Toward Measuring Conversation Duration Using a Wristwatch-type Wearable Device

Yuki Komatsu*, Kazuki Shimojo†, Yuuki Nishiyama† and Kaoru Sezaki†

*Ritsumeikan Keisho Senior High School, Hokkaido, Japan, †The University of Tokyo, Tokyo, Japan

{yuukin,sezaki}@iis.u-tokyo.ac.jp

OHANASHI: Measuring Conversation Duration Using A Smartwatch

Potential for secondary health effects!!

Assessment of situation

Subsequent development of countermeasures

Smartphone places in various locations

Recognition accuracy might be decreased

On the desk

In the pocket

Originality

Propose system

System for measuring the duration of conversations in daily life by using an off-the-shelf wristwatch-type wearable device

• Detecting conversation events every second
• On-device Machine Learning
• Recording conversation duration
• Practicality in daily life

Evaluation

1. Classification accuracy of the audio classification model
2. Battery consumption of the system

OHANASHI: An overview of our proposed system

A Spectrogram of training data used

A human voice has characteristic waves around 100 to 1000 Hz

Conversations can be detected using ML

Learning Phase

1. Using Audio Feature Print
2. Training 100 times
3. Using Sound Classification model in CreateML (3.0) on macOS

Prediction Phase

1. Using AVAudioEngine
2. Using AudioStreamAnalyzer
3. Measuring working with the Audio
4. Determining whether it is classified (each second)
5. Saving label + Timestamp
6. Local SQLite database

A The training dataset(8K)

Voice(3K)

Male voices (1.9K) + female voices (1.9K)

From Common Voice Corpus

Noise(3K)

White noise

12K samples

B Accuracy of Conversation Detection

<table>
<thead>
<tr>
<th>Condition</th>
<th>Label</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversation</td>
<td>1.00</td>
<td>0.73</td>
<td>0.84</td>
<td>0.81</td>
</tr>
<tr>
<td>Noise</td>
<td>0.94</td>
<td>1.00</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>Average</td>
<td>0.97</td>
<td>0.85</td>
<td>0.90</td>
<td></td>
</tr>
</tbody>
</table>

The average value of the F1-score in silent and noisy conditions was 0.90 and 0.86.

B. Battery Consumption

We monitored changes in the battery level by using Apple Watch series 7, 44mm, GPS model.

- Natural battery consumption with the device on a desk (Baseline)
- With microphone activated conditions (With-mic)
- With the model applied condition (With-mic&sql)

Result: The battery life was Baseline: 54.25 hours, With-mic: 18 hours, With-mic&sql: 15.52 hours

Discussion and Conclusion

• F1-score 0.86

Optimized power consumption

Less than a day (15.52 hours)

The power consumption could be optimized by executing conversation detection periodically at regular intervals and continuing the process when a conversation is detected.

Our system can extend to detect multiple contexts from sounds

- Recognizing people
- Human-generated sounds (e.g., laugh, cough, and crying)
- Environmental sounds
- Emotion

In the future, we plan to integrate context estimation functions onto Ohanashi.
