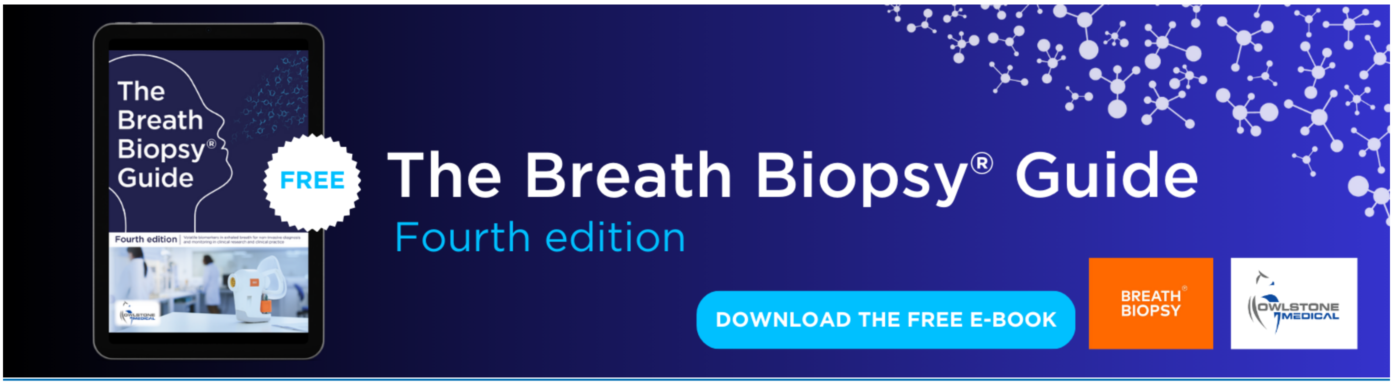
Journal of Neural Engineering

Observation and assessment of acoustic contamination of electrophysiological brain signals during speech production and sound perception



Journal of Neural Engineering

OPEN ACCESS

RECEIVED
22 December 2019

REVISED
19 August 2020

ACCEPTED FOR PUBLICATION
25 August 2020

PUBLISHED
13 October 2020

PAPER

Observation and assessment of acoustic contamination of electrophysiological brain signals during speech production and sound perception

Philémon Roussel[1,2], Gaël Le Godais[1,2,3], Florent Bocquelet[1,2], Marie Palma[1,2], Jiang Hongjie[4], Shaomin Zhang[5], Anne-Lise Giraud[6], Pierre Mégevand[6,7], Kai Miller[8], Johannes Gehrig[9], Christian Kell[9], Philippe Kahane[10], Stéphan Chabardés[11] and Blaise Yvert[1,2,12]

[1] Inserm, BrainTech Lab, U1205, Grenoble, France
[2] University Grenoble Alpes, BrainTech Lab, U1205, Grenoble, France
[3] University Grenoble Alpes, Gipsa-Lab, Grenoble, France
[4] Zhejiang University, Department of Neurosurgery, Hangzhou, People's Republic of China
[5] Zhejiang University, Qiushi Academy for Advanced Studies, Hangzhou, People's Republic of China
[6] University of Geneva, Faculty of Medicine, Department of Basic Neuroscience, Geneva, Switzerland
[7] Geneva University Hospitals, Division of Neurology, Geneva, Switzerland
[8] Mayo Clinic, Department of Neurosurgery, Rochester, MN, United States of America
[9] Goethe University Frankfurt, Department of Neurology and Center for Personalized Translational Epilepsy Research, Frankfurt, Germany
[10] CHU Grenoble Alpes, Department of Neurology, Grenoble, France
[11] CHU Grenoble Alpes, Department of Neurosurgery, Grenoble, France

E-mail: blaise.yvert@inserm.fr

Keywords: brain-computer interface, speech prostheses, decoding, microphonic effect, electrophysiology, ECoG, intracortical recordings, high-gamma,

Abstract

Objective. A current challenge of neurotechnologies is to develop speech brain-computer interfaces aiming at restoring communication in people unable to speak. To achieve a proof of concept of such system, neural activity of patients implanted for clinical reasons can be recorded while they speak. Using such simultaneously recorded audio and neural data, decoders can be built to predict speech features using features extracted from brain signals. A typical neural feature is the spectral power of field potentials in the high-gamma frequency band, which happens to overlap the frequency range of speech acoustic signals, especially the fundamental frequency of the voice. Here, we analyzed human electrocorticographic and intracortical recordings during speech production and perception as well as a rat microelectrocorticographic recording during sound perception. We observed that several datasets, recorded with different recording setups, contained spectrotemporal features highly correlated with those of the sound produced by or delivered to the participants, especially within the high-gamma band and above, strongly suggesting a contamination of electrophysiological recordings by the sound signal. This study investigated the presence of acoustic contamination and its possible source. *Approach.* We developed analysis methods and a statistical criterion to objectively assess the presence or absence of contamination-specific correlations, which we used to screen several datasets from five centers worldwide. *Main results.* Not all but several datasets, recorded in a variety of conditions, showed significant evidence of acoustic contamination. Three out of five centers were concerned by the phenomenon. In a recording showing high contamination, the use of high-gamma band features dramatically facilitated the performance of linear decoding of acoustic speech features, while such improvement was very limited for another recording showing no significant contamination. Further analysis and *in vitro* replication suggest that the contamination is caused by the mechanical action of the sound

[12] Author to whom any correspondence should be addressed.

waves onto the cables and connectors along the recording chain, transforming sound vibrations into an undesired electrical noise affecting the biopotential measurements. *Significance.* Although this study does not *per se* question the presence of speech-relevant physiological information in the high-gamma range and above (multiunit activity), it alerts on the fact that acoustic contamination of neural signals should be proofed and eliminated before investigating the cortical dynamics of these processes. To this end, we make available a toolbox implementing the proposed statistical approach to quickly assess the extent of contamination in an electrophysiological recording (https://doi.org/10.5281/zenodo.3929296).

1. Introduction

The development of brain-computer interfaces (BCI) to restore speech (Guenther *et al* 2009, Brumberg *et al* 2010, Leuthardt *et al* 2011) is a long-term quest that seems within possible reach. Several advances have indeed been made over the past decade regarding the decoding of intracranial brain signals underlying either speech perception (Pasley *et al* 2012, Chan *et al* 2013, Pasley and Knight 2013, Fontolan *et al* 2014, Hyafil *et al* 2015, Yildiz *et al* 2016, Akbari *et al* 2019) or production (Angrick *et al* 2019, Miller *et al* 2011, Bouchard *et al* 2013, Martin *et al* 2014, 2016, Mugler *et al* 2014, 2018, Herff *et al* 2015, 2019, Cheung *et al* 2016, Chartier *et al* 2018, Anumanchipalli *et al* 2019), and most recent works have tackled with noticeable success the prediction of continuous speech from ongoing brain activity. Because of the difficulty to record from individual neurons with microelectrodes inserted in human speech areas (Bartels *et al* 2008, Kennedy *et al* 2011, Tankus *et al* 2012, Chan *et al* 2013), most of speech decoding studies use field potential signals in the high-gamma frequency range, which typically covers frequencies from 70 to 200 Hz. However, a recent study in patients implanted primarily for limb motor BCI purposes, indicate that speech could also be decoded from intracortical multiunit activity recorded in the hand knob motor cortex, a region not previously described to encode speech production (Stavisky *et al* 2019).

A noticeable feature of acoustic speech signals is the fundamental frequency f_0 of the human voice, which corresponds to the vibrational source of speech produced by the vocal folds in the larynx and further modulated by the vocal tract to produce the variety of speech sounds. The fundamental frequency depends on the size of the vocal folds and typically falls around 125 Hz for males and 215 Hz for women (Small 2012). The high-gamma frequency band and the range of the fundamental speech frequency thus generally overlap. At frequencies above f_0, the acoustic content of speech is further characterized by the harmonics of the fundamental frequency, which typically span frequencies overlapping those of unit and multiunit neural activity.

Here, we analyzed human electrocorticographic (ECoG) and intracortical recordings during speech production and perception as well as a rat microelectrocorticographic (μ-ECoG) recording during sound perception. We found that electrophysiological signals may often be contaminated by spectrotemporal features of the sound produced by the participant's voice or played by the loudspeaker. This contamination seems to result from a microphonic effect at the level of the cables and connectors along the recording chain, affecting the range of high-gamma frequencies and above. These findings suggest that care should be taken to exclude the presence of such artifacts when investigating cortical signals underlying speech production and perception.

2. Methods

2.1. Human recordings

2.1.1. Participants

The present study was conducted as part of the Brainspeak clinical trial (NCT02783391) approved by the French regulatory agency ANSM (DMDPTTECH/MM/2015-A00108-41) and the local ethical committee (CPP-15-CHUG-12). It is primarily based on electrophysiological recordings obtained in 3 patients at the Grenoble University Hospital: a 42-year-old (P2) and a 29-year-old (P3) males undergoing awake surgery for tumor resection, and a 38-year-old female (P5) implanted for 7 days as part of a presurgical evaluation of her intractable epilepsy. These three participants gave their informed consent to participate in the study.

We further included datasets obtained by four other centers in China, Germany, Switzerland and the USA. Firstly, a 22-year-old male participant (CN) suffering from intractable epilepsy requiring surgical treatment was recorded at the Second Affiliated Hospital of Zhejiang University. These procedures were followed from the guide and approved by the Second A liated Hospital of Zhejiang University, China. Participant CN gave written informed consent after detailed explanation of the potential risks of the research experiment. Secondly, two additional datasets were acquired at Frankfurt University. A 33-year-old bilingual Russian and English-speaking male patient suffering from a left frontal anaplastic astrocytoma (D1) and a 36-year-old native German speaking female suffering from an anaplastic glioma (D2) were recorded. Participants D1 and D2 gave written informed consent after detailed explanation of the potential risks of the research experiment, which were approved by the ethics committee of the medical faculty of Goethe University (GZ 310/11). Thirdly, two patients with drug-resistant epilepsy were recorded extra-operatively at Geneva University Hospitals: a 49-year-old woman (CH1) and a 50-year-old man (CH2). Both gave written consent to participate in speech production and processing experiments, which were approved by the local ethics committee (*Commission cantonale d'éthique de la recherche*, Geneva, Switzerland). Finally, datasets were recorded from three patients, a woman aged 18 (US1), a man aged 42 (US2), and a man aged 21 (US3), at the University of Washington (Seattle, WA, USA). All three patients participated in a purely voluntary manner, after providing informed written consent, under experimental protocols approved by the Institutional Review Board of the University of Washington (#12 193). Patient data was anonymized according to IRB protocol, in accordance with HIPAA mandate (Miller 2019).

2.1.2. Electrophysiological recordings

Brain activity from participants P2, P3, D1, and D2 was recorded during awake surgery in the operating room before tissue resection. For participant P2, a 256-electrode array (PMT Corp., USA) was positioned after opening the skull and the dura matter over the left sensorimotor cortex and the tumor (figure 1). Ground and reference electrodes were integrated on the back side of the array and maintained wet using compresses soaked with saline. The 16 electrodes' pigtails were connected to eight 32-channels Cabrio Connectors (Blackrock Microsystems, USA) connected by shielded cables to two front-end amplifiers (FEA, Blackrock Microsystems, USA) for amplification and digitalization at 10 kHz. The digitized signals were then transmitted by an optic fiber to two synchronized Neural Signal Processors (NSP, Blackrock Microsystems, USA) interfaced with a computer. For participant P3, a 96-channel intracortical Utah microelectrode array (UEA, Blackrock Microsystems, USA) was inserted in the pars triangularis of Broca's area (figure 1), at a location that was subsequently resected to access the tumor for its removal. The pedestal serving as ground was screwed to the skull. Two wires with deinsulated tips were inserted below the dura, and one was used as reference. The electrodes were connected via a Patient Cable (Blackrock Microsystems, USA) to a FEA where signals were digitized at 30 kHz and further transmitted through an optic fiber to a NSP. Participant D1 received left perisylvian electrocorticography using an 8 × 12 electrode grid and two 2 × 8 electrode grids with 5 mm spacing (Ad-Tech Medical, USA). The ground electrode was the subgaleal needle electrode P4 and data were referenced against a frontocentral subgaleal needle electrode Fz and sampled at 5 kHz (four amplifiers connected to two headboxes; BrainAmp MR plus amplifier, BrainProducts, Germany). Participant D2 received left perisylvian electrocorticography using an 8 × 12 grid electrode (Ad-Tech Medical, USA). The ground electrode was the subgaleal needle electrode C4 and data were referenced against a frontocentral subgaleal needle electrode Fz and sampled at 2048 Hz (two amplifiers 64 channels each, one headbox per amplifier, Micromed, Italy).

