

Differentiation of speech-induced artifacts from physiological high gamma activity in intracranial recordings

3 Alan Bush^{1,2}, Anna Chrabaszcz³, Victoria Peterson^{1,2}, Varun Saravanan^{1,4}, Christina Dastolfo-
4 Hromack⁵, Witold J. Lipski⁶ and R. Mark Richardson^{1,2}

- 5 1. Brain Modulation Lab, Department of Neurosurgery, Massachusetts General Hospital,
6 Boston, MA, 02114, USA.

7 2. Harvard Medical School, Boston, MA, 02115, USA.

8 3. Department of Psychology, University of Pittsburgh, Pittsburgh, PA, 15260, USA.

9 4. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Boston,
10 MA, 02139, USA.

11 5. University of Pittsburgh, Department of Communication Science and Disorders, Pittsburgh,
12 PA, 15260, USA.

13 6. University of Pittsburgh, Department of Neurological Surgery, Pittsburgh, PA, 15260, USA.

14 Key words: Intracranial recordings, Deep Brain Stimulation, Electrocorticography, Local Field
15 Potentials, Intraoperative research, Speech, High gamma, Artifact

16 Corresponding author: Alan Bush (alan.bush@mgh.harvard.edu)

17 **Abstract**

18 There is great interest in identifying the neurophysiological underpinnings of speech production.
19 Deep brain stimulation (DBS) surgery is unique in that it allows intracranial recordings from both
20 cortical and subcortical regions in patients who are awake and speaking. The quality of these
21 recordings, however, may be affected to various degrees by mechanical forces resulting from
22 speech itself. Here we describe the presence of speech-induced artifacts in local-field potential
23 (LFP) recordings obtained from mapping electrodes, DBS leads, and cortical electrodes. In
24 addition to expected physiological increases in high gamma (60-200 Hz) activity during speech
25 production, time-frequency analysis in many channels revealed a narrowband gamma
26 component that exhibited a pattern similar to that observed in the speech audio spectrogram.
27 This component was present to different degrees in multiple types of neural recordings. We
28 show that this component tracks the fundamental frequency of the participant's voice, correlates
29 with the power spectrum of speech and has coherence with the produced speech audio. A
30 vibration sensor attached to the stereotactic frame recorded speech-induced vibrations with the
31 same pattern observed in the LFPs. No corresponding component was identified in any neural
32 channel during the listening epoch of a syllable repetition task. These observations demonstrate
33 how speech-induced vibrations can create artifacts in the primary frequency band of interest.
34 Identifying and accounting for these artifacts is crucial for establishing the validity and
35 reproducibility of speech-related data obtained from intracranial recordings during DBS surgery.

36 **Introduction**

37 Invasive brain recordings performed in awake patients undergoing clinically indicated
38 neurosurgeries provide a unique opportunity to study speech production with better spatial and
39 temporal precision than noninvasive neuroimaging methods. One of the advantages of
40 intracranial recordings is a much higher signal-to-noise ratio (SNR) and larger measurable
41 frequency range. This allows examination of frequency bands above 70 Hz that are typically
42 unattainable with noninvasive methods due to volume conduction effects and a sharp
43 attenuation in power at higher frequencies when passing the skull (Mukamel and Fried, 2012).
44 Many assume that a higher SNR in the intracranial recordings makes them less susceptible to
45 artifacts frequently observed in noninvasive recordings, such as electro-myographic artifacts
46 due to eye, jaw, lip and tongue movement (Flinker et al., 2018; Lachaux et al., 2003; Llorens et
47 al., 2011). Comprehensive quantitative examination of the quality of the signal and identification
48 of potential sources of artifacts in intracranial recordings, however, have received very little
49 attention. Several types of artifacts found in scalp EEG have been described in intracranial
50 recordings, such as eye movement artifacts in fronto-anterior regions (Ball et al., 2009), and
51 facial and mouth movement artifacts in electrodes close to temporal muscles (Otsubo et al.,
52 2008). This suggests that movement artifacts can contaminate intracranial LFP recordings
53 acquired to study of the neural control of speech production.

54 Neural activity in the high gamma frequency band (60-200 Hz) tracks specific features of
55 speech perception and production. For example, increased power in the high gamma frequency
56 range has been observed in the superior temporal gyrus in response to auditory stimuli (Crone
57 et al., 2001; Hamilton et al., 2018; Mesgarani et al., 2014), and in Broca's and motor cortices
58 during speech production (Edwards et al., 2010; Flinker et al., 2015; Mugler et al., 2018). High
59 gamma activity recorded from the Rolandic cortex (pre and postcentral gyri) has been shown to
60 track articulatory and/or acoustic features of speech sounds (Bouchard et al., 2013; Cheung et

61 al., 2016; Chrabaszcz et al., 2019; Conant et al., 2018). Some recent advances have even
62 made it possible to reconstruct speech from the brain's activity in the high gamma band
63 (Anumanchipalli et al., 2019; Martin et al., 2019), pointing at its potential utility for brain-
64 computer interfaces (BCI) to develop speech prostheses. Contamination of the neural signal
65 with audio acoustics therefore is a potential barrier to decoding the true electrophysiological
66 correlates of speech production.

67 Given the impact that speech dysfunction can have in patients with movement disorders,
68 and the fact that the role of subcortical regions in speech production are not well understood, we
69 recently developed a strategy to simultaneously record from the cortex and the subcortical
70 implantation target during DBS surgery. With the patient's consent, it is possible to temporally
71 place an electrocorticography (ECoG) electrode strip on the surface of the brain, a technique
72 that has been used safely in over 500 patients (Panov et al., 2017; Sisterson et al., 2021). Here,
73 we report the systematic identification and quantification of speech-induced artifacts in several
74 types of intracranial electrophysiological recordings obtained using a speech production task
75 performed during DBS implantation surgery. We show that this artifact is caused by mechanical
76 vibrations induced by the produced speech, and that it can also be found in a 'blank' headstage
77 pin not connected to any electrode. The results presented in this study encourage careful
78 assessment of possible audio-induced artifacts in intracranial recordings obtained during
79 speech production research. Additionally, we provide suggestions for data collection and
80 analysis that may reduce the potential for false discoveries.

81 Materials and Methods

82 Participants

83 Participants were English-speaking patients with Parkinson's disease (21M/8F, age: 65.6 ± 7.1
84 years, duration of disease: 6.1 ± 4.1 years) undergoing awake stereotactic neurosurgery for

85 implantation of DBS electrodes in the subthalamic nucleus (STN). Dopaminergic medication
86 was withdrawn the night before surgery. All procedures were approved by the University of
87 Pittsburgh Institutional Review Board (IRB Protocol #PRO13110420) and all patients provided
88 informed consent to participate in the study.

89 **Behavioral task**

90 Participants performed a syllable repetition task intraoperatively. Subjects heard consonant-
91 vowel (CV) syllable triplets through earphones (Etymotic ER-4 with ER38-14F Foam Eartips)
92 and were instructed to repeat them. Triplets were presented at either low (~50dB SPL) or high
93 (~70dB SPL) volume using BCI2000 (Schalk et al., 2004) presentation software. The absolute
94 intensity was adjusted to each participant's comfort level, keeping the difference between low-
95 and high-volume stimuli fixed at 25dB SPL. Participants were instructed to repeat the low-
96 volume syllable triplets at normal conversation level and the high-volume triplets at increased
97 volume, "as if speaking to someone across the room". The syllables were composed of a
98 combination of either of the 4 English consonants (/v/, /t/, /s/, /g/) and either of the 3 cardinal
99 vowels (/i/, /a/, /u/), resulting in 12 unique CV syllables. A total of 600 triplets were constructed
100 which were divided among 5 presentation lists, with each session composed of 120 triplets. No
101 syllable was repeated within a triplet. Syllable position and phoneme occurrence within a triplet
102 and intensity level (low or high) within each session were balanced. The produced audio was
103 recorded with an PRM1 Microphone (PreSonus Audio Electronics Inc., Baton Rouge, LA, USA)
104 at 96 kHz using a Zoom-H6 portable audio recorder (Zoom Corp., Hauppauge, NY, USA). The
105 average number of trials per session was 106 (range: 14-120). The duration of the utterances
106 was of 1.3 ± 0.3 s (mean \pm standard deviation pooled across subjects), with a median within-
107 subject standard deviation of 0.14 s.

108 **Neural recordings**

109 As part of the standard clinical DBS implantation procedure, functional mapping of the STN was
110 performed with microelectrode recordings (MER), acquired with the Neuro-Omega recording
111 system (Alpha-Omega Engineering, Nof HaGalil, Israel) using parylene-insulated tungsten
112 microelectrodes (25 µm in diameter, 100 µm in length). LFPs were recorded from stainless steel
113 macroelectrode rings (0.55 mm in diameter, 1.4 mm in length) located 3 mm above the tip of the
114 microelectrode. The microelectrodes were oriented on the microtargeting drive system using
115 three trajectories of a standard cross-shaped Ben-Gun array with 2 mm center-to-center
116 spacing (Central, Posterior, Medial). The MER and LFP signals were recorded at a sampling
117 rate of 44 kHz. The neural signal was referenced to the metal screw holding one of the guide
118 cannulas used to carry the microelectrodes. Prior to initiating MER mapping, one or two
119 subdural electrocorticography (ECoG) strips with 54 or 63 contacts each (platinum 1 mm disc
120 contacts arranged in a 3x18 or 3x21 layout, with 3 mm center to center spacing, PMT Cortac
121 Strips models 2110-54-001 and 2011-63-002) were placed through the standard burr hole.
122 ECoG strips were targeted to the left superior temporal gyrus (covering the ventral sensorimotor
123 cortex) and left inferior frontal gyrus. Signals from ECoG electrodes and DBS leads were
124 acquired at 30 kHz with a Grapevine Neural Interface Processor equipped with Micro2 Front
125 Ends (Ripple LLC, Salt Lake City, UT, USA). Recordings were referenced to a sterile stainless-
126 steel subdermal needle electrode placed in the scalp, approximately at the location of Cz in a
127 standard EEG montage.

