ANN cell body	The LSTM cells are fed the processed MFCCs and output their cell states to the CTC function. They are a subset of cells called RNNs that have 'temporal memory' which allow them to excel at tasks like sequence labelling and speech recognition [10]. 2 layers and 256 hidden units allowed the LSTM to achieve an acceptable level of abstraction without sacrificing too much run time.
CTC- Output and Cost	CTCs are widely used in training LSTMs and RNNs, where they excel in cases like speech recognition, where timing is variable[11]. They output a probability distribution over all the possible labels.
RMSProp – Optimizer	RMSProp is based on the Adagrad optimizer, which adapts its learning rate based on how frequent parameters are being updated. RMSProp is different in which it tries to tackle Adagrad's radically diminishing learning rates[11]. An initial learning rate of 1e-4 and γ, which controls the rate of decay, was set to be 0.9.
Greedy Decoder - CTC decoder	Because CTC models its outputs on a sequence of probability distributions of all possible labels, a Greedy Decoder was implemented. Compared to the Beam Search algorithm, this provides a substantial boost in runtime at the cost of a small loss in accuracy [13].

4.2.4 ANN Data Flow

There are 3 main steps that data go through when running data through an ANN.

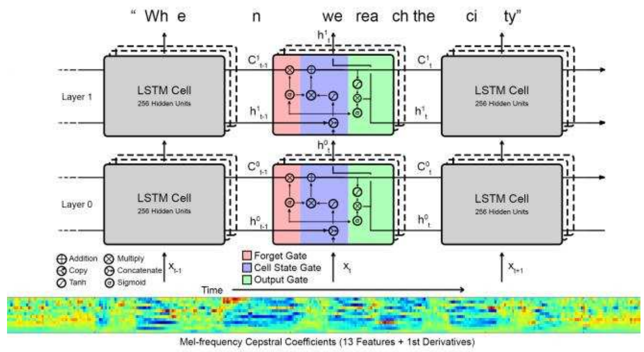
1) Pre-processing

2) Running the ANN

a) Run each timestep and concatenate it with the previous timestep's output

b) Forget Gate - Decide how much of the previous cell state to keep.

c) Cell Update Gate - Update the current cell state.



3) Getting the output - Find the most likely set of sequence labels using CTC Greedy Decoder

5. Character Sequence to Haptic Feedback

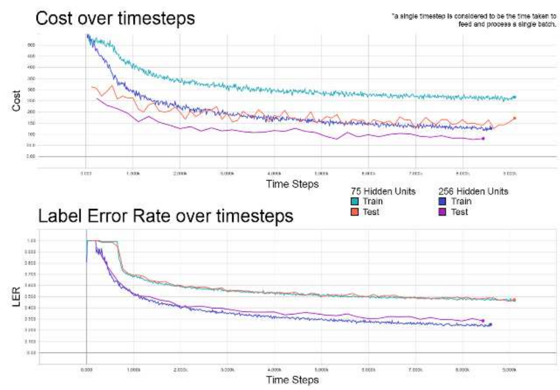
Braille was chosen to present haptic feedback devices due to it being small, easy to learn, and easy to implement. As a proof of concept, a 2x3 solenoid array was used to display tactile Braille. The character sequences are encoded into Braille using a short Python script, then sent to the GPIO pins. Additional Darlington transistors were also used to provide required voltage and current to drive the solenoids.

6. Data and Results

6.1 ANN Performance Metrics

Two different metrics were used to measure the performance of the ANN over the training period. Cost - defined as the negative natural log of the probability to correct classification of a label in every possible search path [11] and Label Error Rate - defined as the total Levenshtein Edit distance over total characters of a batch [6][11]. The ANN achieved an accuracy of 74.77% on the training set and 71.50% on the test set after training for approximately 37 epochs or 44 hours and 10 minutes. This is comparable to Google's Speech Recognition API in 2013, which claims to have achieved an accuracy of 77.00% [14].

6.2 Performance Trends



Over the training period, there were a few notable observations.

1) The model underfits the data, showing high bias and low variance. The complexity of the ANN was not enough to capture the level of abstraction.
2) The improvements to both the LER and cost functions decrease exponentially. Increased training would not have yielded wildly better results.

7. Conclusion

A prototype speech to Braille embedded device has been successfully created with 2 main technologies in mind:

- ANN-based speech recognition using the Tensorflow framework
- Self-built solenoid-driven haptic feedback system to display tactille braille

Accuracy could be improved in real-world applications by adding more random noise to the training process [15]. In the future, a smartphone application could be developed to move the processing away from the Raspberry Pi. The hardware aspect served to be a functional prototype and demonstrated everything that the end model needed to. Overall, this project was successful in accomplishing the original goal of creating an alternative method of receiving communication for the deaf that is independent of visual cues.