Brain activity from participants P5, CN, CH1, CH2, US1, US2, and US3 was recorded in subchronic condition at the hospital. Participant P5 was implanted with a 72-electrode ECoG array (PMT Corp., USA) covering a large portion of her left hemisphere as well as a 4-electrode strip (PMT Corp., USA) over the left ventral temporal lobe and a 96-electrode UEA inserted in the left ventral sensorimotor cortex (figure 1). An electrode of the strip was used as the reference and another as the ground. The transcutaneous pigtails of the ECoG grids were connected to PMT pigtail adaptors and then to two headboxes (64-Channel Splitter Box, Blackrock Microsystems, USA) through individual touch-proof connectors. The headboxes were connected to a FEA linked to a NSP. The transcutaneous pedestal of the UEA was screwed to the skull and connected to a Cereplex-E headstage (Blackrock Microsystems, USA) ensuring signal amplification and digitization before transmission to a second NSP through a digital hub. For these intracortical recordings, the reference was a wire deinsulated at its tip and inserted below the dura, and the ground was the pedestal. Data from both electrode arrays was sampled at 30 kHz and recorded on the two synchronized NSPs. Participant CN was implanted with a 32-electrode clinical subdural ECoG grid (HuakeHesheng, China) in his right sensorimotor cortex for clinical monitoring and localization of his seizure foci (figure 1). The clinical electrodes were platinum electrodes with a diameter of 4 mm (2.3 mm exposed) spaced every 10 mm, implanted for seven days. The configuration and location of the electrodes, as well as the duration of the implantation, were determined by clinical requirements. An electrode of the ECoG grid was used as the reference and another as the ground. The transcutaneous pigtails of the ECoG grids were connected to HuakeHesheng adaptors. The adaptors were connected to a headbox linked to a NSP via an FEA, like for P5. The signals from the ECoG grid were sampled at 30 kHz. Participant CH1 was implanted with grids and strips of subdural electrodes over the right cerebral hemisphere (124 recording sites; Ad-Tech Medical, USA). Participant CH2 was implanted with depth electrodes through stereotaxic surgery in both cerebral hemispheres (213 recording sites; Dixi Medical, France). EEG signals for CH1 and CH2 were referenced to a subdermal wire electrode inserted at the vertex, digitized at 2048 Hz and recorded with a BrainQuick LTM system (Micromed, Italy).

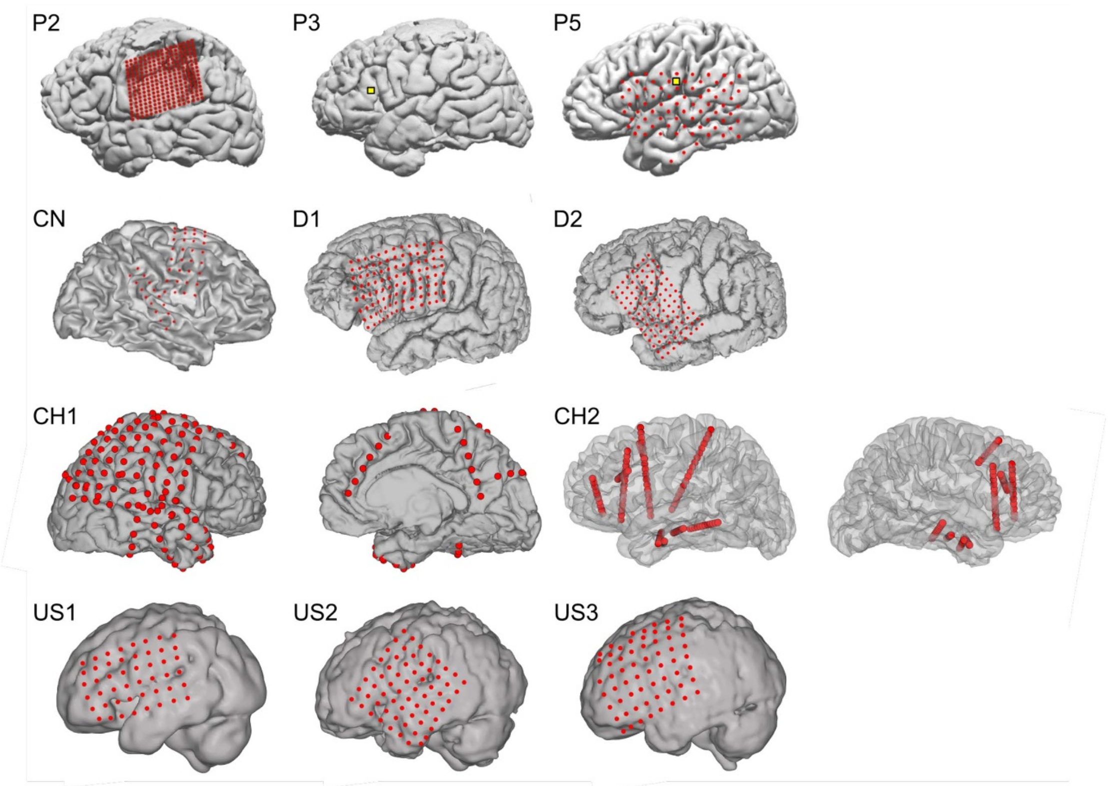


Figure 1. Electrode positions for the different human participants considered in this study. Red: ECoG electrodes (all but P3). Yellow: intracortical Utah array (P3 and P5).

The Seattle participants all had subdural platinum electrode arrays (Ad-Tech Medical, USA) implanted over the left hemisphere in 6 × 8 (US1) and 8 × 8 (US2 and US3) rectangular arrays. These electrodes had 4 mm diameter (2.3 mm exposed), 1 cm inter-electrode distance, and were embedded in silastic. These data were recorded with Synamps2 amplifiers (Compumedics Neuroscan, USA) in parallel with clinical recording, sampled at 2 kHz (US1) or 1 kHz (US2 and US3). The data were exported from the amplifiers to the BCI2000 software environment on a separate laptop, using a TCP/IP protocol (Schalk *et al* 2004).

2.1.3. Audio recordings

In case of speech production tasks, the participant speech was recorded along with his or her neural data. For participants P2, P3 and P5, a microphone (SHURE Beta 58 A) was positioned at about 10–20 cm from the mouth. The signal was amplified using an audio interface (Roland OCTA-CAPTURE) and digitized by one of the NSPs, at the same rate and synchronously with the neural data (see figure 2(a)). For participant CN, the played sentences were realigned with the neural signals using triggers indicating the start of the stimuli. D1's and D2's voices were recorded using a custom-made microphone (BrainProducts, Germany) connected to two bipolar EMG channels of the headboxes. For participants CH1 and CH2, the patient's produced speech or the sound delivered by the stimulus presentation portable computer positioned in front of the patient was captured with a battery-powered microphone (TCM160, AV-Leader Corp, Taiwan) and fed to an available analog input in the EEG amplifier through a custom-made jack-to-touchproof cable. For US1, US2 and US3, speech was recorded using a Logitech USB microphone (Logitech, Lausanne, Switzerland) placed approximately 20 cm from the patients' mouth, input to a USB port on a laptop separate from the amplifiers, where sound sample indices were logged into the BCI2000 programming environment (Schalk *et al* 2004).

2.1.4. Task and stimuli

All but CH2 participants performed an overt speech production task. Participants P2, P3 and P5 were asked to read aloud short French sentences, which were part of a large articulatory-acoustic corpus acquired previously (Bocquelet et al 2016b) and made freely available (https://doi.org/10.5281/zenodo.154083). Participant P5 also took part in a protocol involving speech perception, where she was exposed to the sound of computer-generated vowels delivered by a loudspeaker positioned about 50 cm on her left. This second dataset involving P5 also contains speech production segments as she was interacting with the experimenters. The first (speech production) and the

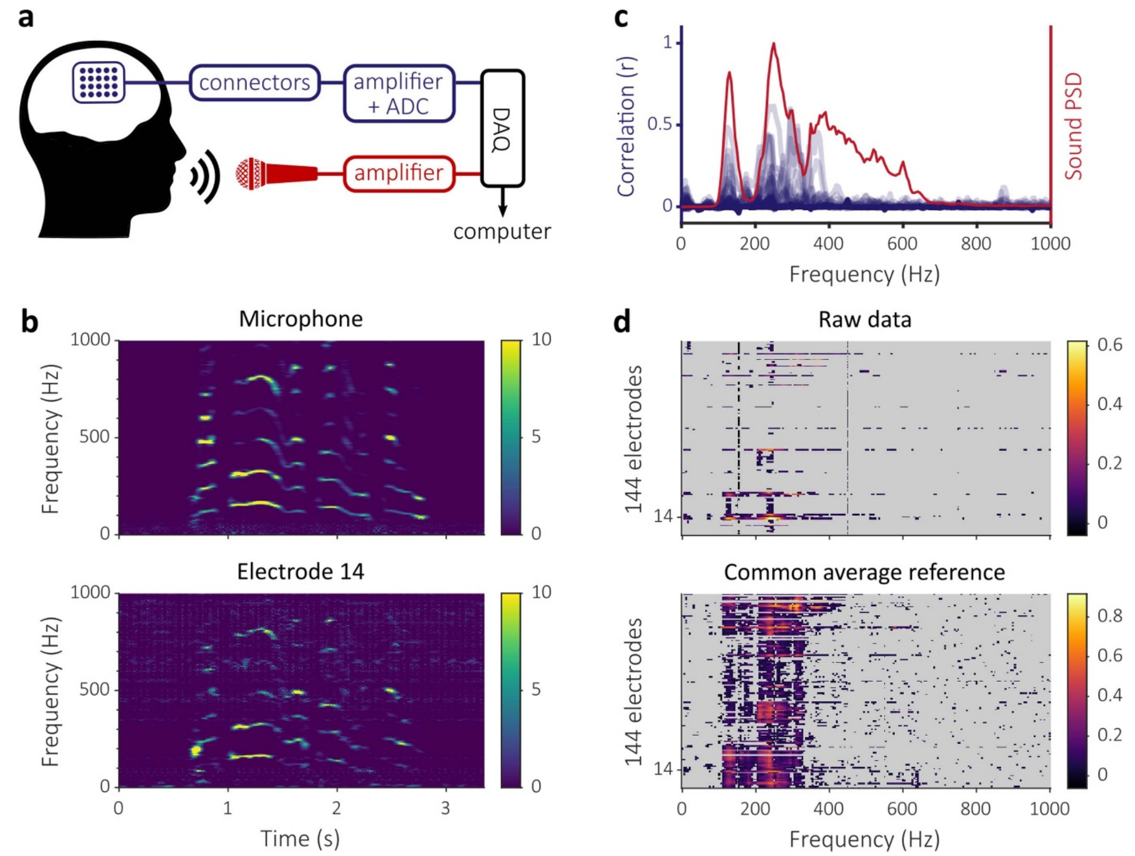


Figure 2. Correlation between voice and ECoG signals during speech production in participant P2. (a) Schematic representation of the recording setup, including neural (blue) and audio (red) data streams. The analog-to-digital conversion (ADC) of the audio signal is done in the data acquisition system (DAQ) whereas it occurs in the FEA for neural signals (see section 2.1.2 for more details). (b) The upper and lower graphs show the z-scored spectrograms of the microphone and of electrode 14, respectively. The succession of stable striped patterns and transient states is typical of human speech. (c) Each blue curve represents, at all frequency bins, the value of the correlation coefficients between the spectrogram of one electrode signal and the spectrogram of the audio signal. The red curve represents the mean PSD of the audio signal (a.u.). (d) Heat maps representing the correlation coefficients between audio and neural data across electrodes and frequency bins. Correlation coefficients not statistically significant are displayed in grey. The upper and lower graphs show the results when using raw neural data and neural data after common average reference, respectively.

second (speech perception and production) datasets are labeled as sessions a and b, respectively. Participant CN was asked to listen and repeat aloud individual sentences of an ancient Chinese poem. Each block consisted of four sentences and each sentence lasted between 2 and 5 s. There were six blocks in total, three from the morning and three from the afternoon of the same day. The two sessions are labeled as a.m. and p.m. Participants D1 and D2 performed a sentence repetition task which consisted of an auditory presentation of pre-recorded sentences and their repetition following a visual go signal. The task and stimuli are described in (Gehrig *et al* 2019). Participant CH1 performed a speech production task, where she had to repeat a written word after a 2 s delay, and a speech processing task, where she heard fragments of a presidential discourse. Participant CH2 performed a speech perception task, where he heard fragments of movie soundtracks that contained speech. Participants US1, US2 and US3 performed a simple verb-generation task, where nouns (approximately 2.5 cm high, and 8–12 cm wide) were presented on a screen approximately 1 m from the patient, at the bedside. The patient's task was to say a verb that was connected to the noun: for example, if the cue read 'ball', the patient might say 'kick', or if the cue read 'bee', the patient might say 'fly'. In between each 1.6 s cue was a blank-screen 1.6 s interstimulus interval (task and stimuli described in further detail in (Miller *et al* 2011)).