128 Two or three sessions of the syllable repetition task were performed by each subject
129 when the microelectrode was positioned at different depths within the STN. The clinical setup
130 evolved during the collection of this dataset between subjects 22 and 23, transitioning from a
131 traditional stereotactic frame to the use of robotic stereotactic assistance (Faraji et al., 2020).

132 This change resulted in modification of the stereotactic frame's mechanical coupling to the
133 electrode micro-drive.

134 After the functional mapping phase patients were implanted with one of the DBS leads
135 models shown in Table 1:

- 136 ● Medtronic 3389: Platinum/Iridium, 4 cylindrical macroelectrodes, contact length 1.5 mm,
137 1.27 mm in diameter, 0.5 mm axial electrode spacing (Medtronic, Minneapolis, MN,
138 USA).
- 139 ● St. Jude 6172 short: Platinum/Iridium, directional lead with two central rings split in three
140 contacts, 8 contacts total, length 1.5 mm, axial spacing 0.5 mm, lead diameter 1.27 mm,
141 outer tubing material polycarbonate urethane (Abbott Neuromodulation, Austin, TX,
142 USA).
- 143 ● Boston Scientific DB-2202-45: Platinum/Iridium, directional lead with two central rings
144 split in three contacts, 8 contacts total, length 1.5 mm, axial spacing 0.5 mm, lead
145 diameter 1.3 mm, outer tubing material polyurethane (Boston Scientific Neuromodulation
146 Corp, Valencia, CA, USA).

147 After the leads were successfully implanted, a final session of the speech task was performed
148 by some participants, providing LFP data from DBS leads. A summary of recording types and
149 the corresponding acquisition specifications is provided in Table 1. A schematic illustration of
150 the intraoperative intracranial recording setup is shown in Figure 1.

151 **Vibration sensor**

152 A shielded piezoelectric vibration sensor (model SDT1-028K, TE Connectivity Company) was
153 fixed to the stereotactic frame using a sticky pad for the duration of one experiment. This sensor
154 was selected for its flat transfer function at frequencies above 30 Hz. The signal from the sensor
155 was captured without amplification as an analog input to the Grapevine system

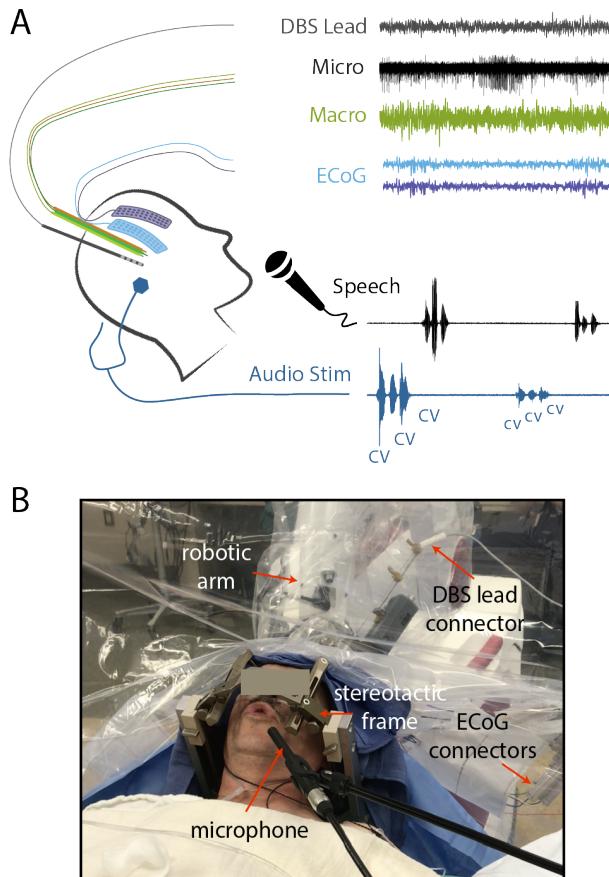
Data type	Acquisition system	Model/Manufacturer	Sampling rate	Number of recordings
DBS Lead	Grapevine	Medtronic 3389; St. Jude 6172 short; Boston Scientific DB-2202-45	30 kHz	4 or 8
Micro	Neuro-Omega	Alpha Omega	44 kHz	3
Macro	Neuro-Omega	Alpha Omega	44 kHz	3
ECoG	Grapevine	PMT 2110-54-001; PMT 2011-63-002	30 kHz	54 to 126
Presented audio	Grapevine	BCI2000	30 kHz	1
Produced audio	Grapevine	PRM1 PreSonus Microphone; Zoom-H6 audio recorder	96 kHz	1
Vibration sensor	Grapevine	SDT1-028K, TE Connectivity Company	1 kHz	1

156 **Table 1.** Acquisition system specifications for neural and physiological signals.

157 **Electrode localization**

158 DBS electrodes were localized using the Lead-DBS localization pipeline (Horn and Kühn, 2015).
159 Briefly, a preoperative anatomical T1-weighted MRI scan was co-registered with a postoperative
160 CT scan. The position of individual contacts was manually identified based on the electrode
161 artifact present in the CT image and constrained by the geometry of the implanted DBS lead.
162 This process rendered the coordinates for the leads in each subject's native space. Based on
163 the position of the lead and the known depth and tract along which the lead was implanted in
164 each hemisphere, the positions of the micro and macro recordings from the functional mapping
165 were calculated using custom Matlab scripts (github.com/Brain-Modulation-Lab/Lead_MER).
166 The position of the ECoG strips was calculated from an intraoperative fluoroscopy image as
167 described in (Randazzo et al., 2016). Briefly, the cortical surface was reconstructed from the
168 preoperative MRI using the FreeSurfer image analysis suite

169 (http://surfer.nmr.mgh.harvard.edu/). A model of the skull and the stereotactic frame was
170 reconstructed from the intraoperative CT scan using OsiriX neuroimaging viewing tool
171 (<https://www.osirix-viewer.com/>). The position of the frame's tips on the skull and the implanted
172 DBS leads were used as fiducial markers. The models of the pial surface, skull and fiducial
173 markers were co-registered, manually rotated and scaled to align with the projection observed in
174 the fluoroscopy image. Once aligned, the position of the electrodes in the ECoG strip was
175 manually marked on the fluoroscopy image and the projection of those positions to the convex
176 hull of the cortical surface was defined as the electrodes' locations in the native brain space.
177 The coordinates were then regularized based on the known layout of the contacts in the ECoG
178 strip (github.com/Brain-Modulation-Lab/ECoG_localization). All coordinates were then
179 transformed to the ICBM MNI152 Non-Linear Asymmetric 2009b space (Fonov et al., 2011)
180 using the Symmetric Diffeomorphism algorithm implemented in the Advanced Normalization
181 Tools (Avants et al., 2008). Anatomical labels were assigned to each electrode based on the
182 HCP-MMP1 atlas (Glasser et al., 2016) for cortical electrodes, and the Morel (Niemann et al.,
183 2000) and DISTAL (Ewert et al., 2018) atlases for subcortical electrodes.



184

185 **Figure 1: Intraoperative intracranial recording setup and speech task.**

186 **A.** Schematic representation of intracranial electrodes and syllable triplet repetition task. During
187 the mapping phase of the DBS implantation surgery ECoG strips were temporarily placed
188 through the burr hole, allowing simultaneous recording of cortical and subcortical LFPs and
189 MER from the STN. Participants were instructed to repeat aloud CV syllable triples at a volume
190 matching the auditory stimuli. **B.** Photograph of a participant performing the speech task.

191 **Phonetic coding**

192 Phonetic coding of each participant's produced speech was performed by a trained team of
193 speech pathology students. Using a custom Matlab GUI ([github.com/Brain-Modulation-
194 Lab/SpeechCodingApp](https://github.com/Brain-Modulation-Lab/SpeechCodingApp)), onset and offset times, IPA transcription, accuracy and disorder
195 characteristics of each produced phoneme were coded. Speech onsets were marked based on
196 acoustic evidence of key speech features. Voice segments, such as vowels, were marked at
197 the first visible glottal pulse, indicating the onset of vocal fold vibration. Unvoiced phonemes
198 were marked based on the characteristic noise features for that phoneme. For example, the

199 rapid increase of high frequency energy denoted the onset of plosive consonants. A broadband
200 spectrogram was utilized to maximize visualization of key speech features. All coders completed
201 a course in speech science to ensure knowledge of important speech features and were trained
202 in coding procedures by a speech language pathologist.

203 **Electrophysiological data preprocessing**

204 Data processing was performed using custom code based on the FieldTrip (Oostenveld et al.,
205 2011) toolbox implemented in Matlab, available at (github.com/Brain-Modulation-Lab/bml).
206 Recordings from the Grapevine, Neuro-Omega and Zoom-H6 systems were temporally aligned
207 based on the stimulus and produced audio channels using a linear time-warping algorithm.
208 Continuous alignment throughout the entire recording session was achieved with sub-
209 millisecond precision. Data was low-pass filtered at 250 Hz using a 4th order non-causal
210 Butterworth filter, down-sampled to 1 kHz and stored as continuous recordings in FieldTrip
211 datatype-raw objects in mat containers. This frequency range is well-suited for analyses in the
212 canonical frequency bands normally used to explore cognitive functions. All annotations,
213 including descriptions of each session (duration, type of subcortical recording, depth of the MER
214 recordings), details of the electrodes (active time intervals, channel names, coordinates in
215 native and MNI space, anatomical labels), phonetic coding at the phoneme, syllable and triplet
216 level, and times of stimulus presentation were stored in annotation tables.

