

Sound representation in higher language areas during language generation

Lorenzo Magrassi^{1,2,3}, Giuseppe Aromataris⁴, Alessandro Cabrini⁴, Valerio Annovazzi-Lodi⁴, Andrea Moro³

1) Neurosurgery - Dept. of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Italy 2) IGM-CNR, Pavia, 3) Neurocognition and Theoretical Syntax Research Center (NeTS), Institute of Advanced Study IUSS Pavia, Italy 4) Dept. of Electrical, Computer, and Biomedical Engineering, University of Pavia, Italy

Submitted to Proceedings of the National Academy of Sciences of the United States of America

ABSTRACT: How language is encoded by neural activity in the higher level language areas of humans is still largely unknown. We investigated if the electrophysiological activity of Broca's area correlates with the sound of the utterances produced. During speech perception the electric cortical activity of the auditory areas correlates with the sound-envelope of the utterances. In our experiment we compared the electrocorticogram recorded during awake neurosurgical operations in Broca's area and in the dominant temporal lobe with the sound-envelope of single words vs. sentences read aloud or mentally by the patients. Our results indicate that the electrocorticogram correlates with the sound-envelope of the utterances starting before any sound is produced and even in the absence of speech when the patient is reading mentally. No correlations were found when the electrocorticogram was recorded in the superior parietal gyrus an area not directly involved in language generation or in Broca's area when the subjects were executing a repetitive motor task without any linguistic content with their dominant hand. The distribution of supra-threshold correlations across frequencies of cortical activities varied whether the sound-envelope derived from words or sentences. Our results suggest that the activity of language areas is organized by sound when language is generated before any utterance is produced and heard.

Language | Sound envelope | Broca's area | Morphosyntax | Temporal lobe

An important aspect of human language is speech production although language may be generated independently from sound as when one writes or thinks. However, introspection seems to suggest that our thoughts resound in our brain much as if we were listening to an internal speech, yielding the impression/illusion that sound is inseparable from language.

When human subjects listen to utterances, the neural activity in the superior temporal gyrus is modulated to track the envelope of the acoustic stimulus. The correlation between the power envelope of the speech and the neural activity is maximal at low frequencies (2-7 Hz, theta range) corresponding to syllable rates and becomes less precise at higher frequencies (15-150 Hz, gamma range) corresponding to phonemes rates (1). Entrainment of neural activity to the speech envelope in auditory regions has allowed recognition of the phonetic features heard during speech perception (2-5) and even the reconstruction of simple words (6). The above evidence indicates that during listening, speech representation in the auditory cortex and adjacent areas of the superior temporal lobe reflects acoustic features directly related to linguistically defined phonological entities such as phonemes and syllables. The relationship of specific patterns of sound amplitude and frequency to all similar patterns experienced over phylogeny and individual ontogeny is then responsible for sound perceptions (7). Moreover, as for subjects listening to natural speech, spatio-temporal features of the acoustic trace representative of the sound of the listened words have also been detected in the neural activity of cortical areas outside the superior temporal gyrus (8).

Is sound representation essential during the generation of linguistic expression before the implementation of the motor program for speech or is the neural activity in higher language areas completely independent of it?

