Article

Whole-brain dynamics of articulatory, acoustic and semantic speech representations

Check for updates

Maxime Verwoert [1] ✉, Joaquín Amigó-Vega [1,2], Yingming Gao [3], Maarten C. Ottenhoff[1], Pieter L. Kubben[1,4] & Christian Herff [1] ✉

Speech production is a complex process that traverses several representations, from the meaning of spoken words (semantic), through the movement of articulatory muscles (articulatory) and, ultimately, to the produced audio waveform (acoustic). In this study, we identify how these different representations of speech are spatially and temporally distributed throughout the depth of the brain. Intracranial neural data is recorded from 15 participants, across 1647 electrode contacts, while overtly speaking 100 unique words. We find a bilateral spatial distribution for all three representations, with a more widespread and temporally dynamic distribution in the left compared to the right hemisphere. The articulatory and acoustic representations share a similar spatial distribution surrounding the Sylvian fissure, while the semantic representation is more widely distributed across the brain in a mostly distinct network. These results highlight the distributed nature of the speech production neural process and the potential of non-motor representations for speech brain-computer interfaces.

Neuroimaging studies have shed light on various components of language representation in the brain for many years. Recently, the representation of semantic information has been more comprehensively mapped thanks to the aid of large language models[1-3]. Conversely, the definitions of Broca's area, traditionally associated with speech production, and Wernicke's area, traditionally linked to language comprehension, have been up for debate due to emerging evidence that challenges their classical boundaries and functions[4-9]. These findings, along with results from more data-driven analyses and neural data with high temporal resolution from deeper brain areas, suggest that the canonical language map[10,11] may be changing.

Different components of speech have been mapped to specific, but also many shared regions in the brain. The left inferior frontal gyrus (IFG), left precentral gyrus, left superior temporal gyrus (STG), left fusiform gyrus, right STG, and bilateral supplementary motor area (SMA) have been identified as "core" cortical language regions[10]. Speech production studies have shown strong left-lateralization, while speech perception appears to be more bilateral[1,10]. More specifically, the ventral "comprehension" stream, involving auditory and temporal cortices, is thought to be organized bilaterally, while the dorsal "articulatory" stream, involving mostly frontal and motor cortices, is thought to be strongly left-hemisphere dominant[11]. However, more recent evidence suggests that even the articulatory

representation of speech production occurs bilaterally[12,13], as well as other speech components[4]. Moreover, speech studies have largely overlooked potentially important deeper cortical and subcortical areas such as the insula[14], basal ganglia[15,16], thalamus[17], and hippocampus[18].

Traditionally, studies investigating speech, or language in general, have employed meticulously designed hypothesis-driven approaches to focus on isolated aspects of language, often in speech perception. Advances in Natural Language Processing and Deep Neural Networks now enable data-driven approaches that can more efficiently extract complex features of acoustic speech, especially in naturalistic settings[1,2,19]. These methods complement traditional approaches by providing deeper insights into the complex dynamics of spoken language and have also facilitated progress in the development of speech neuroprostheses, or, brain-computer interfaces (BCIs). Speech BCIs aim to restore communication in individuals who have lost their ability to speak by decoding neural signals and transforming it into a reconstruction or classification of speech features[8,20-25].

One of the approaches to decoding speech is through articulatory features. These features represent the movements of articulatory organs (such as the larynx, tongue, and lips). Previous studies have mapped the neural representation of speech articulators to areas such as the ventral sensorimotor cortex[26-28] and the dorsal precentral gyrus[29]. Another

[1]Department of Neurosurgery, Mental Health and Neuroscience Research Institute, Maastricht University, Maastricht, The Netherlands. [2]Computer Science Department, Gran Sasso Science Institute, L'Aquila, Italy. [3]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China. [4]Academic Center for Epileptology, Kempenhaeghe/Maastricht University Medical Center, Maastricht, The Netherlands. ✉e-mail: m.verwoert@maastrichtuniversity.nl; c.herff@maastrichtuniversity.nl

approach is at the acoustic level, the sounds that are produced through movement of the articulatory muscles. Spectrotemporal features can be extracted from recorded speech audio to obtain an acoustic representation of speech. These features have been used to decode and synthesize speech from the inferior frontal and motor cortex[30], the STG[31] and from a combination of multiple distributed regions[32,33]. The articulatory and acoustic representations have both, separately, already led to successful BCI communication in an individual with paralysis due to a brain-stem stroke using a device on the sensorimotor cortex[34]. Combining these two approaches has the potential to further improve BCI performance[35,36]. While these studies have decoded speech from motor or acoustic representations of speech, individuals with diseases that cause impairments in the creation of motor representations, such as with aphasia, or in the generation of audible speech, such as with locked-in syndrome, might benefit from the decoding of more semantic representations of speech. The semantic representation captures the underlying meaning of words or sentences. Through the aid of large language models, it has been shown that we are able to capture such semantic representations as well, using both functional magnetic resonance imaging[1,3] (fMRI) and intracranial electrophysiology[2,19,37]. These studies found semantic representations spread widely across the brain. The question remains whether these different speech representations share a similar or distinct network of brain regions.

While exploring these different speech representations is crucial, understanding the strengths and limitations of the methods used to study them is equally important. fMRI is a useful method to study language representations as it is non-invasive and can capture whole brain dynamics, however, it suffers from various sources of noise (e.g., head motion, loud scanner noise) and is particularly sensitive to the motion artifacts induced by speaking. Furthermore, fMRI is lacking in temporal resolution, which is crucial for certain aspects of speech. Intracranial electrophysiology, on the other hand, is characterized by its high temporal resolution, although the spatial resolution depends on the specific electrode device and implantation planning. Stereo-electroencephalography (sEEG), specifically, is routinely used for functional mapping and to determine the seizure onset zone in epileptic patients[38]. The sEEG electrodes are shafts with individual contacts that reach subcortical areas and are typically implanted in many different locations at once. In the current study, we analyze sEEG data from multiple participants with distinct and overlapping electrode locations which allows for a high overall spatial resolution, a brain-wide coverage, together with high temporal resolution. This is ideally suited to study the widespread process of speech production.

Language thus appears to be encoded in a wide variety of brain regions, although the methods used to study it also vary in terms of speech features,

analyses and technology. We aimed to provide an overview of how multiple speech representations are neurally distributed throughout the brain, using similar analyses on the same high-resolution data. We focus on the three representations of the speech production process that are most commonly used for the purpose of speech BCIs[25], namely articulatory, acoustic and semantic representations. We find that all three speech representations are present in both hemispheres, although more spatially spread and temporally dynamic in the left hemisphere. The semantic representation is more widespread than the acoustic and articulatory representations, which are predominantly overlapping. The results shed further light on the overall speech production network, including deeper regions, and are discussed in the context of speech neuroprostheses.

Results

We utilized data-driven approaches to extract and compare multiple speech features from the same exact data in a continuous manner. We employed an AAI Recurrent Neural Network[39] to estimate the movements and positions of articulatory organs during speech production to generate an articulatory representation, a Fourier transform to capture auditory properties of spoken speech as an acoustic representation and word embeddings from a Word2Vec auto-encoder[40] for a semantic representation (Fig. 1). Next, to examine the neural distribution of these speech representations, we correlated the neural activity of each electrode channel with the extracted articulatory and acoustic features and use a linear regression encoding model to estimate neural activity across speech trials from the extracted semantic embeddings. We then identified which channels and points in time encoded semantic information and which correlated with articulatory or acoustic features significantly above chance level. Significant channels were identified by estimating a distribution of chance correlation coefficients through random permutations and circular shifts, setting a strict significance threshold ($\alpha = 0.05$) using max-t correction, and determining channels as significant if their correlations exceeded this threshold in any time-frame for each of the representations separately. These relatively simple analyses enable a clear interpretation of the spatial and temporal dynamics of these speech representations, focusing on uncovering overlapping and distinct anatomical patterns involved in speech production rather than optimizing correlational strength or predictive power.

