Brain–machine-interface device translates internal speech into text

For patients affected by speech disorders, brain–machine-interface (BMI) devices could restore their ability to verbally communicate. In this work, we captured neural activity associated with internal speech — words said within the mind with no associated movement or audio output — and translated these cortical signals into text in real time.

This is a summary of:
Wandelt, S. K. et al. Representation of internal speech by single neurons in human supramarginal gyrus. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-024-01867-y (2024).

The project

Speech is a fundamental aspect of human communication, but many neurological conditions (such as amyotrophic lateral sclerosis and cerebral brain lesions) can impair the ability of patients to speak. Brain–machine interfaces (BMIs) offer a promising approach to restoring communication in patients with loss of speech. By capturing neural activity related to speech from the cortex, BMIs can translate neural signatures into words or sentences. Speech decoders that are able to decode vocalized speech[1], attempted speech[2] (speech attempts from participants who have difficulties articulating words) and mimed speech[3,4] (mouth movement without vocalization) have been developed; however, the few decoders developed for internal speech have generally had low decoding accuracy and have not been implemented in real time. The lack of behavioural output in internal speech, and differences in cortical activation pattern and reduced associated neural signatures compared to vocalized speech, make decoding internal speech challenging. However, we hypothesized that internal speech might modulate single unit activity in the posterior parietal cortex (PPC), as we previously demonstrated that neurons in the PPC modulate during vocalized speech[5].

The discovery

We recorded the action potentials of individual neurons located in the supramarginal gyrus (SMG) — a region of the PPC — in two participants with tetraplegia who were chronically implanted with microelectrode arrays. With a high level of spatial and temporal resolution, we captured cortical activity while these participants performed inner speech and vocalized speech. During each trial of an experimental task, the participants were cued with either an auditory cue (sound) or a written word (text). Our dictionary of text cues was composed of eight words, consisting of six lexical words and two pseudowords (words without semantic meaning). Following a brief delay period, the participants were instructed to say the word internally and, after a second delay, to physically vocalize the word.

We observed that SMG neurons are significantly modulated during both internal and vocalized speech production (Fig. 1). After recording the data (offline analysis), we were able to predict which word participant 1 was internally saying with a classification accuracy of up to 72% for the eight words — only slightly less accurate than the decoding of vocalized speech (Fig. 1a). We found strong shared neural representations between internal speech and vocalized speech, which suggests that SMG activity during vocalized speech is more attributable to language-related processes than to motor signals. We developed a real-time decoder that was trained only on internal speech representation, and demonstrated an average classification accuracy of 79% using 16 trials per word for training (online analysis) (Fig. 1b). In the second participant, classification accuracies were lower but still significantly above chance, which suggests that there is variability in the decoding efficacy of internal speech within the SMG depending on implant location. Overall, these results suggest that the PPC is a prime location for implants to enable internal speech decoding.

Future directions

Previous speech BMIs have focused on decoding vocalized, attempted or mimed speech, which depend on muscle or primary motor activity in the cortex. The achievement of decoding an entirely internal process is particularly promising for patients in a locked-in state who are unable to generate motor movements. This technology provides a potential avenue for restoring communication in these patient populations.

A caveat of our study is the small number of participants, as well as the variance in decoding results across different participants. We observed excellent classification results in one participant; however, the results were not as robust in the second. These findings suggest that language-related activity is highly distributed across the PPC, and that there is a need for the use of functional magnetic resonance imaging (fMRI) to guide precise implant location in future studies.

Our proof-of-concept study demonstrates the feasibility of a high-performing BMI for decoding internal speech. The next steps involve scaling the vocabulary to increase the functionality of the device. One promising approach could be to decode phonemes, as this has proven effective in the decoding of attempted speech.

Sarah Kim Wandelt & David Bjånes
California Institute of Technology,
Pasadena, CA, USA.