Results

Cortical activity correlate with the sound envelope of the words

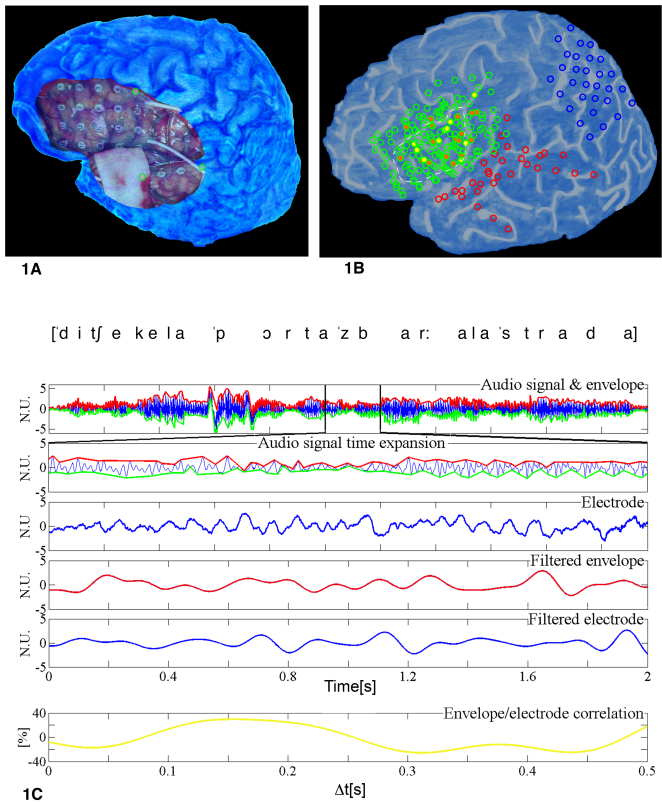
After obtaining approval from the institutional ethics committee, we retrospectively analyzed the electrocorticographic activity (ECoG) and concomitant sound tracks recorded from the dominant frontal and temporal lobes of native Italian-speakers using high-density surface multielectrode arrays (HDM) during awake neurosurgical operations. In Fig.1 we show an example of the typical HDM positioning on the frontal and temporal lobes (Fig. 1a) and a map of the positions of all electrodes in all patients involved in the study (Fig. 1b). We calculated the cross-correlation between the sound envelope of the words and sentences read aloud by the patients and the corresponding ECoG traces of the prefrontal electrodes (Fig. 1c) along with its periodogram, computed by the fast Fourier transform (FFT). This correlation cyclically increased with a frequency coherent to the pace of reading (Fig. 2a). On the contrary, when we compared the sound envelope of the same utterances to the corresponding ECoG activity recorded in the superior parietal gyrus, an area not involved in language generation showing weak fMRI signal during discourse comprehension (9), no periodic variations in correlation amplitudes were found (Fig. 2b). As a further control, during the same surgical session, in 3 patients we also compared the sound envelope of the words and sentences read aloud in

Significance

the results of our experiments show that a special representation of sound is actually exploited by the brain during language generation even in the absence of speech. Taking advantage of data collected during neurosurgical operations on awake patients we cross-correlated the cortical activity in frontal and temporal language areas of a person reading aloud or mentally with the envelope of the sound of the corresponding utterances. In both cases cortical activity and the envelope of the sound of the utterances were significantly correlated. This suggests that, in hearing people, sound representation deeply informs generation of linguistic expressions at a much higher level than previously thought This may help design new strategies to help people with language disorders like aphasia.

Reserved for Publication Footnotes

137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204



Magrassi et al. Fig.1

Fig. 1. a. Reconstruction of the surface of the cerebral cortex of one of the patients from magnetic resonance images obtained before surgery. An intraoperative picture of the HDMs was projected over the reconstruction at the corresponding intraoperative position. Electrodes that were partially or completely covered by the dura or a cottonoid used to stabilize the HDM are indicated by green circles. **Fig. 1b.** The position of electrodes for all participants recorded from the dominant left hemisphere were mapped on the outline of the surface of the cerebral cortex shown in fig.1A. Green circles represent electrodes in the dominant frontal lobe, yellow filled green circles are electrodes positioned over the areas where stimulation induced speech arrest in all trials, orange filled green circles indicate electrodes located in areas where stimulation induced speech arrest in at least one trial. Red circles represent electrodes in a separate HDM located over the dominant left temporal lobe, blue circles indicate the position of electrodes located over the superior parietal gyrus. The white dashed line outlines the anatomical Broca's area. **Fig. 1c.** From top to bottom: phonetic transcription of the sample Italian sentence (dice che la porta sbarrà la strada: says-3rd-sing. that the door blocks the way "s/he says that the door blocks the way" [dɪtʃe ke la 'pɔrta 'zbar:a la 'strada]); (Audio signal & envelope) its acoustic waveform with the corresponding envelopes (positive and negative) outlined in red and green respectively; (Audio signal time expansion) an expanded acoustic waveform and envelopes of the section comprised among the two vertical lines; (Electrode) the corresponding ECoG trace recorded from one electrode positioned in the area of speech arrest; (Filtered envelope) the corresponding positive envelope (red trace) of the 2750-3250 Hz audio band; (Filtered electrode) the corresponding ECoG filtered in a selected frequency band (2-8 Hz). The x-axis of all the above signals represent time in seconds (Time [s]). (Envelope/electrode correlation) the resulting cross-correlation of the filtered positive sound envelope and filtered ECoG, $\Delta t[s]$: time differential in seconds between the ECoG and the audio trace, in the present example the correlation maximum entirely contained in the 500ms window occurs when the ECoG anticipates the start of sound of about 170 ms.

