Vol.2021-MBL-100 No.1
Vol.2021-UBI-71 No.1
Vol.2021-CDS-32 No.1
Vol.2021-ASD-21 No.1
2021/9/2

ManWaKei: Measuring the Amount of Speech and Investigating its Effect on Stress

Taku Yamane   Yin Chen   Tadashi Okoshi   Jin Nakazawa

Abstract: Due to the recent COVID-19 pandemic, a lot of people are now required to stay at home and/or work remotely. This is thought to be causing a significant decrease in vocal communication between people, which is leading to increased stress between people. However, quantitative researches to back this up has not been done much yet. In this research project, we created a system called the ManWaKei system that quantitatively measures the amount of a user ’ s speech using voice recognition transcription technology, in order to investigate the correlation between speech and stress from a quantitative standpoint. We carried out a 3-week experiment with 5 participants, collecting data about the conversations they made and the stress they experienced. Through the experiment, we were able to spot some signs of correlation between the amount of speech and stress, however there are too many factors that we still need to take into count, and improvements to the system is crucial for future researches.

Keywords: Speech, Vocalization, Stress, Personality

1. Introduction

The world today is being severely affected by the COVID-19 epidemic that started in 2019. Over 180 million people have been infected and over 4 million have lost their lives from it, and this number is still continuing to grow.[1]

Due to this, in many countries including Japan, people were put into quarantine: restricted and/or recommended not to go outside unless necessary, and many schools and workplaces have started the implementation of remote work. In April 2021, 56.6% of companies in Tokyo were actively implementing remote work, [2] and in June 2020, 60.1% of Japanese universities were carrying out lessons fully remotely, with an additional 30.2% partially implementing remote work. [3]

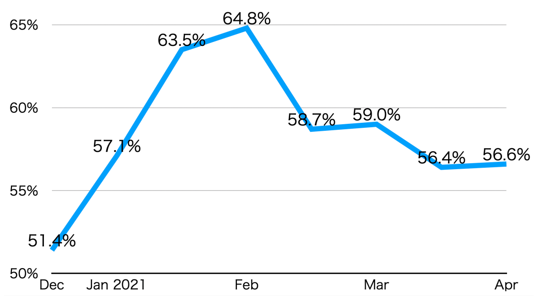

Fig. 1 Change in Percentage of Companies employing Remote Work in Tokyo

One of the direct drawbacks of this is that people now have a much less chance of meeting people and talking to them directly.

Speaking with someone has long been believed to be an important process in coping with stress, [4] [5] and the decrease of such communication is thought to be causing a significantly negative effect on people's stress.

[1] Keio University, 5322 Endo, Fujisawa, Kanagawa 252-0882, Japan

On top of this, social isolation is becoming a serious problem, especially amongst young people, [6] and psychiatrists are calling out to people to be careful of being isolated, and are recommending people to try their best to communicate with others.[7]

Although the correlation between speech and stress has been discussed for a long time, it was difficult to quantitatively prove this since the means to measure the amount of speech were almost non-existent in the past. However, with the current technology, and by taking advantage of the current remote work environment, it is now possible to objectively and quantitatively measure the amount of speech with relative ease, which opens up a new possibility of investigating the correlation between speech and stress.

The main objective of this research is to quantitatively investigate the correlation between the amount of speech and stress.

To achieve this goal, we will create a system that measures a user's amount of speech. By combining this with preexisting methods of measuring quantitatively stress levels, we aim to collect data both about the user's amount of speech and their stress levels, allowing us to investigate the correlations between them.

2. Related Works

2.1 Stress Detection from Voice Information

The pitch of a person's voice has long been considered to have a strong connection to the person's mental state, such as stress, and numerous studies have been carried out. [8] A notable project is one that a team at Fujitsu has developed a system that can recognize a user's instantaneous stress level by analyzing the tone, pitch and volume of the user's voice. [9] This project is similar to our project in how it deals with speech and stress, but it is created in order to detect instantaneous stress levels according to details of the user's voice, whereas our project aims to investigate the amount of speech of the user and draw comparisons between those results and the user's longterm stress levels.

2.2 Voice Recognition Transcription Technology

When talking about researches on speech, we must mention the evolution of voice recognition transcription technology. Voice recognition transcription technology is a type of system that uses technologies like deep learning to allow a computer system to indicate words said by a person, and transcribe them into text format. Development of this technology has been continuing for decades, and researches on their speed / accuracy has also been done numerous times. [10] Nowadays, numerous different companies such as Google [11] and IBM [12] are developing technologies of similar sorts, and are offering them as part of their service.