Differential distribution of speech representations

For all three speech representations (articulatory, acoustic and semantic), there are significant channels present in both hemispheres (Fig. 2A). Whereas the total number of significant channels is similar between the three representations (Fig. 2B), there are relative differences between the two
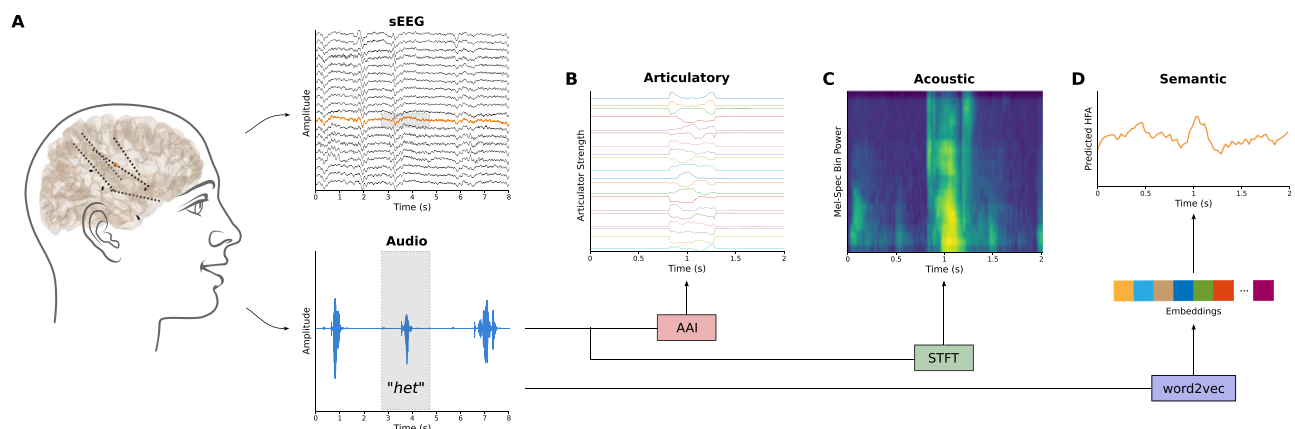

Fig. 1 | Extraction of the three speech representations. A The dataset contains synchronized sEEG and audio data during speech production. Highlighted is one single channel's neural signal (also in orange) and the audio signal of one word. **B** The articulatory trajectories ($N = 20$) were estimated using an acoustic-to-articulatory inversion (AAI) LSTM neural network. **C** The acoustic representation in the form of mel-spectrograms was calculated using short-time Fourier transform (STFT) with $N = 23$ filterbanks. **D** The semantic embeddings (160 dimensional) were extracted for each word using a word2vec model and a linear regression was subsequently trained to predict the neural timeseries. All data presented in this figure were extracted from sub-06.
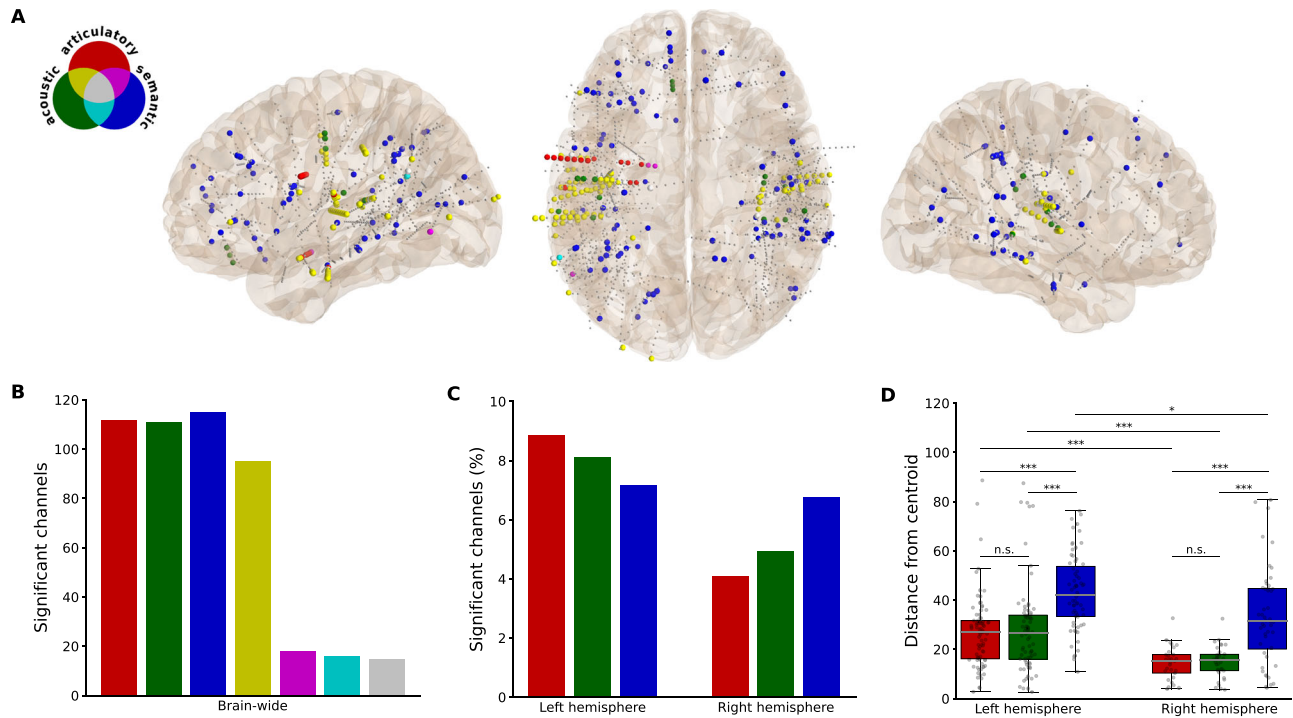
Fig. 2 | Differential spatial distribution of speech representations. A Locations of channels significantly correlated with each representation uniquely (red—articulatory, green—acoustic, blue—semantic) or with overlapping representations (colors are disseminated in the Venn diagram). Smaller gray dots represent non-significant channels. **B** Total number of significant channels for each representation and for the overlap between representations. **C** Percentage of significant channels for each representation, separated by hemisphere. **D** The Euclidean distance between each significant channel and a centroid for each representation, calculated per hemisphere and presented in boxplots. In the left hemisphere, there were $n = 83$ articulatory, $n = 76$ acoustic and $n = 67$ semantic significant channels. In the right hemisphere, there were $n = 29$ articulatory, $n = 35$ acoustic and $n = 48$ semantic significant channels. The horizontal gray line represents the median distance in each category. The error bars represent the range of distances within 1.5 times the interquartile range from the first and third quartiles. Significant differences between categories were tested using two-sided Mann–Whitney U tests (***$p < 0.001$, *$p < 0.05$, Bonferroni-corrected).

hemispheres (Fig. 2C). Since there were a greater number of channels implanted in the left hemisphere (937) than in the right hemisphere (710), we looked at the percentage of significant channels within the hemispheres. Percentage-wise, there are more significant channels in the left hemisphere for the articulatory and acoustic representations. While the articulatory and acoustic representations are largely overlapping in their spatial distribution (74.22% out of all articulatory and/or acoustic channels), there appears to be a stronger tuning for articulatory features relative to acoustic features in the left hemisphere and the opposite, a stronger tuning for acoustic features relative to articulatory features, in the right hemisphere (Fig. 2C). The articulatory-acoustic overlap can be seen in yellow in Fig. 2A, B. The semantic representation, on the other hand, has little overlap with the other two representations (16.52% overlap out of all semantic channels).