EXPERT OPINION

"This manuscript presents convincing evidence for the encoding of covert and overt speech by the ensemble activity of single units in the left SMG. Even if this result could not be replicated in the second participant, this study is important because it is to my knowledge the first achievement of a real-time speech brain–computer interface based on single unit recordings in the SMG." **Blaise Yvert, The Grenoble Institute of Neuroscience, Inserm and Univ. Grenoble Alpes, Grenoble, France.**

FIGURE

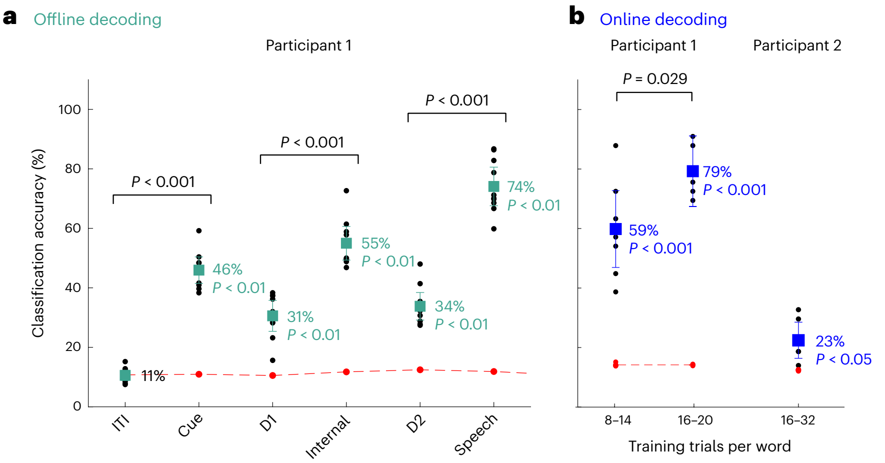

Fig. 1 | Internal and vocalized speech are significantly decodable both offline and online. The experiment consisted of an intertrial interval (ITI), a cue phase, a short delay (D1), a non-verbal internal speech phase (internal), a second delay (D2) and a vocalized speech phase (speech). Word classification was performed individually for each task phase and session day (black dots). Significance was defined as mean accuracies above the 97.5 percentile of a null distribution, which was calculated by shuffling the labels of the words 100 (offline) or 1,000 (online) times. The red dots indicate the mean of the null distribution. **a**, Offline internal and vocalized speech decoding achieved average classification accuracies of 55% and 74%, respectively, in the first participant. **b**, Online, classification accuracies improved significantly with more training trials for participant 1, averaging 79% using 16–20 trials per word. For participant 2, online decoding averaged 23%. © 2024, Wandelt, S. K. et al. CCBY 4.0.

BEHIND THE PAPER

During my PhD in the R. Andersen laboratory at Caltech, which focused on grasp representation in the PPC, we came across an intriguing sequence of observations. While cueing participants with tetraplegia with images of hand positions to perform grasp motor imagery, we observed high neuronal modulation in the SMG during cue presentation. To understand the cognitive behaviour that underlies this cue-evoked neural activity, we introduced audio and written cues as a control. Observing high neuronal modulation during these processes as well, we suspected that the neurons we were studying might be involved in language processes. This unexpected observation shifted our attention from motor imagery to exploring the role of the SMG in vocalized speech. Under the mentorship of D.B., our most significant achievement was successfully demonstrating real-time internal speech decoding. **S.K.W. & D.B.**

REFERENCES

1. Makin, J. G., Moses, D. A. & Chang, E. F. Machine translation of cortical activity to text with an encoder–decoder framework. *Nat. Neurosci.* **23**, 575–582 (2020).
   An article that presents high-accuracy vocalized speech decoding into text using electrocorticogram data.
2. Moses, D. A. et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N. Engl. J. Med.* **385**, 217–227 (2021).
   An article that presents high-accuracy attempted speech decoding into text using electrocorticogram data.
3. Willett, F. R. et al. A high-performance speech neuroprosthesis. *Nature* **620**, 1031–1036 (2023).
   An article that presents high-accuracy attempted and mimed speech decoding into text using multielectrode arrays.
4. Metzger, S. L. et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* **620**, 1037–1046 (2023).
   An article that presents high-accuracy attempted and mimed speech decoding into text, audio and avatar movement using electrocorticogram data.
5. Wandelt, S. K. et al. Decoding grasp and speech signals from the cortical grasp circuit in a tetraplegic human. *Neuron* **110**, 1777–1787 (2022).
   An article that presents grasp and vocalized speech decoding from the SMG.

FROM THE EDITOR

"When I saw this work posted as a preprint, I immediately found it fascinating, both for the novelty of the findings and for the technological advance. This proof-of-concept study of high-performance decoding of internal speech will certainly be of great interest to researchers working to push the boundaries of BMIs and other therapeutic devices for people who can no longer speak." **Giacomo Ariani, Associate Editor, *Nature Human Behaviour*.**