one trial, with the ECoG activity recorded from the dominant prefrontal cortex in a further trial where the patients were silent. In this trial instead of reading, they pushed a button with the hand contralateral to the dominant hemisphere each time they saw a

slide showing the drawing of a finger pushing a button. Slides with the pushing finger were randomly alternated with black slides at a rate comparable to the rate of the slides displaying linguistic items. Under this experimental condition, we did not find any periodic variations in correlation amplitudes (Fig. 2c). Finally, we compared the ECoG activity of patients reading mentally with the sound envelope obtained while they were reading aloud. The words and the sentences the patients were reading mentally, were the same and they were presented at the same pace as when they were reading aloud. In this context the correlation cyclically increased with a frequency coherent to the pace of reading even in the complete absence of sound (Fig. 2d).

In order to increase the temporal resolution of the study of the cross-correlation of the sound envelope with the ECoG, the ECoG and the sound traces were sliced into periods corresponding to the single utterances of each subject. Periods containing the same utterances in a single trial were put in register and averaged. The average duration of the audio trace of any utterance for each subject was relatively constant with relative standard deviations always below 10% (min. 5.11% , max 9.64%) of the average duration of the utterance. Furthermore, the sound resulting from the average of the single utterances of a subject was always understandable. Strengthening our findings in longer unselected time periods, when we considered all electrodes, the average value of the cross-correlation of the averaged ECoG and the envelope of the acoustic trace of words and sentences was 36.86 (S.D. + 15.64) well-above the background (patient being silent and not reading mentally) maximal value, i.e. 20.00. Both in the lateral inferior prefrontal cortex containing Broca's area and in the superior temporal gyrus, the amplitude of correlations during reading was maximal (44.72, S.D. + 17.12) in the theta range but present (29.07, S.D. + 8.61) also in the gamma range (Fig. 3a). These findings are similar to those described in the superior temporal gyrus during listening (1, 6).

Random shuffling of the ECoG abrogates the correlation

When we randomly shuffled segments obtained from the ECoG, the amplitude of the cross-correlation with the corresponding unshuffled sound envelope decayed exponentially with the inverse of segment duration (Fig. 3b). On average the amplitude of the cross-correlation reduced to half of the original when 90ms segments were randomly shuffled. This, together with the negative results of the experiments on long correlations obtained from the superior parietal gyrus or from the frontal lobe when the patient was involved in a motor task, confirms that the cross-correlation between the sound envelope and the corresponding ECoG is not due to interference or random fluctuations of the signals.

Cortical activity of areas close to Broca's area also correlate with the sound envelope of the linguistic expressions generated

Comparing the mean value of cross-correlation among ECoG activities recorded in each explored sites of the dominant frontal and temporal lobes with the envelope of the sound of the words and sentences pronounced, we found correlation values above background in many electrodes located in the dominant frontal and temporal lobes close but outside the anatomically defined Broca's and Wernicke areas (Fig. 3c). Our results are in line with previous results derived from stimulation mapping of language areas during neurosurgery (10, 11) showing that location of sites whose stimulation could interfere with language varied widely from patient to patient and were not limited to the anatomically defined Broca's and Wernicke areas. We did not find statistically significant differences (t-test unpaired $p = 0.158$) between electrodes in the frontal lobe where cortical stimulation induced speech arrest during naming (average 36.52 S.D. + 15.42) with respect to those where the stimulation did not interfered with language activity (average 36.85 S.D. + 15.68). When we compared electrodes located over the anatomically defined Broca's