2.3 Measuring Stress

Measuring stress is a very difficult problem, and many researches about it are still ongoing. The measurement of stress can be divided widely into two ways: subjective measurement and objective measurement. [13]

2.3.1 Subjective Measurement

As the word suggests, subjective measurement mainly involves questioning the user about how they personally feel, and takes that as a direct measurement. It can also be said that it is more of a qualitative way of measuring stress. The most simple method of taking a subjective measurement of one's stress, is by asking them directly whether they feel stressed or not. The simpleness of the question makes the reliability of the results slightly questionable, but stress by its nature is a vague concept, and differs heavily depending on the individual, so it could be considered as the most simple yet most effective means of objectively measuring one's stress.

On the other hand, in order to increase the reliability of such subjective measurements, many questionnaires and models have been created. [14] These involve a large number of more complex and less obvious questions for the user to answer, decreasing the level of bias that could be included in the answer, allowing the testers to derive a more reliable measurement for stress. However, as the questionnaire itself contains a large number of questions, it is simply time consuming, and may appear to be stressful itself to answer, especially repeatedly.

2.3.2 Objective Measurement

On the other hand, objective measurement is measuring the user's stress quantitatively. [15] Sanae Fukada [13] explains that the investigation of biological samples such as blood, saliva, hair, nails and breast milk, or investigation of physiological indicators such as blood pressure, heart rate, heart rate variability, autonomic nervous functions, sweating and blinking are some of the ways of making objective measurements of stress. For example, Garmin's smartwatches such as "Garmin Venu" [16] measures the user's heart rate, heart rate variability, breath rate or blood oxygen levels, in order to assess the physiological state of the user, and especially uses the value of the heart rate variability to calculate the user's stress level.

It is said that the lower the heart rate variability is, the more stress the user is experiencing. The correlation between the two have been supported through numerous researches, such as the one carried out by Kim et al in 2018. [17] However, we must keep in mind that since the model for showing the correlation between the user's heart rate variability and their stress is a comparatively new concept, it is still far from complete, especially when it comes to giving a quantitative score for one's stress.

Alongside the development of high-end smartwatches capable of precise measurements, smartphone apps suited for easy and accessible stress measurements are also being developed. [18]

Most of these apps employ a similar method: they use the light and camera of the smartphone to record the color change of the user's fingertip, and derives the heart rate and heart rate variability of the user from it, allowing it to calculate the stress level of the user. The advantage of this is that anyone can use the system without having to purchase or wear a dedicated device, but it also has a shortcoming that since the user has to have their finger on their smartphone's camera in order to take measurements, it cannot continuously take measurements for a long period of time.

3. Problem Awareness

3.1 Measuring Speech

A crucial part of this research is quantitatively measuring the amount of the user's speech. We decided to use the word count of the user's speech as our unit of measurement. In order to measure the word count of a conversation accurately, we decided that the aforementioned voice recognition transcription technology would be the most suitable way.

Also, considering how the pitch of one's voice is also considered as a significant factor in investigating one's speech, we decided to implement a function in the system that allows us to measure and record the pitch of the user's voice as well. We will be using a Python library called pyworld in order to make these measurements.

3.2 Measuring Stress

For this research, we decided to use both of these methods and take two measurements for the user's stress, in order to balance the reliability of the data and convenience of collecting the data.

For the subjective measurement, we decided to use a very simple questionnaire, simply asking the user's subjective level of stress in a scale of 1 to 10, giving us a simple yet powerful measurement of their stress.

For the objective measurement, we decided that using a stress checking smartphone app would be a very simple, accessible, and yet considerably reliable way of taking an objective measurement of stress. We will be using the aforementioned "StressScan" app. [18]

3.3 Taking Personality into Count

As mentioned in the previous chapter, it can be easily assumed that people with different personality traits may feel differently about speaking, and may differ in how they are stressed or relieved from stress due to conversations. Considering such cases, we will be using the Big Five Model[19] to determine the user's personality traits, and make use of them in our analysis. We will have the participants take a short Japanese version of the Big Five Model Personality test online.[20]

Here, we took all of the problems and considerations above together to create an overview of our research project, in order to clarify what factors we are dealing with and what kind of system we have to create. Fig. 2 shows how each of the factors are connected in this research.