8. Appendix

8.1 References

1. WHO. (n.d.). “Deafness and Hearing Loss.” World Health Organization, World Health Organization, www.who.int/mediacentre/factsheets/fs300/en/.
2. Bauman, Niel. (2011). “Speechreading (Lip-Reading).” Center for Hearing Loss Help, hearinglosshelp.com/blog/speechreading-lip-reading/.
3. *“Cochlear Implants.”* (2017). American Academy of Otolaryngology-Head and Neck Surgery www.entnet.org/content/cochlearimplants.
4. Center for Devices and Radiological Health. (2017)“Cochlear Implants - Benefits and Risks of Cochlear Implants.” U S Food and Drug Administration Home Page, Center for Devices and Radiological Health, www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/ImplantsandProsthetics/CochlearImplants/ucm062843.htm.
5. Lecun, Y., Bottou, L., Orr, G. B., & Müller, K. (1998). Efficient BackProp. Lecture Notes in Computer Science Neural Networks: Tricks of the Trade, 9-50. doi:10.1007/3-540-49430-8_2
6. Leeuwen, D.A., & Niedek, T.V. (2016). Phonetic Classification in TensorFlow.
7. Forsberg, Markus. (2003). Why is speech recognition difficult.
8. Sak, Haşim & Senior, Andrew & Rao, Kanishka & Beaufays, Françoise. (2015). Fast and Accurate Recurrent Neural Network Acoustic Models for Speech Recognition.
9. Povey, D. (n.d.). LibriSpeech ASR corpus. Dataset, www.openslr.org/12/
10. Panzner, M., & Cimiano, P. (2016). Comparing Hidden Markov Models and Long Short Term Memory Neural Networks for Learning Action Representations. Lecture Notes in Computer Science Machine Learning, Optimization, and Big Data, 94-105

11. Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist temporal classification. Proceedings of the 23rd International Conference on Machine Learning - ICML 06. doi:10.1145/1143844.1143891
12. Ruder, Sebastian. (2016). “An Overview of Gradient Descent Optimization Algorithms.”, Ruder.io, ruder.io/optimizing-gradient-descent/.
13. Tensorflow. (n.d.) “Tensorflow Documentation - tf.nn.ctc_beam_search_decoder” , https://www.tensorflow.org/api_docs/python/tf/nn/ctc_beam_search_decoder
14. Bohouta, Gamal & Këpuska, Veton. (2017). Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx). Int. Journal of Engineering Research and Application. 2248-9622. 20-24. doi: 10.9790/9622-0703022024.
15. Geiger, J.T., Rigoll, G., Schuller, B.W., Weninger, F., & Zhang, Z. (2014). Robust speech recognition using long short-term memory recurrent neural networks for hybrid acoustic modelling.

8.2 Bibliography

N., & A., M. Neural Networks and Deep Learning. (1970, January 01), Web-book http://neuralnetworksanddeeplearning.com/

Deng, L., & Liu, Y. (2017). Deep learning in natural language processing. Singapore: Springer Singapore.

Hope, T., Lieder, I., & Resheff, Y. S. (2017). Learning TensorFlow: A guide to building deep learning systems. Sebastopol: OReilly Media.

Form 4.1A: Participation of Humans
Low Risk - Approval

Youth Science Canada
Sciences jeunesse Canada

This form certifies that a low risk project involving the participation of humans is in full compliance with Policy 4.1.1.1 *Participation of Humans in Research – Low Risk*

Region	Greater Vancouver			
Project Title	Augmenting hearing capabilities of the deaf with Natural Language Processing			
	First Name	**Last Name**	**Email**	**Phone**
Student 1	Jacky	Zhao	jacky.zhao2019@stgeorg	778.956.8798
Student 2				
Chair, RSF[1]	Jessica	Zhang	chair@gvrsf.ca	778.807.8820
Chief Judge, RSF	Leonard	Foster	leonard@gvrsf.ca	N/A
Adult Supervisor	Kathryn	Murray	kmurray@stgeorges.bc.ca	604.224.1304 x3756

Request for Ethics Approval

I have visited and understood the web site of the National Ethics Committee:
http://ethics.youthscience.ca/
I have read and understood Policy 4.1.1.1 *Participation of Humans in Research – Low Risk*

April 29th 2018

Signed - Student 1 | Date | Signed - Student 2 | Date

Ethics Approval

I certify that this Low Risk Project involving Human Participation is in full compliance with Policy 4.1.1.1 *Participation of Humans in Research – Low Risk*, and that it is eligible for the Regional Science Fair and the Canada Wide Science Fair.