To quantify the spatial distribution of the results (Fig. 2D), we calculated distances between the location of significant channels and their centroid within representation and hemispheres. A Kruskal–Wallis test was conducted to compare the distances between the articulatory, acoustic, and semantic representations in the left and right hemispheres. There was a significant main effect between the groups in both the left ($H(2) = 48.42$, $p < 0.001$, $\eta^2 = 0.21$) and the right ($H(2) = 31.16$, $p < 0.001$, $\eta^2 = 0.27$) hemisphere. Post-hoc two-sided Mann–Whitney U tests revealed significant differences between articulatory and semantic distances (Left: $U = 1095$, $p < 0.00$, $r = 0.52$; Right: $U = 264$, $p < 0.001$, $r = 0.52$) and between acoustic and semantic distances (Left: $U = 1108$, $p < 0.001$, $r = 0.49$; Right: $U = 323$, $p < 0.001$, $r = 0.52$), but not between articulatory and acoustic distances (Left: $U = 3189$, $p = 1.0$, $r = 0.00$; Right: $U = 487$, $p = 1.0$, $r = 0.03$). Another three two-sided Mann–Whitney U tests were performed to compare the distances between hemispheres within the representations. These tests revealed significantly greater distances in the left hemisphere for all three representations (Articulatory: $U = 1889$, $p < 0.001$, $r = 0.43$; Acoustic:

$U = 2019$, $p < 0.001$, $r = 0.42$; Semantic: $U = 2139$, $p = 0.02$, $r = 0.28$). All post-hoc tests were strictly corrected using the Bonferroni correction for multiple comparisons ($N = 9$).

These results indicate that the semantic representation is more widely distributed across the brain than the other two representations, in both hemispheres. Importantly, this holds true within individuals (8/10 participants with significant channels in all three representations). The articulatory-selective channels were spread farther than the semantic ones for only one individual and the acoustic-selective channels for two individuals, both within the left hemisphere (Supplementary Fig. 2). The results also indicate that all representations are more widely distributed across the left than across the right hemisphere. This difference could potentially be due to the sampling, since the electrode coverage is not exactly mirrored between the hemispheres (along with more contacts in the left hemisphere overall). Therefore, we additionally calculated the distance between all recorded channels and their centroid in each hemisphere. This analysis did not reveal a statistically significant difference between the distances of the left (38.47 ± 16.69) and the right (37.25 ± 17.57) hemisphere overall ($U = 350849$, $p = 0.06$). The difference in the distribution of the representations between hemispheres is therefore not likely due to the electrode coverage alone. Together, these findings indicate distinct and widespread processing of the meaning of words.

Anatomical contributions to speech representations

Since the overall distribution covers many different anatomic areas, including contacts located in white matter tracts (80 out of the 224 (35.7%) overall significant channels), we next looked at how specific regions-of-interest (ROIs) may contribute differently to the speech representations (Fig. 3). Here, we provide a qualitative examination of the data to describe general patterns within and across ROIs. We find the most consistent
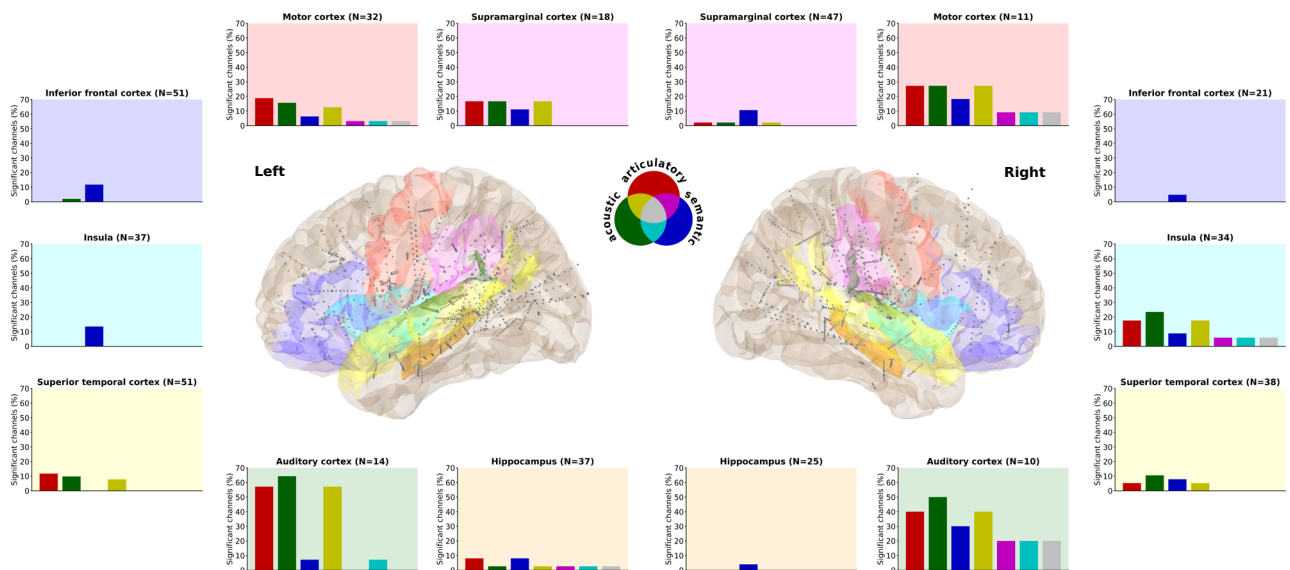
Fig. 3 | Anatomical contributions to speech representations. Seven language-related anatomical areas within each hemisphere are highlighted. The percentage of significant channels (out of the total number of channels in that area) for each representation and their overlap is displayed in the corresponding bar plot. Background colors in the bar plots match and indicate the regions in the brain plots where all channels are depicted in grey, the significant ones are slightly larger in diameter.

contributions from auditory and motor regions, with a preference for acoustic features in the auditory cortex and a slight preference for articulatory features in the motor cortex. Interestingly, these regions were strong contributors in both hemispheres. However, the semantic representation appears to be stronger in the right hemisphere in these primary regions specifically.

The semantic representations are present in nearly all the ROIs, although to a relatively small degree. It is the only representation we find in the IFG, the left insula and the right hippocampus. The left hippocampus also has a small articulatory and even smaller acoustic representation. The right insula, on the other hand, has stronger articulatory and acoustic representations. There are weak representations in the superior temporal cortices, relative to the auditory regions. Finally, the supramarginal cortex appears to be involved specifically in the left hemisphere.

Temporal dynamics of speech representations
To explore the brain-wide temporal dynamics, including channels that are not represented in ROIs, we evaluated significant channels per time-point or time-frame for a qualitative temporal analysis and grouped these results into three larger time segments of equal length for a qualitative spatial analysis. The time-frames represent neural features that were temporally shifted to be prior to speech alignment (−200 until −50 ms), surrounding speech alignment (−50 until +50 ms) and after speech alignment (+50 until +200 ms) for the articulatory and acoustic representations. For the semantic representation, the two-second speech trial time-points were equally divided. Since speech onset latency is roughly between 500 and 600 ms[41], the middle segment is, similar to the articulatory and acoustic middle segment (−50 to +50 ms), expected to involve most of the word production. Spatially, we find a relatively stable pattern in the right hemisphere for the articulatory-acoustic representations and there appears to be a posterior to anterior pattern in the left hemisphere for the semantic representation (Fig. 4A). Notably, there is one articulatory-selective electrode shaft in the left precentral or motor cortex specifically prior to and after speech onset and one acoustic-selective electrode shaft in the left postcentral or sensory cortex specifically after speech onset.

