

Commentary: Neuroprosthetic Speech: Pragmatics, Norms, and Self-Fashioning

KAROLA KREITMAIR

While the technical challenges involved in neural speech prosthetics remain formidable, strides are being made to develop devices that allow individuals with expressive aphasia to communicate with the world around them. Such technology would permit patients who have suffered a stroke or who are in a locked-in state to communicate by circumventing the articulatory system and translating thoughts directly into generated speech. Through electrophysiological monitoring, combined with algorithmic processing and computer-generated phonetic production, individuals would be able to make their inner thoughts public.

Before delving into more substantive considerations regarding this technology, it is worth addressing some obvious concerns. One such concern is that this technology may not be able to distinguish between inner thoughts that are intended to remain private and ones that are meant to be externalized. The externalization of thought that is either not intended or ready for externalization would be a considerable burden to individuals employing neuroprosthetic speech (NPS).¹ This would likely render the technology less attractive to users, who might perceive it as a violation of the integrity of their private inner worlds, which healthy individuals take for granted. However, technological solutions may be found to solve this problem. One option may be to identify an electrophysiological signal that users could employ to trigger the externalization of thoughts. For example, they could think of a particularly salient mental image or sentence, similar to how Apple AI uses the word “Siri” or Amazon AI the word “Alexa,” in order for a particular functionality to spring into action. Alternatively, there could be an electrophysiologically detectable dimension of thought that users could learn to modulate depending on whether thoughts are intended to remain internal or be made external. For instance, there could be ‘forcefulness’ or ‘emphaticness’ that could mark certain thoughts for externalization. This would obviously require considerable training on the part of the user and the algorithm, and may be technically out of reach for a long time, but is a possibility that would have the advantage of being a potentially organic way for users to control their output. In addition to a way for users to signal that they wish to externalize a thought, there may need to be a mentally triggerable ‘kill-switch,’ such that if the NPS system is externalizing thought that the user does not intend, she can interrupt it.

What these preliminary considerations illustrate is the complexity of the NPS undertaking. NPS technology must grapple with hundreds of thousands of years of coevolution of brain, language, environment, and culture, in order to produce something that approaches the same dexterity of ordinary speech. Whether NPS technology can be integrated into the existing structures of language use or will require a radical redesign of how, at least in certain circumstances, language operates remains to be seen. In order to build on the work done in this volume by Stephen Rainey et al.² in exploring these new developments, I consider three issues of central importance to the instantiation of NPS technology in the future: pragmatics, norms, and self-fashioning.

Pragmatics

Rainey et al.³ discuss pragmatics as it pertains to the interconnected activities of language production on the part of the speaker and language comprehension on the part of the listener. As Rainey et al. state, Gricean pragmatics holds that a speaker tailors her linguistic output in accordance with the *cooperative principle* such that the semantic content she expresses is relevant to the listener.⁴ Thus, for instance, if Janet asks Patricia, “Do you know what time it is?,” we expect Patricia to answer, “4pm,” or whatever the time may be, rather than, “yes,” or “no.” This is because Paul Grice’s *maxim of relation* demands that Patricia interprets Janet’s question as a request for information about the state of the world, rather than a request for information about whether Patricia possesses a certain mental state. This shows how the semantics of the sentences we utter underdetermine what we *mean* or *intend* by uttering those sentences. This distinction is sometimes referred to as the distinction between ‘what is said’ and ‘what is meant,’ as well as the distinction between the *semantic content* of a sentence versus the *implicature* of the sentence.