205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272

273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340

341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408

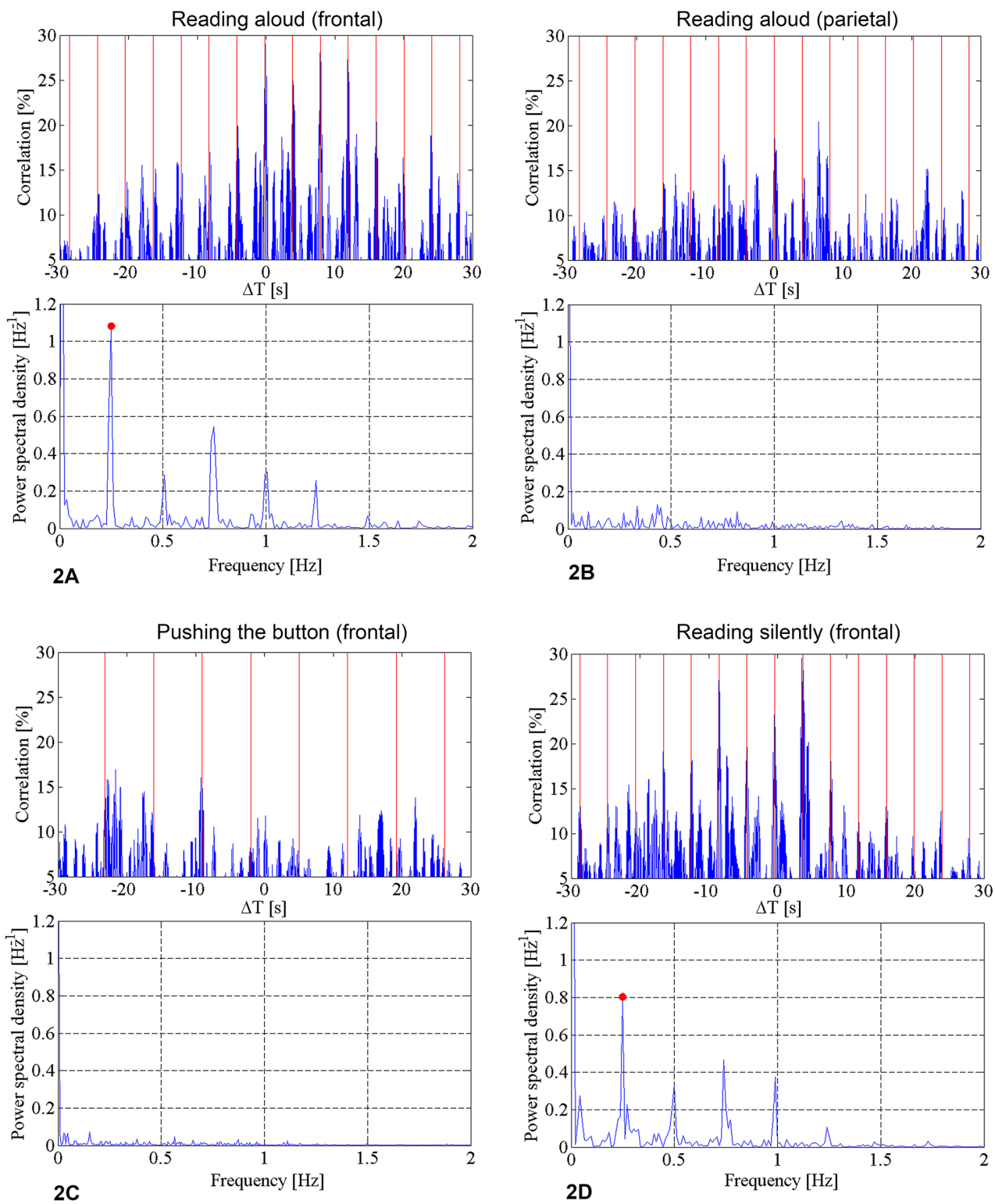


Fig. 2. d. Same patient, same pace of presentation of the linguistic items and same electrode as in 2a but here the patient reads mentally 12 of the sentences he previously read aloud. The trace represents a period of 50 s read in a completely silent mode. Variations in cross correlation between the ECoG recorded while the patient was reading mentally and the envelope of the audio recorded in a previous trial when the patient was reading aloud the same linguistic items, follow the presentation time (4 s) of the linguistic items (red vertical lines). The periodogram below shows a dominant frequency of variations of the correlation of 0.25 Hz (red dot), corresponding to a period of 4 s that, as in 2A, equals the rate of presentation of the linguistic items.