217 An automatic data cleaning procedure was used to remove segments of data with
218 conspicuous high-power artifacts. First, a 1 Hz high-pass 5th order non-causal Butterworth filter
219 was applied to remove low frequency movement related artifacts. The power at frequencies in
220 different canonical bands (3 Hz for δ , 6 Hz for θ , 10 Hz for α , 21 Hz for β , 45 Hz for γ_L and 160
221 Hz for γ_H) was estimated by convolution with a Morlet wavelet with a width parameter of 9. For
222 each frequency, the maximum power in 1-s time bins was calculated, log-transformed and the
223 mean (\bar{x}) and standard deviations (σ) of the distribution were estimated using methods robust to

224 outliers. A time bin was defined as artifactual if its maximal log-transformed power in any band
225 exceeded $\bar{x} + 3\sigma$ for that band. Note that this resulted in thresholds at least 10-fold higher than
226 the mean power. Channels were entirely discarded if more than 50% of the time-bins were
227 classified as containing artifacts. Blocks of channels sharing a head-stage connector were
228 entirely discarded if more than 50% of those channels were defined as artifactual.

229 For each trial, 3 different behavioral epochs were defined: stimulus presentation or
230 listening epoch - the 1.5-s long window during which syllable triplets were presented auditorily;
231 speech production epoch - the variable time during which subjects repeated the syllable triplet;
232 baseline epoch - a 500-ms time window centered between speech offset of one trial and
233 stimulus onset of the following.

234 **Electrophysiology data analysis**

235 **Time-frequency plots.** Time-frequency analyses for neural and audio data were performed
236 using the Short Time Fourier Transform (STFT) method with a 100 ms Hanning window and a
237 frequency step of 2 Hz, based on multiplication in the frequency domain as implemented by
238 FieldTrip. Trials were aligned to speech onset and Z-scored relative to a 500-ms baseline
239 epochs included in every trial. Frequencies up to 250 Hz were used for this analysis as that
240 covers the canonical frequency bands normally used to explore cognitive functions.

241 **Spectrogram correlation analysis.** To index the degree of similarity between the audio
242 spectrogram and the time-frequency spectrogram of a neural channel calculated by the STFT
243 method, a correlation between these two matrices was calculated. This correlation was
244 computed as the normalized sum of the element-by-element products of the two matrices.

245
$$r = \frac{\sum_m \sum_n (A_{m,n} - \bar{A})(B_{m,n} - \bar{B})}{\sqrt{\sum_m \sum_n (A_{m,n} - \bar{A})^2 \times \sum_m \sum_n (B_{m,n} - \bar{B})^2}}$$
 (Equation 1)

246 where A represents the audio spectrogram and B the time-frequency spectrogram of data at a
247 given neural channel.

248 **Coherence analysis.** Phase relationship between the audio signal and the neural signal at a
249 given channel was quantified using a metric of inter-trial phase consistency (ITPC) (Cohen,
250 2014). First, the audio and the neural signals were band-pass filtered between 70 and 240 Hz
251 using a 5th order non-causal Butterworth filter. This frequency range contains the fundamental
252 frequency of the voice in humans and the narrowband component studied in this work. A notch
253 filter was applied to remove line noise and its harmonics. For each individual epoch of interest
254 (noted by the index e), and neural channel (X) a complex value φ_e was calculated following

$$255 \quad \varphi_e = \frac{1}{|X_e| |A_e|} \sum_k S_{e,k} A_{e,k} \quad (\text{Equation 2})$$

256 where $S_e = X_e + iH(X_e)$ is the analytic signal, $H(X_e)$ is the Hilbert transform of the neural data
257 for the epoch e , and A_e is the audio signal for that epoch. The sum is taken for every sample k
258 within the epoch. $|X_e|$ and $|A_e|$ are the Euclidean norms of the neural and audio signals for
259 epoch e . The absolute value and phase of φ_e represents the degree of correlation and phase
260 relationship between the neural and the audio channels for epoch e . If there is inter-trial phase
261 consistency, the mean value of φ across trials (noted as $\langle\varphi\rangle$) will be significantly different from 0.
262 To quantify this, the ITPC index was defined as

$$263 \quad ITPC = \frac{|\langle\varphi\rangle|}{StdErr(\varphi)} \quad (\text{Equation 3})$$

264 where the standard error of φ is defined as $StdErr(\varphi) = \sqrt{\sum_e |\varphi_e - \langle\varphi\rangle|^2 / N}$, and the sum is
265 taken over the N epochs considered for the neural channel of interest. Note that this metric was
266 calculated independently for the speech production, listening and baseline epochs.

267 **Significance threshold for coherence index.** To define a threshold of significance for the
268 coherence index, a Monte Carlo simulation was performed to obtain the distribution of

269 coherence index under the null hypothesis of no consistent phase relationship between the
270 neural and the audio channels. To this end, a random time jitter uniformly distributed between -
271 100 and 100 ms was applied to the neural data before calculating each φ_e value. 22000
272 independent randomizations were calculated using data from all electrodes in the dataset. This
273 analysis resulted in a significant threshold for the coherence index of 3.08 that corresponds to
274 the 99.99 percentile of the distribution. Thus, a coherence index greater than 3.08 suggests
275 significant correlation of inter-trial phase consistency between the audio and the neural signal.

276 **Objective assessment of acoustic contamination.** A recent study by Roussel et al. 2020,
277 comparing intracranial recordings collected from human subjects during speech perception and
278 production at five different research institutions found that spectrotemporal features of the
279 recorded neural signal are highly correlated with those of the sound produced by the
280 participants or played to participants through a loudspeaker. The method proposed by Roussel
281 et al. for quantifying the extent of acoustic contamination in an electrophysiological recording
282 was applied to the data using the open source toolbox developed by the authors (Roussel et al.,
283 2020) (<https://doi.org/10.5281/zenodo.3929296>). Briefly, the method correlates the power of the
284 neural data and audio across different frequency bins, thus creating a correlation matrix for
285 every combination of frequencies of the two channels. Significantly higher correlation
286 coefficients at matching frequencies of the two channels (i.e. on the diagonal of the matrix),
287 compared to non-matching frequencies, is considered to be evidence of acoustic contamination.
288 A permutation test is used to determine the significance threshold. The method was applied to
289 data from individual electrodes, and the False Discovery Rate was adjusted according to
290 (Benjamini and Hochberg, 1994).

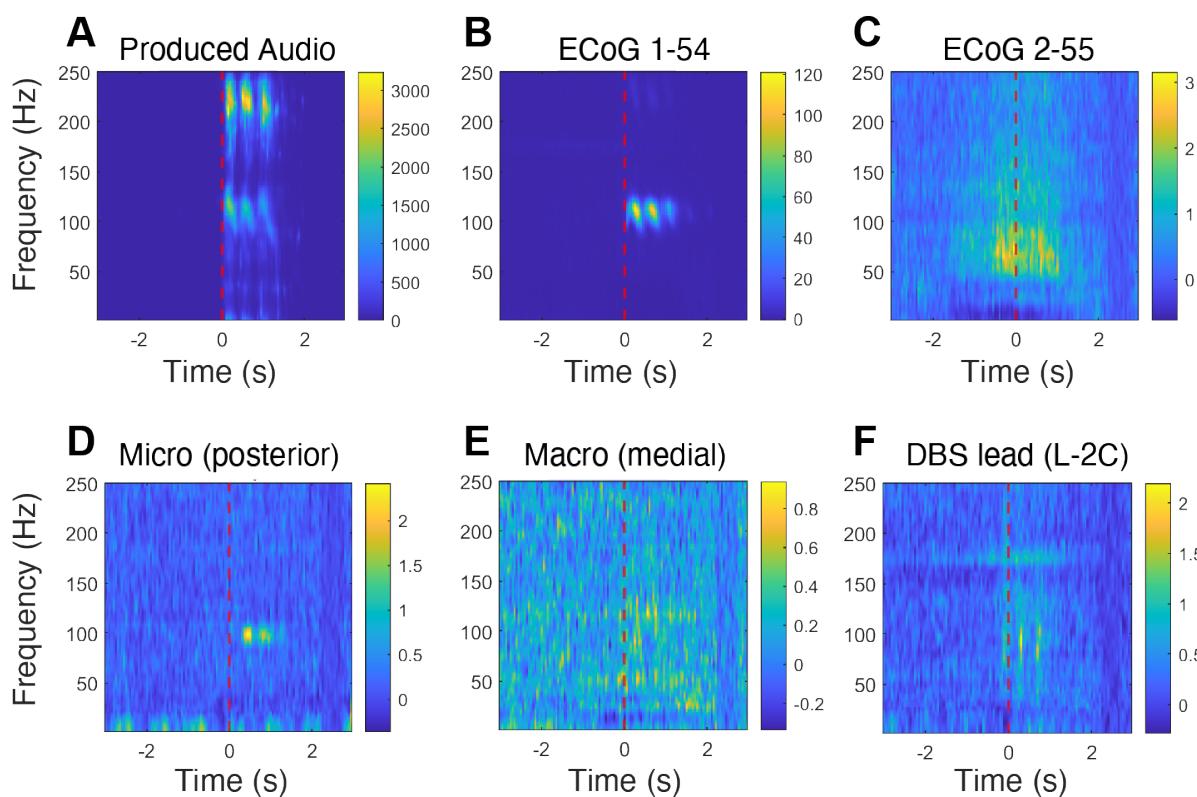
291 **Testing the spatial distribution of coherence over the brain.** We used hierarchical
292 bootstrapping (Saravanan et al., 2020) to test whether any particular brain region displayed an

293 average coherence that was significantly different from the rest of the brain. Specifically, we
294 built a null distribution of average coherence across the brain by calculating the mean
295 coherence across the brain accounting for subjects as one hierarchical level but ignoring brain
296 regions within subjects. We then calculated the average coherence for the 15 brain regions (as
297 parcellated by the MMP1 atlas) that had coverage in 10 or more subjects, with 3 or more
298 electrodes per subject. Each distribution was again built using subjects as the hierarchical level
299 and electrodes were restricted to those within that particular brain region. Bootstrapping was
300 performed weighing each subject by the number of electrodes present in that region. The
301 distributions were then all compared to the null and the significance threshold was adjusted
302 using an FDR correction for 15 comparisons. Resampling number ($N_{\text{bootstrap}}$) was set to 10,000
303 for all bootstrap samples.

