**PROPOSAL FOR SPEECH DATABASE RECORDING**

horizontal line

# Introduction

The aim of this work is to record an emotional speech database in New Zealand English, named as UOAEMO-DB. This proposal explains the plan for the recording and the methods and database to be adopted for the process. This is further extended by the inclusion of the recording of the EGG signals as well, along with the speech signals.

## Participants:

The purpose of the database is to collect emotional speech database with a good coverage in terms of emotions and also adequate data from both male and female speakers. Professional artists will be hired for the speech production and can be expected to deliver stimulated intense emotions during the whole recording process.

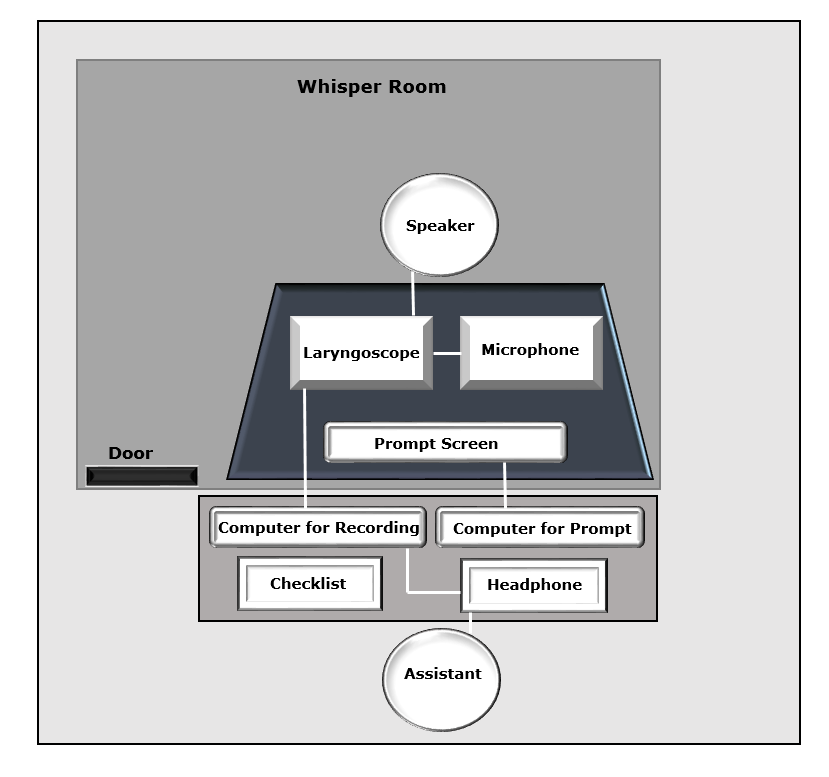
Proposed speakers - 4 ( male : 2 , female : 2), within the age range of 20~50.

**Recording device and setup:**

The recording requires the following devices:

1. Condenser microphones with good quality.
2. A Microphone calibration device.
3. Laryngoscope -for EGG recordings.
4. Headphones for feedback.
5. A soundproof environment.
6. A recording device.
7. Computer screen for prompts.
8. A system for checking each waveform being recorded for clipping, popping and other errors.

The following is a rough diagram for the setup that can be followed during the recording process.



**Prompts for recording:**

The prompts of the text to be recorded is to be displayed to the speaker on a computer screen. This will be done using macros and powerpoint. There will be options to change to the previous or next prompts, as required by the procedure.

Since the recording involves emotions, and the speakers have to be stimulated with these emotions, some context and pictures pertaining to the required emotions will be given to the speakers prior to the recording of sentences on each emotion.

**Emotional coverage:**

The emotions proposed to be covered in the database are mainly five most frequently encountered ones:

1. *Happness*
2. *Sadness*
3. *Anger*
4. *Neutral*
5. *Excitement*

Along with this, certain sub-emotions or variants of these main emotions will also be covered:

1. *Questions/Interrogative*
2. *Apolegetic*
3. *Encouragement/Reassuring*
4. *Concerned*
5. *Assertive*
6. *Anxiety*

This will include the coverage of 5 main emotions and 6 variants. These variants have been chosen due to their relevance with respect to Human Computer Interactive technology (eg: healthcare robot)

**Sampling Rate:**

The sampling rate of the recording is set to be 44 kHz.

**Database size:**

The proposed database will have the voice from 4 speakers - 2 male and 2 female.

There will be 15 common sentences for different main emotions(15\*5 sentences).

There will be 10 specific sentences for each sub-emotion(10\*6 sentences).

These sentences will be recorded twice in two separate sessions, to generalize the possible time related mental variation and to ensure that the required emotion intensity level is consistent and that it is repeatable.

So, the total number of sentences for each person will be :

5 (main emotions) \* 2 (repetitions) \* 15 (sentences) +

6 (sub emotions) \* 2 (repetitions) \* 10 (sentences) = 150 + 120 = 270 sentences per person.

Total = 4 (number of speakers) \* 270 = **1080 sentences.**

**Memory requirement:**

An average 44KHz sampled waveform will need a memory requirement of : 502060 bytes

So, total size of database will be: 1080 (total sentences) \*502060 =**520 Megabytes.**

**Recruitment and Payment:**

4 professional speakers will be hired for the task.