The amount of significant channels dynamically changes across articulatory and acoustic time-frames, especially in the left hemisphere (Fig. 4B). In the left hemisphere, most significant channels were found 200 ms prior to speech onset, after which the number quickly declines to a level similar to that of the right hemisphere at around +50 ms. In contrast, in the right hemisphere, the number of significant channels is relatively stable until it drops after +50 ms. Notably, in the left hemisphere, the peak for the acoustic representation is one time-frame later (−150 ms) than for the articulatory representation (−200 ms) and there is a switch around 50 ms prior to speech onset where the acoustic representation becomes more prominent than the articulatory representation. We find similar patterns in most of the ROIs (Supplementary Fig. 3), although the right hemisphere insula stands out. Here we find a steep increase in the articulatory- and acoustic-selective channels until around speech onset, which drops more slowly afterwards for the acoustic representation (Fig. 4B).

The largest number of significant channels is roughly around the middle of the speech trial for the semantic representation (Fig. 4C). The channels fluctuate across time and locations and between hemispheres. The spatial results, as mentioned before, indicate that the locations fluctuate from more posterior to more anterior in the left hemisphere, although this is not apparent in the right hemisphere (Fig. 4A). Finally, for appropriate comparison we also highlight the right insula for the semantic representation, which is also involved relatively late during or after speech production (Fig. 4C).

Discussion
In this work, we explored the brain-wide spatial distribution and temporal dynamics of articulatory, acoustic and semantic speech representations. The results indicate more widespread and temporally dynamic contributions from the left compared to the right hemisphere, in all representations. This is in line with the broad literature on the left-hemisphere dominance in speech production[10,11]. However, we do note strong contributions from the right hemisphere as well, especially from the primary auditory and motor regions as well as from the right insula. Thus, the results also align with the notion of bilateral involvement in speech production[12,13]. The temporal dynamic results indicate that the majority of articulatory- and/or acoustic-selective channels are active well before speech onset. This finding is especially pronounced in the left hemisphere, even in auditory regions. This could reflect the coordination between different areas for speech preparation[10,11,42], which has similarly been found for speech comprehension[43]. Interestingly, the neural patterns were temporally stable in the right hemisphere. This activity could potentially reflect feedback control, for which the right hemisphere has previously been implicated[42].
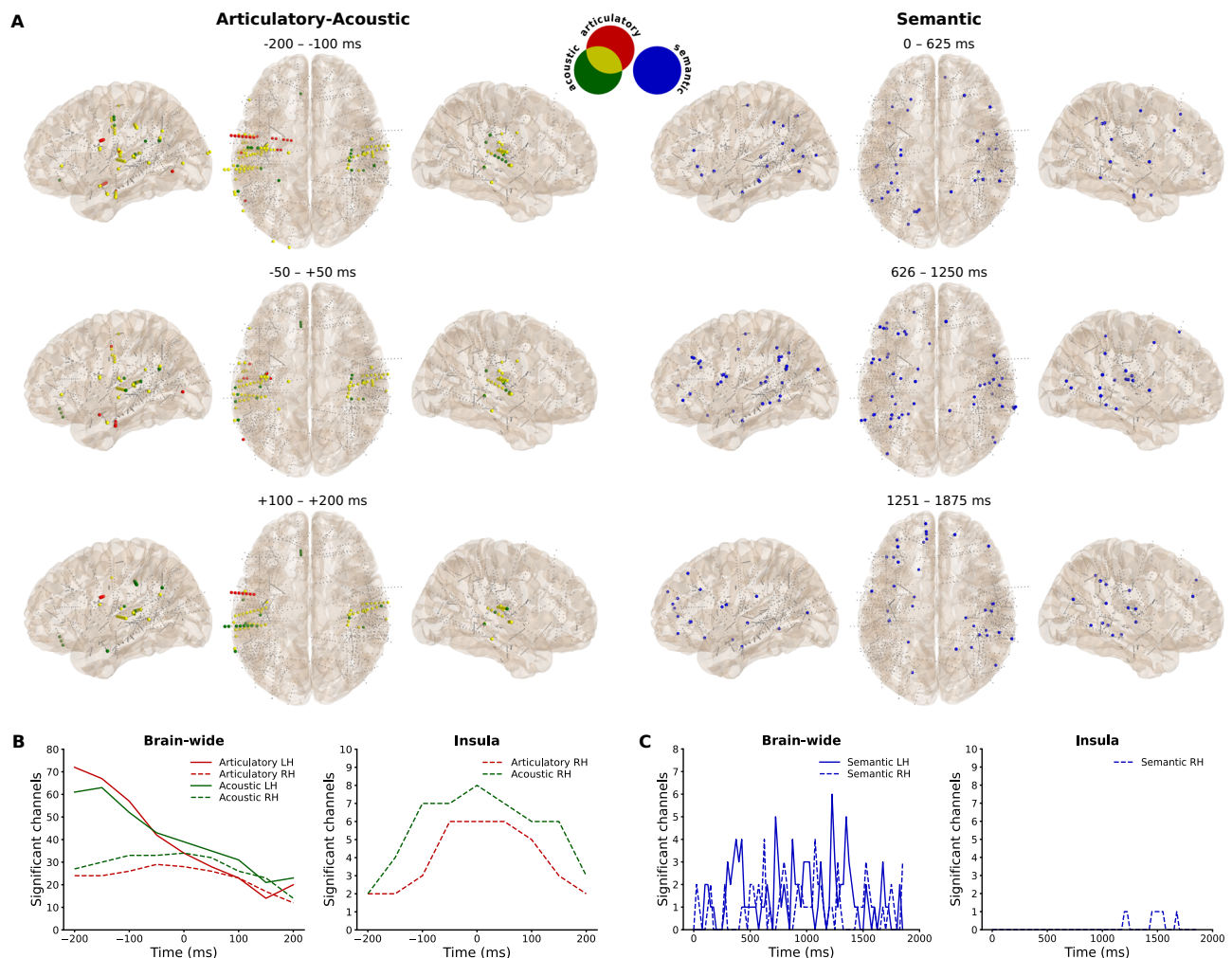
Fig. 4 | Temporal dynamics of speech representations. A Significant channel locations with colors representing the representation and the articulatory-acoustic overlap, divided into equal time segments. Smaller gray dots represent non-significant channels. **B** Number of significant articulatory- and acoustic-selective channels across time-frames from all (brain-wide) contacts and from the right insula. **C** Number of significant semantic-selective channels across time-points from all (brain-wide) contacts and from the right insula.

The articulatory and acoustic representations share a similar neural distribution, most prominently within and surrounding the Sylvian fissure. A high degree of overlap between these features can be expected due to their natural correlation, and has similarly been found in earlier studies using fMRI[44], sEEG[13] and with a chronic implant on the somatosensory cortex[34]. The semantic representation, on the other hand, has a widespread neural distribution, which is mostly distinct from the articulatory-acoustic network, and consistent with a previous fMRI study[44] on speech comprehension. This indicates that semantics may indeed operate in distinct networks. However, it is worth noting that there was more overlap between the representations in the right hemisphere, especially in the auditory cortex.