304 Results

305 **Narrowband high gamma component in neural recordings during speech production**
306 We averaged spectrograms time-locked to the speech onset across trials for each audio and
307 neural channel. Representative examples of the averaged spectrograms for each recording type
308 are provided in Figure 2. Figure 2A shows the spectrogram for the produced audio, in which the
309 fundamental frequency (F0) of the participant's voice (at around 120 Hz) and its first harmonic
310 (at around 240 Hz) can be easily discerned as an increase in power at the corresponding
311 frequencies. The 3 peaks of power in the spectrogram at different times correspond to the three
312 produced syllables. A similar narrowband component occurring around the frequencies of the
313 participants' F0 also appeared in some electrodes during the speech production epoch. For
314 example, this narrowband component can be readily observed in the time-frequency plot for one
315 ECoG electrode shown in Figure 2B, but not for another ECoG electrode from the same subject
316 shown in Figure 2C. Thus, while both electrodes show an increase in speech-related gamma

317 activity, the increase in gamma activity in the electrode in Figure 2B is remarkably similar to the
318 narrowband component in the audio spectrogram (Figure 2A), overlapping with it in frequency,
319 time, and overall shape. Similar narrowband components in the high gamma frequency range
320 were also identified in some of the LFP recordings extracted from the micro electrodes (Figure
321 2D), macro electrodes (Figure 2E) and from the DBS leads (Figure 2F). Thus, this narrowband
322 speech-related component can be observed in different electrophysiological recordings
323 collected simultaneously with different acquisition devices.



324

325 **Figure 2: A speech-related narrowband component in the high gamma frequency range**
326 **was observed across different types of neural recordings.**

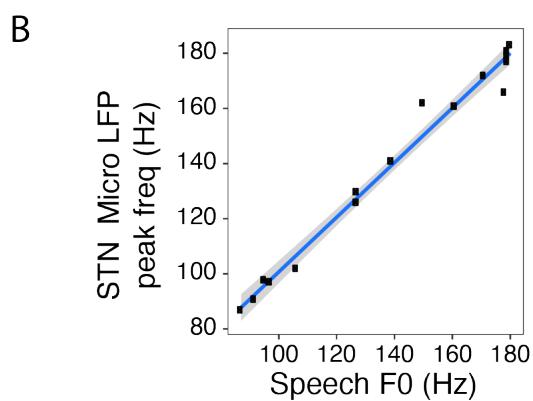
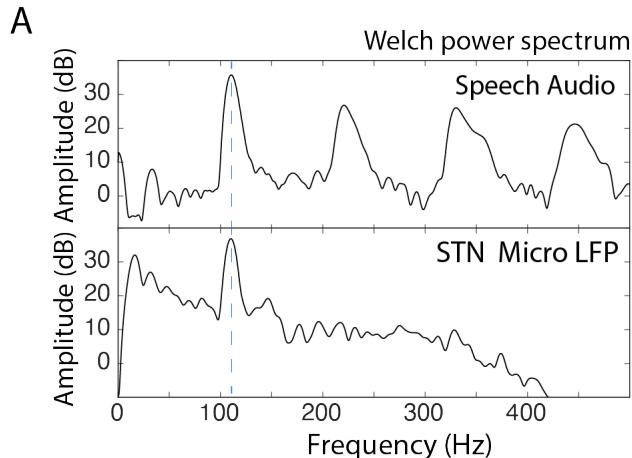
327 Time-frequency plots for the audio and different neural channels from subject DBS3014. The
328 red vertical dashed line represents speech onset time, which was used to time-lock the data
329 across trials. **A.** Average spectrogram of the produced speech audio, z-scored to baseline. **B.**
330 ECoG contact 1-54, showing a prominent narrowband high gamma component during the
331 speech production epoch. **C.** ECoG contact 2-55, showing a broadband activation during
332 speech production (note that this activity begins before speech onset). **D.** Spectrogram for LFP
333 signal extracted from the posterior micro electrode targeting the left STN. **E.** Spectrogram for

334 LFP signal from the medial macro ring targeting the left STN. **F**. Spectrogram for the left DBS
335 lead contact.

336 **The narrowband high gamma component correlates with the fundamental frequency of**
337 **the voice**

338 Because of the striking similarity between the neural data and speech spectrogram, we set out
339 to quantify the identified narrowband high gamma component and its relationship with the
340 produced audio further, using different analytical approaches. First, we asked whether the
341 observed overlap in the frequency range between the narrowband component and the produced
342 speech audio was consistent across participants. To this end, we calculated the Welch power
343 spectrum of the audio and neural data during the speech production epochs (Figure 3A),
344 identified at which frequencies peak power within the F0 range (70-240 Hz) occurs in both types
345 of spectra, and correlated the obtained frequency values. For each subject, we used data from
346 single trials of the LFP signal extracted from one of the micro channels of the subcortical
347 mapping electrodes, selected for having a prominent narrowband high gamma component.

348 As can be seen in Figure 3B, there is a strong correlation (Spearman's $\rho=0.98$, p -value
349 $< 10^{-6}$, intercept = 1.4 ± 5.1 Hz, slope = 0.99 ± 0.4) between the fundamental frequency of the
350 voice and the peak frequency of the narrowband component across participants. Furthermore,
351 the relation not only is linear, but also has a slope not significantly different from 1, meaning that
352 the frequency of peak high gamma power in the neural data corresponds to the fundamental
353 frequency of the voice.



354

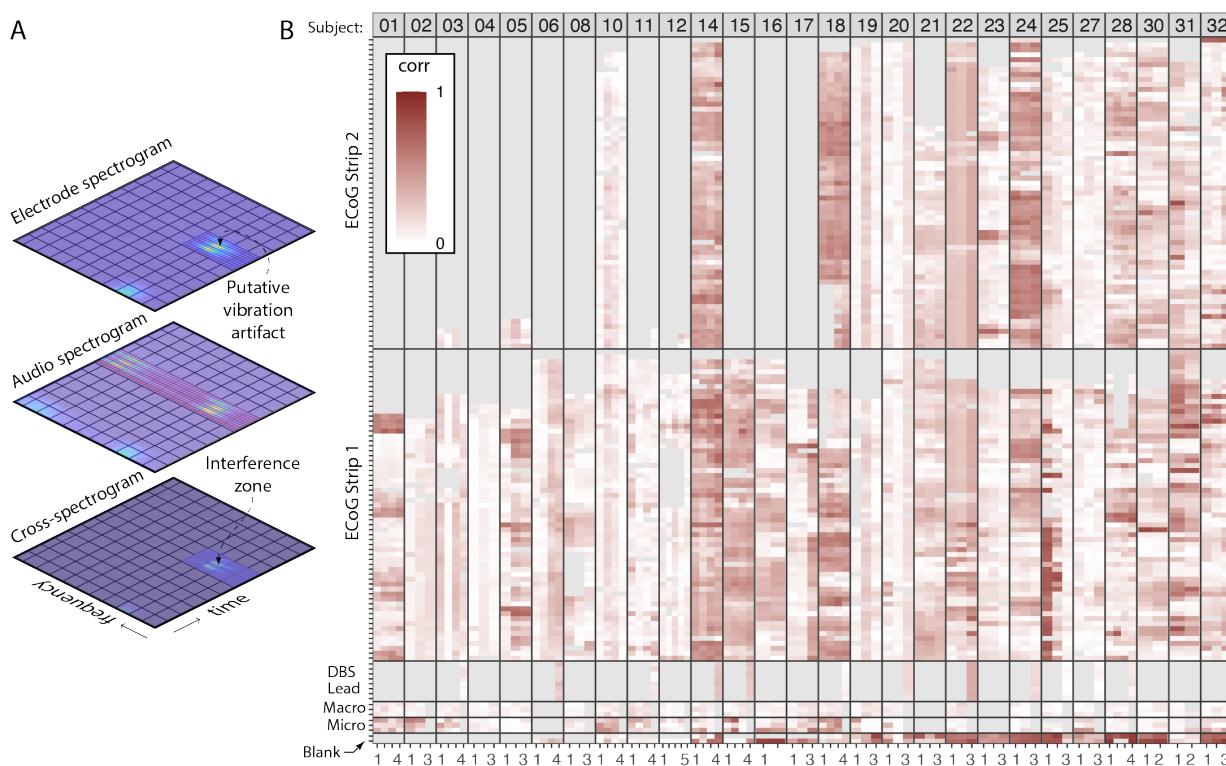
355 **Figure 3: The frequency of peak power of the high gamma narrowband component**
356 **correlates with the fundamental frequency of the voice across participants.**

357 A. Welch power spectrums for the produced speech audio (top) and the LFP signal extracted
358 from the micro tip of the subcortical mapping electrode by low-pass filtering at 400 Hz (bottom).
359 The vertical dashed blue line represents the frequency of peak power identified in each
360 spectrum. B. Correlation of the frequency of peak power in the range of 70-200 Hz between the
361 audio and neural data. Blue line represents the best linear fit to the data; gray ribbon represents
362 the confidence interval of the fit.

363 **Spectrogram cross-correlation between audio and neural data**

364 To further characterize the distribution of this narrowband component across electrodes and
365 recording sessions, we developed two different and complementary measures of similarity
366 between the neural signal and the audio signal. The first metric consists of correlating the time-
367 frequency spectrogram of a neural channel across times and frequencies with the spectrogram

368 of the produced audio. A schematic representation of the method is shown in Figure 4A. For
369 each subject, the correlation coefficient between the spectrogram of the produced audio and the
370 spectrogram of neural recordings was calculated (Equation 1). We found widespread correlation
371 between spectrograms obtained from intracranial recordings and spectrograms of the produced
372 audio, reaching correlation coefficients of up to 0.8 (Figure 4B). The strength of correlations was
373 variable across subjects, recording sessions, and electrodes. Also note that strong correlations
374 were present in different types of intracranial recordings, albeit with a varying consistency.