2.2. Rat recording

In order to consider data recorded in a different condition, we also analyzed an electrophysiological recording obtained over the left auditory cortex of a ketamine (90 mg/kg)-xylazine (2 mg kg^{-1}) anesthetized 600 g adult Sprague Dawley rat using a 64-electrode micro-ECoG array (E64-500-20-60-H64; NeuroNexus Inc. USA). This data was obtained in compliance with European (2010–63–EU) and French (decree 2013–118 of rural code articles R214–87 to R214–126) regulations on animal experiments, following the approval of the local Grenoble

Table 1. Assessment of the presence or absence of contamination for 20 datasets from 5 different research centers. The durations reported refer to the time segments kept for analysis after the data selection step (detailed in section 2.4.1). For all datasets, the estimated risk to wrongly consider the existence of contamination (P, see section 2.4.5) is reported in the last column. Note that this value depends on the frequency range considered in the contamination matrix (indicated in the second to last column). Contaminated datasets are marked with an asterisk ($P < 0.05$). Their contamination matrices are presented in figures 7, 8 and 9.

Center	Participant	Electrodes	Task (session)	Duration (s)	Frequency range (Hz)	*P*
Grenoble	P2	ECoG	production	600	75–1000	$<10^{-4}$ *
					0–200	$<10^{-4}$ *
	P3	UEA	production	257	75–1000	$<10^{-4}$ *
	P5	ECoG	production (a)	374	0–200	0.34
			production (b)	133	75–1000	7.0×10^{-4} *
			perception (b)	312	75–1000	$<10^{-4}$ *
		UEA	production (b)	141	75–1000	0.11
			perception (b)	319	75–1000	0.46
	Rat	μ-ECoG	perception	600	75–2500	$<10^{-4}$ *
Hangzhou	CN	ECoG	perception (a.m.)	40	75–1000	0.54
			perception (p.m.)	42	75–1000	$<10^{-4}$ *
Frankfurt	D1	ECoG	production	98	75–1000	0.013 *
		ECoG	perception	105	75–1000	0.94
	D2	ECoG	production	180	75–1000	0.46
		ECoG	perception	151	75–1000	$<10^{-4}$ *
Geneva	CH1	ECoG	production	109	75–1000	0.98
			perception	88	75–1000	0.85
	CH2	ECoG	perception	497	75–1000	0.99
Seattle	US1	ECoG	production	268	75–400	0.91
	US2	ECoG	production	133	75–400	0.46
	US3	ECoG	production	383	75–400	0.50

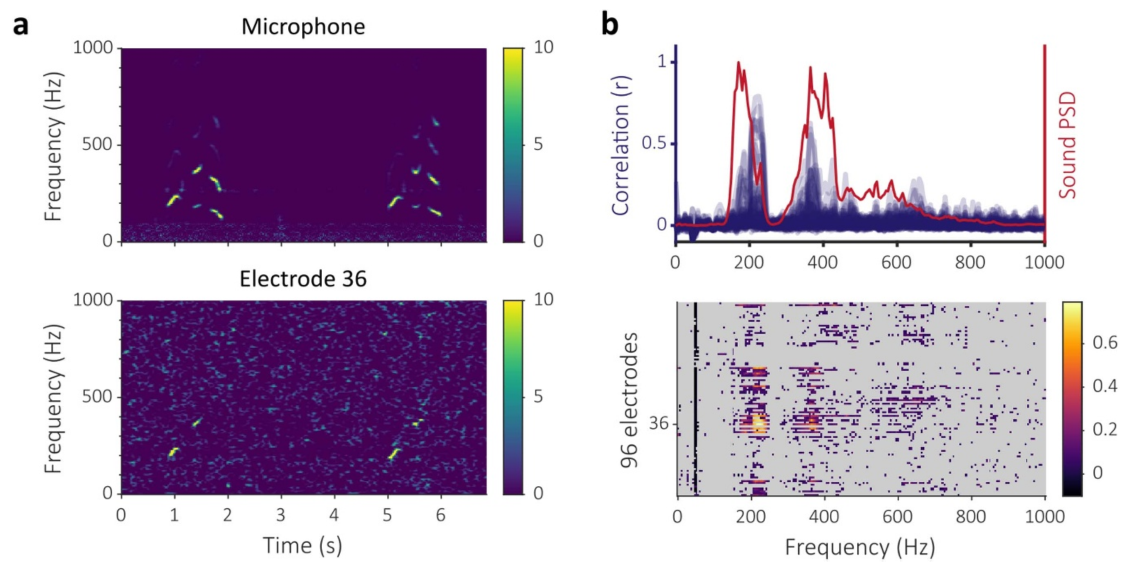


Figure 3. Correlations between voice and intracortical signals during speech production in participant P3. (a) The upper and lower graphs show the z-scored spectrograms of the microphone and electrode 36, respectively. (b) On the upper graph, each blue curve represents, at all frequency bins, the value of the correlation coefficients between the spectrogram of one electrode signal and the spectrogram of the audio signal. The red curve represents the mean PSD of the audio signal (a.u.). The lower panel represents a heat map of the correlation coefficient between audio and neural data for all electrodes and frequency bins. Correlation coefficients not statistically significant are displayed in grey.

ethical committee ComEth C2EA–12 and the ministry authorization 04815–02. A bone screw was used for the ground and a stainless-steel wire inserted below the skin ahead of Bregma was used for the reference. Signals were acquired using the RHD2000 acquisition system and two 32-channel RHD2132 headstages (Intan Technologies, USA). To avoid any possible crosstalk inside the Intan acquisition system, the sounds delivered to the rat were recorded on an independent CED Micro1401 (Cambridge Electronic Design, UK). Both acquisition devices were interfaced and synchronized by the Spike2 software with the IntanTalker module (CED programs) and signals were digitized at 33.3 kHz. The time jitter

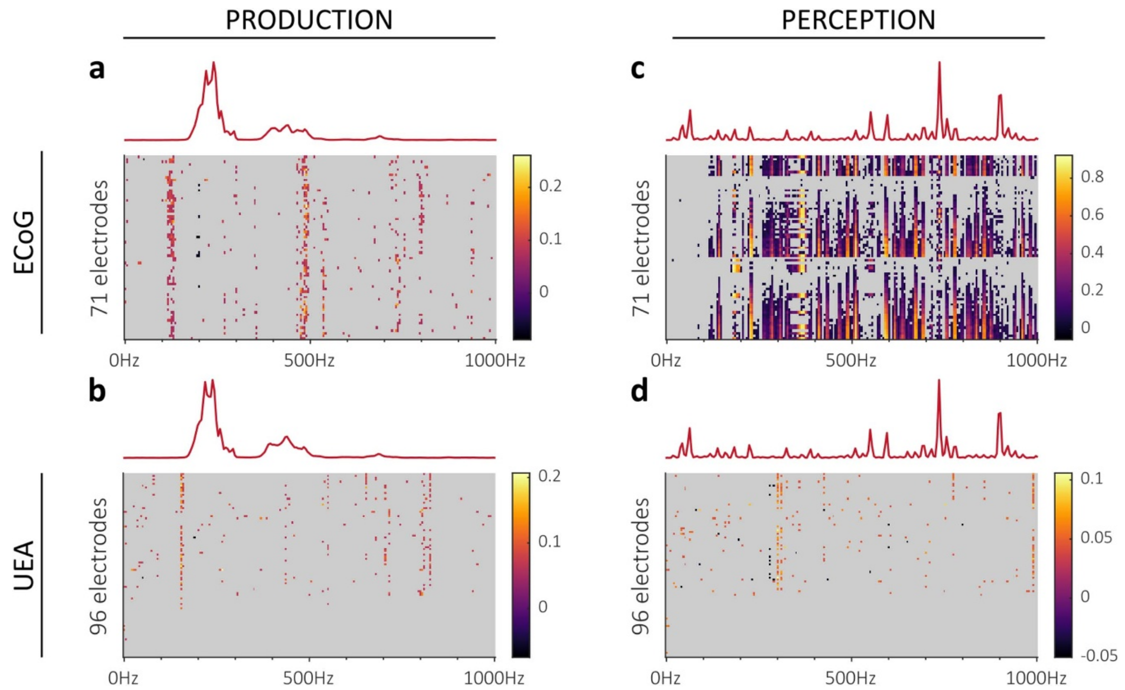


Figure 4. Spectrogram correlations between sound and neural data observed in the recording of participant P5 during session b. The results are presented depending on the experimental condition (speech production or perception) and the type of electrophysiological measure (ECoG and UEA). The heat maps show the value of the correlation coefficient with the spectrogram of the sound for each electrode and each frequency bin. The correlation coefficients that are not statistically significant are displayed in grey. The red curves indicate the mean PSD of the sound (a.u.) recorded during the experiments. The speech perception experiment used computer-generated vowels, peaks appearing in the mean PSD are the fundamental frequency and first harmonics of the vocal synthesizer. (a, b) Results for the speech production condition using ECoG and UEA data, respectively. (c, d) Results for the speech perception condition using ECoG and UEA data, respectively.

between sound and neural signals was checked to be below 2 ms.

Pure tones (3 ms rise, 167 ms plateau and 30 ms fall times) with frequencies ranging from 0.5 to 16 kHz were presented with pseudo-random interstimulus intervals of 1.8–2.2 s. Sounds were delivered at about 80–90 dB SPL in open field configuration using a MF1-S speaker (Tucker Davis Technology Inc. USA). The three lowest tone frequencies that were further considered in the present study were 0.5, 1 and 2.5 kHz.

2.3. In vitro recordings in PBS solution

In order to further identify the origin of acoustic contamination, an *in vitro* setup was used. A 24-electrode ECoG array (DIXI Medical SAS) was placed in a plastic container filled with 1X phosphate-buffered saline (PBS). Two of the electrodes were used as the ground and reference electrodes, respectively. All electrodes were plugged into a clinical headbox (64-Channel Splitter Box, Blackrock Microsystems, USA) connected by shielded cables to a FEA and a NSP used for human recordings. A microphone was placed close to the plastic container. The audio data was acquired using the same hardware as for human P2-P3-P5 recordings (see section 2.1.3). Data was acquired at 30 kHz. Sounds were delivered either through a loudspeaker (M-Audio BX5-D2 loudspeaker used with participant P5) located about 1–2 m from the electrodes and recording chain, or very locally using a MF1-S speaker (Tucker Davis Technology Inc. USA) mounted in a closed-field configuration and sending sounds through a small plastic tube (3 mm outer diameter). A plastic box with a removable lid, soundproofed with cotton fiber insulation, was used to reduce sound propagation between the loudspeaker and the content of the box. For *in vitro* experiments, 20 pure tones lasting 4 s, with frequencies ranging from 25 to 975 Hz (spaced every 50 Hz), were played four times with 2 s of silence between sounds.

2.4. Data processing

2.4.1. Data selection

All recordings were visually inspected. For participant P2, 112 electrodes were removed due to several loose connections at the level of the Cabrio Connectors. For participant P5, 1 ECoG electrode showing saturating noise was removed.

Whenever the audio recordings contained data from more than one source (stimuli, participant's voice, experimenter's voice), a single source was studied at a time by manually annotating and excluding the segments featuring the other source(s). For all recordings, segments containing high-power transient noises were excluded from further analysis. To do so, the signal of each channel was detrended (using a 500 ms moving average) and positive and

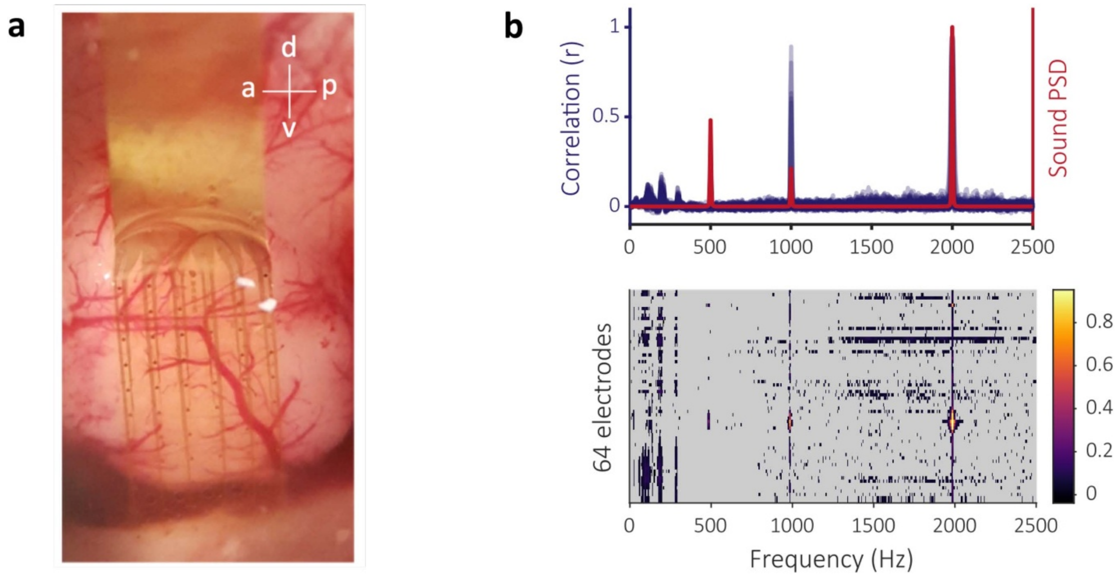


Figure 5. Correlations between sound and µ-ECoG spectrograms during pure tones perception in an anesthetized rat. (a) Photograph of a µ-ECoG grid positioned over the left auditory cortex of a rat. Directions: a = anterior, p = posterior, d = dorsal, v = ventral. (b) On the upper panel, each blue curve represents, at all frequency bins, the value of the correlation coefficients between the spectrogram of one electrode signal and the spectrogram of the audio signal. The red curve represents the mean PSD of the audio signal (a.u.). The lower panel represents a heat map of the correlation coefficient between audio and neural data for all electrodes and frequency bins. Correlation coefficients not statistically significant are displayed in grey.

negative thresholds with an absolute value of 5 times the median absolute deviation were used to detect potential high-power noises. A sample was considered as noisy when at least 10% of the channels reached their threshold. A 500 ms window was also excluded around each noisy sample. The total time durations that were kept after these data selection steps are indicated in table 1.

