DOI: 10.3217/978-3-99161-050-2-189

Intracortical voice synthesis neuroprosthesis to restore expressive speech to an individual with ALS

M. Wairagkar[1]*, N. S. Card[1], T. Singer-Clark[1], X. Huo[1], C. Iacobacci[1], L. M. Miller[1], L. R. Hochberg[2], D. M. Brandman[1+], S. D. Stavisky[1+]

[1]*University of California, Davis, CA, USA;* [2]*Massachusetts General Hospital/ Brown University/ Harvard Medical School/ Providence VAMC, Boston, MA, USA*

*UC Davis, Davis, CA, USA. E-mail: m.wairagkar@ucdavis.edu, + co-senior authors

Introduction: Brain-computer interfaces (BCIs) have enabled people with speech loss due to neurological disease to communicate by decoding their neural activity into text[1-3]. However, text-based communication falls short of restoring crucial aspects of human speech such as prosody, intonation and voice feedback. In this study, we present a "brain-to-voice" neuroprosthesis that instantaneously synthesizes voice with continuous closed-loop audio feedback by decoding intracortical neural signals of a man with ALS. We also decode paralinguistic features of speech from neural activity, enabling the participant to modulate intonation and pitch of his BCI-synthesized voice.

Materials and Methods: A 45-year-old man 'T15' with ALS and severe dysarthria (unable to speak intelligibly) participated in the BrainGate2 clinical trial. Four microelectrode (Utah) arrays with a total of 256 electrodes placed in the precentral gyrus recorded intracortical signals as T15 attempted to speak. The absence of ground truth speech from T15 posed a major challenge for building an instantaneous voice neuroprosthesis. We overcame this hurdle by generating target speech time-aligned with neural activity as a proxy to T15's intended speech[4]. We developed a real-time causal neural decoding pipeline with a Transformer-based brain-to-voice model to predict spectral and pitch speech features from neural activity. A vocoder synthesized voice every 10 ms and played it aloud (**Fig.1a**). Separate paralinguistic decoders ran simultaneously to detect changes in intonation and pitch. We then modulated the synthesized voice in closed-loop, allowing T15 to ask a question, emphasize words, and "sing" melodies.

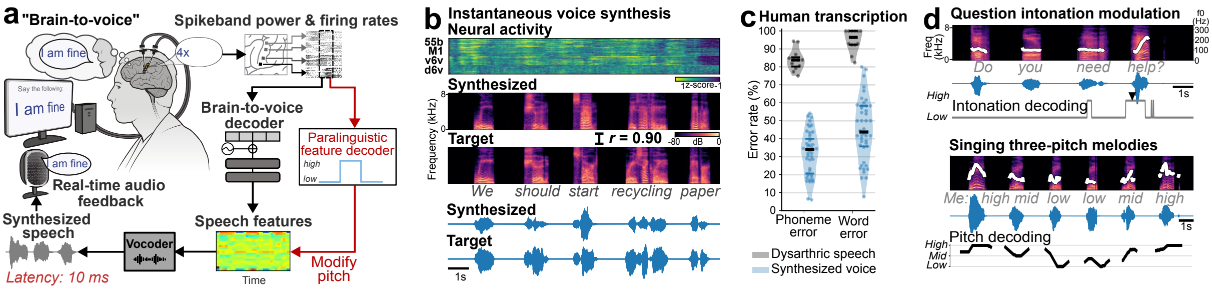

Fig. 1: ***Closed-loop expressive voice synthesis BCI. (a)*** *Brain-to-voice decoder pipeline for instantaneous voice synthesis.* ***(b)*** *An example of causally synthesized speech from neural activity, which matches the target speech with high fidelity.* ***(c)*** *Intelligibility of synthesized speech- median phoneme and word error rates obtained via open transcription by human listeners.* ***(d)*** *Example trials of closed-loop modulation of paralinguistic speech features- question intonation modulation (top) and three-pitch melody (bottom). White trace shows synthesized pitch.*

Results: The brain-to-voice BCI causally synthesized voice nearly-instantaneously with high accuracy (Pearson r=0.89±0.04 with target speech across 40 Mel-frequencies) (**Fig.1b**). Human listeners performed an open transcription of the synthesized voice with a median phoneme error rate of 34.0% and word error rate of 43.8% (in contrast to 96.4% word error rate for T15's residual speech) (**Fig.1c**). Thus, the brain-to-voice BCI vastly improved T15's intelligibility. T15 used this BCI to produce flexible unrestricted vocalizations (including with a neural decoder that approximated his own voice). He could expressively modulate his BCI-voice to ask a question (accuracy: 90.5%), emphasize words (accuracy: 95.7%), and sing short melodies with three different pitch levels (**Fig.1d**).

Conclusion and Significance: These results demonstrate an unprecedented quality of expressive brain-to-voice synthesis, advancing the potential of BCIs to fully restore naturalistic speech.

Acknowledgments and Disclosures: We thank T15 and his family for their invaluable contribution. M. W., D.M.B., S.D.S.: Inventors on IP owned by UC Davis; L.R.H: Consultant to Neuralink, Synchron, Axoft, Precision Neuro, Reach Neuro; D.M.B: Surgical consultant to Paradromics; S.D.S: Inventor on IP owned by Stanford University licensed to Blackrock Neurotech and Neuralink, advisor to Sonera.

References:

[1] Card, NS et al. An accurate and rapidly calibrating speech neuroprosthesis. *New Engl. J. Med.* 391, 609–618 (2024).
[2] Willett, FR et al. A high-performance speech neuroprosthesis. *Nature* 620, 1031–1036 (2023).
[3] Metzger, SL et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* 620, 1037–1046 (2023).
[4] Wairagkar, M et al. Instantaneous voice synthesis neuroprosthesis. *bioRxiv*, 10.1101/2024.08.14.607690 (2024)

Published by Verlag der Technischen Universität Graz