They will be paid on an hourly basis, from the Press account and CARES seed funding grants.

The expected time for each session is 4-5 hours.

So the total estimated time will be 5 (hours) \* 4 (speakers) \* 2 (sessions on separate dates) = **40 hours.**

**Ethics Approval:**

The ethics approval for the task will be submitted to the university ethics committee for approval if necessary.

**Recording Context:**

The following is the list of sentences to be recorded:

For the five main emotions:

1. Tom b**ea**ts that f**ar**mer.
2. John l**au**ghs like your f**a**ther.
3. The s**ee**d is bur**ie**d in d**ee**p.
4. Taylor likes st**ew**ed asian f**oo**d.
5. The l**or**d swims in the s**ea**.
6. Jack vi**ew**s an **ar**t p**ie**ce.
7. C**ar**l l**ea**ps int**o** a j**ee**p.
8. Linda **a**sks f**or** m**ore** d**ar**ts.
9. Find your b**oo**t in this sh**u**te.
10. W**a**ter h**ar**ms the new b**or**n boy.
11. I have not s**ee**n my t**oo**th.
12. Work h**ar**d or y**ou** l**o**se.
13. Jim s**aw** the p**or**t.
14. They should st**ar**t to t**al**k.
15. Sound the h**or**n if y**ou** n**ee**d m**ore**.

**a**10 **u**: 10 **o**: 10 **i**: 10

The same 15 sentences will be recorded for each of the 5 main emotions.

For the sub emotions, the following sentences are proposed to be included:

* ***Apolegetic***

|  |
| --- |
|  |
| 1. ***I owe you an apology.*** 2. ***Forgive me, but I must leave now.*** 3. *Tom beats that farmer.* 4. *John laughs like your father.* 5. *The seed is buried in deep.* 6. *Taylor likes stewed asian food.* 7. *The lord swims in the sea.* 8. *Jack views an art piece.* 9. *Carl leaps into a jeep.* 10. *Linda asks for more darts.* 11. *Find your boot in this shute.* 12. *Water harms the new born boy.* 13. ***Your results are unbelievable.*** 14. ***We were away a year ago.*** 15. ***How are you doing?*** |
|  |

* ***Encouragement/Reassuring***

1. ***Thank you for your help!***
2. ***You should be proud of yourself!***
3. *Tom beats that farmer.*
4. *John laughs like your father.*
5. *The seed is buried in deep.*
6. *Taylor likes stewed asian food.*
7. *The lord swims in the sea.*
8. *Jack views an art piece.*
9. *Carl leaps into a jeep.*
10. *Linda asks for more darts.*
11. *Find your boot in this shute.*
12. *Water harms the new born boy.*
13. ***Your results are unbelievable.***
14. ***We were away a year ago.***
15. ***How are you doing?***

* ***Concerned***

1. **It seems like you are taking a long time.**
2. **I know how frustrating it can be.**
3. *Tom beats that farmer.*
4. *John laughs like your father.*
5. *The seed is buried in deep.*
6. *Taylor likes stewed asian food.*
7. *The lord swims in the sea.*
8. *Jack views an art piece.*
9. *Carl leaps into a jeep.*
10. *Linda asks for more darts.*
11. *Find your boot in this shute.*
12. *Water harms the new born boy.*
13. ***Your results are unbelievable.***
14. ***We were away a year ago.***
15. ***How are you doing?***

* ***Assertive***

1. ***Please try to do your exercises next time.***
2. ***I will think about it and get back to you.***
3. *Tom beats that farmer.*
4. *John laughs like your father.*
5. *The seed is buried in deep.*
6. *Taylor likes stewed asian food.*
7. *The lord swims in the sea.*
8. *Jack views an art piece.*
9. *Carl leaps into a jeep.*
10. *Linda asks for more darts.*
11. *Find your boot in this shute.*
12. *Water harms the new born boy.*
13. ***Your results are unbelievable.***
14. ***We were away a year ago.***
15. ***How are you doing?***

* ***Anxious***

1. ***I am so worried about my exam results.***
2. ***I am not sure how severe the wound is!***
3. *Tom beats that farmer.*
4. *John laughs like your father.*
5. *The seed is buried in deep.*
6. *Taylor likes stewed asian food.*
7. *The lord swims in the sea.*
8. *Jack views an art piece.*
9. *Carl leaps into a jeep.*
10. *Linda asks for more darts.*
11. *Find your boot in this shute.*
12. *Water harms the new born boy.*
13. ***Your results are unbelievable.***
14. ***We were away a year ago.***
15. ***How are you doing?***

**Post-Processing Work:**

Each of the recorded sentences have to be checked for errors during recording. This includes:

* Clipping effects
* Removing unwanted pauses before and after the recording.
* Dividing the recording into sentences(if needed).
* Separate the EGG channel from the voice channel.

**Perception tests:**

It is proposed to conduct a post-recording perception test, to rate the quality and naturalness of the emotions recorded.

This is proposed to be conducted in the form of online and offline surveys with the sound waveforms being played to the people.

They will be asked to rate the recordings based on:

* Intensity of emotion
* Naturalness
* Quality of the recordings