2.4.2. Data pre-processing

A built-in analog band-pass filter was applied to the data recorded with the NSP (0.3–2500 Hz for 10 kHz sampling rate and 0.3–7500 Hz for 30 kHz sampling rate). Common average reference (CAR) was applied on the recording of participants P3 to lower the influence of the intrinsic spatial correlation of LFPs stemming from the close spacing of the electrodes of the Utah array, and of participant D1 for whom the reference electrode was itself highly contaminated making in turn all channels highly contaminated. CAR was also applied on participant P2 recording to analyze its effect as shown in figure 2(d). In these cases, the average neural signal, computed on the signals of all selected electrodes, was subtracted to each electrode signal. To center audio signals, a moving average was computed over 1 s windows and subtracted.

2.4.3. Spectrogram computation

In the present study, a spectrogram refers to the time-varying power spectral density (PSD) computed over a recording channel. For all analyses, spectrograms of neural and audio data were computed using short-time Fourier transform with 200 ms time windows (after Hamming windowing). The window overlap was chosen to obtain spectrograms sampled at 50 Hz. The mean sound PSD (or mean power spectrum) of a recording was computed by averaging the audio spectrogram over all time samples (after selection like described in section 2.4.1). For display purposes, the spectrograms in figures 2(a), (b) and 10(c) were computed with higher frequency and time resolutions. These spectrograms were also z-scored within each frequency bin using artifact-free data segments containing the displayed extracts.

2.4.4. Audio-neural correlations

For all recordings, the correlations between the neural and the audio spectrograms were computed. For each channel, the sample Pearson correlation coefficient r between the power amplitudes across time of the channel and audio signals was computed for all possible pairs of frequency bins, resulting in an audio-neural correlation matrix. Correlations corresponding to the same frequency bin between the two signals (i.e. the diagonal of the correlation matrix) are further termed audio-neural correlations. For each value of r, a p-value was computed using Student's t-test to test the null hypothesis that $r = 0$. For audio-neural correlations, the statistical significance of each correlation coefficient was determined with respect to a Bonferroni adjusted significance level $\alpha = 0.05/N$ where N was the number of frequency bins times the number of channels in the recording.

2.4.5. Objective assessment of contamination

To determine whether a dataset is contaminated, we developed a specific statistical approach. First, a contamination matrix was built by computing the

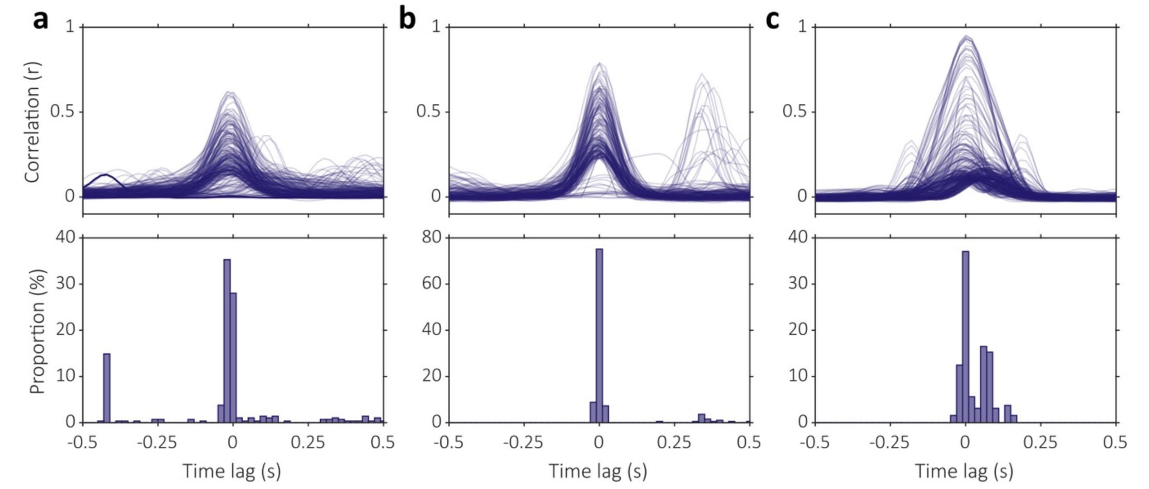


Figure 6. Audio-neural cross-correlations were computed with lags between −500 and +500 ms applied to the audio spectrogram (positive lags = delaying audio). All the frequency bins between 75 and 1000 Hz were considered. In the top graphs, each curve represents, for a given channel and a given frequency bin, the cross-correlation between the audio and the neural spectrogram. For each recording, only the 1% curves reaching the highest correlation values were kept. In each bottom graph, a histogram represents the distribution of time lags that maximize correlation for the curves in the corresponding top graphs. The histogram bins have a width of 20 ms. (a) Participant P2 ECoG recording during speech production. (b) Participant P3 UEA recording during speech production. (c) Rat micro-ECoG recording during sound perception.

maximum of the correlation matrices across all electrodes. Each maximum was thus computed separately for each element of the matrix (i.e. each pair of frequency bins). Then, we evaluated the values on the diagonal of the contamination matrix in relation to the rest of the matrix. The mean value on the diagonal was computed to obtain a contamination index. To evaluate the statistical significance of this original index, a distribution of surrogate index was then built by computing the mean diagonal 10 000 times on as many shuffled versions of the contamination matrix. Each shuffled matrix was built by randomly shuffling either the lines or the columns. Shuffling only one dimension at a time favors the preservation of the values on the diagonal in case of horizontal or vertical patterns in the original matrix. The original contamination index was finally compared to the distribution of the surrogate ones. The proportion P of surrogate indices that were superior to the original index was considered as the risk taken when rejecting the null hypothesis that no contamination exists (in other words P was considered as the probability of being wrong when considering that a contamination exists).

2.5. Neural decoding

ECoG data from participants P2 and P5 (session a) were used to predict acoustic mel-cepstral coefficients of overt speech produced by the participants. Both participants were visually presented with a series of short sentences or vowel sequences written on a screen positioned about 50–100 cm in front of them, and asked to repeat them overtly. The number of sentences was 118 for participant P2 and 150 for participant P5, corresponding to an overall duration of 230 and 329 s of speech, respectively. The participants' speech audio signals were decomposed into 25 mel-cepstral coefficients using the SPTK toolkit (*mcep* function). Spectrograms of the ECoG data were computed as described in paragraph 2.4.3 but at a rate of 100 Hz. Neural features were the spectrogram amplitudes in 10 Hz bands (i.e. 0–10, 10–20, … 190–200 Hz) and the band-pass filtered time domain LFP signal (between 0.5 and 5 Hz). Two sets of neural features were considered, a first one where only features below 90 Hz were used and a second one where all features up to 200 Hz were used. A feature selection process was applied to keep only the features that were significantly modulated during speech production with respect to silence intervals, as assessed by Welch's *t*-test with a Bonferroni adjusted significance level $\alpha = 0.05/N$ where N is the number of electrodes times the number of candidate features. The resulting number of selected features was 3147 out of 5376 for P2 and 1115 out of 1512 for P5. These selected features were normalized and decomposed using PCA (both transformations were based on training sets). The first 50 (for the 0–90 Hz feature set) or 100 (for the 0–200 Hz feature set) components were used as the final set of features. A linear model was then used to map these neural features onto the mel-cepstral trajectories using 10-fold cross-validation. Each mel-cepstral sample was decoded using a 200 ms window of neural activity centered on the time of this sample. Chance decoding level was assessed by repeating the whole procedure after shuffling and time-reversing the mel-cepstral trajectories of the different sentences (truncation was applied to match sentence durations).

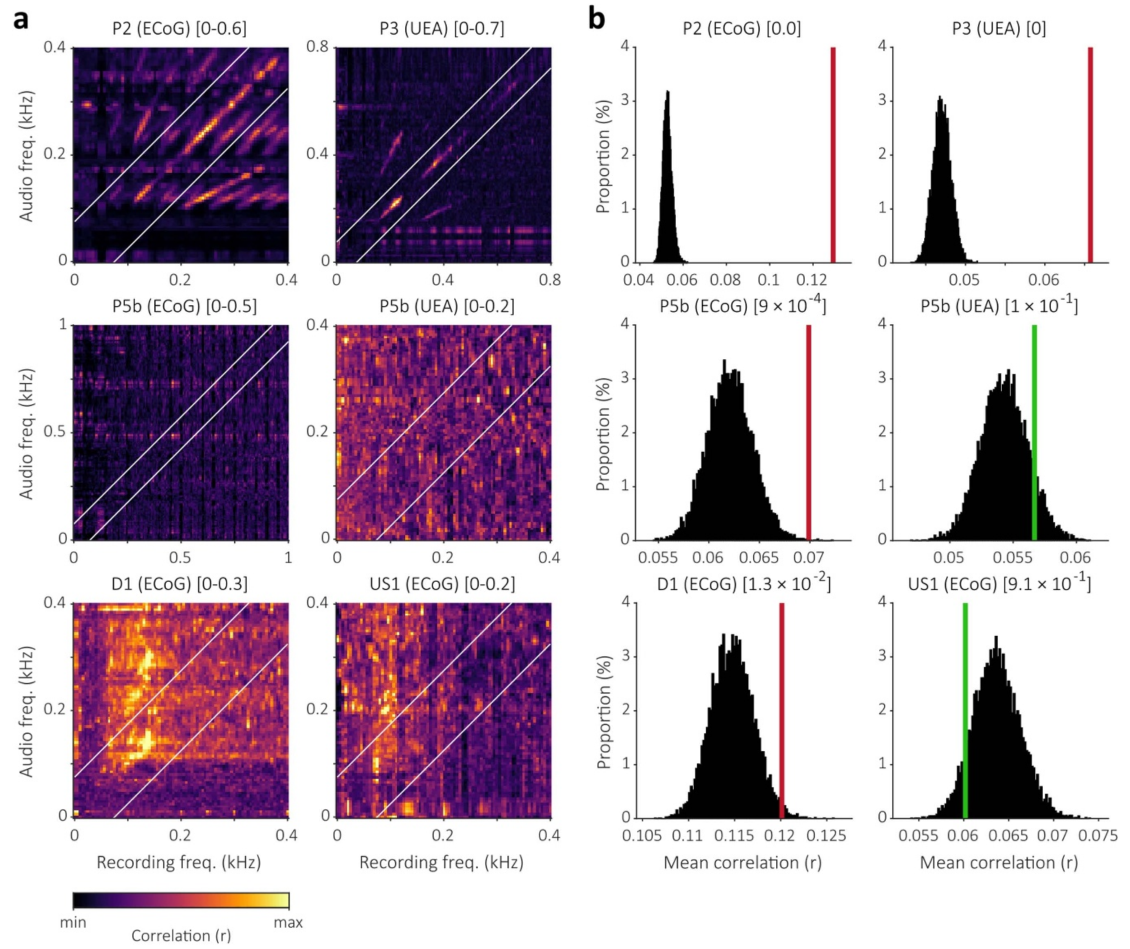


Figure 7. Objective assessment of acoustic contamination of 6 speech production datasets. (a) Audio-neural contamination matrices of each datasets. Each heatmap represents a recording. The color scale min and max values are indicated within brackets in the title of each heatmap (the max values were computed discarding frequencies below 75 Hz, which do not contain any speech information, so that the colormaps were not influenced by 50 Hz or 60 Hz line noise). The white lines are a visual aid to assess the presence of high correlations on the diagonal, they are parallel to the diagonal of the matrix and cross the X and Y axes at 75 Hz. (b) Statistical assessment of contamination for the same datasets. The mean of the diagonal of the contamination matrix (vertical colored bar, red when statistically significant, green when not) is compared to the distribution of such values in 10 000 shuffled contamination matrices (see section 2.4.5 for details). The estimated risk to wrongly consider the existence of contamination (P) is shown in square brackets for each dataset.