The semantic representation encompassed temporal, parietal and frontal regions. This distribution is in line with previous studies investigating semantic processing in speech comprehension[1–3,44], indicating that the same semantic network is engaged in both speech comprehension and production. A key difference lies in the semantic context, whereas earlier studies involved listening to sentences or stories, our work focused solely on speaking isolated words without a broader semantic context. We therefore show that it is possible to uncover semantic processing in speech production, even with limited semantic context, a small amount of data, and without an explicit semantic task. The distributed results align with the 'spokes' part of the hub-and-spoke theory of semantic cognition[45]. They position the mediating 'hub' at the bilateral anterior temporal lobe, which unfortunately was not sampled sufficiently in our recordings to comment on

further. In contrast to the articulatory-acoustic representations, there was a similar relative number of semantic-selective channels between the hemispheres, although we noted a more widespread and dynamic temporal pattern in the left hemisphere, moving from early posterior to later anterior regions. This may be explained by previous findings that the left hemisphere primarily represents language-mediated semantic knowledge, whereas the right hemisphere encodes more perceptually based sensory-motor conceptual representations[46]. This might also explain the relatively larger contribution in the motor and auditory regions in the right compared to the left hemisphere. Future research may explore whether the temporal pattern could additionally vary depending on the input domain (visual vs. auditory) and its corresponding processing hierarchy.

The inferior frontal cortex, also termed Broca's area, was almost exclusively and weakly (to a similar degree as other regions-of-interest) correlated with the semantic representation. This region was previously found to be specifically active for overt word generation and not for simple (meaningless) overt sound reading[47] and may be more involved in the higher-order executive control of semantic knowledge than its representation per se[45]. More involvement might be observed when more semantic context needs to be processed, i.e., in word or sentence generation. The lack of clear articulatory and acoustic representations in this region further shows its indirect connection to the low-level articulatory and acoustic features of speech, limiting its use for an articulatory-based speech neuroprosthesis[8].

The superior temporal cortex showed low selectivity for any speech representation, in contrast to very high selectivity in the auditory regions. Whereas the STG is known to play a key role in speech comprehension[48], its role in speech production is less clear. The posterior portions of the STG and superior temporal sulcus (STS) have been more specifically implicated in speech production[42]. However, this portion may be masked by the more anterior regions of our relatively large region-of-interest (ROI) and was partially (the 'planum temporale') included in the auditory cortex ROI. On the other hand, the STG has been found to entail intermediate acoustic-to-semantic sound representations that neither acoustic nor semantic models could account for[49]. This may similarly be the case for the opposite direction (semantic-to-acoustic representations), which could explain the relatively weak correlations with our representations. Another study found a special role of the right posterior STS (which may be closer to our auditory cortex ROI) in matching auditory expectations with spectrotemporal processing from auditory feedback during speech production and propose its involvement in the internal representation of speech[50]. For a speech neuroprosthesis, the question remains whether the auditory cortex remains involved when speech is only imagined instead of overtly produced. The temporal dynamics results indicate that even these regions, especially in the left hemisphere, are active well before speech onset. Other studies have found speech-related information in auditory regions even with imagined speech[50–52]. Together, these results suggest that the activity may indeed persist even when no sound is produced, which is important for the development of speech neuroprostheses for individuals who are unable to produce sounds.

Interestingly, the right insula showed relatively strong involvement in the speech production process, particularly for the acoustic and articulatory representations. While this area seems an interesting candidate for a speech neuroprosthesis[53], the temporal dynamics, revealing most involvement around speech onset rather than before, suggest that it may not be a good predictor of speech. The insula is an integration hub linking many different cortical and subcortical regions and its activity may reflect any of a large number of functions[54]. However, since distinct regions of the insula have been shown to play different roles in speech and language processing[14,55], its potential may still need to be explored in further detail.

The hippocampus has previously been shown to be involved in language processing[18], more specifically, in navigating semantic representations[56]. However, we do not see a large contribution from the hippocampi in our results. This may, in part, be due to the fact that we did not explicitly probe semantic relations or memory, as only isolated words were produced as opposed to sentences or longer narratives. Additionally, our participants may have abnormal hippocampal activity due to their epilepsy[57]. Another recent sEEG study did find a large contribution from the hippocampus in decoding the production of vowels[58], however, this was in combination with micro-wires. They did not find good decoding on the sEEG high-frequency local field potential on its own, which is similar to our work. Alternatively, previous findings linking the hippocampus with language were found in low-frequency activity[59], which we did not include in our current work.

The supramarginal cortex was particularly involved in left hemisphere. Similarly, stimulation of this region was recently found to cause speech arrest with a short latency[60]. Both words and pseudowords were also recently found to be decodable with both overt and covert (imagined) speech using chronic implants on the left supramarginal gyrus[24]. This region is thought to potentially encode phonetics, however, they also found it to be activated during a visual imagination strategy and, separately, for decoding hand gestures. We also found all three speech representations in this region, together suggesting that it may serve as a potential area for a flexible and multi-purpose neuroprosthesis. However, they also noted a large difference between their two subjects in terms of decoding and anatomical folding of the region, suggesting the importance of individually localizing the specific area related to speech. While we did not find much representation other than semantics in the right hemisphere, it has to be noted that the specific electrode locations within the region, as well as in any other ROI, may differ between the hemispheres.

Finally, although exploring the white matter goes beyond the scope of the current paper, a large percentage (35.7%) of significant channels were located in white matter. This suggests that we may be able to tap into these deeper communication highways directly.

Conclusion

The articulatory and acoustic speech representations share a similar bilateral neural distribution surrounding the Sylvian fissure, most prominently in the left deeper auditory regions prior to speech onset. The semantic representation is spatially distinct from the others in a widespread frontal-temporal-parietal network, similar to what is seen in speech comprehension. Whereas the primary auditory and motor regions show bilateral involvement, the insula was particularly correlated in the right hemisphere and the supramarginal cortex particularly in the left hemisphere. In combination, our results shed further light on the distributed and widespread networks involved in the speech production process, which might be utilized in the development of speech neuroprostheses based on differential representations of speech.

Methods

Participants

This study includes 15 Dutch-speaking participants (6 female, 9 male, age 36.9 ± 13.7 (range 16–60) years) who were implanted with sEEG electrodes to investigate the onset zone of their epilepsy. They agreed to participate on their own volition and provided written informed consent. Electrodes were implanted purely based on clinical needs. The Institutional Review Boards of both Maastricht University (METC 2018-0451) and Epilepsy Center Kempenhaeghe approved the study. All ethical regulations relevant to human research participants were followed.

Task

The task consisted of 100 unique Dutch words, individually presented on a screen for 2 s. During this time, the participant read and spoke the word out loud once. This was followed by a 1-s inter-trial interval with a fixation cross. Of note, the data of sub-10 only contains 95 words due to a technical issue. The choice of a single-word paradigm was motivated by the goal to generate highly interpretable results and to minimize complexity. Full sentences would add further semantic dependencies that are beyond the scope of this study. Additionally, single words provide an optimal framework for the articulatory-to-acoustic inversion model, that is used to generate the articulatory features, by reducing the complexity of coarticulation effects and anticipatory gestures present in continuous speech.

Data acquisition

The neural recordings were obtained using either one or two Micromed SD LTM amplifiers (Micromed S.p.A., Treviso, Italy), each equipped with 64 channels and operating at a sampling rate of 1024 Hz or 2048 Hz. Subsequently, the signal sampled at 2048 Hz was downsampled to 1024 Hz. The sEEG electrode shafts (Microdeep intracerebral electrodes; Dixi Medical, Beçanson, France) featured a diameter of 0.8 mm, a contact length of 2 mm, and an inter-contact distance of 1.5 mm. The number of contacts on a given electrode shaft varied, ranging from 5 to 18, with an overall count of implanted shafts ranging between 5 and 19 per participant. The data was recorded using ground and reference electrodes typically located in white matter regions that did not show epileptic activity. For audio data acquisition, the onboard microphone of the recording laptop (HP Probook) was utilized at a sampling rate of 48 kHz. The data was collected using our custom-made open-source data collection tool T-REX[61], which utilizes LabStreamingLayer[62] to ensure synchronization among neural, audio, and stimulus data (Fig. 1A).

