Multimodal sensing for Subvocal speech recognition for Silent speech interfaces in future AR glasses

Melinda Anderson (Meta Platforms Technologies, LLC), Thomas Lunner (Meta Platforms Technologies, LLC), Ananta Narayanan Balaji (Meta Platforms Technologies, LLC), Morteza Khaleghimeybodi (Meta Platforms Technologies, LLC)

Summary

Figure 1: Silent speech interfaces for future AR glasses and hearables embedded with EEG, surface EMG electrodes. The glasses and hearable can be used independently or can also be used together.

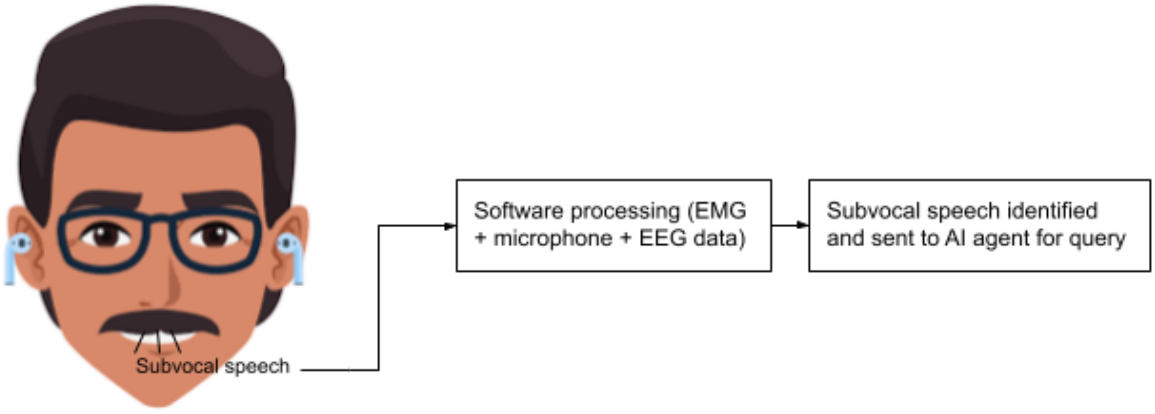

Figure 2: The user wears AR glasses/hearables and utters a silent speech (subvocal speech) query for the AI agent. Ear muscles and facial muscles are activated when the user utters silent speech. Software processing of the acquired microphone, sEMG and

EEG electrodes reveals the uttered silent speech which may be sent to an AI agent for query processing.

Subvocal speech occurs while someone reads a book or talks to oneself or murmurs. Unlike whispering where closer ones can hear speech, such voices cannot be heard by others nearby. Subvocal speech recognition is usually done with electromyograms (EMG) that are unique for each speaker. EMG electrodes are placed on the neck and facial muscles to record muscle activations while making subvocal speech. In addition, EEG electrodes can also be placed on the head to capture electrical signals generated by the brain when subvocal speech has been made. These electrical signals can then be interpreted using signal processing techniques for inferring speech.

An audio system utilises EMG and EEG electrodes placed at different locations on the inner side of future AR glasses to record muscle activity and brain activity when the user makes a subvocal speech. In some embodiments, in-ear EEG, EMG and microphone signals are used, as in-ear sensors can also measure muscle activations, brain activity and vibrations being made when a user makes a silent subvocal speech. By utilizing the data available from such multimodal sensors distributed on the glasses and/or hearables, the audio system can make use of data-driven ML/DL models to reproduce speech from the EEG, EMG, microphone signals collected from the glasses as well as hearables.

Example utilization scenarios

(a) subvocal speech recognition with AR glasses alone (b) subvocal speech recognition with hearables alone and (c) subvocal speech recognition with both AR glasses and hearables. Such speech reproduction can be useful for the following applications:

1. Silent commands/wake words for future AR/VR user interfaces and smartphones.
2. Unlike existing voice assistants like Siri, alexa etc. which require explicit loud voice being made to answer user queries (even in public), this speech interface just requires the users to produce subvocal speech which doesn’t require lip as well as mouth movements.
3. It will serve as a communication tool for the hard-of-speaking specially disabled people.