375

376 **Figure 4: Spectrogram correlation between audio and neural data.**
377 **A.** A schematic representation of the spectrogram correlation approach. The electrode
378 spectrogram (top) is multiplied by the audio spectrogram (middle) to give the cross-spectrogram
379 (bottom). **B.** Correlation coefficients (Equation 1) for each electrode (represented on the y-axis),
380 and each recording session (represented on the x-axis). Panels are defined by subject and
381 electrode type. Tiles in gray correspond to electrodes not present in the montage or which were
382 removed after the artifact rejection.

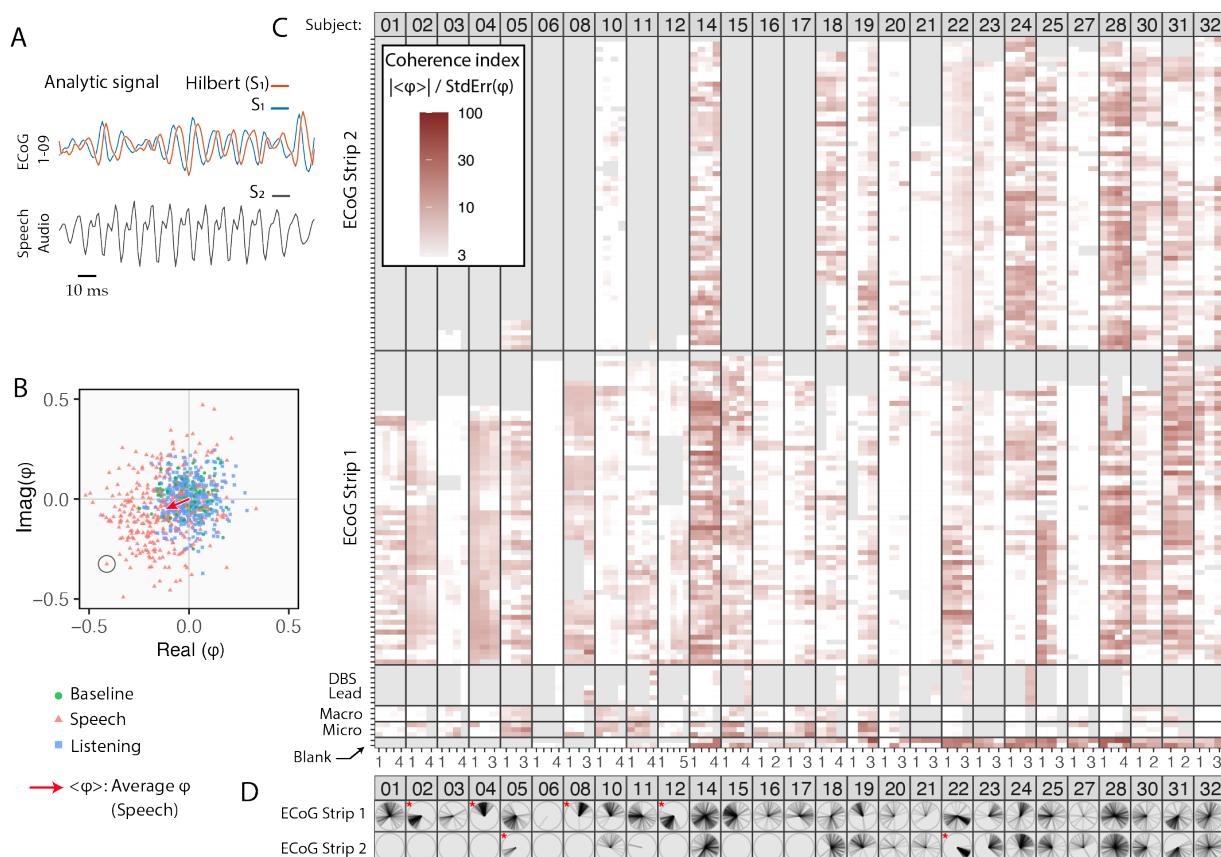
383 **Consistent phase relationship between audio and neural data**

384 The method described above compares similarity of time-frequency resolved *power* between
385 neural and audio data but does not take into account the phase information of the signals.
386 Therefore, we developed a complementary metric to quantify the *phase* relationship between
387 the neural data and the produced audio, that is, a measure of coherence or inter-trial phase
388 consistency (ITPC) with the produced audio. This metric is based on dot products of the analytic
389 signal of the neural channel and the produced audio, as schematized in Figure 5A. For each
390 epoch, a complex value representing the magnitude of the correlation and the phase
391 relationship of the neural signal with the audio is obtained (Equation 2). If the mean of these
392 complex values is significantly different from zero, this serves as evidence of a consistent phase
393 relationship between the neural data and the produced speech (Figure 5B). We used the
394 absolute value of the mean of these complex values, normalized by their standard error, as an
395 index of coherence (Equation 3). This method is computationally efficient and well suited for
396 narrowband signals as the one we are characterizing (see details in Methods Section). We
397 calculated a significance threshold using a Monte Carlo simulation in which we applied random
398 time jitters before calculating the coherence index (see details in Methods Section).

399 As observed in Figure 5C, there is widespread coherence across many subjects and all
400 electrode types during the speech production epoch. Around 50% of the analyzed electrodes
401 show significant coherence with the audio during the speech production epoch (Table 2).
402 Importantly, no significant coherence with the speech audio was observed during the baseline
403 or listening epochs (Table 2). Coherence of the neural signals with the stimulus audio in all
404 behavioral epochs, including the listening epoch, was negligible (Table 2). This suggests that
405 only the process of producing speech sounds was contaminating the neural signal.

406 The distribution of the coherence indices across electrodes for each recording session
407 can be classified as i) having no or little coherence between neural channels and the produced
408 speech (e.g. subjects 06, 03, 21); ii) having homogenous coherence across electrodes, (e.g.

409 subjects 02, 04); and iii) having heterogeneous coherence across electrodes (e.g. subjects 14,
410 28). While for the homogeneous case both the amplitude of the coherence ($|<\varphi>|$) and the
411 phase relationship with the produced speech are similar across the neural electrodes, in the
412 heterogeneous recording sessions these values may change substantially across electrodes
413 (Figure 5D).



414

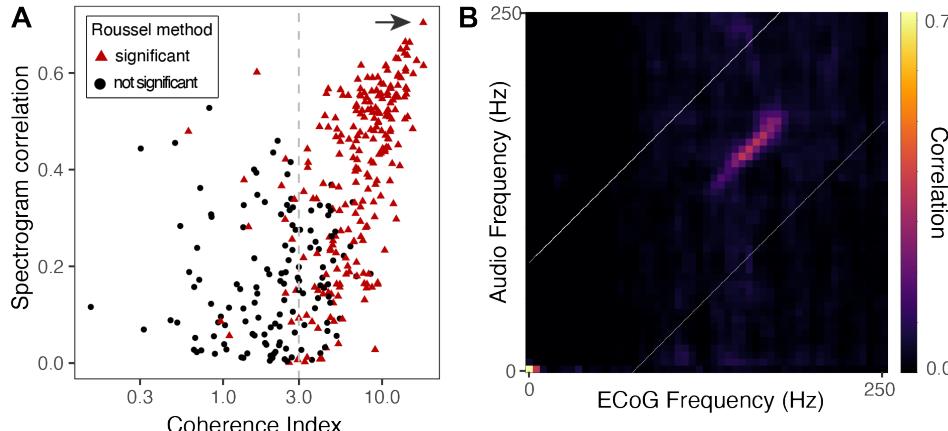
415 **Figure 5: Inter-trial phase consistency (ITPC) reveals significant coherence of neural data**
416 **with the produced speech audio.**

417 **A.** An ITPC index is derived from multiplying the band-pass filtered analytic neural signal (blue
418 and orange lines for real and imaginary parts) with the produced speech signal (black line) using
419 the internal product, resulting in a complex value φ for each trial. **B.** An example φ values
420 plotted on the complex plane (real part of φ on the x axis and imaginary part on the y axis) for
421 three behavioral epochs: baseline (green circles), listening (blue squares), and speech
422 production (red triangles). The ITPC index is related to the average value of $<\varphi_e>$,
423 represented as a red arrow for the speech production epoch. For the baseline and listening
424 epochs the $<\varphi_e> = 0$, indicating no phase consistency between the neural and the audio
425 data. The circled point in the bottom left quadrant of the plot corresponds to the traces shown in

426 panel A. **C.** ITPC indices from the speech production epoch plotted for each electrode, recording
427 session, and subject. Each row corresponds to an electrode; each column - to an individual
428 session; panels are defined by subject and data type. Tiles in gray correspond to electrodes that
429 were not present in the montage or which were removed after artifact rejection. White tiles
430 represent electrodes that do not show significant coherence indices **D.** Phase consistency of the
431 neural data with the produced audio across contacts of the ECoG strips. In each polar plot, the
432 angle between the radial lines and the horizontal axis represents the phase relation with the
433 produced audio of an ECoG electrode. Red asterisks mark ECoG strips with homogeneous
434 coherence, defined as those with at least 90% of the electrodes' phases within 90° of each
435 other.

436 **The speech acoustic component can be detected in the neural data by different
437 quantification methods**

438 The spectrogram correlation index (Figure 4) and the coherence index (Figure 5) are highly
439 correlated with each other, as can be observed in Figure 6A (Spearman's $\rho = 0.7$, $p < 1e-6$). In
440 applying the method proposed by Roussel et al. (2020), the power of the neural data was
441 correlated with the audio across different frequency bins to create a correlation matrix (Figure
442 6B). High correlation coefficients on the diagonal of this matrix indicate acoustic contamination
443 (see details in the Methods Section). Electrodes classified as contaminated based on this
444 method are indicated by red triangles in Figure 6A. As can be observed, these points tend to
445 have high spectrogram correlation index and coherence index above the significant threshold
446 (Figure 6A). This result suggests that all three methods have similar sensitivity in identifying
447 speech-related artifacts in our dataset.