area (average 36.99 S.D. + 15.75) independently from the result of cortical stimulation, versus electrodes in the frontal lobe that

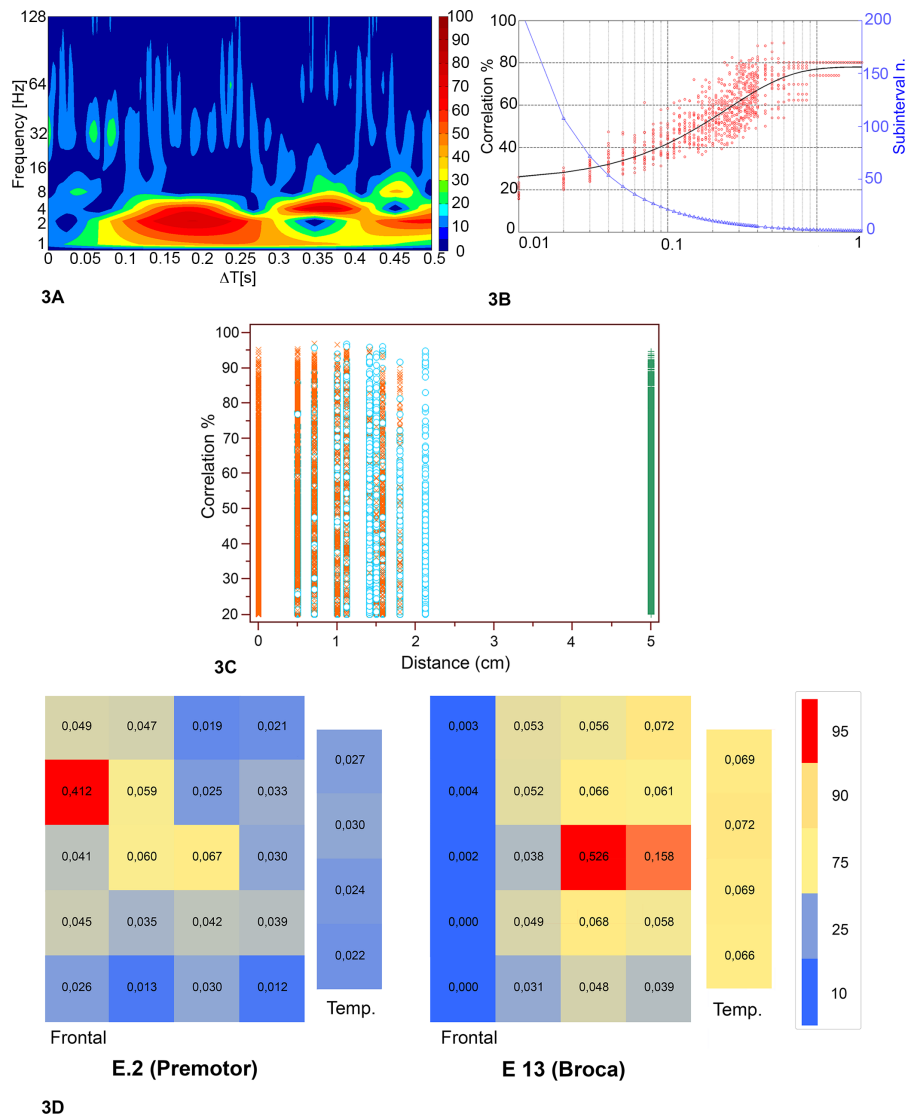


Fig. 3. a. Cross-correlation as a function of frequency of the ECoG and the speech envelope, same sentence as in fig. 1c. Data have been separately normalized on the different frequency bands. Notice that maximal correlation values are within the theta band (2-8Hz). Frequency [Hz]: frequency of the ECoG in Hz, ΔT : time differential in seconds between the ECoG and the audio trace of overt speech, ECoG response always anticipates sound. **Fig. 3b.** Plot showing the dependency of the amplitude of cross-correlation (Correlation %) of the ECoG and the speech envelope (left y axis) on the correct linear order of the original signal. When we randomly shuffled subintervals of decreasing length obtained by cutting the audio and ECoG trace of the same sentence represented in 1C and 3A we obtained an exponential decay of the correlation amplitude that halved for an interval duration of approximately 0.09 s. Red circle represents the correlation value obtained of each shuffling, the black line fits the mean values of the correlation amplitude obtained in all independent shuffling for each subinterval length. Subinterval n.: the number of possible subintervals for each interval duration (right y axis) is shown by blue triangles. Duration (s): length of subintervals in seconds. **Fig. 3c.** Amplitude of the cross-correlation above background maximal value (20%) of the ECoG and the speech envelope as a function of the linear distance (cm) of the recording electrodes from the electrodes overlaying the area where electrical stimulation induced speech arrest (0). Temporal electrodes indicated by green "+" were all referred to the average distance of 5cm. Electrodes located over the anatomical Broca's area are indicated by red "x", frontal electrodes outside anatomical Broca's area are indicated by cyan circles. All data (90.504) obtained in all patients at all frequency of audio and ECoG bands are represented. No significant differences in the distribution of the correlation amplitudes are visible between electrodes located over the speech arrest area and electrode outside it. Please notice that electrodes located over the superior parietal gyrus are not represented here. **Fig. 3d.