TUS
Technological University of the Shannon:
Midlands Midwest
Ollscoil Teicneolaíochta na Sionainne:
Lár Tíre Iarthar Láir

TUS Research

Evaluating performance of commercial automatic speech recognition systems for speakers with dysarthria

Sirajum Munir Fahim, Niall Murray, Ronan Flynn
Faculty of Engineering & Informatics, Technological University of the Shannon: Midlands Midwest

ORCID

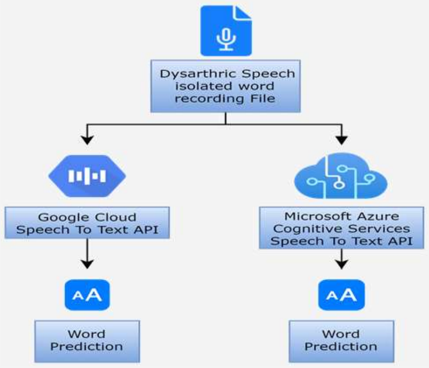
INTRODUCTION

Dysarthria refers to motor neuron speech disorders that limit the ability to control muscle groups used for speech production. Dysarthric speech has reduced intelligibility due to being slurry, breathy and slow compared to typical speech. The modern advancements of automatic speech recognition (ASR) and the increase of speech-controlled systems raise the question on their inclusiveness for speakers with dysarthria.

This study evaluated the performance of two popular commercial ASR systems with dysarthric speech and compared the performance with state-of-the-art dysarthric speech ASR systems.

METHOD

Data: Universal Access Database for Dysarthric Speech (UASpeech) [1] corpus containing isolated word recording of 15 dysarthric speakers are used for the study. The speakers have varied level of intelligibility.


Dysarthric Speech isolated word recording File

Google Cloud Speech To Text API

Microsoft Azure Cognitive Services Speech To Text API

Word Prediction

Word Prediction

Figure 1. Flowchart of gathering recognised words from Google and Microsoft Speech to Text ASR systems

API: The speech recordings are sent to the Google Cloud and the Microsoft Azure Cognitive Services speech-to-text APIs. The APIs take the speech recordings as input and attempt to recognise the speech by predicting text as output.

$$\text{Word Error Rate (WER)} = \frac{\text{Insertions + Deletions + Substitutions}}{\text{Number of words in reference transcript}}$$

WER: Word Error Rate (WER) is calculated by comparing predicted texts with the original transcript. Any substitution, deletion or insertion to the original transcript counts as an error.

Acknowledgement
This study was funded by AIT Presidents Doctoral Scholarship

RESULTS

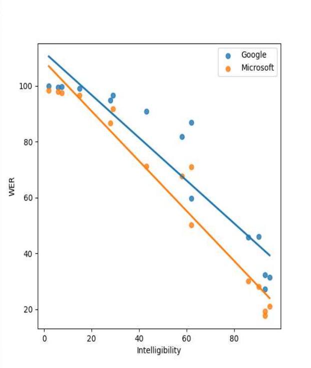


Figure 2. Word Error Rate (WER) vs speaker intelligibility for Google and Microsoft Speech to Text API

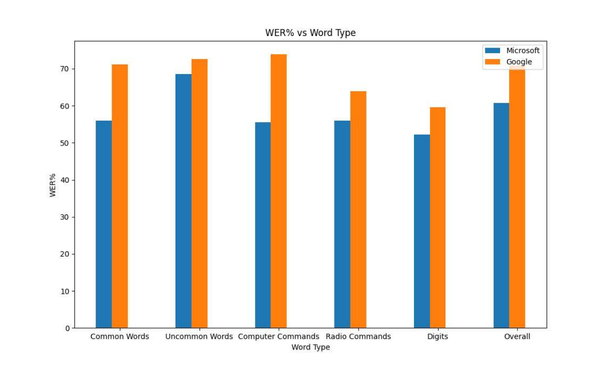


Figure 3. Word Error Rate (WER) for different types of words on UASpeech test block

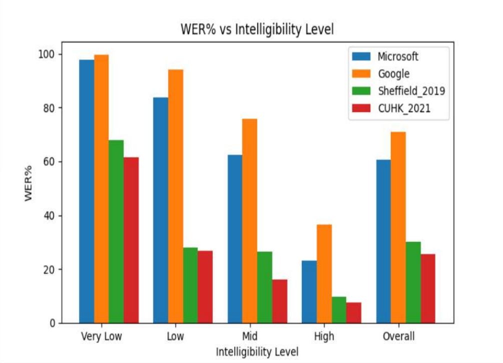


Figure 4. Word Error Rate (WER) vs speaker intelligibility performance of Google and Microsoft and two state-of-the-art ASR systems [2], [3] trained on dysarthric speech

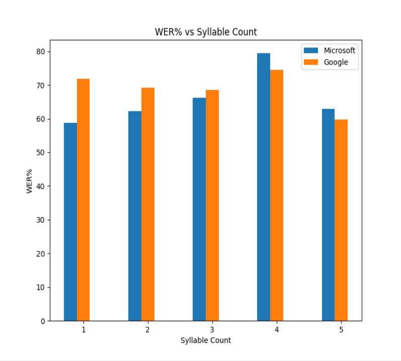


Figure 5. Word Error Rate (WER) for words with different syllable count

DISCUSSION

The study found that the performance of commercially available ASR systems are very poor for speakers with dysarthria. ASR systems trained on dysarthric speech [2], [3] significantly outperform typical speech systems.

The results show that the Google and Microsoft's ASR systems almost completely failed to recognise speech from very low to low intelligibility groups with dysarthria. Performance improved for moderate and high intelligibility levels but still with high word error rate. Shorter and common words achieved higher accuracy than uncommon words. Microsoft's system outperformed Google's system on every intelligibility level of dysarthric speech.

State-of-the-art dysarthric speech ASR systems [2], [3] that are trained on dysarthric speech are significantly more accurate than these two commercially available systems trained on typical speech.

Recognition accuracy consistently worsens on the Microsoft ASR system for words with higher syllables up to four syllables. Interestingly, for five syllable words, the recognition accuracy increases for the Google and Microsoft ASR systems.

References

[1] Heejin Kim, Mark Hasegawa-Johnson, Adrienne Perlman, Jon Gunderson, Thomas Huang, Kenneth Watkin, and Simone Frame. Dysarthric Speech Database for Universal Access Research. Technical report, 2008
[2] Feifei Xiong, Jon Barker, and Heidi Christensen. Phonetic Analysis of Dysarthric Speech Tempo and Applications to Robust Personalised Dysarthric Speech Recognition. In ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, volume 2019-May, pages 5836– 5840. Institute of Electrical and Electronics Engineers Inc., May 2019. ISSN: 15206149.
[3] Shansong Liu, Mengzhe Geng, Shoukang Hu, Xurong Xie, Mingyu Cui, Jianwei Yu, Xunying Liu, and Helen Meng. Recent Progress in the CUHK Dysarthric Speech Recognition System. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29:2267–2281, 2021. Conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing.