CoughMotion: A CNN model that uses IMU and audio data for cough-detection

SIDHARTH GUPTA, University of Toronto, Canada

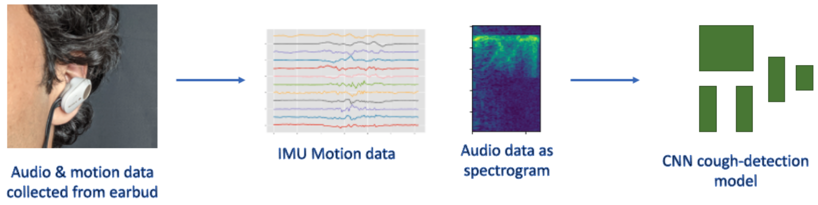
ALEX MARIAKAKIS, University of Toronto, Canada

Many wireless earbuds contain motion sensors that collect Inertial Movement Unit (IMU) data. This data can define various head positions, and it is commercially used to adjust audio playback from head movement. We investigate an alternative use for this data, and explore its utility in cough detection. Specifically, we use a set of earbuds that collect both audio data and IMU data from the head, and visualize how a cough creates a distinct signal in the audio and IMU domain. Next, we describe a theoretical CNN model that intakes audio and IMU data in two separate branches, and outputs a cough detection prediction. With this work being a preliminary presentation of ideas, we do not present results from this CNN model – but rather its architecture, hyperparameters, and the kinds of experiments that can be done. Overall, cough-detection is a task with great significance in health-monitoring, and we hope to provide ideas on how new earbud technology can be leveraged to help.

CCS Concepts: • **Machine learning → Convolutional neural network architectures**; • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: convolutional neural networks, IMU data, cough-detection, hyperparameters

ACM Reference Format:
Sidharth Gupta and Alex Mariakakis. 2021. CoughMotion: A CNN model that uses IMU and audio data for cough-detection. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn


Audio & motion data collected from earbud
IMU Motion data
Audio data as spectrogram
CNN cough-detection model

1 INTRODUCTION

Cough-detection has broad applications, from general health monitoring to disease diagnosis. It is used to detect overall lung health [3], and also to diagnose diseases such as pertussis and COVID-19 [4, 7]. In addition, cough detection can be scaled and deployed to many users around the globe who may not have close access to a physician. However, there are some existing problems with auditory cough detection – one of which is ignoring coughs that come from someone other than a primary user. For tasks in health monitoring, this is an important feature, and we believe that movement data from the head can help prevent these kinds of coughs from being falsely detected.

2 RELATED WORK

2.1 Cough Detection

Many cough-detection models use machine learning to transform a cough audio clip into a latent embedding that can be classified. CNNs, LSTMs, and transformers have all been used in cough-detection to achieve state-of-the-art performance [5, 9]. One CNN model in particular that has been used is VGGish by Google, which is a VGG16 augmented model that intakes an audio-clip as a mel-spectrogram [2]. For the task of diarized cough detection, which tries to learn who a cough belongs to from a set of candidates, an investigation has been done using multitask learning [8]. This investigation achieves 82% accuracy when classifying amongst four coughers, which outperforms human evaluators on average by 9.82%. We believe that adding IMU data can be extra information that helps a model more accurately detect coughs for one specific person.

2.2 IMU and Audio Deep Learning Models

The combination of IMU and audio data has been used to increase the detection accuracy of activity recognition system. One study for detecting shots in racquet sports uses IMU and audio data from a smart watch, by processing IMU features with random forests, and processing audio data with dense neural networks [6]. A more recent study has been done called GestEar, which processes IMU and audio data in separate branches of a CNN [1]. GestEar has tried processing IMU data using classical machine learning models, but found that 1D CNNs evaluate best. The research found that using both sources of data led to the best performing model with the least false positive rate. To the best of our knowledge, no such study has been done applying IMU data to cough detection. However, the studies mentioned here give plausible evidence as to why combining IMU data with audio can help boost performance.

2.3 Earphone Sensing

3 DESIGN OF COUGHMOTION

3.1 CNN Model Structures

3.1.1 Transfer learning. We will now describe our overall model architecture. We start by pre-processing the audioset cough clips into mel-spectrograms, and then pre-training the VGGish model without IMU data. Our audio VGGish model in particular takes in a size (48, 64, 1) spectrogram, has four convolutional & max-pool layers, and four dense layers. Each convolutional layer uses a (3, 3) size kernel, with stride 1, and same padding. Each max-pool layer is of size (2, 2) with stride 2. The first layer creates 64 channels, the second 128 channels, the third 256 channels, and the fourth 512 channels. The last layer flattens into a size 3072 vector, which gets fed into the series of dense layers. The first two dense layers are of size 4096, and the last two are of sizes 128 and 1.