Electrode localization

A custom version of the img_pipe[63] Python package was used to localize electrodes alongside Freesurfer. The processes involved the co-registration between a pre-implantation T1-weighted anatomical magnetic resonance

imaging (MRI) scan and a post-implantation computerized tomography (CT) scan, manual identification and inspection of contacts, and a non-linear warping to an average brain (MNI152) which was used only for visualisation purposes and the distribution metric. Anatomical labels were extracted using the Fischl atlas[64] for subcortical areas and the Destrieux atlas[65] for cortical areas. There were 1647 recorded electrode contacts (937 left hemisphere, 710 right hemisphere), besides the reference and ground contacts, spread across 128 unique anatomical labels (Supplementary Fig. 1).

Neural data processing

The neural data was re-referenced to an electrode shaft reference, in which the average of all other electrodes on a shaft is subtracted from each electrode on that shaft. In each channel, the neural signal underwent filtering to isolate broadband high-frequency activity (70–170 Hz) using an IIR Butterworth bandpass filter with a filter order of 4. To mitigate interference from the 50 Hz line noise, two IIR bandstop filters with a filter order of 4 were implemented. These filters were applied in both the forward and backward directions to eliminate any potential phase-shifts. Subsequently, by computing the absolute of the Hilbert transform, we extracted the absolute amplitude envelope of the high-frequency signal for further analysis. We segmented the resulting signals in 50 ms windows with a 10 ms frameshift for articulatory and acoustic representations to capture the complex dynamics of speech. Since neural activity precedes (in case of production) or succeeds (in case of perception) behaviour, we shifted the neural and audio data ranging between -200 and $+200$ ms in steps of 50 ms. These time-shifted neural signals were added as 'timeframe' features for the articulatory and acoustic data analyses, representing temporal context to elucidate temporal dynamics. The neural signal was segmented into larger windows for the semantic representation than for the other two representations, since semantic information is expected to have slower dynamics. This data was segmented in 200 ms windows with a 25 ms frameshift instead.

Feature extraction and analysis

Articulatory representation. The acoustic-to-articulatory inversion (AAI) pipeline[39] was used to estimate articulatory trajectories from the audio data. We first processed the audio by manually silencing background noises unrelated to speech using Audacity. It was not necessary to apply sophisticated filtering to separate speech from noise, since noises did not temporally coincide with speech utterances. We then modified the experimental markers to ensure that the markers indicating the start and end of the word trials always included the entire audio of the trial's word. The modification was necessary as participants sometimes initiated speech late and did not finish within the allotted trial time. Three words in total were not spoken at all and therefore discarded from further analysis in the pipeline (Supplementary Table 2). The markers were further adjusted to ensure a crucial 100 ms of silence at both the beginning and end of the utterances. This step is essential as the quality of the reconstructed articulatory trajectories is susceptible to the acoustic properties of utterances, especially the initial and final segments, particularly for plosive sounds.

The AAI pipeline contains a set of machine learning models pre-trained using 16 kHz audio segments. Thus, we decimated our signal from 48 kHz to 16 kHz, after applying an anti-aliasing filter. We then cut the audio at the redefined markers and created segments for each word. The AAI pipeline also required an estimated syllable count for each word, which we manually defined. Following the methodology described by Gao et al.[39], the system first extracts acoustic features from the speech signal, including mel-frequency cepstral coefficients and voicing information. These features are then processed through a pre-trained Long Short-Term Memory (LSTM) neural network that estimates the underlying articulatory movements. Using this methodology, we extracted a fixed set of 30 time-series related to the articulation process for each utterance. These 30 parameters are integral to the model, it includes eight parameters that are fixed values for a potential subsequent articulatory-to-audio conversion, and one that is a

binary estimation of speech versus silence ('subglottal pressure'). The remaining 21 parameters represent the trajectories of different articulatory muscles during the utterance production. However, one ('vocal fold lower displacement') was exactly the same as another ('vocal fold upper displacement') and thus discarded. The 20 remaining articulatory trajectories (Supplementary Table 1) and the 'subglottal pressure' time-series were decimated to match the neural signal's sampling rate. At this stage, the dataset was at the word level, meaning that we had a set (approximately 300 samples) of trajectories and neural data for each uttered word. The pipeline failed to generate proper trajectories for 20 word trials in total across all of the participant's data (Supplementary Table 2), these trials were discarded from further analysis. The sets of trajectories and neural data were concatenated back into continuous signals (Fig. 1B).

We correlated the articulatory trajectories with the neural data from each channel and their time-shifted neural features (time-frames) using Pearson correlation to find the neural articulatory representation. This was calculated in 10 folds to match the tenfold cross-validation in the semantic pipeline. The folds were averaged to obtain the final score per channel, time-frame and articulator. The results from the articulators were pooled together to form one articulatory representation (described further in 'Significant channels').

Acoustic representation. The audio signal was downsampled to 16 kHz, after applying an anti-aliasing filter, and we extracted the speech spectrogram using the Short-Time Fourier Transform. This was done in 50 ms windows with a 10 ms frameshift to ensure an alignment between the audio and neural features. The resulting spectrogram was compressed into a log-mel representation with 23 triangular filter banks (Fig. 1C). We then correlated, also in 10 folds, each mel-bin ($N = 23$) separately with each single channel and their time-shifted neural features (time-frames) using Pearson correlation. The folds and mel-bins were averaged to calculate the final score for each channel and time-frame.

Semantic representation. As opposed to the other two representations, the semantic representation was analyzed on a trial-based level. The whole speech trial (2 seconds) was assigned to the cued word. Semantic embeddings (160 dimensional numerical representations of words capturing their semantic meaning) were extracted for each Dutch word using a pre-trained *word2vec* model[40]. The word2vec model was chosen for its simplicity and suitability for single-word representations, avoiding the contextual complexity of more modern models that are unnecessary for single words. The embeddings were available for 85 of the 100 words, thus the 15 word trials without embeddings were discarded for each participant. As opposed to the previous two representations, we chose to use an encoding model for the semantic representation. A linear regression was trained to predict the neural timeseries for each recorded channel individually from the semantic embeddings, in a tenfold cross-validation (Fig. 1D). Linear regression was chosen for its interpretability and suitability for the relatively small number of word trials, allowing us to capture the linear relationship between the semantic features and neural activity. Performance was evaluated using the Pearson correlation between the original and the predicted timeseries per window (time-point) across word trials for each single channel. This procedure evaluates temporal variability in the response to spoken words rather than measuring how well the average response is captured. This procedure closely follows previous studies on semantic representations[2,37].

Statistics and reproducibility

Significant channels. Randomized baseline performance was calculated for each of the representations individually. We applied a random circular shift procedure for the articulatory and acoustic representations. This procedure consisted of generating a random timepoint, between 10% and 90% of the data, on which the neural features were swapped. Then, the swapped data was correlated with the held-out 'subglottal pressure' time-series (articulatory) or each of the mel-bins (acoustic) and

repeated 1000 times to estimate a distribution of chance correlation coefficients for every channel within every participant. The correlations across mel-bins were averaged before saving. Random permutations of the embedding vectors and re-training of the model were used to estimate a distribution of chance correlations for the semantic representation, this was also repeated 1000 times.