In ordinary speech, pragmatics is applied to the semantics of a sentence to determine what is meant by that sentence. This is called *far-side pragmatics*. It is why when one responds to the question, “What do you think of your sister’s new boyfriend after your night out together?” with, “He’s punctual,” the questioner learns that the speaker does not think highly of the new boyfriend, despite the semantic content of her answer not containing any words that expressed that. But there is also *near-side pragmatics*, which is pragmatics that needs to be worked out to even get at the semantics of the sentence. For instance, when Joan makes the utterance, “I’ll take your Queen,” we need Gricean pragmatics to determine whether Joan is saying, “I’ll take your playing card with a ‘Q’ on it,” or, “I’ll take Queen Elizabeth II.” That is because the *maxim of relation*, i.e., the requirement to be relevant, will help us determine whether Joan is talking about a card, perhaps because she is involved in a game of Hearts, or about Queen Elizabeth II, perhaps because she is a captain picking teams for a game of Capture the Flag. This form of pragmatics comes prior to implicature, because it is necessary to determine what the very semantic content of the sentence is.

As with ordinary speech, NPS will involve both far- and near-side pragmatics. However, with NPS, there is a further complexity. Rainey et al.⁵ identify a causal chain from neural activity to verbal output (see *Figure 1*) in NPS. At some point along this preexternalization chain, there will need to be a pragmatic filter to ensure that the verbal output adheres to Gricean maxims. We might call this *super-near-side pragmatics*. In ordinary speech, processing of thought occurs cognitively before it is externalized. But with NPS speech, much of this cognitive work will be replaced by artificial intelligence (AI), including computer algorithms that employ machine learning in order to synthesize speech and generate output.⁶ This is

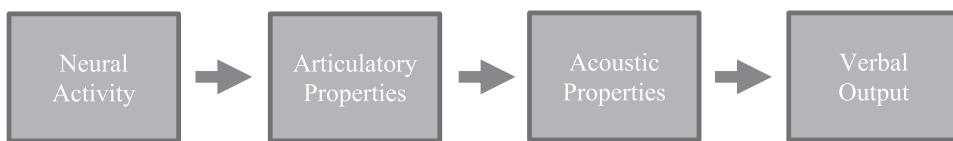


Figure 1. Causal chain from neural activity to verbal output.

necessary because any electrophysiological signal from the cortex is likely to vastly underdetermine the appropriate phonetic output, meaning powerful machine learning will be required to draw on a large corpus of previous instances of speech (from this user or others) to generate what Rainey et al. call 'semantically accurate' utterances. The reason why this sort of prespeech optimization amounts to *Gricean* pragmatics, is because determining the semantically accurate sentence depends in large parts on the context of the utterance and by extension, the class of sentences that are relevant in such a context.

An example may help to illustrate this point. Let us assume, for the sake of argument, that the pragmatic filter falls between the stages of acoustic properties and verbal output. At the stage of acoustic properties, there will be a number of candidate phonemes that can be derived from the articulatory properties of the previous step. This is simply because phonemes exist so close in the phonetic space that one may be mistaken for another.⁷ Thus, in a particular instance, there may be underdetermination as to whether the identified articulatory properties are specifying the phoneme /s/ (the beginning sound of 'sack') or /θ/ (the beginning sound of 'Thursday'). This difference can mean the difference between saying, "Are you sinking?" or, "Are you thinking?." If there is such phonetic underdetermination, the only way for the NPS AI to determine whether the user is trying to say "Are you sinking?" or, "Are you thinking?" is to infer it from the pragmatic context, i.e., apply super-near-side pragmatics. But it is not clear how such AI can do that. AI can glean some information from the surrounding lexical context of the utterance, for example whether the sentence in question follows, "How's the water?" or, "How's the Borges short story?." But there are instances where no such lexical context will be provided, or where it too will underdetermine which is the accurate sentence. Context is frequently nonverbal and depends on an individual's surroundings, her emotions, or who the people are who are present. This is not something, however, that NPS algorithms have access to. All they can glean from is the corpus of prior speech events.

What this points to is just how different NPS is from ordinary speech. While NPS algorithms only have access to internal pragmatic constraints, ordinary speech makes use of a range of external ones. Here, generating utterances is not (just) translating articulatory properties into semantic output. It is optimizing that output so that it coheres dynamically with both internal and external pragmatic constraints. Ordinary speech is embodied, embedded, and enactive, drawing on a rich fabric of nonverbal features to produce meaningful utterances.⁸ Moreover, speaking is a dynamic process, in which sentences are formed as they are spoken. Often the environment in which one is embedded or the action one is performing shapes the sentence one is uttering, sometimes even as one is uttering it. This is, for instance, what happens with gesturing, where speakers replace or precisify words with motion. As Rainey et al. state,⁹ it is also not the case that speakers form fully actualized sentences in their minds before these are uttered. If this were the case, we would never see the phenomenon of a speaker realizing she cannot think of the word for something in the middle of uttering a sentence.