3. Results

3.1. Observation of acoustic contamination in neural recordings

3.1.1. Correlation between ECoG and sound signals during speech production

We observed strong correlations between ECoG and sound spectrograms in participant P2 during speech production. Participant P2's brain activity was recorded with an ECoG grid while he was reading sentences aloud. Simultaneously, a microphone was used to capture the sound of his voice (see figure 2(a)). Figure 2(b) shows a portion of the z-scored spectrograms of the sound signal (top) and of an electrode of the ECoG grid (bottom). In this example, the ECoG signal shows a very similar spectrotemporal structure as that of the sound. The time-frequency patterns observed are consistent with human speech and are unlikely to be brain activity.

We quantitatively assessed this phenomenon by computing the correlation between the power of the signal within each frequency bin of each electrode signal with that of the sound signal. As shown in figure 2(c) and in the top of figure 2(d), correlations up to 0.6 could be observed depending on the electrode. Up to 370 Hz, the strongest correlations were observed at frequencies most present in the sound signal, and in particular between 115 and 145 Hz, which corresponded to the range of the fundamental frequency of the subject's voice. Above 370 Hz, correlations were low even at frequencies for which the power of the speech signal remained high. As shown in figure 2(d), the correlations between sound and ECoG spectrograms were still present and even

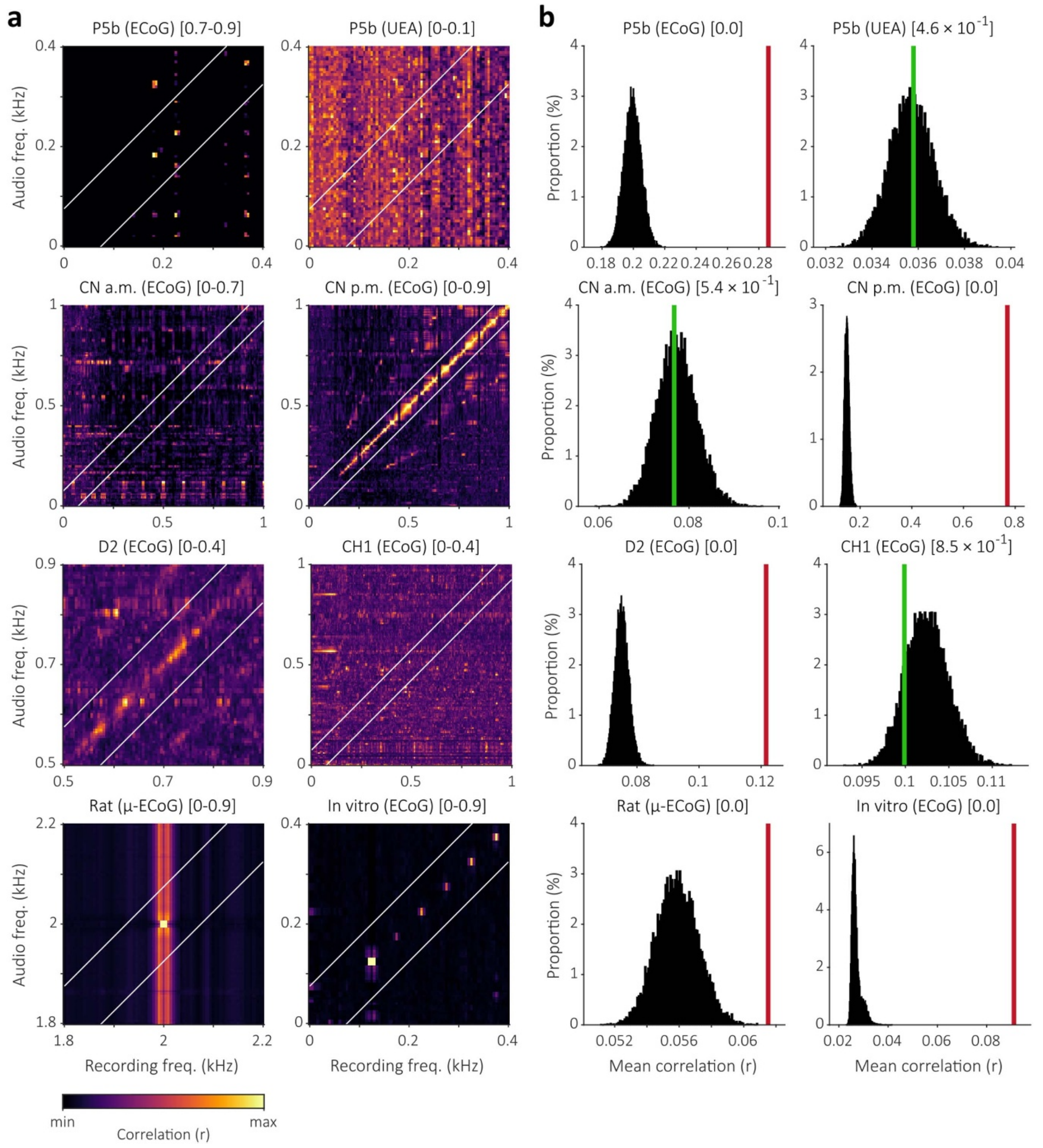


Figure 8. Objective assessment of acoustic contamination of 8 sound perception datasets. (a) Audio-neural contamination matrices of each datasets. Each heatmap represents a recording. The color scale min and max values are indicated within brackets in the title of each heatmap (the max values were computed discarding frequencies below 75 Hz, which do not contain any speech information, so that the colormaps were not influenced by 50 Hz or 60 Hz line noise; for P5 ECoG we set the min value to 0.7 to increase visibility of diagonal correlations). The white lines are a visual aid to assess the presence of high correlations on the diagonal. They are parallel to the diagonal of the matrix and cross the X and Y axes at 75 Hz. (b) Statistical assessment of contamination for the same datasets. The mean of the diagonal of the contamination matrix (vertical colored bar, red when statistically significant, green when not) is compared to the distribution of such values in 10 000 shuffled contamination matrices (see section 2.4.5 for details). The estimated risk to wrongly consider the existence of contamination (P) is shown in square brackets for each dataset.

exacerbated after common average re-referencing of the ECoG signals.

3.1.2. Correlation between intracortical and sound signals during speech production

In P3 recording, we further observed statistically significant correlations between the spectrograms of intracortical signals recorded using a Utah array and that of the produced speech signal. Figure 3(a) shows a portion of the z-scored spectrograms of the subject's voice (top) and of one electrode of the array (bottom). The spectrogram of the selected micro-electrode clearly shows spatio-temporal features also observed in the sound spectrogram (between 200 Hz and 400 Hz). Statistically significant correlation coefficients up to 0.7 were observed, with peaks falling in the range of frequencies where the sound signal showed high power (figure 3(b)). Noticeably, correlations between intracortical and sound signals during speech production were much weaker in participant P5 (figure 4(b)).

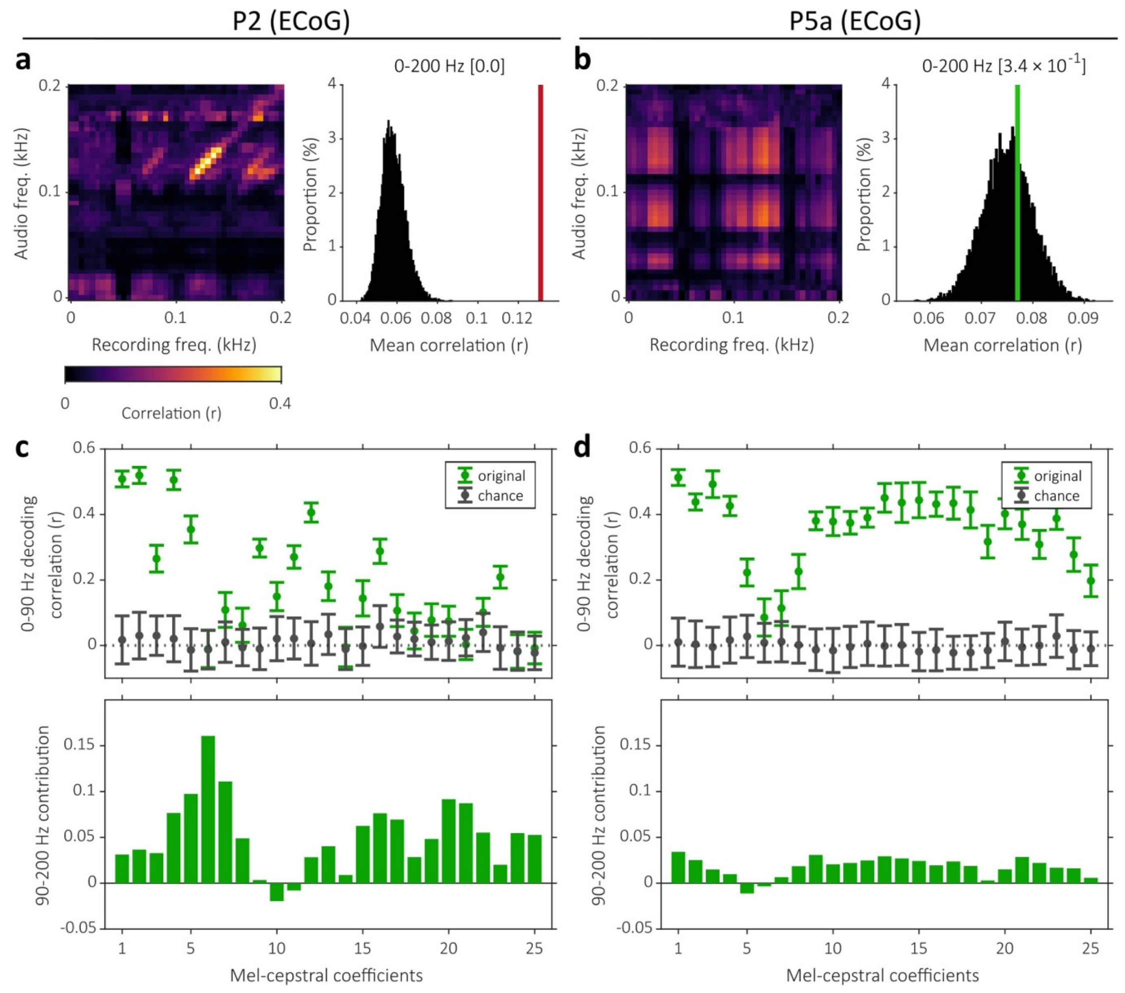


Figure 9. Linear decoding of speech features using either contaminated or not contaminated neural recordings. (a)–(b) Contamination matrix between 0 and 200 Hz (left) and corresponding statistical assessment of contamination (right) for ECoG datasets from participants P2 (panel a) and P5 (panel b), contamination P value is indicated between square brackets (P2 contaminated, P5 not contaminated). (c)–(d) Decoding performances for participants P2 (panel c) and P5 (panel d) of models using neural features from 0 to 90 Hz (top graphs) and the change in performance for models using neural features from 0 to 200 Hz (bottom graphs). The lower graphs can be interpreted as the contribution of features of the 90–200 Hz range.

3.1.3. Correlation between electrode and sound signals during sound perception

Statistically significant correlations between electrode and sound signals were not only present during speech production as reported above, but also during sound perception. This phenomenon was observed in human and animal recordings using different recording instrumentations. Participant P5 participated during session b in a paradigm where artificially synthesized speech sounds were presented to her through a loudspeaker positioned on her left. Brain activity was recorded from both ECoG electrodes and intracortical microelectrodes. The sound produced by the loudspeaker was also simultaneously recorded. During this session, the segments when the participant was speaking to the experimenter were also separately studied. Performing the same analysis as in the previous section, we found that ECoG signals during speech production segments were not as highly correlated with the participant's voice (figures 4(a) and (b)). The highest correlations were found in the ECoG recording, within the frequency range of the voice's first harmonic (figure 4(a)). The neural recordings showed strong correlations with the perceived sound signal, with peaks up to 0.9 (figure 4(c)). As observed in the recordings from P2 and P3, frequencies showing strong correlations were mostly found in the bands that concentrate most of the sound power. By comparison, the spectrograms of intracortical signals were poorly correlated with that of the sound (figure 4(d)).