The audio VGGish model makes up one branch in our CNN architecture. There are two more branches – one processing IMU gyroscope data, and one processing IMU accelerometer data; though both branches are structured the same. Each intakes a size (50, 1, 3) input, where the three channels correspond to the x, y, z data. A series of two or three 1D convolutional & 1D max-pool layers are then used, followed by a flatten. Then, the two IMU branches are connected into their own FC layer. Finally, the output of this layer is concatenated into the output of one of the audio FC layers, which gets processed into the final prediction. The architecture for the IMU branches is stated more vaguely here, because there are many hyperparameters to consider, which will be discussed in Section 3.1.2. A high-level algorithm box of our model can be found in Figure 1.

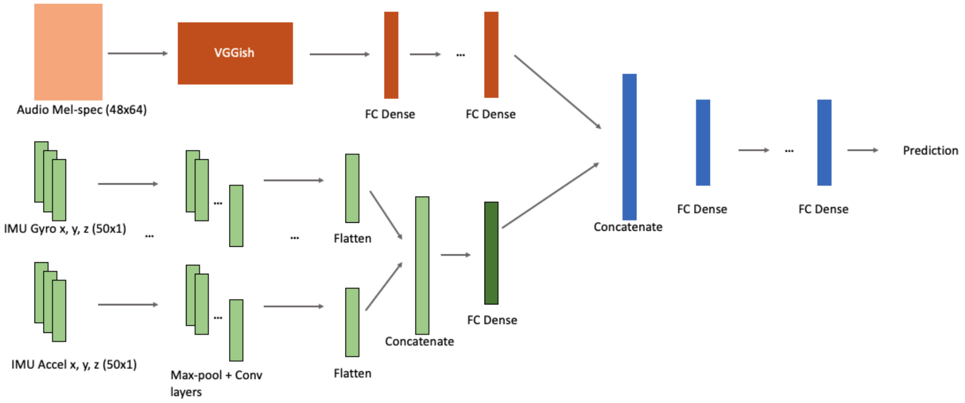


Fig. 1. High-level algorithm box of our CNN architecture

3.1.2 *Hyperparameters on the fundamental structure.*

In order to evaluate the effect of adding IMU data, we fix the audio VGGish model with the hyperparameters mentioned in [2] and perform a hyperparameter search on the IMU branches. Specifically, we parameterize the number of layers, number of IMU channels, location of concatenation, dropout coefficients, and skip connections. In Table 1, we present a plausible set of hyperparameters to search from in our described model. Note that the number of channels in the last IMU layer control the channels in the preceding layers, as they will increase in powers of two.

Number of layers	Number of channels in last IMU layer	Concatenation location	Skip connection locations	Dropout coefficients
2	32	Audio FC-1	IMU Conv-2	0.1
3	64	Audio FC-2	IMU Conv-3	0.2
		Audio FC-3	IMU FC-1	0.3
				0.5

Table 1. Various configurations of hyperparameters to search

4 DATA COLLECTION

In this preliminary proof-of-concept study, we evaluate coughs of different types, rotations, and movements from one participant. We achieve promising preliminary results, which motivates expanding the study to a larger census size.

4.1 User study protocol

A modified version of the Sony WF-1000XM3L in-ear headphones are used to collect both audio and IMU data. The headphones have an MPU 9250 chip inside that reads IMU data, and the modification connects that chip to an Arduino board which we use to collect data. In addition, the microphone and audio components of the headphone are connected to a handheld audio recording device. The audio and IMU recording devices are synced by an action of the head lightly (but quickly) hitting a table, which creates a sudden jump in both signals at the same time. These signals are aligned at this jump, and then partitioned from there.

The participant executed 90 different types of coughs and movements, in a stationary sitting position. The coughs are varied across head pitch, yaw, degree of inhale, and single or double cough. The signals for these coughs are visualized in Figure 2.

We use data augmentation to increase the utility of our data during training. Each audio sample is modified with a random chop, random gain, and random added noise. The IMU samples are also mutated with the same location of chop, and their own degree of random gain.

