TRANSLATING THOUGHT TO SPEECH:

GIVING VOICE TO THE SILENCED

Author: **Christian Keszthelyi**

*What if you could speak without ever moving your mouth? Just picture a word, and a machine says it for you. At the University of Malta, AI graduate **Jeremy Farrugia** is working to make that a reality.*

Jeremy Farrugia's research, supervised by Prof. Alexiei Dingli, explores a deceptively simple idea – using thought alone to communicate. The project is part of a broader initiative **THINK** has observed across UM departments focused on assistive technologies designed to support people with profound physical disabilities. Such individuals remain fully aware and cognitively capable, but face significant barriers to communication and movement.

Farrugia's method, speech imagery, is the act of internally repeating a word without actually vocalising it. It's the kind of mental behaviour most of us do daily, almost subconsciously. However, when combined with AI and neuroimaging, this silent whisper may one day become a lifeline.

The project relies on electroencephalography (EEG), a non-invasive method that measures brain activity using sensors placed on the scalp. Figure 1 shows the EEG device used in Farrugia's study – a commercially available headset capable of recording from 14 channels across the head.

To help decode where these signals come from, Figure 2 illustrates the electrode placement based on the international 10–20 system – a standardised map of the scalp used in EEG. This placement helps capture faint electrical patterns that diffuse outward from neuron activity deep inside the brain.

'EEG provides a low-resolution view of brain activity,' Farrugia tells **THINK**. 'We don't get detailed 3D scans like with fMRI, but we do get enough of a signal to detect broad patterns in brain activity, and that's often all we need.'

20 WORDS, 84% ACCURACY

To train the AI, participants were shown a word on-screen and asked to repeat it silently in their minds. These ranged from simple affirmations, such as 'yes' and 'no', to request verbs like 'help', 'put', 'give', and 'come'.

The resulting data was fed into a hybrid model combining convolutional neural networks (CNNs) and transformers – AI architectures adept at recognising spatial and temporal patterns in EEG signals. 'CNNs helped us process the signals as 2D grids while the transformers allowed the model to spot long-range dependencies within the brain's activity,' Farrugia explains. The initial results have been promising.

Figure 1: The EEG headset used to record speech imagery signals.
Estreito, Z., Le, V., Harris Jr., F. C., & Dascalu, S. M. (2024). Evaluating an Elevated Signal-to-Noise Ratio in EEG Emotion Recognition. *International Journal of Software Innovation, 12*(1), 1–15. https://doi.org/10.4018/IJSI.333161

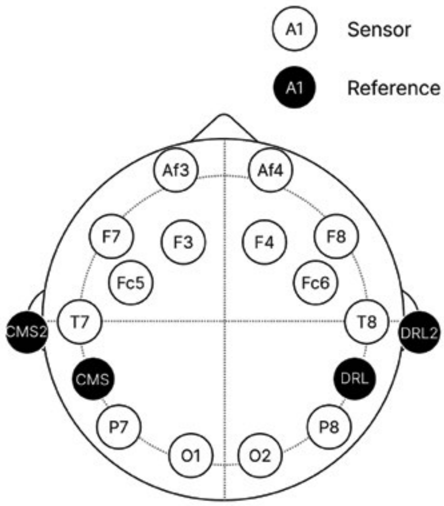

Figure 2: Electrode positions on the scalp used to capture brain activity.
Image courtesy of Jeremy Farrugia

The system could classify imagined words with an average accuracy of 84%. This is a significant margin above the 5% chance level and, crucially, just beyond the 80% threshold identified in assistive tech literature as the baseline requirement for practical use.

WHAT THE BRAIN SAYS

Figures 3 and 4 visualise this brain activity in full colour. Figure 3 shows the average EEG activity patterns for each word. Notice how different each map looks – every word produces a unique 'neural fingerprint' of the imagined speech.

Meanwhile, Figure 4 illustrates the individual variations in participants' brains. 'There's some overlap in the frontal cortex, but beyond that, each brain lights up in different places,' Farrugia notes, moving on to state that 'it's one of the reasons this task is so hard.'

SHIFT TO THE RIGHT

Curiously, despite language typically being predominantly processed in the temporal lobe of the brain's left hemisphere, Farrugia's study observed dominant activity in the right hemisphere during the experiment. At first, this seemed like an error. However, further research revealed a fascinating explanation – repetition shifts speech to automatic, right-brain processes, much like chanting, counting, or prayer.

'After repeating a word like "help" thirty times, you're no longer actively constructing speech,' explains Prof. Alexiei Dingli. 'You're simply retrieving it – like a form of linguistic muscle memory.'

This observation may carry more profound implications beyond neuroscience. In many religious practices, such as reciting mantras with Buddhist prayer beads or praying the Christian rosary, repetition plays a central role. This cognitive shift – from the brain's language-planning areas to more automatic processes – may help disassociate the words from their literal meaning. That dissociation, in turn, might enable practitioners to enter a meditative state, engaging the brain in a manner distinct from ordinary speech.