The significance threshold ($\alpha = 0.05$) was set at the largest chance correlation across channels (articulatory and acoustic) or across channels and time windows (semantic) within each participant. Thereby, we used the max-t correction to correct for multiple comparisons[66]. The 'subglottal pressure' time-series was chosen in the articulatory pipeline as it distinguished overall speech from silence and yielded the strictest threshold. Any channel that had a correlation above the articulatory significance threshold in any of the articulators and time-frames was deemed significant for the articulatory representation. Thus, the results from individual articulators were pooled together to form one articulatory representation. The channel was significant if the average correlation across mel-bins for that channel was above the acoustic threshold in any of the time-frames for the acoustic representation. For the semantic representation, a channel was deemed significant if any of the time windows were above the semantic threshold. These calculations are provided in the code along with the pre-processed data to facilitate reproducibility.

Distribution metric. We took the coordinates (x, y, z) in MNI space from all significant channels within each hemisphere to calculate the geometric centroid for each representation. We then calculated the Euclidean distance between each channels' coordinates and the centroid within each representation. Kruskal–Wallis and post-hoc two-sided Mann–Whitney U tests (Bonferroni-corrected for nine repeated tests) were used to calculate the statistical difference in distribution (i.e., distances) between the three representations within the hemispheres and between the left and right hemisphere within representations. The Euclidean distances were additionally calculated between all recorded channels and their centroid within each hemisphere, to rule out differences due to sampling alone, and tested using a two-sided Mann–Whitney U test. These non-parametric tests were chosen as they do not assume normally distributed data. Effect sizes were calculated with eta squared for the Kruskal–Wallis main effects and the rank-biserial correlation for the Mann–Whitney U tests.

Regions-of-interest
To find out how the speech representations are distributed within and across language areas, we did a ROI analysis in six language-related cortical areas (motor cortex, auditory cortex, superior temporal cortex, inferior frontal cortex, supramarginal cortex and the insula) and one subcortical area (hippocampus). The cortical areas included both gyri and sulci, a full description of which anatomical labels from the Destrieux/Fischl atlas we chose to include in each ROI can be found in the supplementary material (Supplementary Table 3). Individual channels were selected based on the anatomical labels in individual space. We calculated the percentage of significant channels for each representation and their overlap out of the total number of channels present in that region. Other relevant regions mentioned in the introduction (the fusiform cortex, SMA, postcentral cortex and other subcortical areas) were not sufficiently sampled to be included as ROIs.

Temporal dynamics
To leverage the temporal resolution of our recordings, we analyzed the temporal dynamics of the results. The number of significant channels and their correlations were evaluated within each time-point (semantic) or time-frame (articulatory and acoustic). Since the semantic analysis is trial-based, the time-points go from 0 ms (word cue) to 1875 ms (end of shortest trial). Since the articulatory and acoustic analyses are based on the entire recording, the time-frames represent the misalignment between the neural and target data from −200 ms to +200 ms (as described in 'Neural data

processing'). The same statistical procedures were used as in the main spatial analysis, except that significant channels were now evaluated per time-point or time-frame rather than across them.

Acoustic contamination
Electrophysiological recordings can potentially be contaminated by the acoustics of overt speech, which can be detected with a significant correlation in the spectral energy between the recorded neural and audio data[67]. None of the 15 participants had a significant correlation (all $p > 0.01$), indicating a chance of less than 1% of acoustic contamination.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The processed data is publicly available at https://doi.org/10.17605/OSF.IO/QZWSV[68].

Code availability
The code to analyze the processed data is openly available at https://doi.org/10.5281/zenodo.14887361[69]. It requires Python version 3.6 or above and the numpy, scipy, scikit-learn and pandas packages.

References
1. Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
2. Goldstein, A. et al. Shared computational principles for language processing in humans and deep language models. *Nat. Neurosci.* **25**, 369–380 (2022).
3. Tang, J., LeBel, A., Jain, S. & Huth, A. G. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat. Neurosci.* **26**, 858–866 (2023).
4. Tate, M. C., Herbet, G., Moritz-Gasser, S., Tate, J. E. & Duffau, H. Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited. *Brain* **137**, 2773–2782 (2014).
5. Flinker, A. et al. Redefining the role of broca's area in speech. *Proc. Natl. Acad. Sci. USA* **112**, 2871–2875 (2015).
6. Gajardo-Vidal, A. et al. Damage to broca's area does not contribute to long-term speech production outcome after stroke. *Brain* **144**, 817–832 (2021).
7. Andrews, J. P. et al. Dissociation of broca's area from broca's aphasia in patients undergoing neurosurgical resections. *J. Neurosurg.* **138**, 847–857 (2022).
8. Willett, F. R. et al. A high-performance speech neuroprosthesis. *Nature* **620**, 1031–1036 (2023).
9. Binder, J. R. Current controversies on wernicke's area and its role in language. *Curr. Neurol. Neurosci. Rep.* **17**, 1–10 (2017).
10. Indefrey, P. The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* **2**, 255 (2011).
11. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
12. Cogan, G. B. et al. Sensory–motor transformations for speech occur bilaterally. *Nature* **507**, 94–98 (2014).
13. Thomas, T. M. et al. Decoding articulatory and phonetic components of naturalistic continuous speech from the distributed language network. *J. Neural Eng.* **20**, 046030 (2023).
14. Oh, A., Duerden, E. G. & Pang, E. W. The role of the insula in speech and language processing. *Brain Lang.* **135**, 96–103 (2014).
15. Chrabaszcz, A. et al. Subthalamic nucleus and sensorimotor cortex activity during speech production. *J. Neurosci.* **39**, 2698–2708 (2019).

16. Thibault, S. et al. Tool use and language share syntactic processes and neural patterns in the basal ganglia. *Science* **374**, eabe0874 (2021).

17. Klostermann, F., Krugel, L. K. & Ehlen, F. Functional roles of the thalamus for language capacities. *Front. Syst. Neurosci.* **7**, 32 (2013).

18. Covington, N. V. & Duff, M. C. Expanding the language network: direct contributions from the hippocampus. *Trends Cogn. Sci.* **20**, 869–870 (2016).

19. Cai, J., Hadjinicolaou, A. E., Paulk, A. C., Williams, Z. M. & Cash, S. S. Natural language processing models reveal neural dynamics of human conversation. Preprint at *bioRxiv* 2023–03 (2023).

20. Bocquelet, F., Hueber, T., Girin, L., Chabardès, S. & Yvert, B. Key considerations in designing a speech brain-computer interface. *J. Physiol.-Paris* **110**, 392–401 (2016).

21. Rabbani, Q., Milsap, G. & Crone, N. E. The potential for a speech brain–computer interface using chronic electrocorticography. *Neurotherapeutics* **16**, 144–165 (2019).

22. Moses, D. A. et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N. Engl. J. Med.* **385**, 217–227 (2021).

23. Luo, S., Rabbani, Q. & Crone, N. E. Brain-computer interface: applications to speech decoding and synthesis to augment communication. *Neurotherapeutics* **19**, 263–273 (2022).

24. Wandelt, S. K. et al. Representation of internal speech by single neurons in human supramarginal gyrus. *Nat. Hum. Behav.* **8**, 1136–1149 (2024).

25. Silva, A. B., Littlejohn, K. T., Liu, J. R., Moses, D. A. & Chang, E. F. The speech neuroprosthesis. *Nat. Rev. Neurosci.* **25**, 473–492 (2024).

26. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional organization of human sensorimotor cortex for speech articulation. *Nature* **495**, 327–332 (2013).

