Predicting speaking status using full 9 Degrees Of Freedom Inertial Measurement Unit (IMU) data

Mark Groenendijk[1]
Responsible professor: Hayley Hung[1] , **Supervisor: Stephanie Tan**[1]
[1]EEMCS, Delft University of Technology, The Netherlands
maagroenendijk@student.tudelft.nl, H.Hung@tudelft.nl, S.Tan-1@tudelft.nl

Abstract

The goal for this paper is to find out what the smart badge provided by the Social Perceptive Computive Lab (SPCL) group is and what it contains. The sensors that are used in the smart badge are the Accelerometer, Gyroscope and Magnetometer. The main question of this paper is ”What is the benefit of using full 9-DOF IMU data in predicting speaking status, as opposed to using only accelerometer signals?”. The three senors all contribute in their own way and complement each other to give an estimate about the speaking status. The ability to estimate the speaking status using the smart badge opens up the potential for analyzing more about the social aspects of people without the need to record what they are saying.

1 Introduction

Nowadays, smart wearable devices are not something new. Every smartphone from Samsung to Apple from Microsoft to what else is available contains an Inertial Measurement Unit (IMU). Phones are not the only devices carrying an IMU, the IMU sensors are used widely in different movable applications [1]. The IMU's first use was in aircraft navigation [20] and other large devices. This was because of the large size it had and the power that was required. Since recent years, the micro-electromechanical system (MEMS) IMU is the most common IMU used because of its size and the low cost of creation [5].

The IMU carries multiple sensors that are important for measuring the acceleration, the angular rate of rotation [14] and the magnetic field. There is not a single sensor that can measure all of these simultaneously. To measure these three aspects, the IMU must have an accelerometer, a gyroscope and a magnetometer respectively. All three sensors measure data in three directions, thus creating nine Degrees of Freedom (DOF).

The Socially Perceptive Computing Lab (SPCL) has developed a smart badge that contains the IMU with the accelerometer, gyroscope and magnetometer. This smart badge can be worn around the neck to gain data. In earlier studies, a single body-worn accelerometer was able to estimate different types of social interaction such as speaking, laughing or gesturing [4]. However, this could still be improved by using the two other mentioned sensors: the gyroscope and magnetometer. Therefore, the main research question is: ”What is the benefit of using full 9-DOF IMU data in predicting speaking status, as opposed to using only accelerometer signals?”
The SPCL has built upon and finetuned pre-existing work from the MS-G3D [1]. The MS-G3D is a PyTorch implementation of ”Disentangling and Unifying Graph Convolutions for Skeleton-Based Action Recognition” [10]. The goal of this implementation was to overcome limitations of previous methods. These earlier approaches treated human joints as a set of independent features, and they modelled the spatial and temporal joint correlations through either hand-crafted [18; 19] or learned [3; 8; 16] aggregations of these features [10]. However, human joints also have connections with each other and these relations are not detected by the previously mentioned methods. These relations are better captured by representing joints as nodes and edges for their connectivity. This will look like a skeleton representation of the human body.

The SPCL takes the gathered data, which includes the data from the three sensors, from the smart badge. This data is used to generate a dataset that can be split up, trained and used to measure the average loss, area under the curve (AUC) and the accuracy[2]. With this knowledge, one can predict the speaking status of people wearing such a device [2] based on the measurement mentioned earlier.
There are multiple reasons to predict the speaking status with an IMU instead of simply using a video and audio-recording. The first and foremost is privacy [17]. To record a video of a person one must have signed a form stating it is actually okay to do that. However, when working with IMU's everything is anonymous. Nobody knows who the person was wearing the IMU. This makes running experiments a lot easier and faster. Another reason for using the IMU over video and audio-recording for predicting speaking status is that movements in a conversation are important. Talkers in

[1] https://github.com/kenziyuliu/MS-G3D
[2] https://github.com/josedvq/MS-G3D

conversations spontaneously assimilate facial expressions, postures, pronunciation and speech rates even when they are specifically instructed not to do this [7]. Non-verbal communication is thus a big part of the conversation with physical communication being the most used form [13]. So, the motion data provides an insight into the social aspect of a conversation.

