1 Description of the recording setup

In the data collection, speech and EGG [1] signals were recorded from 50 speakers, comprising 25 male and 25 female speakers, by varying vocal intensity. The age range was 21 to 31 years for female speakers and 20 to 38 years for male speakers. For collecting the speech signals, DPA 4065-BL headset condenser microphone was used, and for collecting EGG signals, EG2-PCX2 electroglottograph [1] was used. The other equipment used during the recordings were a calibrator, a sound card, and a laptop with Audacity program.

During the recordings, the microphone was set at a distance of 5 cm from the center of the lips, and electrodes positions were adjusted using the EGG device's electrode placement indicator. Both the signals were sent over an RME Babyface sound card. They were recorded using the Audacity program with a sampling frequency of 44.1 kHz. After the setup and calibration, speakers were asked to produce specified four intensity categories (soft, normal, loud, and very loud) in sequential order.

1.1 Speaking tasks

The data collection process was divided into two sessions (Session-1 and Session-2) for each speaker. Session-2 was a repetition of Session-1, and the tasks were identical in both sessions. Each session had two tasks. The first task (hereafter referred to as Task-1) was a production of individual sentences, and the second task (hereafter referred to as Task-2) was a production of two different paragraph readings. All the tasks performed by each speaker in both sessions were saved in a single audio file per speaker in the Audacity program. A detailed explanation of each task is described below.

- **Task-1**: In this task, each speaker was asked to recite the 25 given sentences (see Table 1) in the four intensity categories. These sentences were selected from the TIMIT [2] database [2]. The sentences were selected based on the number of words (ranging from 3 to 7 words) per sentence by aiming at sentences of different duration. The process of reciting was like producing the "Sentence-1" in soft; "Sentence 1" in normal; "Sentence-1" in loud; "Sentence-1" in very loud; and so on up to "Sentence-25" in a sequential order. In between each intensity level, the speaker was allowed to take a natural pause.
- **Task-2**: As in Task-1, each speaker was asked to recite two different paragraphs (hereafter referred to as para) in four intensity categories. The first paragraph was taken from a weather forecast excerpt [3] (see para-1 in Table 2), and the second one was a continuous excerpt from a novel *"The Call of the Wild by Jack London"* [4] (see para-2 in Table 2). The first paragraph was identical for all the speakers, whereas the second paragraph was different for all the speakers. The second paragraph was a continuous excerpt from the novel from the 1^{st} speaker to the 50^{th} speaker. As in Task-1, the process of reciting was like producing the "para-1" in soft; "para-1" is normal; "para-1" in loud; "para-1" in very loud; then "para-2" in soft; "para-2" is normal; "para-2" in loud; and "para-2" in very loud (in a sequential order). Between each intensity level, the speaker was allowed to take a natural pause.

[1] EGG is a device that measures the degree of contact between vocal folds while producing speech.
[2] A particular database which provides the speech data required for acoustic-phonetic studies.

Table (1) Sentences used for Task-1 data collection.

Sentences list extracted from TIMIT
We think differently.
He spoke soothingly.
They despised foreigners.
That is your headache.
Nevertheless, it's true.
Leave me your address.
Come home right away.
Turn shaker upside down.
He makes me uncomfortable.
Did you eat yet?
Did anyone see my cab?
Push back up and repeat.
Hope to see you again.
This was easy for us.
Are you looking for employment?
Guess the question from the answer.
Orange juice tastes funny after toothpaste.
They all like long hot showers.
How do they turn out later?
Who is going to stop me?
All nut kernels are rich in protein.
Don't plan meals that are too complicated.
They often go out in the evening.
It was time to go up myself.
Birthday parties have cupcakes and ice cream.

1.2 Speech segmentation using target-induced labeling

For the segmentation, each raw speech file was imported into the Audacity software. The raw file consisted of two-channel data (speech and EGG). Thus, the first five seconds corresponded to the calibration signal, followed by a speaker's actual speech recording in both tasks and sessions. Then, as shown in Figure **??**, the recorded data was labeled manually into *target-induced* categories using the Audacity software. The target-induced labeling refers to categorizing each recorded speech signal as either soft, normal, loud or very loud according to the speaking task, which the speaker used in the production of the corresponding signal.

The target-induced labeling used for segmenting the files is as follows:
"spNN_SN_XXXX_XXXX", where

- Characters 1-4 indicate the speaker number. For instance, 'sp10' refers to speaker number 10.
- Characters 5-6 indicate the sessions ('s1' refers to Session-1 and 's2' refers to Session-2).
- Characters 7-10 indicate the tasks in Task-1 or in Task-2. 'senX' refers to the sentences in Task-1 where X ranges 1-25. 'paraX' refers to the paragraph in Task-2 where X=1 refers to paragraph-1 and X=2 refers to paragraph-2.
- Characters 11-18 indicate the target-induced categories, i.e., soft, normal, loud, or very

Table (2) Paragraphs used for Task-2 data collection.

Paragraphs extracted for Task-2
para-1: "*Weather forecasting is the application of science and technology to predict the state of the atmosphere for a future time and a given location. Human beings have attempted to predict the weather informally for millennia, and formally since at least the nineteenth century. Weather forecasts are made by collecting quantitative data about the current state of the atmosphere and using scientific understanding of atmospheric processes to project how the atmosphere will evolve.*" [3]
para-2 (sample): "*Buck did not read the newspapers, or he would have known that trouble was brewing, not alone for himself, but for every tide-water dog, strong of muscle and with warm, long hair, from Puget Sound to San Diego. Because men, groping in the Arctic darkness, had found a yellow metal, and because steamship and transportation companies were booming the find, thousands of men were rushing into the North land. These men wanted dogs, and the dogs they wanted were heavy dogs, with strong muscles by which to toil, and furry coats to protect them from the frost.*" [4]

loud.

Example: "sp1_s1_sen1_soft" refers to speaker number 1, session number 1, sentence number 1 produced in the soft intensity class.

After the labeling, the speech files (108 files per session) of each speaker were exported. Overall, as the number of speakers was 50, a total of 5400 files per session were created for all categories. The statistics of the number of speech files per session created using segmentation are provided in Table 3.

Table (3) Number of speech files created after target-induced labeling per session.

Target-induced categories	**Task-1**	**Task-2 para-1**	**Task-2 para-2**
Soft	1250	50	50
Normal	1250	50	50
Loud	1250	50	50
Very loud	1250	50	50
Total	5000	200	200

1.3 Description of files in database

There are five subfolders, namely *calibrationtone_female, calibrationtone_male, Task1, Task2.1* and *Task2.2*. Among these folders, *calibrationtone_female* and *calibrationtone_male* folders contain calibration tone signal per speaker. Further, the task folders (*Task1, Task2.1* and *Task2.2*) contain separate subfolders regarding the sessions, gender, and target-induced categories.

References

[1] "Eg2-pcx2 electroglottograph home page.." http://www.glottal.com/Electroglottographs.html. Accessed: 2021-08-30.

[2] J. S. Garofolo, "Timit acoustic phonetic continuous speech corpus," *Linguistic Data Consortium, 1993*, 1993.

[3] "Weather forecasting excerpt." https://en.wikipedia.org/wiki/Weather_forecasting. Accessed: 2021-06-30.

[4] "The Call of the Wild by Jack London." https://www.gutenberg.org/ebooks/215. Accessed: 2021-06-30.