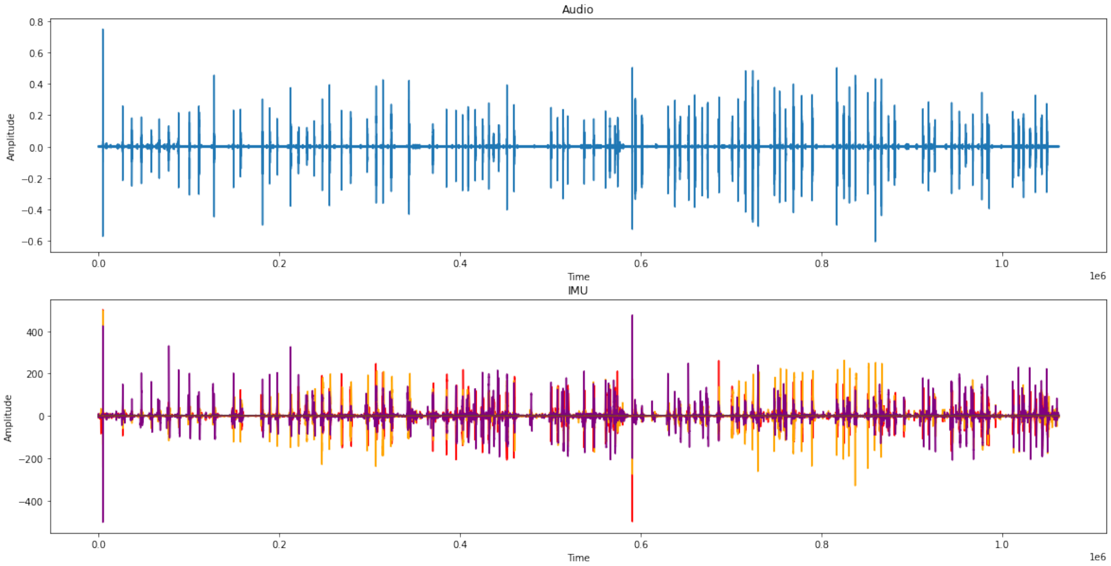


Fig. 2. Audio data (top signal) and IMU data (bottom signal) for 90 cough samples; the sudden signal impulses are coughs.

5 DISCUSSION

As shown in Figure 2, there certainly seems to be a visible association with IMU and audio data. This association, however, can be attributed to limitations in this study, as the coughs were collected by one person in a stationary sitting position. More work does need to be done to handle more organic cases, such as coughing while walking, standing, driving, etc. Regardless, the association we see in a stationary position shows promise that the IMU layers in a cough-detection machine learning model can help in the task. To better visualize how the model learns from the IMU layers, we propose two suggestions: one is to visualize the IMU weight distributions, and the other is to visualize a t-SNE or PCA projection of the IMU embeddings after training. The former method will give us insight on how the IMU layers change during training, and can tell us how significantly they are used in predictions. The latter can help visualize how the latent learned representation of the IMU data looks like, and if any clusters exist across cough IMU datapoints and other IMU datapoints.

6 CONCLUSION

We propose a preliminary algorithm box, experiment design, and results for the idea of using IMU data to aid the task of cough detection. We start with a state-of-the-art auditory cough-detection model — VGGish from Google — and augment it by adding two additional branches to process IMU gyroscope and accelerometer data. We present a set of hyperparameters for these additional branches, and explain how weight distributions and PCA plots can help visualize IMU weights during training and predictions. We also describe the hardware that is used to collect the IMU and audio data, and present findings from a single-person data collection study. A total of 90 coughs were collected, and an association with IMU and audio data for coughs is seen visually. Although this study is still limited and preliminary, the initial results are promising, and provide motivation for a larger study to be done.

ACKNOWLEDGMENTS

We would like to thank Matthew Whitehill from the University of Washington, for kindly providing code from his WhoseCough study for us to work off of. We would also like to thank Yuntao Wang from Tsinghua University, for providing us with unique headphone hardware that allowed us to collect both IMU and audio data.

REFERENCES

[1] Vincent Becker, Linus Fessler, and Gábor Sörös. 2019. GestEar: combining audio and motion sensing for gesture recognition on smartwatches. In *Proceedings of the 23rd International Symposium on Wearable Computers*. 10–19.

[2] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron Weiss, and Kevin Wilson. 2017. CNN Architectures for Large-Scale Audio Classification. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://arxiv.org/abs/1609.09430

[3] Eric C. Larson, Mayank Goel, Gaetano Boriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N. Patel. 2012. SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. ACM, New York, NY, USA, 280–289. https://doi.org/10.1145/2370216.2370261

[4] Renard Xaviero Adhi Pramono, Syed Anas Imtiaz, and Esther Rodriguez-Villegas. 2016. A Cough-Based Algorithm for Automatic Diagnosis of Pertussis. *PLOS ONE* 11, 9 (09 2016), 1–20. https://doi.org/10.1371/journal.pone.0162128

[5] Hasib-Al Rashid, Arnab Neelim Mazumder, Utteja Panchakshara Kallakuri Niyogi, and Tinoosh Mohsenin. [n.d.]. CoughNet: A Flexible Low Power CNN-LSTM Processor for Cough Sound Detection. ([n. d.]).

[6] Manish Sharma, Akash Anand, Rupika Srivastava, and Lakshmi Kaligounder. 2018. Wearable audio and IMU based shot detection in racquet sports. *arXiv preprint arXiv:1805.05456* (2018).

[7] Neeraj Sharma, Prashant Krishnan, Rohit Kumar, Shreyas Ramoji, Srikanth Raj Chetupalli, Nirmala R., Prasanta Kumar Ghosh, and Sriram Ganapathy. 2020. Coswara — A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis. *Interspeech 2020* (Oct 2020). https://doi.org/10.21437/interspeech.2020-2768

[8] Matt Whitehill, Jake Garrison, and Shwetak Patel. 2020. Whosecough: In-the-Wild Cougher Verification Using Multitask Learning. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 896–900.

[9] Neo Wu, Bradley Green, Xue Ben, and Shawn O'Banion. 2020. Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case. arXiv:2001.08317 [cs.LG]