This report focuses on the difference between using an IMU that only generates accelerometer data as opposed to an IMU generating nine degrees of freedom data. The main question of this research is: ”What is the benefit of using full 9-DOF IMU data in predicting speaking status, as opposed to using only accelerometer signals?”. With the subquestions: ”How do the gyroscope and magnetometer complement the accelerometer” and ”How does one predict the speaking status and what is the benefit of knowing this?”.

2 Methodology

To answer the main research question, data was given that was previously collected by the Socially Perceptive Computing Lab. But what is the purpose of estimating the speaking status? When a person knows who is speaking they know they should listen to that specific person and give them their concentration of listening to receive the (important) information. This helps with inferring the speaking turns. For humans, this is very natural and you do not stop and think about this question.
However, this is not the case for intelligent agents/ devices. If an agent can perceive when and how long a person is talking or perhaps predict when a person is likely to talk, this agent might be able to become part of the social interaction. So knowing the speaking status in a conversation is very important for the social aspect of intelligent agents.

The given data was collected by running an experiment where 49 participants had a small social gathering whilst wearing the IMU sensor around their necks. This data was then used in a machine learning script programmed in Python using Pycharm. This script takes files that contain skeleton data for the recorded subjects, annotated from top-down videos of the interaction space from when the experiment is recorded. Together with this data, the data from the accelerometer and later on the gyroscope and magnetometer is inserted. These two together will create a skeleton that replicates the action a person does whilst wearing a smart badge. Examples of these skeletal representations can be seen in figures 1 and 2.

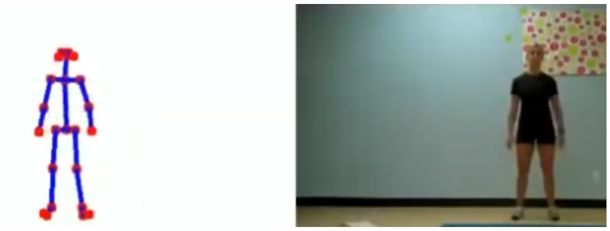

Figure 1: Skeletal representation of standing still. [9]

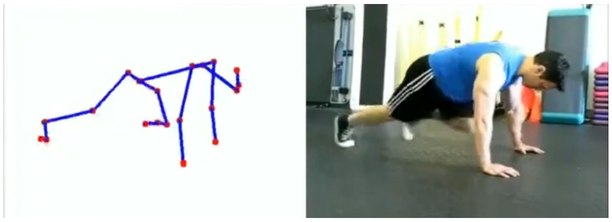

Figure 2: Skeletal representation of the mountain climbing exercise. [9]

The reason to do this training is to evaluate whether or not the data is valuable for the research and if so, use this knowledge to create a better understanding of the problem. With this data, one can determine an estimation of the speaking status. However, these results from the given data were not enough to answer all the sub-questions and the main research question.
A literature study combined with training and evaluating the data to gather results was necessary for this research project. Therefore, the literature study started by creating a good visual about what the accelerometer, gyroscope and magnetometer are recording individually.
General information about the sensors and previous research about accelerometers [12], gyroscopes and magnetometers explained how they work individually. Then the question about how the three sensors work together as one Inertial Measurement Unit (IMU) was next. The answer for this was given in already existing research papers [6; 15]. These researches discovered a way to use the gravity vector measured by the accelerometer to calibrate the gyroscope and align the magnetometer [15]. So, after this was done information was gathered via research on the data and the literature research was accomplished. All the outcomes were combined in this paper and reviewed by the peers within the project group. After the peer reviews, the paper was improved and completed.

3 Experimental Setup and Results

To know what the benefits are of using full 9-DOF IMU data we first need to know what information the data is giving per sensor and what to expect from it. Then with this data run the experiment first using accelerometer data only and then using data from all three sensors combined to see if there are any differences in the result.