3.4 Research Overview

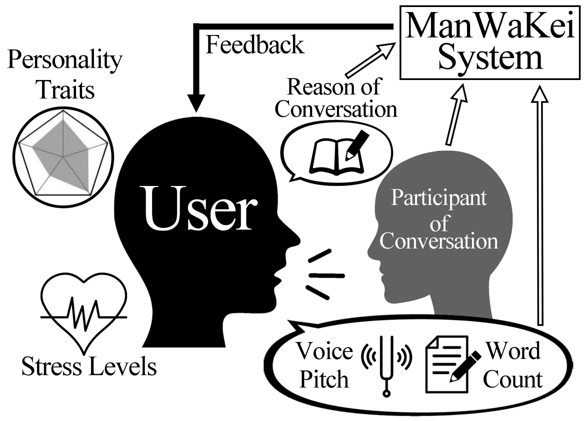

Fig. 2 Research Overview Diagram

The system we intend to create is a system that can quantitatively measure factors of the user's speech. The voice pitch and the word count are the two factors we intend to measure from the user's voice. We will be using the IBM Watson Speech to Text API to count the word count, and the voice pitch will be measured using the pyworld Python library's sound analysis.

The two metadata, the participant of conversation and reason of conversation are important factors as well. It would be best if we could estimate this by analyzing the content of the user ' s speech, but this is both technically difficult and is violating the user's privacy. So, we decided that we ask the users to fill in designated text boxes in the user interface to manually input these two factors.

For the stress levels, we will be taking both the subjective and objective measurements. The subjective measurements will be taken through a self-assessed stress level survey, on a scale of 1 to 10. The objective measurements will be taken using the aforementioned "Stress Scan" App on the participant's smartphones.

Finally, for the personality traits, we will be asking each of the participants to answer a personality check based on the Big-Five Personality Model, which will allow us to know the participant's 5 personality traits, which we can later use for analysis.

4. ManWaKei System

In this chapter, we propose the ManWaKei system, a system capable of collecting numerous types of data about a user's conversation through their computers.

4.1 Overview of the ManWaKei System

The ManWaKei System is a Python program that works on Mac OS computers. It was developed on Python 3.8.3. It is capable of recording a user's conversation, and collects various data about the conversation through analyzing the recording and asking the users to fill in some metadata.

The system is designed specifically for a remote work environment, and is only expected to be used in an environment in which only the main user's voice can be heard.

The quantitative data the system can collect are the word count of the user's speech and the user's average voice pitch during the conversation.

The metadata the system collects are: the person/people the user is having a conversation with (participant of conversation) and the reason of conversation.

Fig. 3 is the system architecture diagram of the system.

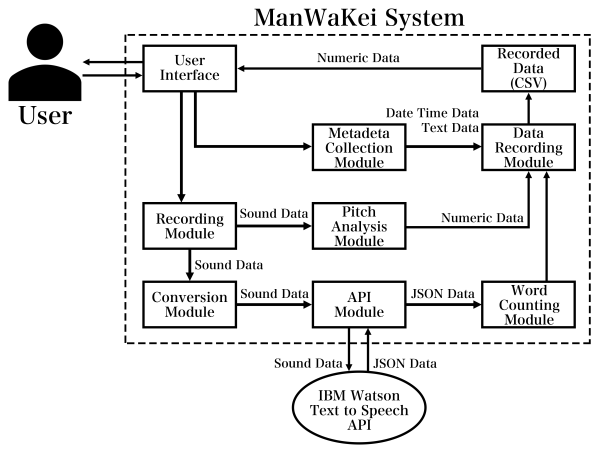

Fig. 3 System Architecture

4.2 Functions of ManWaKei

This section describes the functions of the ManWaKei system, the reasons why they are required in the system, and how specifically they were implemented.

4.2.1 Recording of Conversation

The ManWaKei system first records the user's conversation, and subsequently processes the recording to take measurements. Originally, we aimed to create a system that takes measurements of the conversation in realtime, but there were numerous technical difficulties, so we decided to build the system this way. Consequently, this gave us more freedom in taking measurements.

The recording is done through the user's computer's built-in microphone, and is stored in a wav format. The recording data is immediately deleted after all the measurements have been taken, since it contains a large amount of private information.