NPS operates on Cartesian principles: speech depends only on properties of the mind plus machine learning. Ordinary speech, on the other hand, is dynamically embedded in the rich, interconnected fabric of the world, including body, environment, and action. This makes it unlikely that NPS will ever reach the complexity and dexterity of ordinary speech. What it does suggest is that new norms will be necessary to elevate NPS speech to the level of robust communication.

Norms

Speech is not just the production of semantically meaningful utterances. It is the participation in a complex system of human norms and practices that permit communication and the performance of actions. Speech involves *speech acts*; we can make promises, give consent, refuse, declare two people married, and lie. The system of norms in place that enables such speech acts relies on certain assumptions about the mechanics of speech. For example, usually we hold someone accountable for what she says. Of course, we allow that people sometimes misspeak or muddle up their words. But if Olive says, "Teresa's new haircut makes her look like a meerkat," Teresa might rightly feel insulted, even if Olive later states, "I didn't mean that." It is part of the norms of speaking that we hold Olive accountable for her meerkat statement, and even deduce from the statement that she is rude. Of course, if Olive explains that she actually intended her statement as a compliment, because meerkats, in her opinion, are creatures of great grace, we might revise our assessment of her. But note that we still hold her accountable for having said what she said. We do not allow her to simply disown the statement as though she had not uttered it.

With NPS things are different. Rainey et al. identify the useful concept of *authenticity of speech*. NPS speech is authentic, Rainey et al suggest, if the speaker reflectively endorses the utterance as being 'what she intended' in that particular instance. Thus, if an NPS user utters a sentence, but then does not endorse it, likely because the technology failed and did not actually express the *semantically accurate* sentence, then we should not hold her accountable for that utterance. This shows how the norms of ordinary speech and those of NPS come apart. Presumably, in the context of NPS, we must be more forgiving than in the Olive-and-Teresa context. At the same time, however, we do not want to be forgiving to the point that acts such as lying become impossible for NPS users, if they are, for example, permitted to perpetually retroactively withdraw endorsement from their statements. This illustrates why it is that we will require a new set of norms that will have to be established to accommodate the new realities of NPS, allowing it to fulfil the same or similar functions as ordinary speech.

Such a shift in norms suggests that NPS will not become a *restoration* of lost speech, but rather a *replacement*.¹⁰ But that does not mean that we cannot develop sophisticated norms around NPS that allow users the full set of communicative and speech act abilities. Thus, while *speech* may only be replaced for NPS users, *communication* may well be restored. Particular care will have to be given to the intermediary phases, in which NPS will not fit within the norms of ordinary speech but will not yet have established norms of its own. As with halfway technologies, where technology traps individuals in a suboptimal state between sickness and recovery, we must be wary of trapping NPS users in a state of suboptimal norms and practices, where full communication is technically achievable, but adherence to existing norms of ordinary speech prevents it.

Self-Fashioning

A further issue to consider is how the production and perception of speech will affect NPS users. As noted by Rainey et al., ordinary speech involves a feedback effect, in which we learn what we believe, in part, from what we say. We often try

out ideas by uttering them and then retroactively agree with ourselves or not. Moreover, as stated above, often it is not even the case that a sentence is fully formed in the mind before it is uttered, but rather that it is formed as it is being uttered. In these ways, we use speech to learn about ourselves, construct our identities, and *fashion ourselves*.