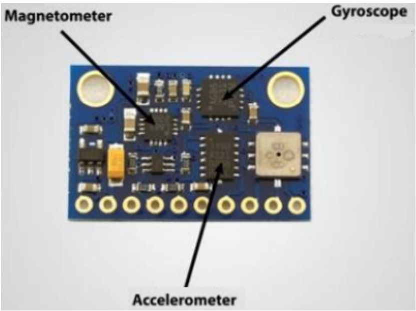

Figure 3: Arduino board with the accelerometer, gyroscope and magnetometer. [11]

3.1 Sensors

Accelerometer

The accelerometer is a device that measures, as the name says, the acceleration. More specific the vibration of a motion of a structure or in this case a person. An accelerometer consists of multiple parts to calculate the acceleration. It consists of a suspended mass, fixed plates and polysilicon springs. A visual representation of an accelerometer is displayed in figure 4.

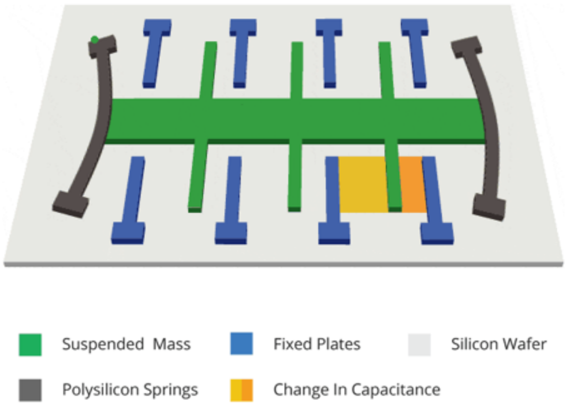

Figure 4: Animated representation of an accelerometer. [11]

The force used by the person causes the mass to shift around. This mass will then create a bigger or smaller gap between itself and the fixed plates which is called the change of capacitance. This change of capacitance can be measured and used to calculate the force on the springs. When the force is calculated and you know the mass you can use Newton's law: $F = m * a$ to calculate the acceleration.

Gyroscope

The gyroscope measures angular rate using the Coriolis effect. When a person or object is moving in a direction and an external angular rate has applied a force will occur, which will cause perpendicular displacement of the person/object. In figure 5 there is a visual version of the gyroscope where the mass will cause a change in capacitance just like with the accelerometer which can be measured and will correspond to a particular angular rate.
The gyroscope is able to detect three types of rotations which means extra degrees of freedom. Each type is a rotation around a different axis in a three dimensional space:

1. Pitch
2. Roll
3. Yaw

Magnetometer

The magnetometer works differently than the accelerometer and the gyroscope. Where the accelerometer and gyroscope work with a change of capacitance the magnetometer works with an output of electrical quantity. The magnetometer measures the earth magnetic field by using the Hall effect. The Hall effect is when an electrical flow is going over a conductive plate it goes straight from one to the other side of the plate. If there is a magnetic field near the plate this would disturb the straight flow causing the electrons to deflect to one side of the plate. When using a voltage meter on the plate it would show a difference in voltage. This change of voltage is caused by the Lorentz force. A particle of charge x moving with a velocity v in an electric field E and a magnetic field B experiences a force of:

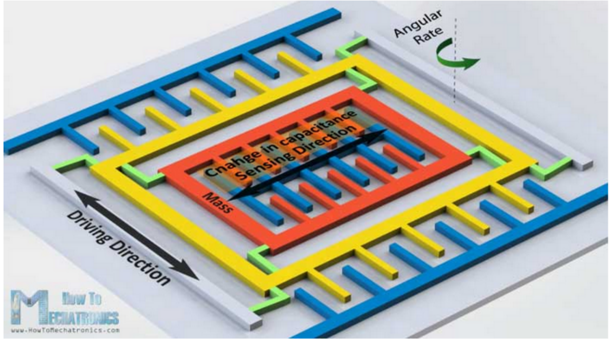
Angular Rate
Change in capacitance Sensing Direction
Mass
Driving Direction
How To Mechatronics
www.HowToMechatronics.com