448

449 **Figure 6: The spectrogram correlation index and the coherence index show strong**
 450 **correlation with each other and high consistency with the Roussel et al. method's**
 451 **outcome.**

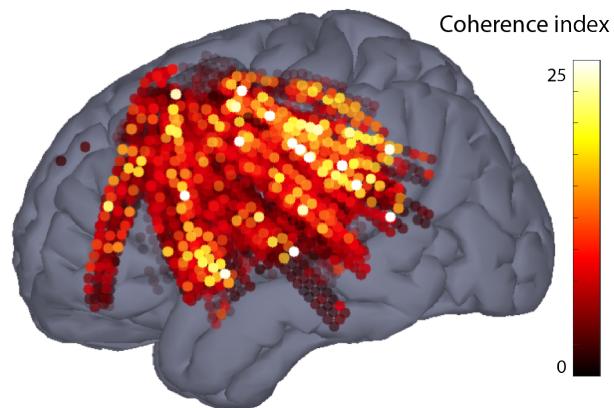
452 **A.** Relationship between the spectrogram correlation index and the coherence index for ECoG
 453 electrodes for one representative subject (DBS3024). Red triangles indicate electrodes with
 454 significant 'acoustic contamination' as assessed by the method developed by Roussel et al.,
 455 2020. **B.** Detection of acoustic contamination by the Roussel method is based on the cross-
 456 frequency correlation matrix, which indicates degrees of correlation of power across time for
 457 different frequencies of the neural and audio data. The shown matrix corresponds to ECoG
 458 electrode 2-31, indicated by an arrow in panel A.

Audio	Produced speech audio				Stimulus audio			
	Coherence		Roussel	Coherence		Roussel		
Method	baseline	listening	speech	all	baseline	listening	speech	all
ECoG	0.2±0.1%	0.2±0.1%	52±6%	54±5%	0.4±0.3%	0.2±0.1%	1.4±0.7%	2.2±0.5%
Macro	0%	0%	42±8%	25±5%	0%	0%	0%	4±2%
Micro	0%	0%	56±7%	96±7%	0%	0%	0%	0.6±0.8%
DBS Lead	0%	0%	47±10%	59±10%	6±6%	7±6%	10±8%	12±8%
Blank	10±4%	12±5%	65±8%	71±6%	2±2%	15±8	3±2%	12±7%

459 **Table 2.** Percentage (\pm standard error) of electrodes showing significant similarity between
 460 neural data and the produced speech audio or stimulus audio, as established by the coherence
 461 method and the Roussel et al.'s method. The coherence method was run independently for the
 462 baseline, listening and speech production epochs. The Roussel et al. 2020 method used the
 463 entire duration of the session. Hierarchical bootstrapping was used to estimate the standard
 464 error.

465 **Coherence indices for cortical recording sites are independent of anatomic location**

466 Next, we asked if there is any spatial structure of the coherence index within the cortex. To this
467 end, we performed a hierarchical bootstrapping analysis accounting for the nested nature of the
468 data (electrodes within anatomical regions within subjects) and found that none of the cortical
469 regions (as defined by the Multi Modal Parcellation 1 in Glasser et al., 2016) show values of
470 coherence significantly higher than the average distribution over the entire brain (see methods
471 for detail). All cortical electrodes with their corresponding coherence values are plotted on a
472 standard brain in Figure 7. Electrodes with high coherence with the produced audio do not
473 cluster on any specific neuroanatomical region.



474

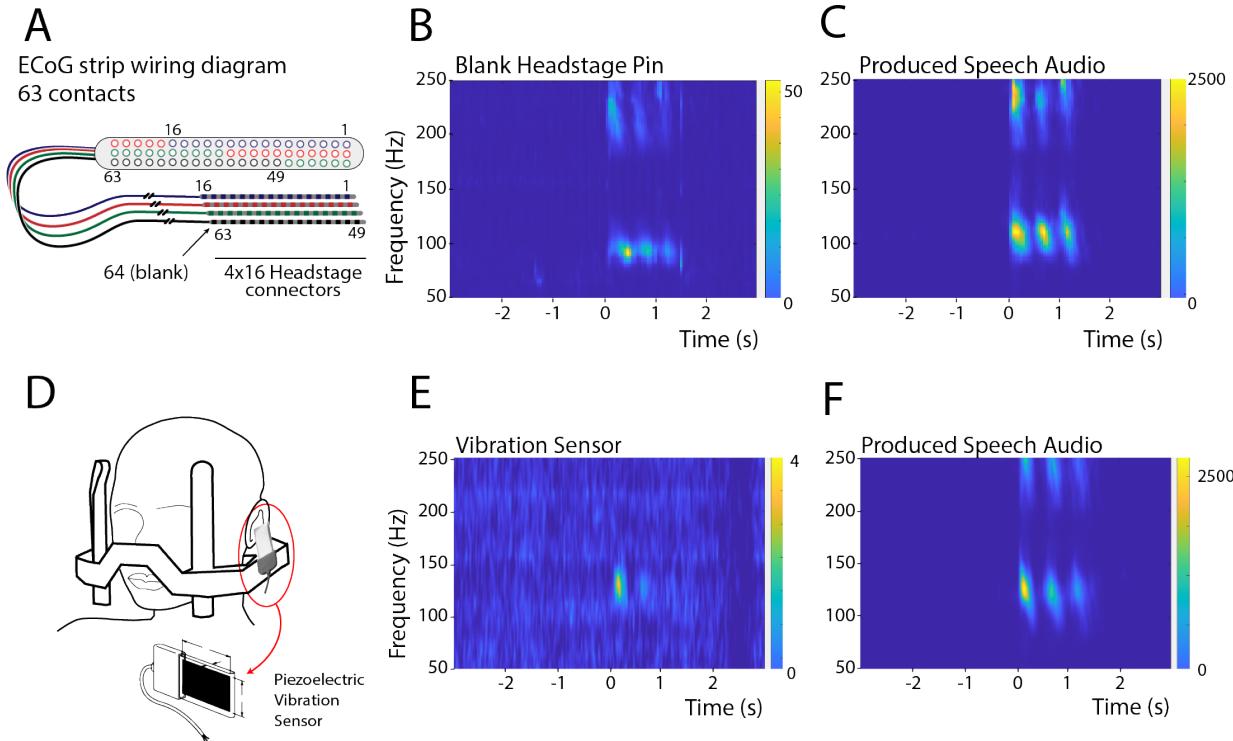
475 **Figure 7: Increased Inter-trial Phase Consistency is not specific to any cortical region.**
476 Localizations of cortical ECoG electrodes for the entire subject cohort ($n = 29$) plotted in MNI
477 space (MNI152 Nonlinear Asymmetric 2009b, (Fonov et al., 2011)). The color of the points
478 represents the average coherence between neural and produced audio data for that electrode.

479 **Speech-related vibrations can be detected in non-neural data**

480 The experiments described above demonstrated that some neural recordings show time-
481 frequency patterns similar to the produced audio recordings. This suggests that there is
482 contamination of the electrophysiological neural data with the speech audio signal. If this is the
483 case, we expect to find the same kind of contamination in "blank" electrodes not in contact with
484 the brain. Most of the ECoG strips used in our experiments contained 63 contacts laid out in a 3
485 x 21 arrangement (Figure 8A). These contacts were connected to the amplifier's front-end

486 through 4 cabrio-type connectors, each containing 16 pins. This resulted in the last pin of the 4th
487 cabrio connector (#64) not connected to any electrode on the brain. Instead, the wire connected
488 to this pin runs the length of the cable and ends within the silicon rubber matrix of the ECoG
489 strip. We recorded the signal from this "blank" pin as it provides a control for all non-neural
490 sources of noise affecting the neural signal. Using the same time-frequency method as before,
491 we analyzed the recorded signal from this blank ECoG headstage pin. As can be observed in
492 Figure 8B, a narrowband component similar to that observed in some neural recordings is also
493 present during speech production in the recordings from the blank headstage pin. It is
494 interesting to note that the frequency of peak power recorded from this pin is slightly lower than
495 the fundamental frequency of the voice (Figure 8C), although the component has the same
496 timing and overall pattern (note the power increase around the first harmonic of the F0). As
497 shown in the bottom row of the Figure 5C, labeled 'Blank', signals from the blank headstage
498 pins show significant coherence with the produced speech audio in the same recording
499 sessions that showed strong coherence with the produced speech audio in other electrode
500 types (see also Figure 4B and Table 2).

501 These results strongly suggest that the source of the observed narrowband component
502 is not neural. Among possible sources of this component are speech-induced vibrations of the
503 stereotactic frame, cables, connectors and/or acquisition chain. To address this question, we
504 attached a piezoelectric vibration sensor to the stereotactic frame (Figure 8D) during data
505 collection in one subject (subject DBS3029). The recorded signal from the piezoelectric sensor
506 was analyzed in the same way as other types of recordings. A strong narrowband component
507 was observed at the time corresponding to the production of the first syllable, in the same
508 frequency range as the fundamental frequency of the voice (Fig. 8E). The attenuated power
509 observed during the time of the production of the second and the third syllables might be due to
510 the fact that in this particular participant speech intensity decreased across the produced
511 syllable triplet, as seen in the audio spectrogram in Figure 8F.