4.2.2 Word Count Measurement

The purpose of the ManWaKei system is to measure, record and notify the user of how much they have spoken, thus the measurement of word count in the user's speech is a crucial part of the system.

As mentioned previously, the ManWaKei system makes use of a voice recognition transcription service in order to obtain the word count of the user's speech. The system makes use of the IBM Watson Speech to Text API. [12] It sends the recording of the user's conversation to the server, and receives a transcribed text data as a json response. Then it reformats the json response so it only includes the text content, and simply measures the length of the text to determine the word count. The transcribed text data is immediately deleted after the word count has been measured, since, as was the case with the recording data, it contains a large amount of private information.

The IBM Watson Speech to Text API has a size limit for the file size it accepts. In order to overcome this, the ManWaKei system splits the recording into several smaller sound files using the AudioSegment library.

Also, as we are testing this system with Japanese users, the system uses the Japanese version of the IBM Watson Speech to Text API. By using a different version and slightly editing the counting function, the ManWaKei system can be easily used for different languages as well. Along with this, keep in mind that Japanese words are more commonly counted in characters rather than words, so all measurements will be made in terms of characters.

4.2.3 Pitch Measurement

Pitch is considered to be a significant factor in investigating the correlation between vocalization and stress, thus we decided to measure the average pitch of the user's voice in each conversation.

The measurement of pitch is using pyworld library. [21] Pyworld is a python wrapper version of the WORLD Vocoder, [22] which is a system created in order to support the analysis and merging of music files. Pyworld has a function to analyze a given sound file and extract a list of values representing the time and pitch of the given sound file. We will be using this to calculate the average pitch of the recording.

4.2.4 Metadata Collection

In order to investigate the correlation between communication and stress, it is crucial to understand who and why the user is having a conversation with.

Recognizing this through analyzing the recording is extremely difficult, and is not the focus of this research. Therefore, for the current edition of the ManWaKei system, we decided to simply ask the users to fill in two text boxes corresponding to the two metadata we intend to collect, whenever they make a recording.

4.2.5 User Interface

Bearing in mind that the ManWaKei system is replicating the pedometer, the user interface must be simple to use and concise in its presentation. The interface was implemented using the built in tkinter Python library. It contains the "Start Record" button for starting recording, two text boxes used for collecting the necessary metadata, and an indicator showing the user's daily word count. After the "Start Record" button is pressed, the same button turns into the "Stop Record" button. Through this design, the user's operation of the system is kept very simple and concise, while maintaining the necessary functionality. Also, there are two buttons used for uploading the collected data.

Fig. 4 is a screenshot of the user interface of the ManWaKei system.

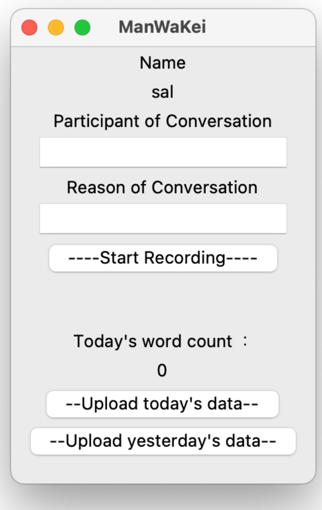

Fig. 4 Default User Interface

4.2.6 Data Storage

In the case of personal use, with the current functionalities it has, the system only needs to keep track of the user's word count for that day. However, for experiment purpose, the data must be stored correctly, and in a format that is easy for later investigation.

The system creates a dedicated csv file for each day, and appends the necessary measurement data onto it after each measurement. The csv file includes the user's ID in order to make the later investigation easier.

4.2.7 Data Transmission

As mentioned before, this function is unneeded for personal use, but for experimental purposes, it is necessary.

When the user chooses to upload the data for the day, the system uploads it to a dedicated Google account's Google drive.

5. Experimentation and Evaluation

In this chapter, we will firstly explain the details of the experiment we carried out. Then, we will explain the results of the experiment, and explore what kinds of correlations we could possibly find from them. After that, we will share some of the opinions we heard from the participants through our surveys, as part of a qualitative analysis of the project. Finally, we will discuss the quality of the results we obtained from the experiment, and consider how they could have been improved.

5.1 Overview of the Experiment

The purpose of this experiment is to record and compare the users' amount of speech and their stress levels, in order to investigate the correlations between the two.

5.2 Research Questions

Our proposed research questions are the following:

- Is there a quantitative correlation between stress and the amount of speech?
- Does personality effect this correlation?
- What other values could possibly be relate to the amount of speech and stress?