Figure 5: Gyroscope animated. [11]

$$F = E_x + v_x * B$$

This formula states that the electromagnetic force on a charge x is a combination of two things namely:

- A force in the direction of the electric field E in proportion to the magnitude of the field and the quantity of charge.
- A force at right angles to the magnetic field B and the velocity v of the charge, in proportion to the strength of the field, the charge, and the velocity

One of the uses for a magnetometer is therefore a miniaturized compass.

3.2 Data training and evaluating

As mentioned in the introduction, the MS-G3D python framework is used with alterations from the Social Perceptive Computing Lab. To get results there are certain steps needed to take:

1. Create a feeder, that temporarily stores that paths, consisting of:
   - The path to the poses which consists of the skeleton data of the recorded subjects of the experiment.
   - The label path to the video segments that consists of data per subject gotten from the smart badge.
2. From the label path load the accelerometer, gyroscope and magnetometer data files per person.
3. use this data from the three sensors and create a dataframe that is either three dimensions for the accelerometer or nine dimensions for combining the sensors. This will generate an output that consists of training, validation and test Pickle and NumPy files.

After these Pickle and NumPy files are generated the script can be trained and used to generate test results. The results will appear as an Area Under Curve (AUC), which represents the performance score of the model, and accuracy. In table 1 there is an average result from the accelerometer data only and the combined results from the three sensors.

	AUC	accuracy
Accelerometer	0.7968	0.737
Combination of sensors	0.730	0.729

Table 1: AUC and accuracy of accelerometer based and fusion of accelerometer, gyroscope and magnetometer based status detection.

These results indicate that using only accelerometer data outperforms when using the combination of data. The accelerometer data is not the big surprise out of these results. However, the results from the combination of the three are disappointing. This is disappointing because theoretically the gyroscope and magnetometer do contribute in creating a vision of how a person is standing and moving. But the results do not correspond with this hypothesis.

4 Conclusions and Discussion

The main question for this report was: ”What is the benefit of using full 9-DOF IMU data in predicting speaking status, as opposed to using only accelerometer signals?”.

The main benefit of using the gyroscope and magnetometer is to be able to detect rotations better. The accelerometer can only measure in linear directions. The main advantage of the gyroscope is that it can detect rotations such as pitch, roll and yaw. And the main advantage of the magnetometer is that it can detect like a compass in what direction the smart badge is pointing by using the earth magnetic field.

So by using an accelerometer on itself you do get enough information about the movement of the person wearing the smart badge to distinguish certain actions and predict speaking status to a certain accuracy depending on the complexity of the script. By adding two extra layers of three dimensions you can get an even better understanding of the direction in which the smart badge is pointing and thus towards who the subject is looking.

The gyroscope and magnetometer do contribute to getting a better understanding of the movement, acceleration and direction of the subject. The results however show the opposite. According to the results, the accuracy of the accelerometer is better than using a fusion of the three sensors and thus have nine degrees of freedom as opposed to only three. These results could mean one or a combination of three aspects :

- One accelerometer with three degrees of freedom is better than adding six more degrees from the gyroscope and magnetometer.
- More data input leads to more noise which could interfere with the results.
- More degrees of freedom and thus more data require a more complex training and evaluating script than there is available right now.

The expected outcome was that adding more data leads to retrieving better results. However, this is not the outcome of the experiment, in contrary the exact opposite happened. This could very well be the case but the literature study concluded that the gyroscope and the magnetometer do contribute in retrieving more information. This means that there is either more noise due to the fact there is more data. Or the script that is used for training and evaluating was not complex enough for nine degrees of freedom since it was build for three degrees originally. It could also be a combination of the two. That could a good starting point for further research within this field.