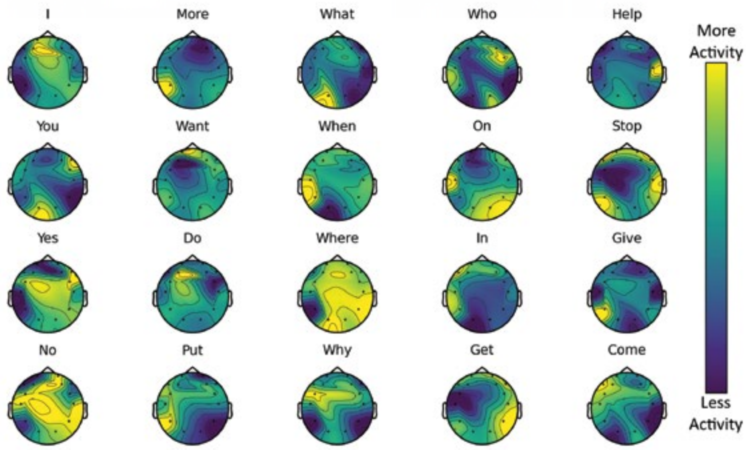

Figure 3: Average brain activation patterns per word, across all participants.
Image courtesy of Jeremy Farrugia

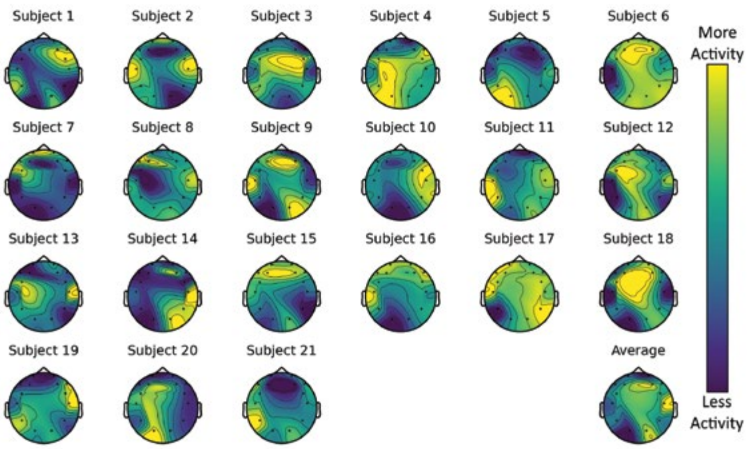

Figure 4: Each subject shows distinct patterns of brain activity during speech imagery.
Image courtesy of Jeremy Farrugia

SPEECH IMAGERY VS. OTHER MENTAL TASKS

To further ground his findings, Farrugia compared his dataset with those from motor imagery and visual evoked potential studies – two well-established paradigms in neuroscience research.

Motor imagery involves imagining physical movements, like clenching a fist or moving a limb, without actually performing them. Visual evoked potential studies, on the other hand, focus on how the brain responds to visual stimuli, such as flashing lights or changing images on a screen.

By contrasting the brain activity observed during speech imagery with these two distinct types of cognitive tasks, Farrugia was able to demonstrate

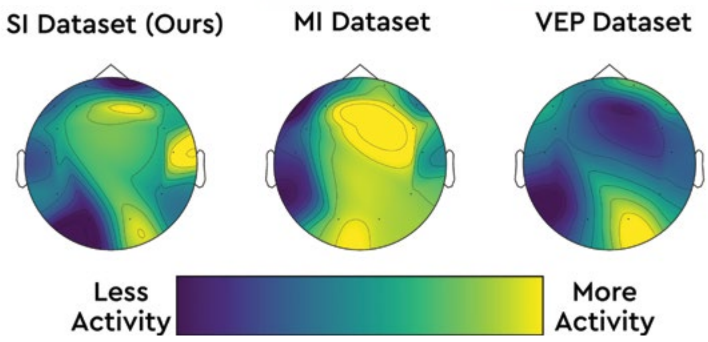

Figure 5: EEG activity from speech imagery, motor imagery, and visual evoked potential studies.
Image courtesy of Jeremy Farrugia

that the patterns elicited by imagined speech are both unique and consistent, confirming that the system was truly detecting the neural signatures of internal speech rather than generic brain noise or unrelated mental processes.

Figure 5 illustrates the differences in EEG activity between these cognitive tasks. 'It validates that speech imagery has its own distinct neural signature,' Farrugia says. 'We're not just picking up noise from blinking or looking at a screen. We're really seeing the act of imagining speech.'

While the research shows promise, Farrugia is careful not to overstate the findings. 'We're a long way from decoding entire sentences,' he says. Scaling up from twenty words to a whole language would require an open-vocabulary approach, where AI can interpolate between similar concepts (big → large → huge). Current models still fall short of capturing the full complexity of natural language.

Instead, Farrugia's system follows a closed-vocabulary model informed by the literature on augmentative and alternative communication. Originally developed for individuals with autism and those who have speech impairments, research in augmentative and alternative communication helps identify the most essential words for basic communication, such as 'want', 'yes', and 'no'.

Jeremy Farrugia and Prof. Alexiei Dingli
Photo by James Moffett

'This isn't about replacing language. It's about giving people a few powerful tools to express basic needs and emotions,' Farrugia says.

THE FUTURE OF THOUGHT-BASED COMMUNICATION

With further development, Farrugia envisions a future where EEG-based speech interfaces seamlessly integrate into everyday smart devices, augmented by technologies such as gaze tracking (read more on this in **Issue 46**, **p. 52**), contextual awareness, and predictive text to enhance expression and nuance.

Imagine someone simply thinking about a word, their gaze confirming intent, and the system responding with fluid, meaningful communication. Take the word 'glass', for instance. It could refer to a drinking vessel or a window pane. On its own, the brain signal does not reveal which one the person intends to mean.

But if their gaze is fixed on a kitchen counter, the system might infer that they are thinking of a cup. If they are looking towards a window, the interpretation shifts. This integration of brain activity and visual context could dramatically enhance the system's ability to resolve ambiguity and respond with accuracy and intent.

That vision may still be years away, but even now in its developing form, this technology does something quietly revolutionary. It offers a pathway – however modest with its twenty basic words – back to connection, agency, and dignity for people with difficulties producing speech, but still have so much to say.

'We're not building minds. We're building bridges,' Farrugia concludes. These bridges may help diminish the gap, moving people from silence to speech, from isolation to interaction, and from imagination to voice.