5.3 Introduction of Evaluation Items

In order to answer our research questions, we will be evaluating the following evaluation items:

- The users' daily word count of speech
- The users' average voice pitch during conversation
- The users' reason and participant of conversation
- The users' objective and subjective stress levels
- The users' personality traits

5.4 Experimental Procedure

The experiment will be 15 weekdays long, and is divided into 3 stages: the Control Experiment, the Solo Vocalization Activity Experiment and the Casual Communication Experiment.

Participants are first asked to answer a pre-experiment survey about their personal information: their sex, age, residence status, etc. Participants are also asked to take an online personality test based on the Big Five model. [20]

Participants are asked to install the ManWaKei system on their personal computers. Whenever they have an opportunity to talk to someone on the computer or by themselves, participants are directed to record their speech using the ManWaKei system's recording function. Before or during this recording, participants are asked to input the metadata about the current recording: the participant of conversation and the reason of conversation. After they have finished recording, the system will analyze the recording and record the results. For privacy reasons, the recorded sound file will be immediately deleted once the analysis is complete, unless an error occurs during the analysis, and the analysis must be redone.

Participants are also directed to measure their objective stress levels using the previously mentioned "Stress Scan" app once a day, preferably at a late time at night, for example, before going to bed.

At the end of each day, or during the next day, participants are asked to use the data transmission function of the ManWaKei system to transmit the collected data from the recordings for the day. Participants are directed to also answer a survey about their day, including their subjective stress levels, objective stress levels that they measured using the "Stress Scan" app, and reports on events they experienced during the day that either caused them to speak outside of the recording environment, and events that the participants think that might have had an impact on their stress levels for the day.

The experiment is further divided into 3 stages. The first stage, is the Control Experiment in which participants simply take measurements. In the second stage, the Solo Vocalization Experiment, participants are asked to take part in solo vocalization activities such as singing, reading books aloud and talking to one's self. In third and final stage, the Casual Communication Experiment participants are asked to contact and talk to people close to them, such as friends. This is in order to compare and investigate what kind of aspects within the user's speech is affecting the user's stress levels.

5.5 Processing Data

Data collected through the submitted result files and surveys will be managed into a spreadsheet for analysis. In this process, each recorded conversation will be manually separated into three types of conversations:

- Casual : Conversations with friends or family members, that do not involve working on school work, jobs or any other kind of content requiring high levels of concentration.
- Formal : Conversations with teachers, classmates, coworkers, etc. that involves working on school work, jobs or any other kind of content requiring high levels of concentration
- Solo : Solo vocalization activities such as reading books aloud, singing, talking to oneself, etc.

This is in order to make the analysis of the data easier.

5.6 Participant Demographics

We originally had about 10 participants who were willing to partake in the experiment. However, due to either technical difficulties or personal lack of time, numerous participants dropped out from the experiment before or during the procedure. As a result, we ended up with only 5 participants with valid results. This is a very small number for the type of experiment we are attempting, but in turn, it has allowed us to look more deeply into each participant.

The following table 1 shows the basic information and personality traits of each participant, collected via the aforementioned pre-experiment survey and the Big Five model personality test. O, C, E, A, N of the table each stand for the five personality traits proposed in the Big Five personality model: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism.

Table 1 Participant Information

P.	Sex	Age	Residence Status	O	C	E	A	N
1	Male	19	Dormitory	78	67	89	78	67
2	Male	18	Dormitory	100	89	34	56	34
3	Male	19	With Parents	100	89	56	100	45
4	Female	20	Alone	89	67	56	67	34
5	Female	22	With Parents	45	35	78	34	100

5.7 Experimental Results

5.7.1 General Participant Behavior

The results showed a huge variance between each participant.

Firstly, the maximum word count per day differs massively depending on the participant. One participant recorded a maximum of only about 5,000 characters per day, where on the other hand, another participant recorded a word count over 30,000 characters for a particular day.

Also, the purpose of conversation was highly dependent on the user as well. One participant had no recordings of casual conversations at all, where on the other hand, another participant had a balanced number of both.

Some participants also responded in surveys that they had part time jobs or lessons in which they talked outside of the recording, which supposedly had a significant effect on the results.