Second, in order to verify that the correlations were not due to our clinical recording system in particular, we performed the same type of analysis on data obtained from an experiment in a rat. The left auditory cortex was recorded using a commercial μ-ECoG grid connected to an Intan neural recording system (figure 5(a)). In this case, pure tones were delivered in an open field configuration. As shown in figure 5(b), we again observed strong correlations between the electrode and sound spectrograms, with sharp peaks at the specific

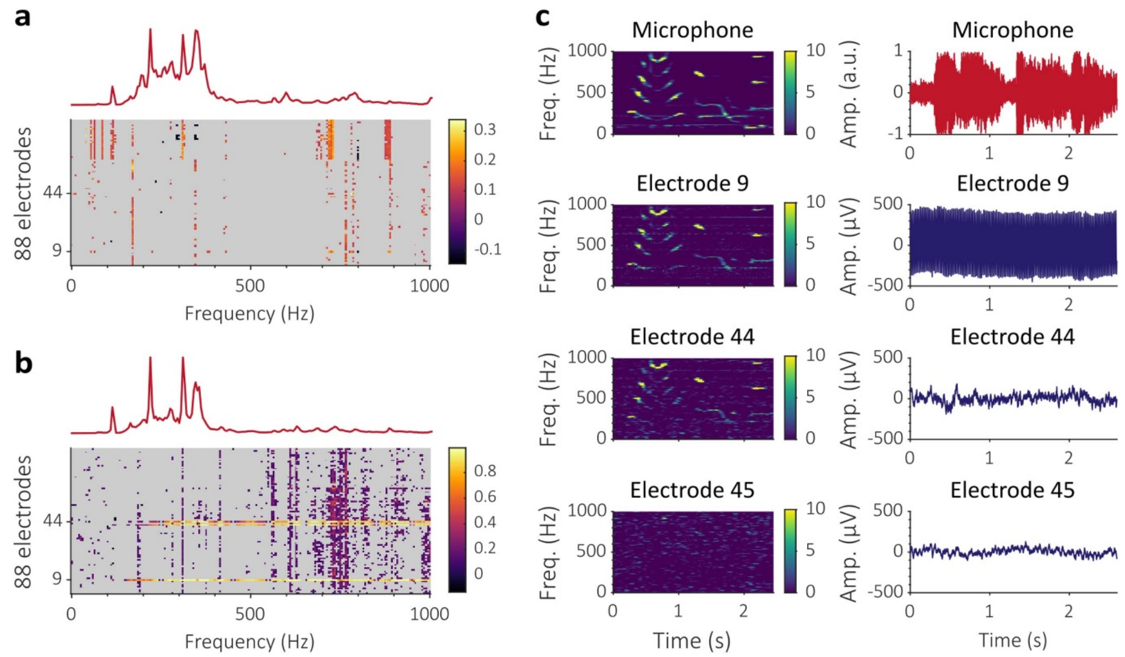


Figure 10. Correlations between sound and ECoG recordings in participant HG during speech perception. (a), (b) Correlation heat maps (as in figure 4) for the morning (panel a) and afternoon (panel b) sessions. (c) Example of 2.5 sec spectrograms (left) and raw signals (right) for the sound that was presented to the participant and signals from 3 electrodes of the grid, one with high 50 Hz power-line noise (electrode 9), and two with standard noise level (electrodes 44 and 45).

frequencies of the pure sound stimuli (500, 1000 and 2000 Hz).

3.1.4. Audio-neural cross-correlations

In the previous sections, examples of highly correlated neural and audio recordings were shown. To further study the nature of this phenomenon, we used the cross-correlation to determine the delay between the two recordings that would maximize their correlation. For each channel and each frequency bin of the neural recording, we computed the correlation with different time lags applied to the audio signal at the same frequency bin. Figure 6 shows that most of the highest correlations are obtained for lags between −30 and +10 ms, in both speech production and sound perception cases. These very small lags suggest that the similar patterns occurring in the audio and neural signals are close to synchronous.

3.1.5. Objective assessment of acoustic contamination

The previous sections show that time-frequency patterns of audio signals are sometimes partially found in neural recordings. These correlations between audio and neural recordings occur at frequencies that correspond to the high-power frequency content of the sound (see sections 3.1.1–3.1.3) and seem to occur almost synchronously in both recordings (see section 3.1.4). These observations suggest a contamination of the electrophysiological measure by the audio signal through a physical phenomenon. This hypothesis of acoustic contamination is supported by the investigations on the origin of the phenomenon and its reproduction in section 3.3.

We developed an approach to assess the presence of contamination. We based this approach on the contamination matrix, which sums up the highest values in the audio-neural correlation matrices of a recording. The diagonal of the contamination matrix shows the audio-neural correlation for a given frequency while the other elements of the matrix represent cross-frequency correlations. Supposing that the contamination phenomenon is linear, we expect that power variations at a given frequency in the audio would cause power variations in the neural recording at the same frequency. Contamination is therefore characterized by high correlations limited to the diagonal. By contrast, when other sources of the electrophysiological signals (actual brain activity, muscle artifacts, motion artifacts) happen to be correlated with sound, the involved frequency bands are typically not exactly the same. In these cases, broad patches, vertical lines and/or horizontal lines of high values are observed in the contamination matrix. As detailed in section 2.4.5, the statistical criterion we propose compares the diagonal of the original contamination matrix to the diagonal of shuffled matrices. This allows to distinguish whether the high correlations are (1) limited to the diagonal, in which case they are significantly higher in the original matrix compared to the shuffled ones or (2) part of a larger patch of high correlations, in which case they are not significantly higher in the original matrix compared to the shuffled ones.

We used this approach to evaluate 20 different speech production and sound perception datasets. Figure 7(a) shows the contamination matrices for

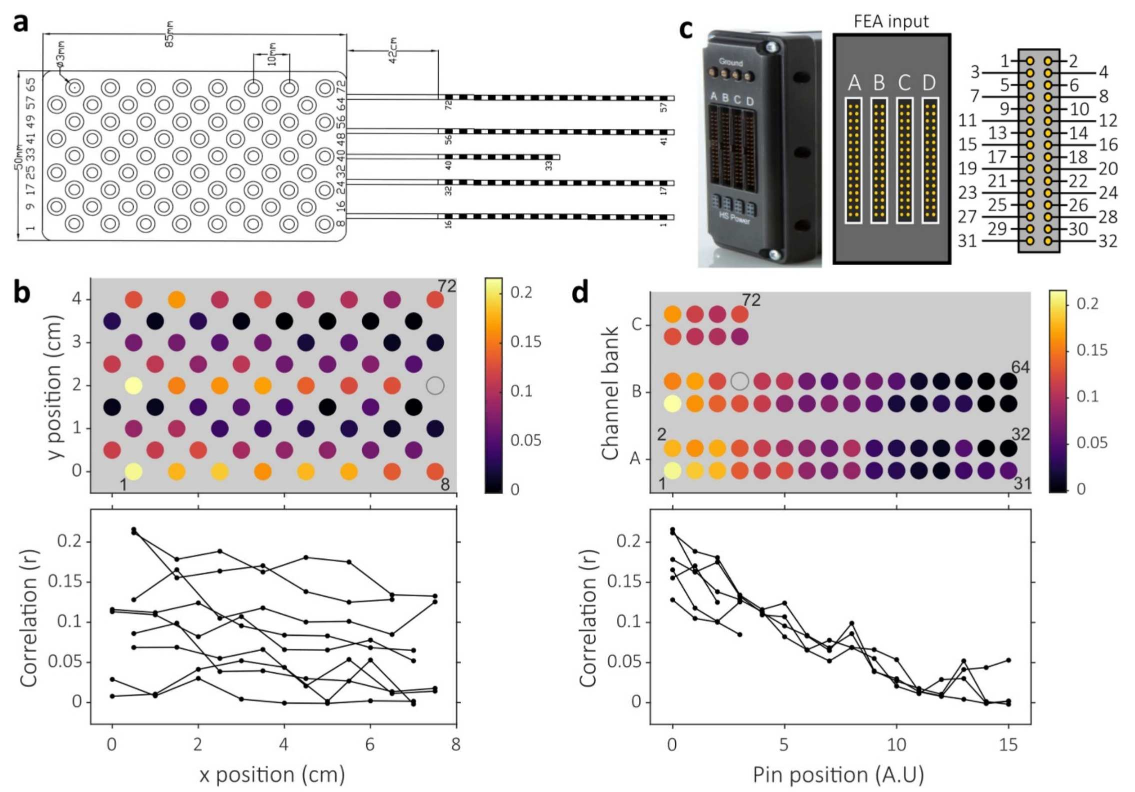


Figure 11. Spatial distribution of acoustic contamination according to the grid and connector layouts (participant P5 ECoG recording during sound perception). (a) Grid layout. (b) Top: The mean audio-neural correlation of each channel is represented as a colored dot at the position of the corresponding electrode on the grid (the empty circle represents a non-selected channel, see section 2.4.2). Bottom: Same values plotted along the rows of the grid. (c) Picture (left) and schematics (middle) of the front-end amplifier displaying four 32-channel connectors, each made of 16 pins (right). (d) Top: The mean audio-neural correlation of each channel is represented this time by a colored dot at the position of the connector pin connecting this channel (the connector layout is rotated 90° anticlockwise with respect to the picture and schematics of panel c). Bottom: Same values plotted along the rows of the connectors.

six speech production datasets and figure 7(b) the corresponding statistical evaluation by randomization. According to our statistical criterion P (see section 2.4.5), four datasets were contaminated (red vertical bars in figure 7(b)) and two were not (green vertical bars). The audio-neural correlations that were observed for participants P2 (figure 2) and P3 (figure 3) can be observed on the diagonal of the contamination matrices (figure 7(a), top row). Very clear and highly statistically significant contaminations appear in these cases. Weaker but still visible and statistically significant contaminations also appear for P5 session b (ECoG) due to frequencies above 200 Hz and for D1 due to frequencies around 100 Hz. In recordings of participants P2, P3 and D1, lines of high correlation can also be observed outside the diagonal. They are due to the correlations between the voice's fundamental frequency and its harmonics.

Figure 8(a) shows the contamination matrices for eight sound perception datasets, five of which are contaminated according to the corresponding statistical evaluation shown in figure 7(b). Contamination-specific patterns appear on the matrices for participants P5 (ECoG) and D2 (ECoG), for the p.m. session of participant CN (ECoG), for the rat recording (μ-ECoG) and for the *in vitro* experiment (ECoG). No contamination was found for participants P5 (UEA), CH1 (ECoG) and the a.m session of participant CN.

The full assessment of all datasets is summarized in table 1.

3.2. Potential influence of contamination on speech decoding

We assessed the potential influence of contamination of electrophysiological signals by sound on the performance of neural decoding to predict acoustic features of produced speech. For this purpose, we considered ECoG data from participant P2 and P5 (session a). As can be seen in figure 9 (panels a–b), the P2 recording was found to be contaminated in the 0–200 Hz range while the P5a recording was not. It should also be noted that the fundamental frequency of both participants exceeds 90 Hz and thus acoustic contamination cannot be observed below. The decoding performances of models using only neural features from 0 to 90 Hz were evaluated (figure 9, panels c–d, top graphs). Models using neural features up to 200 Hz were then evaluated, and compared to the previous one in order to estimate the contribution of 90–200 Hz features (figure 9, panels c-d, bottom graphs). We found that including neural features

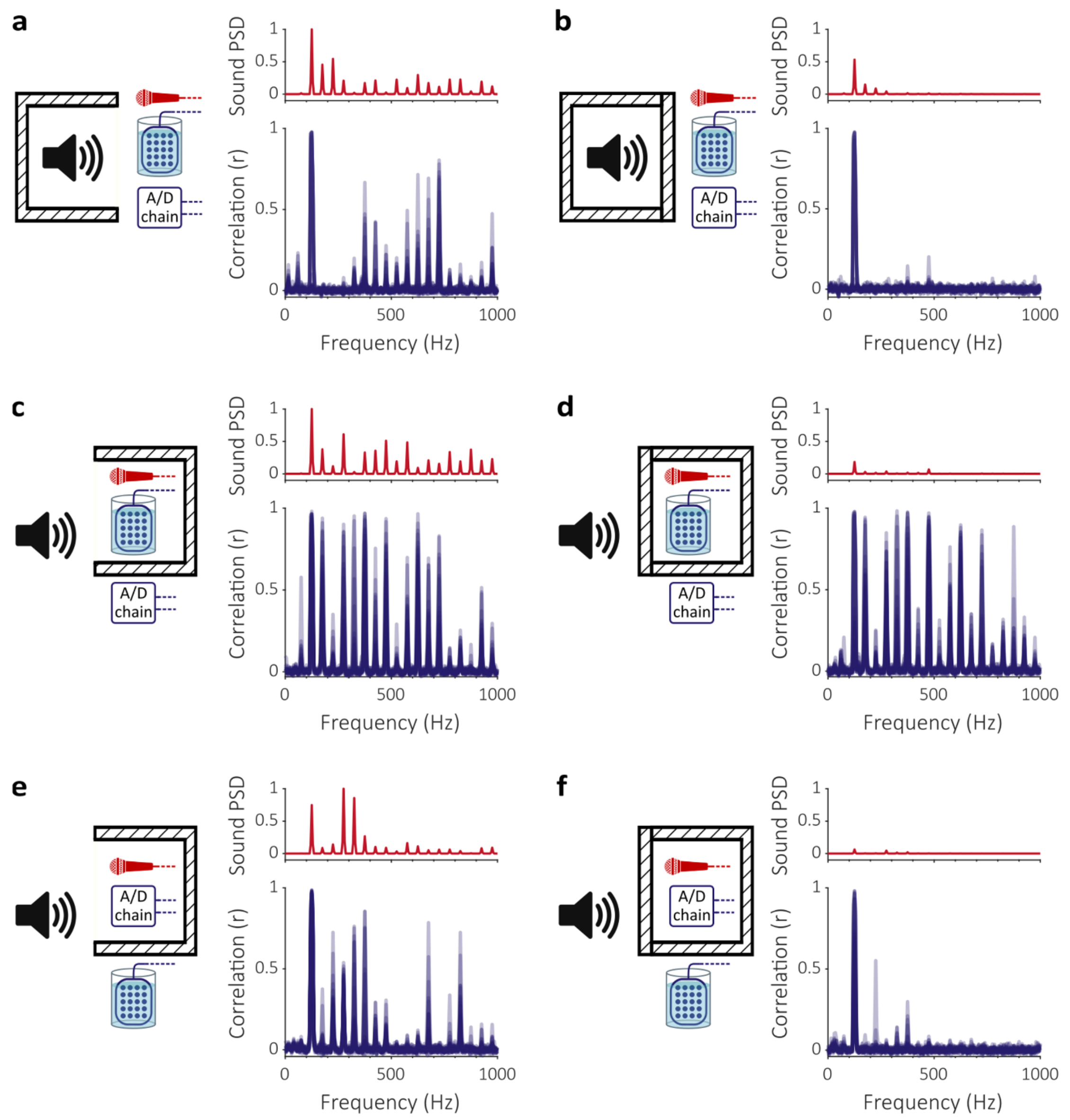


Figure 12. Correlations between sound and ECoG recordings in different *in vitro* experimental configurations. The red graphs show the mean PSD of the sound captured by the microphone. Mean PSD scale (a.u.) is common within each of the 3 panel pairs (a)–(b), (c)–(d) and (e)–(f). In the blue graphs, each curve represents, at all frequency bins, the value of the correlation coefficients between the spectrogram of one electrode signal and the spectrogram of the audio signal. (a) Loudspeaker placed in the open box. (b) Loudspeaker placed in the closed box. (c) Electrodes and microphone placed in the open box. (d) Electrodes and microphone placed in the closed box. (e) Amplification and digitization chain (composed of the cables, adaptors, splitter box and FEA) and microphone placed in the open box. (f) Amplification and digitization chain and microphone placed in the closed box. Within each raw, only the fact that the lid was open or closed changed.