512

513 **Figure 8: The narrowband component is present in non-neural recordings.**

514 **A.** A schematic of the ECoG electrode strip showing the electrode contact and connector layout.
515 4 cables of 16 wires are used for the 63 electrodes, resulting in one headstage pin not
516 connected to any electrode on the brain (marked as contact #64). **B.** Time-frequency plot of the
517 signal recorded from the blank headstage pin (subject DBS3020). **C.** Spectrogram of the
518 produced speech audio for the same subject as in B. **D.** A schematic representation of the
519 montage of the vibration sensor on the stereotactic frame. **E.** Time-frequency plot of the signal
520 recorded from the vibration sensor (subject DBS3029). **F.** Spectrogram of the produced speech
521 for the same subject as in E. In panels B-C and E-F, zero marks the onset of the speech
522 production.

523 **Discussion**

524 We identified the presence of a narrowband high gamma component in the neural signals
525 recorded from patients undergoing DBS implantation surgery that is consistent with a
526 mechanically induced artifact. This component is widespread across many electrode types and
527 was the most prominent feature in many electrode recordings. It occurs almost exclusively
528 during cued speech production epochs and it has spectral and temporal characteristics strikingly

529 similar to the produced speech audio. Indeed, a recent work shows the presence of acoustic
530 contamination on ECoG recordings (Roussel et al., 2020). In the environment of DBS surgery,
531 speech-induced vibrations conducted by the skull and stereotactic frame can impinge on the
532 electrodes and/or signal acquisition chain affecting the recorded signal.

533 Several results support the interpretation that the observed narrowband high gamma
534 component is an artifact due to speech-induced vibrations. First, the frequency of peak power of
535 the narrowband component tracks the fundamental frequency of the voice across participants
536 (Figure 3). Second, the narrowband high gamma component is almost exclusively present
537 during the cued speech production epoch, but not the listening epoch (Table 2). Thirdly, the
538 time-frequency resolved power from the neural recordings strongly correlates with the produced
539 audio spectrogram (Figure 4). Fourth, significant inter-trial phase consistency between the
540 produced audio and the neural data suggests similarities not only across the frequency domain,
541 but also the temporal domain (Figure 5). Fifth, most of the recordings that we classified as
542 contaminated in our analysis were also classified as having acoustic contamination by the
543 recently proposed method in Roussel et al. (2020) (Figure 6 and Table 2). Sixth, there was no
544 significant cortical localization of the ITPC index (Figure 7). Finally, the narrowband high gamma
545 component was also detected in "blank" pins of the headstage connector not connected to any
546 electrode (Figure 8B) and in the recording from a vibration sensor attached to the stereotactic
547 frame (Figure 8E). Taken together, these results strongly suggest the presence of speech-
548 induced vibration artifact.

549 Although we cannot rule out the presence of physiological neural activity with the exact
550 same spectral and timing characteristics as the observed narrowband high gamma component,
551 we favor the interpretation that the narrowband component identified in this work is mainly, if not
552 completely, due to the vibration artifact. It is worth mentioning that broadband gamma activity
553 can be detected in many electrodes (see example in Figure 2C), including some that also show
554 the vibration artifact.

555 These results are in line with a recent report by Roussel et al. (2020), in which the
556 authors show evidence for acoustic contamination of ECoG recordings. The authors analyze
557 multiple-center datasets collected in epilepsy patients in an extra-operative setting, finding that
558 both produced, and stimulus audio can affect ECoG recordings. Our work extends these
559 findings, showing that the same kind of contamination can occur not only in the ECoG
560 recordings, but can also affect other types of invasive neural recordings, such as those collected
561 in the context of DBS implantation surgery. In addition, there are two key differences to note
562 between our results and those of Roussel et al. First, we only observe contamination when the
563 patient is speaking, but not during auditory stimulus presentation. This difference may be due to
564 the use of headphones in our recording setup as opposed to the loudspeaker in the work by
565 Roussel et al., who found that acoustically isolating the speaker reduced contamination of the
566 neural signal. Second, in our recordings the affected signals are those corresponding to the F0
567 and to a lesser extent its harmonics. This could be explained by the fact that the stereotactic
568 frame imposes some additional mechanical constraints, thus changing the way the vibration
569 propagates.

570 Despite the fact that intracranial recordings in general are less prone to artifacts than
571 extracranial recordings, it is recognized that the patient's speech induces vibrations that affect
572 MER. For example, a patent for a new micro-electrode design from AlphaOmega mentions that
573 "*various mechanical noise and vibration, such as motor vibration, motion of the electrode within
574 the tissue or voice of the subject, are detected by the electrode that acts essentially as
575 microphone and is erroneously combined with the neural signal that is being recorded*" (Alpha
576 Omega Technologies, 2020). Therefore, a common practice has been to analyze the recordings
577 only when the patient is not speaking. The introduction of "microphonic free" microelectrodes
578 with improved shielding that reduces the effects of vibrations on the recordings, allows acquiring
579 single unit activity while the patient is speaking (Alpha Omega Technologies, 2020). Besides the
580 clear clinical advantage, this has opened the possibility of studying single unit activity of DBS

581 target structures during speech production. Although the quantifications of single-unit firing can
582 be reliably performed with these electrodes (Lipski et al., 2017, 2018), in light of our results it is
583 clear that mechanical vibrations affect other intracortical recordings including those obtained
584 with macro contacts on the shaft of the micro electrodes, recordings from ECoG strips and
585 recordings from the clinically implanted DBS leads.

586 Interestingly, neural signals with similar characteristics to the narrowband high gamma
587 component described in this work have been reported in the literature. The Frequency Following
588 Response (FFR) in the auditory brainstem is a potential with spectrotemporal features
589 resembling the stimulus audio that reflects sustained neural ensemble activity phase-locked to
590 periodic acoustic stimuli (Bidelman, 2018; Marsh and Worden, 1968; Marsh et al., 1970; Rose et
591 al., 1966). This brainstem response to auditory stimuli can be recorded from scalp electrodes by
592 averaging over hundreds of trials, a technique that has become a powerful diagnostic tool in
593 audiology and neurology known as the Auditory Brainstem Response (Hall, 1992; Jewett et al.,
594 1970). In recent years several works based on scalp EEG, sEEG and MEG have reported
595 cortical FFRs (Behroozmand et al., 2016; Bidelman, 2018; Coffey et al., 2016).

596 In our data, there are two features that argue against the narrowband component being
597 a true FFR. First, it only occurs during speech production and not during auditory stimulus
598 presentation. Second, the narrowband artifact is not localized to the auditory cortex, or to any
599 other cortical region (Figure 7). In light of the results presented in this work and similar results
600 recently published by Roussel et al, it is clear that caution should be taken when interpreting
601 cortical FFR, due to the fact that this signal is exactly what would be expected if there was an
602 artifact (e.g., vibration at electrodes or connectors, electromagnetic induction by speakers,
603 electrical crosstalk in the amplifier or connectors).

604 Other reported sources of artifacts affecting intracranial recordings, including artifacts
605 due to eye blinking (Ball et al., 2009) and other facial muscle contraction (Otsubo et al., 2008),
606 have spectral characteristics distinct from the narrowband component described in this work;

607 they show broadband spectrums that do not match that of the produced audio and are not
608 expected to track the fundamental frequency of the voice or correlate with the power of the
609 speech audio across time.

610 Acknowledging the existence of this speech-induced vibration artifact is an important first
611 step to avoid overinterpreting spurious features of the data. Many important questions related to
612 high-level cognitive processes, including the neural control of speech production, can only be
613 answered by acquiring recordings that are likely to be affected by the described artifact, but
614 several steps can be taken to identify it. First, methods that correlate the audio signal with the
615 neural data can be used to detect the presence of acoustic contaminations. Second, recording
616 from open headstage pins provides a control for non-neural sources affecting the signal. Third,
617 vibration sensors can detect mechanical vibrations along the recording system that might affect
618 the signal. Furthermore, invasive recordings for BCI applications are likely to be affected by
619 speech-induced vibrations, compromising the decoding performance. Therefore, it is necessary
620 to develop source separation methods to remove speech-related artifacts from the neural data
621 in order to reliably quantify underlying neural activity.

622 Identifying and mitigating artifacts in intracranial recordings from awake patients is
623 fundamental to achieving reliable and reproducible results in the field of human systems
624 neuroscience. This, in turn, will lead to an improved understanding of the neural physiology and
625 pathophysiology of uniquely human cognitive abilities.

626 Acknowledgments

627 We would like to thank all participants of this study for their time and effort. We would also like
628 to thank Robert S. Turner for suggestions on data analysis. This work was funded by the
629 National Institute of Health (BRAIN Initiative), through grants U01NS098969, U01NS117836
630 and R01NS110424 to R.M.R.

631 Author Contributions

632 W.J.L. and R.M.R. designed the experiment and performed the recordings. A.B., A.C., C.D.-H.
633 and W.J.L. reconstructed the electrode localizations and preprocessed the electrophysiological
634 data. A.C. and C.D.-H. overviewed the phonetic coding. A.B., A.C., V.P., V.S., C.D.-H. analyzed
635 data. A.B. and A.C. wrote the manuscript. All authors discussed the results and implications and
636 commented on the manuscript at all stages. The authors declare no competing financial
637 interests.

638 References

- 639 Alpha Omega Technologies. (2020). Reduced Larsen Effect Electrode.
- 640 Anumanchipalli, G.K., Chartier, J., and Chang, E.F. (2019). Speech synthesis from neural
641 decoding of spoken sentences. *Nature* **568**, 493–498.
- 642 Avants, B.B., Epstein, C.L., Grossman, M., and Gee, J.C. (2008). Symmetric diffeomorphic
643 image registration with cross-correlation: Evaluating automated labeling of elderly and
644 neurodegenerative brain. *Med Image Anal* **12**, 26–41.
- 645 Ball, T., Kern, M., Mutschler, I., Aertsen, A., and Schulze-Bonhage, A. (2009). Signal quality of
646 simultaneously recorded invasive and non-invasive EEG. *Neuroimage* **46**, 708–716.
- 647 Behroozmand, R., Oya, H., Nourski, K.V., Kawasaki, H., Larson, C.R., Brugge, J.F., Howard,
648 M.A., and Greenlee, J.D.W. (2016). Neural correlates of vocal production and motor
649 control in human Heschl's gyrus. *J Neurosci* **36**, 2302–2315.