5 Responsible Research

This research was done with private data provided by the SPCL. This data was anonymous and not further distributed, therefore it was ethical to do since it is not harmful to any of the subjects that were present when collecting the original data.

To reproduce the experiment that is used in this report one must be a member of the SPCL group or a student working with the SPCL group such that it can access the needed data. This also goes for the scripts used. The MS-G3D[3] script is public and can be altered to the needs of this experiment. The alterations that are made for this research are however private. So reproducing the results can only be done when working with the SPCL group.

References

[1] Norhafizan Ahmad, Raja Ariffin Raja Ghazilla, Nazirah M Khairi, and Vijayabaskar Kasi. Reviews on various inertial measurement unit (imu) sensor applications. *International Journal of Signal Processing Systems*, 1:256–262, 2013.

[2] Laura Cabrera-Quiros, Ekin Gedik, and Hayley Hung. No-audio multimodal speech detection in crowded social settings task at mediaeval 2018. In *Mediaeval 2018 Workshop*, 2018.

[3] Yong Du, Wei Wang, and Liang Wang. Hierarchical recurrent neural network for skeleton based action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1110–1118, 2015.

[4] Hayley Hung, Gwenn Englebienne, and Jeroen Kools. Classifying social actions with a single accelerometer. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’13, page 207–210, New York, NY, USA, 2013. Association for Computing Machinery.

[5] Jack W Judy. Microelectromechanical systems (mems): fabrication, design and applications. *Smart materials and Structures*, 10(6):1115, 2001.

[3] https://github.com/kenziyuliu/MS-G3D

[6] Manon Kok, Jeroen D Hol, Thomas B Schön, Fredrik Gustafsson, and Henk Luinge. Calibration of a magnetometer in combination with inertial sensors. In *2012 15th International Conference on Information Fusion*, pages 787–793. IEEE, 2012.

[7] Nida Latif, Adriano V Barbosa, Eric Vatiokiotis-Bateson, Monica S Castelhano, and KG Munhall. Movement coordination during conversation. *PLoS one*, 9(8):e105036, 2014.

[8] Ce Li, Chunyu Xie, Baochang Zhang, Jungong Han, Xiantong Zhen, and Jie Chen. Memory attention networks for skeleton-based action recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[9] Ken Liu. Ms-g3d demo.

[10] Ziyu Liu, Hongwen Zhang, Zhenghao Chen, Zhiyong Wang, and Wanli Ouyang. Disentangling and unifying graph convolutions for skeleton-based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 143–152, 2020.

[11] Dejan Nedelkovski. What is mems? accelerometer, gyroscope amp; magnetometer with arduino, Oct 2021.

[12] Mark Pedley. Tilt sensing using a three-axis accelerometer. *Freescale semiconductor application note*, 1:2012–2013, 2013.

[13] Deepika Phutela. The importance of non-verbal communication. *IUP Journal of Soft Skills*, 9(4):43, 2015.

[14] D. Piyabongkarn, R. Rajamani, and M. Greminger. The development of a mems gyroscope for absolute angle measurement. *IEEE Transactions on Control Systems Technology*, 13(2):185–195, 2005.

[15] Hamidreza Razavi, Hassan Salarieh, and Aria Alasty. Optimization-based gravity-assisted calibration and axis alignment of 9-degrees of freedom inertial measurement unit without external equipment. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 234(2):192–207, 2020.

[16] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1010–1019, 2016.

[17] Jiaxing Shen, Oren Lederman, Jiannong Cao, Florian Berg, Shaojie Tang, and Alex Pentland. Gina: Group gender identification using privacy-sensitive audio data. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 457–466. IEEE Computer Society, 2018.

[18] Raviteja Vemulapalli, Felipe Arrate, and Rama Chellappa. Human action recognition by representing 3d skeletons as points in a lie group. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 588–595, 2014.

[19] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1290–1297. IEEE, 2012.

[20] He Zhao and Zheyao Wang. Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended kalman filter for data fusion. *IEEE Sensors Journal*, 12:943–953, 2012.