5.7.2 Correlations between Word Count and Stress Levels

In order to investigate the correlation between the measured word count and stress levels, we calculated the correlation coefficients between casual word count and average stress levels, formal word count and average stress levels, solo word count and average stress levels and total word count and average stress levels for each day of each participant. We also calculate the values for the entire group as well. Also, keep in mind that values that are left blank are due to the participant not having a single recording of that conversation type.

Results are shown below in figures 2 to 7.

Table 2 Correlation between daily word count and stress levels : All Participants

X	Y	Correlation Coefficient
Casual Word Count	Average Stress Level	-0.041
Formal Word Count	Average Stress Level	-0.133
Solo Word Count	Average Stress Level	0.124
Total Word Count	Average Stress Level	-0.145

Table 3 Correlation between daily word count and stress levels : Participant 1

X	Y	Correlation Coefficient
Casual Word Count	Average Stress Level	-0.297
Formal Word Count	Average Stress Level	-0.102
Solo Word Count	Average Stress Level	0.105
Total Word Count	Average Stress Level	-0.281

Table 4 Correlation between daily word count and stress levels : Participant 2

X	Y	Correlation Coefficient
Casual Word Count	Average Stress Level	-
Formal Word Count	Average Stress Level	-0.137
Solo Word Count	Average Stress Level	0.173
Total Word Count	Average Stress Level	-0.137

Table 5 Correlation between daily word count and stress levels : Participant 3

X	Y	Correlation Coefficient
Casual Word Count	Average Stress Level	-0.172
Formal Word Count	Average Stress Level	-0.327
Solo Word Count	Average Stress Level	-
Total Word Count	Average Stress Level	-0.386

Table 6 Correlation between daily word count and stress levels : Participant 4

X	Y	Correlation Coefficient
Casual Word Count	Average Stress Level	0.314
Formal Word Count	Average Stress Level	0.010
Solo Word Count	Average Stress Level	-0.339
Total Word Count	Average Stress Level	0.245

Table 7 Correlation between daily word count and stress levels : Participant 5

X	Y	Correlation Coefficient
Casual Word Count	Average Stress Level	0.147
Formal Word Count	Average Stress Level	-0.276
Solo Word Count	Average Stress Level	0.193
Total Word Count	Average Stress Level	0.026

The results of Participant 1 and Participant 3 show a weak yet consist negative correlation between the total word count and average stress levels, meaning that when the total daily word count is high, the average stress level will be low, which supports our hypothesis.

On the other hand, results of Participant 4 indicates otherwise, and shows a positive correlation, meaning that when the word count is high, the average stress level is also high, which is against our hypothesis, and stands out as an irregular.

5.7.3 Correlation between Personality Traits and the Correlation between Word Count and Stress Levels

In carrying out this experiment, we hypothesized that the an individual's correlation between word count and stress levels would vary depending on their personality traits. To be more precise, we hypothesized that individuals who are willing to or/and are open to the idea of talking to people would show a stronger negative correlation between word count and stress levels, since speaking more is likely to act as a way of relieving stress for them.

The following table 8 shows the calculation results. As was the case before, O, C, E, A, N of the table each stand for the five personality traits proposed in the Big Five personality model: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. Also, due to constraints in space, the ”correlation between word count and stress levels” will be written as ”Corr. A”, and the ”correlation between personality traits and the correlation between word count and stress levels” will be written as ”Corr. B”.

Table 8 Correlation between Personality Traits and the Correlation between Word Count and Stress Levels

P.	O	C	E	A	N	Corr. A
1	78	67	89	78	67	-0.281
2	100	89	34	56	34	-0.137
3	100	89	56	100	45	-0.386
4	89	67	56	67	34	0.245
5	45	35	78	34	100	0.026
Corr. B	-0.321	-0.499	-0.100	-0.616	-0.002	

According to these results, we can assume that individuals with higher scores of openness, conscientiousness and agreeableness are likely to have a stronger correlation between personality traits and the correlation between word count and stress levels. This suggests that individuals with a personality of high openness, conscientiousness and agreeableness are more likely to feel more relieved of stress as they talk more. This does instinctively sound valid, but we must keep in mind that these values have been taken from only 5 participants, thus is lacking in reliability at the moment.

5.7.4 Correlation between Average Voice Pitch and Stress Levels

We also calculated the correlation coefficients between the participants' average voice pitch of each day and their stress level of each day. The results are presented in the following figure 9.