With NPS, the speech that is generated relies on machine learning, allowing algorithms to produce utterances that they learn from a vast corpus of prior instances of speech. As already stated, it is likely that ultimately this AI will rely not only on the NPS user's own prior utterances, but will feed off of big data from a plethora of sources. This will give it much more completeness and dexterity. Big data and AI, however, are notoriously vulnerable to algorithmic bias, including racism and sexism,¹¹ as can be seen in the case of Microsoft's chatbot Tay, who went from tabula rasa to racist, sexist Holocaust-denier within 12 hours of big-data-fueled algorithmic learning.¹²

The combination of the self-fashioning function of speech with the algorithmic bias of AI-generated output should give us pause. We can imagine that an individual whose NPS is infiltrated with racist or sexist sentiments may find the experience alienating and might cause her to question her views. Or worse, an individual who finds herself uttering such sentiments may gradually even come to accept them as her own, thus pushing her toward prejudiced positions. Of course, the extent to which this phenomenon will be a concern is largely a question of how effective developers will be at designing unbiased AI,¹³ but there will always be a tradeoff between quantity of data, which facilitates rich speech generation, and selectiveness of data, which can curtail bias.

In conclusion, while the prospect of NPS is exciting and promises empowerment to the disenfranchised, there remain deep questions as to how it can best be implemented in ways that allow users the fullest range of communicative powers while minimizing their exposure to harmful features of the technology. Once we have addressed such questions and ensured that this technology works to restore communication to those who need it, we may even consider further uses for NPS. Rather than using this technology as treatment, we may wonder what forms of enhancement it might enable. For example, NPS could allow individuals, both healthy and not, to communicate without speaking at all, transmitting thoughts to devices anywhere in the world. Such applications are already of interest to the military.¹⁴ Alternatively, this technology might eventually be capable of capturing nonconscious thought and translating that into speech, possibly allowing for applications in the legal context. For now, however, we should focus on addressing the technical and ethical challenges that remain for the treatment application of this technology. Restoring communicative abilities to those who cannot speak would constitute a great benefit to such individuals and their families, and is certainly worthy of pursuit.

Notes

1. Rainey S, Maslen H, Mégevand P, Arnal LH, Fourmeret E, Yvert B. Neuroprosthetic speech: The ethical significance of accuracy, control, and pragmatics. *Cambridge Quarterly of Healthcare Ethics* 2019;28(4): 657–70, at 663.
2. See note 1, Rainey et al. 2019.
3. See note 1, Rainey et al. 2019, at 666.
4. Grice P. Logic and conversation. In Cole P, Morgan J, eds. *Syntax and Semantics 3: Speech Acts*, New York: Academic Press; 1975, at 41–58.

5. See note 1, Rainey et al. 2019, at 661.
6. Akbari H, Khalighinejad B, Herrero JL, Mehta AD, Mesgarani N. Towards reconstructing intelligible speech from the human auditory cortex. *Scientific Reports* 2019;9:874.
7. Lindblom B, Maddieson I. Phonetic universals in consonant systems. In Li C, and Hyman L, eds., *Language, Speech and Mind*. London: Routledge;1988:62–78.
8. Clark A. Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences* 2006; 10:370–4.
9. See note 1, Rainey et al. 2019, at 662.
10. This distinction is introduced by Rainey et al. (see note 1) at 662.
11. Zou J, Schiebinger A. AI can be sexist and racist—it’s time to make it fair. *Nature* 2018; 559:324–6.
12. Garcia M. Racist in the machine: The disturbing implications of algorithmic bias. *World Policy Journal* 2016;33(4):111–7.
13. Teich P. Artificial intelligence can reinforce bias, cloud giants announce tools for ai fairness. *Forbes* September 24, 2108; available at <https://www.forbes.com/sites/paulteich/2018/09/24/artificial-intelligence-can-reinforce-bias-cloud-giants-announce-tools-for-ai-fairness/#7dd5e7059d21> (last accessed 15 May 2019).
14. Drummond K. Pentagon preps soldiers for telepathy push. *Wired* May 14th, 2009; available at <https://www.wired.com/2009/05/pentagon-preps-soldier-telepathy-push/> (last accessed 15 May 2019).