- 650 Benjamini, Y., and Hochberg, Y. (1994). Controlling the false discovery rate: a practical and
651 powerful approach to multiple testing. *Journal of the Royal statistical society: series B*
652 (Methodological), 57(1), 289-300.
- 653 Bidelman, G.M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-
654 following response to speech. *Neuroimage* 175, 56–69.
- 655 Bouchard, K.E., Mesgarani, N., Johnson, K., and Chang, E.F. (2013). Functional organization of
656 human sensorimotor cortex for speech articulation. *Nature* 495, 327–332.
- 657 Cheung, C., Hamiton, L.S., Johnson, K., and Chang, E.F. (2016). The auditory representation of
658 speech sounds in human motor cortex. *Elife* 5, e12577.
- 659 Chrabaszcz, A., Neumann, W.-J., Stretcu, O., Lipski, W.J., Bush, A., Dastolfo-Hromack, C.A.,
660 Wang, D., Crammond, D.J., Shaiman, S., Dickey, M.W., et al. (2019). Subthalamic
661 nucleus and sensorimotor cortex activity during speech production. *J Neurosci* 39, 2698--
662 2708.
- 663 Coffey, E.B.J., Herholz, S.C., Chepesiuk, A.M.P., Baillet, S., and Zatorre, R.J. (2016). Cortical
664 contributions to the auditory frequency-following response revealed by MEG. *Nat
665 Commun* 7, 11070.
- 666 Cohen, M.X. (2014). *Analyzing Neural Time Series Data*. MIT Press.
- 667 Conant, D.F., Bouchard, K.E., Leonard, M.K., and Chang, E.F. (2018). Human Sensorimotor
668 Cortex Control of Directly Measured Vocal Tract Movements during Vowel Production. *J
669 Neurosci* 38, 2955–2966.
- 670 Crone, N.E., Hao, L., Hart, J., Boatman, D., Lesser, R.P., Irizarry, R., and Gordon, B. (2001).
671 Electrocorticographic gamma activity during word production in spoken and sign
672 language. *Neurology* 57, 2045–2053.
- 673 Edwards, E., Nagarajan, S.S., Dalal, S.S., Canolty, R.T., Kirsch, H.E., Barbaro, N.M., and
674 Knight, R.T. (2010). Spatiotemporal imaging of cortical activation during verb generation
675 and picture naming. *Neuroimage* 50, 291–301.
- 676 Ewert, S., Plettig, P., Li, N., Chakravarty, M.M., Collins, D.L., Herrington, T.M., Kühn, A.A., and
677 Horn, A. (2018). Toward defining deep brain stimulation targets in MNI space: A
678 subcortical atlas based on multimodal MRI, histology and structural connectivity.
679 *Neuroimage* 170, 271–282.
- 680 Faraji, A.H., Kokkinos, V., Sweat, J.C., Crammond, D.J., and Richardson, R.M. (2020). Robotic-

- 681 Assisted Stereotaxy for Deep Brain Stimulation Lead Implantation in Awake Patients.
682 Oper Neurosurg 19, 444–452.
- 683 Flinker, A., Korzeniewska, A., Shestyuk, A.Y., Franaszczuk, P.J., Dronkers, N.F., Knight, R.T.,
684 and Crone, N.E. (2015). Redefining the role of Broca's area in speech. Proc National Acad
685 Sci 112, 2871–2875.
- 686 Flinker, A., Piai, V., and Knight, R.T. (2018). Intracranial electrophysiology in language
687 research, in The Oxford handbook of psycholinguistics (2dn ed.). 992–1010.
- 688 Fonov, V., Evans, A.C., Botteron, K., Almlí, C.R., McKinstry, R.C., Collins, D.L., and Group, the
689 B.D.C. (2011). Unbiased Unbiased average age-appropriate atlases for pediatric studies.
690 Neuroimage, 54(1), 313-327.
- 691 Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K.,
692 Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of
693 human cerebral cortex. Nature 536, 171–178.
- 694 Hall, J.W. (1992). Handbook of auditory evoked responses. Allyn & Bacon.
- 695 Hamilton, L.S., Edwards, E., and Chang, E.F. (2018). A Spatial Map of Onset and Sustained
696 Responses to Speech in the Human Superior Temporal Gyrus. Curr Biol 28, 1860–
697 1871.e4.
- 698 Horn, A., and Kühn, A.A. (2015). Lead-DBS: A toolbox for deep brain stimulation electrode
699 localizations and visualizations. Neuroimage 107, 127–135.
- 700 Jewett, D.L., Romano, M.N., and Williston, J.S. (1970). Human auditory evoked potentials:
701 possible brain stem components detected on the scalp. Science, 167(3924), 1517-1518.
- 702 Lachaux, J.P., Rudrauf, D., and Kahane, P. (2003). Intracranial EEG and human brain mapping.
703 J Physiology-Paris 97, 613–628.
- 704 Lipski, W.J., Wozny, T.A., Alhourani, A., Kondylis, E.D., Turner, R.S., Crammond, D.J., and
705 Richardson, R.M. (2017). Dynamics of human subthalamic neuron phase-locking to motor
706 and sensory cortical oscillations during movement. J Neurophysiol 118, 1472–1487.
- 707 Lipski, W.J., Alhourani, A., Pirnia, T., Jones, P.W., Dastolfo-Hromack, C., Helou, L.B.,
708 Crammond, D.J., Shaiman, S., Dickey, M.W., Holt, L.L., et al. (2018). Subthalamic nucleus
709 neurons differentially encode early and late aspects of speech production. J Neurosci 38,
710 5620–5631.

- 711 Llorens, A., Trébuchon, A., Liégeois-Chauvel, C., and Alario, F.-X. (2011). Intra-Cranial
712 Recordings of Brain Activity During Language Production. *Front Psychol* 2, 375.
- 713 Marsh, J.T., and Worden, F.G. (1968). Sound evoked frequency-following responses in the
714 central auditory pathway. *Laryngoscope* 78, 1149–1163.
- 715 Marsh, J.T., Worden, F.G., and Smith, J.C. (1970). Auditory Frequency-Following Response:
716 Neural or Artifact? *Science* 169, 1222–1223.
- 717 Martin, S., Millán, J. del R., Knight, R.T., and Pasley, B.N. (2019). The use of intracranial
718 recordings to decode human language: Challenges and opportunities. *Brain Lang* 193,
719 73–83.
- 720 Mesgarani, N., Cheung, C., Johnson, K., and Chang, E.F. (2014). Phonetic Feature Encoding in
721 Human Superior Temporal Gyrus. *Science* 343, 1006–1010.
- 722 Mugler, E.M., Tate, M.C., Livescu, K., Templer, J.W., Goldrick, M.A., and Slutsky, M.W. (2018).
723 Differential representation of articulatory gestures and phonemes in precentral and inferior
724 frontal gyri. *J Neurosci* 38, 9803–9813.
- 725 Mukamel, R., and Fried, I. (2012). Human Intracranial Recordings and Cognitive Neuroscience.
726 *Annu Rev Psychol* 63, 511–537.
- 727 Niemann, K., Mennicken, V.R., Jeanmonod, D., and Morel, A. (2000). The Morel Stereotactic
728 Atlas of the Human Thalamus: Atlas-to-MR Registration of Internally Consistent Canonical
729 Model. *Neuroimage* 12, 601–616.
- 730 Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open Source
731 Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data.
732 *Comput Intel Neurosc* 2011, 156869.
- 733 Otsubo, H., Ochi, A., Imai, K., Akiyama, T., Fujimoto, A., Go, C., Dirks, P., and Donner, E.J.
734 (2008). High-frequency oscillations of ictal muscle activity and epileptogenic discharges on
735 intracranial EEG in a temporal lobe epilepsy patient. *Clin Neurophysiol* 119, 862–868.
- 736 Panov, F., Levin, E., Hemptinne, C. de, Swann, N.C., Qasim, S., Miocinovic, S., Ostrem, J.L.,
737 and Starr, P.A. (2017). Intraoperative electrocorticography for physiological research in
738 movement disorders: principles and experience in 200 cases. *J Neurosurg* 126, 122–131.
- 739 Randazzo, M.J., Kondylis, E.D., Alhourani, A., Wozny, T.A., Lipski, W.J., Crammond, D.J., and
740 Richardson, R.M. (2016). Three-dimensional localization of cortical electrodes in deep

- 741 brain stimulation surgery from intraoperative fluoroscopy. *Neuroimage* 125, 515–521.
- 742 Rose, J.R., Brugge, J.F., Anderson, D.J., and Hind, J.E. (1966). Phase-locked response to low-
743 frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of*
744 *neurophysiology* 30.4: 769-793.
- 745 Roussel, P., Godais, G.L., Bocquelet, F., Palma, M., Hongjie, J., Zhang, S., Giraud, A.-L.,
746 Mgevand, P., Miller, K., Gehrig, J., et al. (2020). Observation and assessment of acoustic
747 contamination of electrophysiological brain signals during speech production and sound
748 perception. *J Neural Eng* 17, 056028.
- 749 Saravanan, V., Berman, G.J., and Sober, S.J. (2020). Application of the hierarchical bootstrap
750 to multi-level data in neuroscience. Arxiv.
- 751 Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., and Wolpaw, J.R. (2004).
752 BCI2000: a general-purpose brain-computer interface (BCI) system. *Ieee T Bio-Med Eng*
753 51, 1034–1043.
- 754 Sisterson, N.D., Carlson, A.A., Rutishauser, U., Mamelak, A.N., Flagg, M., Pouratian, N.,
755 Salimpour, Y., Anderson, W.S., and Richardson, R.M. (2021). Electrocorticography during
756 deep brain stimulation surgery: safety experience from 4 centers within the national
757 institute of neurological disorders and stroke research opportunities in human consortium.
758 neurosurgery.