from 90 to 200 Hz resulted in an important increase in decoding performance for P2 and only a limited improvement of decoding accuracy for participant P5. This example shows that including a contaminated recording in a decoding study may positively bias the decoding performance and lead to overestimate the contribution of the high-gamma band.

3.3. Possible sources of acoustic contamination

3.3.1. Sound contamination and electrode quality

In the following of this paper, we investigate the possible causes of the contamination observed in neural signals. We first tested whether the level of sound contamination was determined by the quality of the electrode signal. Participant CN was recorded twice on a single day, once in the morning and once in the afternoon. Between the two sessions, the electrodes were disconnected and then reconnected after lunch. Sound contamination was observed only in the afternoon session (figures 7(a) and (b)), which indicates that the contamination was not related in this case to the electrode array and its intracranial environment as those remained unchanged. Moreover, electrodes showing strong contamination in the afternoon session showed very variable signal quality (figure 7(c)). For instance, one electrode with very strong

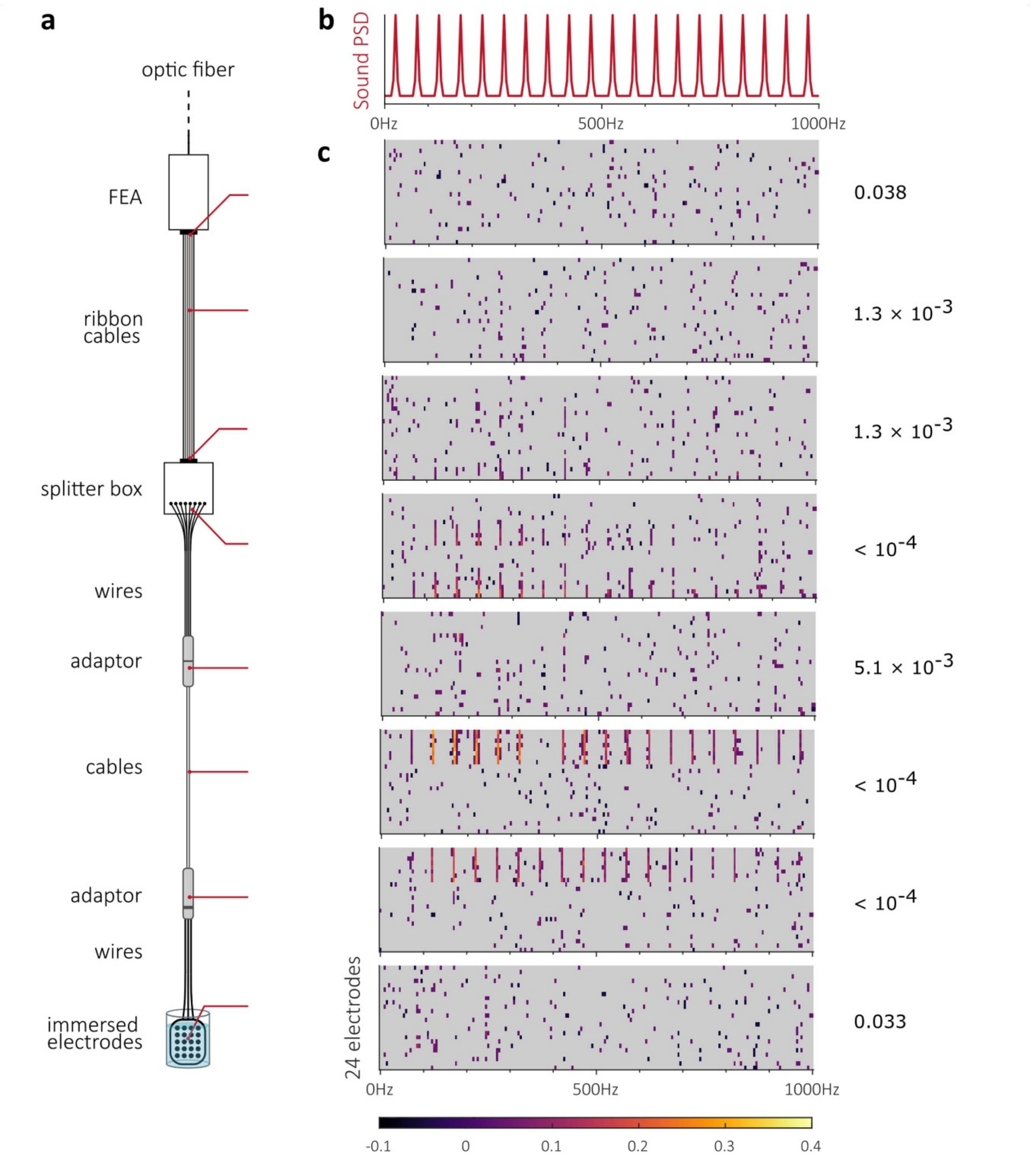


Figure 13. Determination of the location of sound contamination along the recording chain (case of ECoG electrodes). (a) Sounds were delivered focally at different locations (indicated by the red lines) of the recording chain. (b) Mean PSD of the sounds delivered through the speaker. (c) Correlation heat maps for each location of sound delivery (each map is displayed against the corresponding location of sound delivery indicated in panel a). The estimated risk to wrongly consider the existence of contamination (P) is indicated on the right of each heat map.

50 Hz power-line noise (considered as a typical ''bad channel'') showed a strong contamination, while two other channels with no such noise showed an equally strong contamination for one, and very weak or no contamination for the other. These observations indicate that the quality of the signal was not a sufficient predictor of sound contamination.

3.3.2. Electrode versus connector mapping of contamination

The study of how acoustic contamination affects the different channels of a recording in relation to their relative position can provide information about the physics of the phenomenon and its location along the acquisition chain. In most of the recordings available to us, the number of channels that were identified as contaminated was too low to observe spatial effects. However, in P5 ECoG recording during sound perception (session b) we observed that most of the channels were contaminated, with varying intensity (see figure 4(c)). The audio-neural spectrogram correlation coefficients were averaged across all frequency bins to obtain a mean correlation coefficient for each channel. Then these values were mapped either on the grid layout or on the FEA connector layout. As shown in figures 11(a) and (b), no clear spatial cluster of contamination was observed on the grid map. However, as could already be noticed in figure 4(c), it appears that channels 1–32 seem to show a continuous decrease of the contamination level, a pattern that is repeated on channels 32–64. Along the recording pipeline (described in section 2.1.2), the channels are grouped by 32 from the output of the splitter box to the input of the FEA. As shown in figures 11(c) and (d), we observed a very clear spatial organization according to the pin layout of the FEA input connector. The 4 connectors are interfaced with a single adaptor (Amplifier Manifold, Blackrock Microsystems, USA), which is fixed on the FEA case. The fact that higher contamination levels was consistently found on the top pins independently of the socket might be explained by a less tight fixation of the top of the adaptor, possibly causing the microphonic effect.

3.3.3. In vitro evidence of acoustic contamination

Next, we used a reduced experimental setup to determine more in details the cause of the observed correlations (see figure 12). The experiment was designed to verify that the correlations between the sound and the electrode recordings can be obtained without brain activity and to attempt to demonstrate that the correlations originate from the mechanical transmission of sound vibrations. The electrical potentials of ECoG electrodes placed in PBS were recorded while pure tone sounds were played by the same loudspeaker as the one used to present sounds to Participant P5 and with similar intensity. In order to evaluate the intensity of the incident sound, a microphone was placed near the container filled with PBS. A soundproof box was used to insulate either the loudspeaker, the ECoG array, or part of the acquisition chain. The function of the box was to reduce the propagation of sound from the loudspeaker to the devices without substantially interfering with other parameters of the experiment. To determine the impact of sound propagation on the spectrogram correlations, we analyzed the data in open and closed box conditions.

In the first configuration, the loudspeaker was placed in the open box (figure 12(a)). As for *in vivo* experiments, we found that high correlations occurred at some of the frequencies of the sound stimuli. For some electrodes, the value of the correlation coefficient at 125 Hz was larger than 0.9. This result demonstrates that spectrogram correlations similar to those described in sections 3.1–3.3 occur in absence of any brain activity. In the second configuration, the loudspeaker was placed in the closed box (figure 12(b)). The reduction of the power of the incident sound due to the insulation is confirmed by the mean sound PSD (figure 12(b), top). We observed that most of the correlation coefficients also have much lower values (figure 12(b), bottom—compare with figure 12(a), bottom). This result supports the hypothesis of acoustic contamination, i.e. that the spectrogram correlations between sound and electrodes data originate from the mechanical propagation of sound to the neural recording hardware.

In the third and fourth configurations, the electrode array and the microphone were placed in the box but the rest of the acquisition chain was left outside. When the box was left opened (figure 12(c)), we observed high correlations at the frequency of the stimuli, similarly to the previous open box condition (figure 12(a)). The differences of frequency responses visible in the mean sound PSD across the 3 open box conditions can be explained by the modification of the arrangement of the experimental setup. In the last configuration, the box was closed over the electrodes and microphone (figure 12(d)). The sound insulation provided by the box was confirmed by the large reduction of the sound stimuli mean PSD (figure 12(d), top). However, as shown in the bottom graph of figure 12(d), the spectrogram correlations remained largely unaffected by the closing of the lid over the electrode array, contrarily to the previous experiment where the lid was closed over the loudspeaker (figure 12(b)). This suggests that the acoustic contamination of the electrical potential measurement may not only occur at the electrode level but also at other levels of the acquisition chain.

To test this hypothesis, the amplification and digitization chain (A/D chain, composed by the cables, adaptors, splitter box and FEA) was put inside the sound-attenuating box with the microphone. In this case the electrodes in PBS were outside the box. While correlations were high when the box was open (figure 12(e)), they were strongly reduced when the box was closed (figure 12(f)). This further confirmed that the acoustic contamination mainly occurs in the recording chain and not at the electrodes level.

3.3.4. Localization of acoustic contamination along the recording chain

Finally, we aimed at determining where along the recording chain the contamination occurred. The fact that contamination was observed in participant CN in the afternoon but not in the morning session suggests that disconnecting and reconnecting the electrodes to the system could have produced the contamination to occur in the afternoon. To test this more thoroughly, sounds were delivered very locally at different locations along the recording chain connected to electrodes bathed in PBS (figure 13). The statistical criterion P indicated that contamination was found for every location of the sound delivery but with varying intensity, as can be seen on the channel-frequency correlation heat maps (figure 13(c)). Only very weak contamination could be observed when the sound was delivered next to the electrodes in the PBS solution. By contrast, a clear contamination was observed when the sound was delivered against the ECoG grid cables, the pigtail connectors, or the splitter box touchproof connectors. This result is coherent with the idea that contamination is caused by the mechanical vibration of hardware elements, as observed in section 3.3.2 at the level of the FEA input connector in the case of P5 ECoG perception recordings.

4. Discussion

Data considered in this study includes human and animal recordings during speech production and/or sound perception tasks. Using these different setup conditions, we observed statistically significant correlations between the spectrograms of electrophysiological and simultaneously recorded audio signals. These correlations occurred at frequencies most present in the sound signal, thus encompassing the high-gamma range and also frequencies above 300 Hz. This contamination effect was observed in recordings from ECoG and μECoG grids and intracortical micro-electrode arrays, interfaced with different data acquisition systems. The phenomenon was observed in data collected by three out of the five centers worldwide who participated in this study. Thus, this variety of situations suggests that acoustic contamination of neural signals, although not systematic, is a widespread phenomenon.