Table 9 Correlation between Average Voice Pitch and Stress Levels

Participant	Correlation Coefficient
Participant 1	0.417
Participant 2	0.081
Participant 3	0.112
Participant 4	0.080
Participant 5	0.051
All Participants	-0.191

As seen in figure 9, only Participant 1 showed a significant correlation coefficient. However, considering how all the other participants showed a significantly low correlation, and the fact that Participant 1's results also go against our hypothesis, we would like to conclude that direct correlations between a person's average voice pitch and his/her stress level cannot be drawn from the results of this experiment.

5.8 Discussion

5.8.1 Recap of the Results

From the results of this experiment, we could find some small hints of correlations here and there, but it was nowhere near enough for us to make a concrete statement about the correlation between speech and stress, or with personality traits. Although the results are interesting and contains possibilities for future studies, we must say that this experiment on its own has failed to meet its goal.

The biggest reason for our failure is the absolute lack in number of participants. 5 participants were nowhere enough for us to draw meaningful results from.

The number of data per participant was too small as well. 15 days is too short for an experiment of this kind, and the number of measurements we received from each participant for each day was also very low.

There also may have been a problem in how we directed the experiment as well. It may have been better if we had set more clear and strict instructions, especially since we did not have a long time to experiment with, and had to collect a certain amount of data during quickly.

We must also not forget how complex dealing with stress is to start with. The ManWaKei system can only take measurements while the participant is at home and also actively using the system to take measurements. Therefore, it is lacking in the ability to follow and investigate the participant in depth. This is a clear and ultimately inevitable limitation of experiments like this. If we could improve the system so that it can at least take measurements of conversations taking place outside of the house as well, the results may have been a more plausible.

5.8.2 Future Directions

Although the ManWaKei system we designed for this research had a very limited flexibility and had numerous flaws and limitations, alternative and better systems with similar concepts have the possibility of becoming a much more valuable system in the future. Here, we will explore how such systems could be used, and what has to be achieved in order for such applications to be created.

5.8.3 Possible Future Applications

If a system capable of measuring the amount of conversation in a more flexible way can be realized and becomes accessible to people around the world, it would have the possibility of expanding a user's understanding of one's life style into a new level, and may become a new standard of self-driven health care, much like pedometers and smartwatches. As mentioned before, majority of the participants of our experiment showed interest in such system as well, thinking that such a system would help them understand themselves even deeper, and would ultimately connect to more healthy future. In the world today, the interest towards understanding one's own health is continuing to grow, so such system should be of interest for many people around the world.

Not to mention, the development of such system would also expand the possible field of research by a massive amount. Not only can it be used for an advanced version of this project, investigating the correlation between speech and stress in more detail, it could be used in numerous different fields as well. It could be used to quantify the amount of conversation taking place in a specific area, which could be used for city planning or marketing, or it could be used to identify individuals who are communicating more with others, and investigate how they are managing to do so. Making something previously unmeasurable measurable opens up a splendid amount of opportunity.

5.8.4 Identifying the Speaker

In making the system able to be used in a context with more than only the speaker involved, it is crucial for the system to be able to identify the speaker, so that it can differentiate between the user's speech and speech of others. Hardware related solutions could be useful, but a more modern way would be to use deep learning to identify the speaker from the recorded audio. Such technologies are already in development, so once they become more accessible and reliable, there is no doubt that it will expand the possibility of systems like these to the next level.

5.8.5 Taking Word Counts in Smarter Faster Ways

A major problem of the ManWaKei system was how slow its process was on counting the words the user spoke. This was because it had to transcript the entire text once in order to count the words in it. This process could be made faster if the deep learning technology improves and become faster, but it can probably be made even faster if a deep learning model designated for simply counting the words in a given audio is created. In creating an accessible and useful word counting system for daily use, the development of such system must be crucial.

5.8.6 Taking the Contents of the Speech into count

Another major drawback of the ManWaKei system was that it required the users to manually input the participants and reason of conversation. This is far too tiresome for use in a daily situation. If a system can analyze the actual content of the speech, and decide on the purpose / reason of the conversation by itself, it would massively reduce the need of input from the user, making it a much more accessible system. However, such system would have to take extreme care on the user's privacy, so there are numerous difficulties in realizing such a system.

6. Conclusion

We designed a system that can quantitatively measure a user's amount of speech, and used it to analyze the correlation between speech and stress. Due to numerous reasons including lack of participants and technical weaknesses in the system, we could not manage to find a strong correlation from our results. However, some participants showed results that strongly support our hypothesis, so there is potential in our system. If technical improvements are made, it could be used for further and better research, along with being used as a self-driven healthcare tool.