27. Chartier, J., Anumanchipalli, G. K., Johnson, K. & Chang, E. F. Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron* **98**, 1042–1054 (2018).

28. Mugler, E. M. et al. Differential representation of articulatory gestures and phonemes in precentral and inferior frontal gyri. *J. Neurosci.* **38**, 9803–9813 (2018).

29. Stavisky, S. D. et al. Neural ensemble dynamics in dorsal motor cortex during speech in people with paralysis. *Elife* **8**, e46015 (2019).

30. Herff, C. et al. Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices. *Front. Neurosci.* **13**, 469935 (2019).

31. Wang, R., Wang, Y. & Flinker, A. Reconstructing speech stimuli from human auditory cortex activity using a wavenet approach. In *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 1–6 (IEEE, 2018).

32. Martin, S. et al. Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front. Neuroeng.* **7**, 14 (2014).

33. Angrick, M. et al. Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. *Commun. Biol.* **4**, 1055 (2021).

34. Metzger, S. L. et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* **620**, 1037–1046 (2023).

35. Anumanchipalli, G. K., Chartier, J. & Chang, E. F. Speech synthesis from neural decoding of spoken sentences. *Nature* **568**, 493–498 (2019).

36. Sun, P., Anumanchipalli, G. K. & Chang, E. F. Brain2char: a deep architecture for decoding text from brain recordings. *J. Neural Eng.* **17**, 066015 (2020).

37. Herff, C., Verwoert, M., Amigó-Vega, J. & Ottenhoff, M. C. Semantic representations of speech production in intracranial eeg. In *2023 IEEE International Conference on Systems, Man and Cybernetics*, 1–6 (IEEE, 2023).

38. Samanta, D. Recent developments in stereo electroencephalography monitoring for epilepsy surgery. *Epilepsy Behav.* **135**, 108914 (2022).

39. Gao, Y., Steiner, P. & Birkholz, P. Articulatory copy synthesis using long-short term memory networks. In *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung (2020)* (eds. Wendemuth, A., Böck, R. & Siegert, I.) 52–59 (TUDpress, 2020).

40. Tulkens, S., Emmery, C. & Daelemans, W. Evaluating unsupervised Dutch word embeddings as a linguistic resource. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (eds. Calzolari, N. et al.) 4130–4136 (European Language Resources Association, 2016).

41. Ferrand, L. Reading aloud polysyllabic words and nonwords: the syllabic length effect reexamined. *Psychon. Bull. Rev.* **7**, 142–148 (2000).

42. Tourville, J. A. & Guenther, F. H. The diva model: a neural theory of speech acquisition and production. *Lang. Cogn. Process.* **26**, 952–981 (2011).

43. Gwilliams, L., Marantz, A., Poeppel, D. & King, J.-R. Hierarchical dynamic coding coordinates speech comprehension in the brain. Preprint at *bioRxiv* (2024).

44. Kiremitçi, I. et al. Attentional modulation of hierarchical speech representations in a multitalker environment. *Cereb. Cortex* **31**, 4986–5005 (2021).

45. Ralph, M. A. L., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* **18**, 42–55 (2017).

46. Gainotti, G. Why are the right and left hemisphere conceptual representations different? *Behav. Neurol.* **2014**, 603134 (2014).

47. Castellucci, G. A., Kovach, C. K., Howard III, M. A., Greenlee, J. D. & Long, M. A. A speech planning network for interactive language use. *Nature* **602**, 117–122 (2022).

48. Bhaya-Grossman, I. & Chang, E. F. Speech computations of the human superior temporal gyrus. *Annu. Rev. Psychol.* **73**, 79–102 (2022).

49. Giordano, B. L., Esposito, M., Valente, G. & Formisano, E. Intermediate acoustic-to-semantic representations link behavioral and neural responses to natural sounds. *Nat. Neurosci.* **26**, 664–672 (2023).

50. Yamamoto, A. K. et al. A special role for the right posterior superior temporal sulcus during speech production. *Neuroimage* **203**, 116184 (2019).

51. Zhang, W., Liu, Y., Wang, X. & Tian, X. The dynamic and task-dependent representational transformation between the motor and sensory systems during speech production. *Cogn. Neurosci.* **11**, 194–204 (2020).

52. Soroush, P. Z. et al. The nested hierarchy of overt, mouthed, and imagined speech activity evident in intracranial recordings. *NeuroImage* **269**, 119913 (2023).

53. Wu, X., Wellington, S., Fu, Z. & Zhang, D. Speech decoding from stereo-electroencephalography (seeg) signals using advanced deep learning methods. *J. Neural Eng.* **21**, 036055 (2024).

54. Gogolla, N. The insular cortex. *Curr. Biol.* **27**, R580–R586 (2017).

55. Afif, A., Minotti, L., Kahane, P. & Hoffmann, D. Middle short gyrus of the insula implicated in speech production: intracerebral electric stimulation of patients with epilepsy. *Epilepsia* **51**, 206–213 (2010).

56. Duff, M. C., Covington, N. V., Hilverman, C. & Cohen, N. J. Semantic memory and the hippocampus: revisiting, reaffirming, and extending the reach of their critical relationship. *Front. Hum. Neurosci.* **13**, 471 (2020).

57. Chatzikonstantinou, A. Epilepsy and the hippocampus. *Hippocampus Clin. Neurosci.* **34**, 121–142 (2014).

58. Tankus, A. et al. A speech neuroprosthesis in the frontal lobe and hippocampus: decoding high-frequency activity into phonemes. *Neurosurgery*. https://doi.org/10.1227/neu.0000000000003068 (2024).

59. Piai, V. et al. Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proc. Natl. Acad. Sci.* **113**, 11366–11371 (2016).

60. Kabakoff, H. et al. Timing and location of speech errors induced by direct cortical stimulation. *Brain Commun.* **6**, fcae053 (2024).

61. Amigó-Vega, J. et al. The easy and versatile neural recording platform (t-rex): design and development study. *JMIR Neurotech.* **2**, e47881 (2023).

62. Kothe, C. Lab streaming layer (lsl). **26**, 2015. https://github.com/sccn/labstreaminglayer (2014).

63. Hamilton, L. S., Chang, D. L., Lee, M. B. & Chang, E. F. Semi-automated anatomical labeling and inter-subject warping of high-density intracranial recording electrodes in electrocorticography. *Front. Neuroinform.* **11**, 62 (2017).

64. Fischl, B. et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).

65. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **53**, 1–15 (2010).

66. Westfall, P. H. & Young, S. S. *Resampling-Based Multiple Testing: Examples and Methods for p-value Adjustment*, Vol. 279 (John Wiley & Sons, 1993).

67. Roussel, P. et al. Observation and assessment of acoustic contamination of electrophysiological brain signals during speech production and sound perception. *J. Neural Eng.* **17**, 056028 (2020).

68. Neural Interfacing Lab. Whole-brain dynamics of articulatory, acoustic and semantic speech representations [data set]. https://doi.org/10.17605/OSF.IO/QZWSV (2024).

69. Neural Interfacing Lab. Speech representations (version 1) [computer software]. https://doi.org/10.5281/zenodo.14887361 (2025).

Acknowledgements

Author contributions

C.H. and P.K. designed the experiment, C.H, M.C.O., and M.V. collected the data, J.V. and Y.G. generated the articulatory trajectories, M.V. and C.H. ran all other analyses and wrote the manuscript. All authors reviewed the manuscript.

Competing interests

All authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-07862-x.

Correspondence and requests for materials should be addressed to Maxime Verwoert or Christian Herff.

Peer review information *Communications Biology* thanks Giovanni Di Liberto and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editor: Jasmine Pan. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.