April 29th 2018

Signed – Adult Supervisor | Date

[1] Or Designate

Form 4.1A: Participation of Humans Low Risk – Instructions

Youth Science Canada
Sciences jeunesse Canada

1) **Form 4.1A**
This form certifies that a low risk project involving the participation of humans is in full compliance with Policy 4.1.1.1 *Participation of Humans in Research – Low Risk*

2) **Documents**
The following documents must be available to the Adult Supervisor, along with form 4.1A:
   a) The Survey - if your project involves a survey.
   b) Your Letter of Information for every participant in your project.
   c) A blank sample of the Informed Consent form each participant will sign.

3) **Instructions for Participants in the Regional Science Fair.**
   a) Fill in Form 4.1A on a computer, not by hand.
   b) Print the completed Form 4.1A, and then get it signed with all the required signatures.
   c) Make a paper copy of this completed and signed form.
   d) Store the original signed copy of form 4.1A in a safe place.
   e) Take the copy of signed form 4.1A to your Regional Science Fair.
   f) If you did a survey, take these three forms to your Regional Science Fair as well:
      i) The Survey.
      ii) Your Letter of Information that you gave to every participant in your project.
      iii) A blank sample of the Informed Consent form each participant signed before they did your experiment..

4) **Instructions for Finalists selected for the Canada Wide Science Fair**
   a) Upload the completed, but unsigned, electronic copy of form 4.1A on your computer to the Forms area of the CWSF registration site.
   b) If you did a survey, upload these forms as well:
      i) the Survey.
      ii) Your Letter of Information that you gave to every participant in your project.
      iii) A blank sample of the Informed Consent form each participant signed before they did your experiment.
   c) Make two paper copies of the signed and completed form 4.1A.

5) **Paper Copy for You**
Bring one paper copy of the signed form 4.1A to the CWSF and have it available at your project.

6) **Paper Copy for your Delegate**
Give a second paper copy of the signed form 4.1A to your Delegate, and ask that it be brought to the CWSF just in case your copy gets mislaid.

Informed Consent

Letter of Information

Augmenting hearing capabilities of the deaf with Natural Language Processing

1. **Researcher and Advisors**

	First Name	Last Name	Phone	Email
Student 1	Jacky	Zhao	778 956 8798	jacky.zhao2019@stgeorges.bc.ca
Student 2				
Adult Supervisor	Kathryn	Murray	604-726-7064	kmurray@stgeorges.bc.ca
Scientific Supervisor[1]				
School	St. George's School			

2. **Purpose of the Research**

This research will allow me to verify the feasibility of my project, and allow me to optimize certain aspects of my prototype. In particular, the research will go towards optimizing the delay between sequential characters, identifying the learning curve of braille, and determining the sensitivity required to 'read' tactile braille. If you choose to take part in this study, you will be asked to identify braille character sequences from the tactile braille display. In particular, I will be asking questions about difficulties that they experience while learning how to identify braille, and giving feedback.

3. **Benefits from Participating**

You will learn how to read tactile braille and methods of communication of the deaf.

4. **Risks from Participating**

There are no risks associated with this project.

5. **Time Commitment Required**

This will require you to learn to read braille comfortably, and will require anywhere from 30 minutes to a week of your time. You may decide to withdraw at anytime.

6. **Remuneration**

There is no remuneration for participating in this project.

---

[1] Required for Significant Risk Projects only

7. Confidentiality of Data

All data is quantitative and is not connected with your name or your identity. At the Science Fair, all information will be presented anonymously. None of the participants will be identified. The data, which is stored on the computer of the researcher, will be deleted after the Science Fair.

8. Withdrawing from the project

You have the right to withdraw from the project at any time and for any reason. Contact the Adult Supervisor by telephone or email if you would like to do this.

9. Results

A summary of the quantitative results and the analysis of this research will be sent to you by email after the Science Fair is over.

10. Ethics Approval

This research project has been reviewed and approved by the Ethics committee of the Greater Vancouver Regional Science Fair

Informed Consent
Permission Form

Augmenting hearing capabilities of the deaf with Natural Language Processing

Date: ____________________

I have received the *Letter of Information for Informed Consent* for this Science Fair Project, which is mine to keep. All my questions have been answered to my satisfaction, and I agree to participate in this research.

Name of Participant (print): ___________________________________

Signature of Participant: ____________________________________

Name of Parent of Guardian[1]: __________________________________

Signature of Parent of Guardian: ________________________________

Name of person obtaining consent: ______________________________

Signature of person obtaining consent: ___________________________

This Informed Consent Form contains confidential data, and so must be secured at the home of the Adult Supervisor, whose signature on the Human Participation Form confirms that Informed Consent has been obtained.

[1]If the participant is under the age of 18, then a parent or guardian must also give permission by signing this form.

All Informed Consent Forms must be shredded after the project is no longer needed for Science Fairs.