Acknowledgments This work was supported by JST CREST Grant Number JPMJCR19A4 Japan.

References

[1] 日本放送協会. 新型コロナウイルス データで見る感染状況一覧— nhk 特設サイト. https://www3.nhk. or.jp/news/special/coronavirus/data-widget/#mokuji4.
[2] 東京都産業労働局. テレワーク実施率調査結果をお知らせします! 4 月の調査結果. https://www.metro. tokyo.lg.jp/tosei/hodohappyo/press/2021/05/07/10.html, May 2021.
[3] Japan Broadcasting Corporation. 新型コロナウイルス感染症の状況を踏まえた大学等の授業の実施状 況. June 2020.
[4] Atsushi Kawakubo and Takashi Oguchi. 余暇における他者との交流が主観的幸福感および抑うつに及 ぼす影響. ストレス科学研究, Vol. 30, p. 69-76, 2015.
[5] Gotou Nobuyuki. 「話すことによるストレス解消法」にみる「自己開示のメリット」とビジネスでの 活用方法「話すことによるストレス解消法」にみる「自己開示のメリット」とビジネスでの活用方法. https://chitekishisan.com/troubled-consultation-and-self-disclosure/.
[6] 日テレ NEWS24. 「さみしい」コロナで深刻 若年女性の孤独. https://www.news24.jp/articles/2021/ 06/11/04888333.html, Jun 2021.
[7] 新型コロナで深刻「一人暮らしの若者の孤独」専門家が重視する疲労と孤立の解消. https://www. j-cast.com/trend/2021/02/05404457.html, Feb 2021.
[8] Martine Van Puyvelde, Xavier Neyt, Francis McGlone, and Nathalie Pattyn. Voice stress analysis: A new framework for voice and effort in human performance. Frontiers in Psychology, Vol. 9, p.1994, 2018.
[9] Naoshi Matsuo, Shoji Hayakawa, and Shouji Harada. Technology to detect levels of stress based on voice information. Fujitsu Scientific and Technical Journal, Vol. 51, p. 48-54, 2015.
[10] Brian Edward Johnson. The speed and accuracy of voice recognition software-assisted transcription versus the listen-and-type method: a research note. Qualitative Research, Vol. 11, pp. 91-97, Feb 2011.
[11] Google. Speech-to-text: 自動音声認識 - google cloud. https://cloud.google.com/speech-to-text.
[12] IBM. Ibm watson - 概要 - 日本 - ibm. https://www.ibm.com/jp-ja/cloud/watson-speech-to-text.
[13] Sanae Fukuda. Objective and subjective assessment for stress. Kansai University Welfare Sciences,Department of Welafare Sciences, Vol. 10, pp. 23-28, 2016.
[14] Sousuke Miyamoto and Ui Miyoko. 質問紙調査と心理測定尺度: 計画から実施・解析まで. SAIENSU-SHA Co.,Ltd., 2014.
[15] Hikaru Hori, Asuka Katsuki, Kentaro Suga, and Reiji Yoshimura. How to assess objective stress? JJOMT, Vol. 66, pp. 330-334, 2018.
[16] Garmin Ltd. Health science : Garmin. https://www.garmin.co.jp/minisite/garmin-technology/ health-science/.
[17] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. Stress and heart rate variability: A meta-analysis and review of the literature. Psychiatry Investigation, Vol. 15(3), pp. 235-245, 2018.
[18] 人差し指で簡単チェック!ストレスを科学的に測定するアプリ「ストレススキャン」android 版をリリース! iphone 版もさらに進化! https://prtimes.jp/main/html/rd/p/000000004.000020121.html, Jul 2016.
[19] Molly Owens. The big five personality model. https://www.truity.com/book/ big-five-personality-model.
[20] 株式会社ダイレクトコミュニケーション. 心理テスト、詳しい性格・ストレス・恋愛診断が無料で出来ます♪. https://commutest.com/bigfive/.
[21] JeremyCCHsu. Python-wrapper-for-world vocoder. https://github.com/JeremyCCHsu/ Python-Wrapper-for-World-Vocoder.
[22] Masanori Morise. World. http://www.kki.yamanashi.ac.jp/m̃morise/world/, 2013.