Motion artifacts are classically seen in electrophysiological signals. In particular, mechanical vibrations may create variations of biopotential measurements (Luna-Lozano and Pallas-Areny 2010). Such undesired signals may have different origins, including the bending of the electrode wires and the electrochemical changes at the electrode-electrolyte interface induced by small displacements of the electrodes (Michelson *et al* 2018, Nicolai *et al* 2018). Here we observed sound contamination of neural signals in different setups. We could reproduce the phenomenon in a minimal *in vitro* setup, confirming that sound-electrode correlations do not originate from brain activity and arise from the impact of sound vibrations on the acquisition chain. The experiments shown in sections 3.3.1–4 further suggest that in the tested setup, the microphonic effect does not necessarily take place at the electrodes' level, but in the rest of the recording chain. In section 3.3.1, two recordings involving the same participant on the same day show different levels of contamination, which can be attributed to the disconnection of the recording hardware between the two sessions. Analyses of the spatial distribution of contamination in a highly contaminated recording showed that it was coherent with a microphonic effect occurring at the input connection of the amplifier but not at the electrode level (see section 3.3.2). *In vitro* experiments show that isolating the acquisition chain from sound reduces the contamination, as opposed to isolating the electrodes (see section 3.3.3). Focal sound delivery at different locations along the chain showed the acoustic contamination was prominent mainly at the level of cables and connectors. Improving elements composing the recording chain thus appears mandatory to ensure proper and artifact-free neural signal recordings.

Although contamination could occur through a crosstalk between channels within the same hardware acquiring simultaneously sound and neural signals, we excluded such possibility in several of the recordings considered here. In particular contamination was observed in a rat recording while the sound and the neural signals were acquired with two separate hardware (a CED micro1401 for the sound and a RHD Intan system for the neural signals). We observed that there is actually no particular hardware element responsible for the contamination. Rather, the quality of interconnection is critical and should be verified systematically. For instance, when participant P5 was presented with synthetic speech through a loudspeaker, the microphonic effect likely stemmed from the quality of the FEA connector (figure 11). However, for *in vitro* PBS recordings, the contamination was small at the level of the FEA connector and more important at the level of the headbox and cables situated upstream (figure 13). Moreover, the same setup used with different participants (or even the same participant in two different sessions) may sometimes exhibit contamination and sometimes not (as here with participant CN). Thus, a given setup should likely not be granted for clean once and for all. Rather, we suggest that any dataset should be objectively tested for any contamination before being considered for further analysis.

The extent to which the acoustic noise spectrally overlaps with the measured brain activity depends on the nature of the sound and on the studied activity. In the case of ECoG recordings during speech production paradigms (see section 3.1), the overlap between the range of the voice fundamental frequency and the high-gamma band might compromise recording artifact-free signals in this band. As suggested by results in section 3.3, sound stimuli, and by extension any sound during the recording, could contaminate the recorded data in any frequency band, including those of multi-unit activity (see section 3.2). This is all the more important that several studies have reported best decoding performance when using a window for the neural features centered with respect to the current time point of the speech feature to be decoded (Martin *et al* 2014, Chartier *et al* 2018, Herff *et al* 2019, Anumanchipalli *et al* 2019). In such case, contamination occurring at delays inferior to 10 ms (figure 6) could bias the decoding results.

While these results demonstrate what could be seen as a relatively trivial contamination of electrophysiological recordings by surrounding acoustic signals, the implication of the study is important in both the neuroscience and neuro-engineering domains. In particular, the common investigation performed here on several datasets acquired in various research places worldwide (France, China, Germany, Switzerland,

and the United States) suggest that decoding analyses should be performed after having excluded any potential microphonic effect.

Yet and importantly, this report does not question the existence of relevant physiological neural information in high-gamma frequency signals underlying speech production or sound perception. Several groups have shown that spectral features of imagined speech or silent articulation can be predicted to some extend from low or high-gamma signals recorded in participants that are not overtly speaking (Pei *et al* 2011, Ikeda *et al* 2014, Martin *et al* 2014, 2016, Bocquelet *et al* 2016a, Gehrig *et al* 2019, Anumanchipalli *et al* 2019). Also, contamination by mainly the fundamental frequency could be insufficient to explain the decoding performance of sublexical features such as articulatory gestures and phonemes, especially consonants (Chartier *et al* 2018, Mugler *et al* 2018). However, we think it is important for past and future studies assessing the contribution of high-gamma or multiunit activity to speech decoding to make sure that neural signals are free of acoustic contamination in the considered frequency bands.

The purpose of this study is therefore to alert on possible microphonic contamination of neural signals, especially when building decoders of neural activity underlying overt speech production or sound perception. Future developments of speech prostheses should thus build upon these findings. In particular, experimental setups should be improved to become less sensitive to microphonic effects, and signal-processing techniques should be developed to eliminate sound contamination in neural recordings. Meanwhile, data should be carefully tested ahead of further decoding analysis.

5. Code availability

A Matlab toolbox implementing the proposed approach to quickly assess the extent of contamination in an electrophysiological recording is made freely available on Zenodo with the following DOI: https://doi.org/10.5281/zenodo.3929296.

Acknowledgments

This work was supported by the FRM Foundation under grant No DBS20140930785, the French National Research Agency under Grant Agreement No. ANR-16-CE19-0005-01 (Brainspeak), and by the European Union's Horizon 2020 research and innovation program under Grant Agreements No. 732032 (BrainCom), No. 696656 (GrapheneCore1), and No. 785219 (GrapheneCore2), as well as the Swiss National Science Foundation under Grant No. 167836. C K is supported by the German research foundation, Grant KE1514/2-1. K J M is supported by NIH/CTSA/KL2 TR002379 (NCATS) Grant, the Brain and Behavior Research Foundation, and the Brain Research Foundation.

ORCID iDs

Shaomin Zhang https://orcid.org/0000-0001-6311-5946
Pierre Mégevand https://orcid.org/0000-0002-0427-547X
Blaise Yvert https://orcid.org/0000-0002-8850-7935

References

Akbari H, Khalighinejad B, Herrero J L, Mehta A D and Mesgarani N 2019 Towards reconstructing intelligible speech from the human auditory cortex *Sci. Rep.* **9** 874

Angrick M, Herff C, Mugler E, Tate M C, Slutzky M W, Krusienski D J and Schultz T 2019 Speech synthesis from ECoG using densely connected 3D convolutional neural networks *J. Neural. Eng.* **16** 036019

Anumanchipalli G K, Chartier J and Chang E F 2019 Speech synthesis from neural decoding of spoken sentences *Nature* **568** 493–8

Bartels J, Andreasen D, Ehirim P, Mao H, Seibert S, Wright E J and Kennedy P 2008 Neurotrophic electrode: method of assembly and implantation into human motor speech cortex *J. Neurosci. Methods* **174** 168–76

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

Bocquelet F, Hueber T, Girin L, Savariaux C and Yvert B 2016b Real-time control of an articulatory-based speech synthesizer for brain computer interfaces *PLoS Comput. Biol.* **12** 1–28

Bouchard K E, Mesgarani N, Johnson K and Chang E F 2013 Functional organization of human sensorimotor cortex for speech articulation *Nature* **495** 327–32

Brumberg J S, Nieto-Castanon A, Kennedy P R and Guenther F H 2010 Brain-computer interfaces for speech communication *Speech Commun.* **52** 367–79

Chan A M *et al* 2013 Speech-specific tuning of neurons in human superior temporal gyrus *Cereb. Cortex* **10** 2679–93

Chartier J, Anumanchipalli G K, Johnson K and Chang E F 2018 Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex article encoding of articulatory kinematic trajectories in human speech sensorimotor cortex *Neuron* **98** 1042-54.e4

Cheung C, Hamiton L S, Johnson K and Chang E F 2016 The auditory representation of speech sounds in human motor cortex *Elife* **5** 1–19

Fontolan L, Morillon B, Liegeois-Chauvel C and Giraud A-L 2014 The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex *Nat. Commun.* **5** 4694

Gehrig J *et al* 2019 Low-frequency oscillations code speech during verbal working memory *J. Neurosci.* **39** 6498–512

Guenther F H *et al* 2009 A wireless brain-machine interface for real-time speech synthesis *PLoS One* **4** e8218

Herff C, Diener L, Angrick M, Mugler E, Tate M C, Goldrick M A, Krusienski D J, Slutzky M W and Schultz T 2019 Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices *Front. Neurosci.* **13** 1267

Herff C, Heger D, de Pesters A, Telaar D, Brunner P, Schalk G and Schultz T 2015 Brain-to-text: decoding spoken phrases from phone representations in the brain *Front. Neurosci.* **9** 217

Hyafil A, Fontolan L, Kabdebon C, Gutkin B and Giraud A-L 2015 Speech encoding by coupled cortical theta and gamma oscillations *Elife* **4** 1–23

Ikeda S, Shibata T, Nakano N, Okada R, Tsuyuguchi N, Ikeda K and Kato A 2014 Neural decoding of single vowels during covert articulation using electrocorticography *Front. Hum. Neurosci.* **8** 125

Kennedy P, Andreasen D, Bartels J, Ehirim P, Mao H, Velliste M, Wichmann T and Wright J 2011 Making the lifetime connection between brain and machine for restoring and enhancing function *Prog. Brain Res.* **194** 1–25

Leuthardt E C, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z, Solis J, Breshears J and Schalk G 2011 Using the electrocorticographic speech network to control a brain-computer interface in humans *J. Neural. Eng.* **8** 36004

Luna-Lozano P S and Pallas-Areny R 2010 Microphonics in biopotential measurements with capacitive electrodes: *2010 Annual Int. Conf. of the IEEE Eng. Med. Biol. Soc.* 3487-90

Martin S, Brunner P, Holdgraf C, Heinze H-J, Crone N E, Rieger J, Schalk G, Knight R T and Pasley B N 2014 Decoding spectrotemporal features of overt and covert speech from the human cortex *Front. Neuroeng.* **7** 14

Martin S, Brunner P, Iturrate I, Del Millán J R, Schalk G, Knight R T and Pasley B N 2016 Word pair classification during imagined speech using direct brain recordings *Sci. Rep.* **6** 25803

Michelson N J, Vazquez A L, Eles J R, Salatino J W, Purcell E K, Williams J J, Cui X T and Kozai T D Y 2018 Multi-scale, multi-modal analysis uncovers complex relationship at the brain tissue-implant neural interface: new emphasis on the biological interface *J. Neural. Eng.* **15** 033001

Miller K J 2019 A library of human electrocorticographic data and analyses *Nat. Hum. Behav.* **3** 1225–35

Miller K J, Abel T J, Hebb A O and Ojemann J G 2011 Rapid online language mapping with electrocorticography: clinical article *J. Neurosurg. Pediatr.* **7** 482–90

Mugler E M, Patton J L, Flint R D, Wright Z A, Schuele S U, Rosenow J, Shih J J, Krusienski D J and Slutzky M W 2014 Direct classification of all American English phonemes using signals from functional speech motor cortex *J. Neural. Eng.* **11** 035015

Mugler E M, Tate M C, Livescu K, Templer J W, Goldrick M A and Slutzky M W 2018 Differential representation of articulatory gestures and phonemes in precentral and inferior frontal gyri *J. Neurosci.* **4653** 1206–18

Nicolai E N, Michelson N J, Settell M L, Hara S A, Trevathan J K, Asp A J, Stocking K C, Lujan J L, Kozai T D Y and Ludwig K A 2018 Design choices for next-generation neurotechnology can impact motion artifact in electrophysiological and fast-scan cyclic voltammetry measurements *Micromachines* **9** 494

Pasley B N, David S V, Mesgarani N, Flinker A, Shamma S, Crone N E,, Knight R T and Chang E F 2012 Reconstructing speech from human auditory cortex *PLoS Biol.* **10** e1001251

Pasley B N and Knight R T 2013 Chapter 17 - decoding speech for understanding and treating aphasia *Prog. Brain Res.* edn **207** 435–56

Pei X, Barbour D L, Leuthardt E C and Schalk G 2011 Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans *J. Neural. Eng.* **8** 046028

Schalk G, Mcfarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 BCI2000: A general-purpose brain-computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43

Small L H 2012 *Fundamentals of Phonetics: A Practical Guide for Students* 3rd edn (New York: Pearson)

Stavisky S D *et al* 2019 Neural ensemble dynamics in dorsal motor cortex during speech in people with paralysis *eLife* **8** e46015

Tankus A, Fried I and Shoham S 2012 Structured neuronal encoding and decoding of human speech features *Nat. Commun.* **3** 1015

Yildiz I B, Mesgarani N and Deneve S 2016 Predictive ensemble decoding of acoustical features explains context-dependent receptive fields *J. Neurosci.* **36** 12338–50