** Spatial distribution of all the electrodes showing a supra-threshold value (20%) for the correlation between the audio envelope and ECoG simultaneous to the correlation peak of one electrode either located over an area of the frontal cortex whose stimulation did not interfered with language (E2 (Premotor)), or over another area of the cortex whose stimulation induced speech arrest (E13 (Broca)). Each square represent an electrode, colors (red: maximal, blue minimal) are proportional to the frequency of occurrence of suprathreshold correlation values and is indicated by the number in the middle of the square. Bigger rectangles (Frontal) represent electrodes located over the dominant frontal lobe, smaller rectangles (Temp.) represent electrodes located over the dominant temporal lobe. A color scale is provided with the indication of percentiles corresponding to each color. The probability distribution of a suprathreshold correlation changes significantly according to the position of the reference electrode considered. The number of electrodes in the dominant temporal lobe that reach suprathreshold correlation values simultaneously to the electrode located in the area of speech arrest is significantly higher than when the electrode was located on a prefrontal area where speech arrest could not be evoked.

were not (average 36.21 S.D. + 15.52), we found a statistically significant difference in amplitude (t-test unpaired $p < 0.0001$) however this difference is so small compared to the average

fluctuations of background correlation that we do not consider it as biologically relevant. The only robust difference between electrodes located above the areas of speech arrest vs. those that

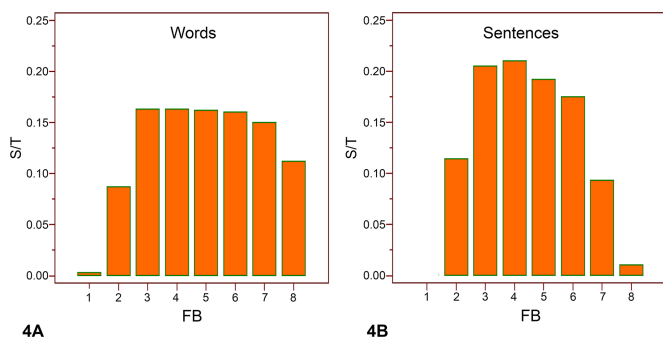


Fig. 4. The cross-correlation of the envelope of the speech sounds with the ECoG above background is differentially distributed in the frequency bands according to the presence of specific morphosyntactic structures (words vs. sentences) contained in the text read by the patients. **Fig. 4a.** Histogram showing the number of electrodes with suprathreshold correlation in each ECoG frequency bands (FB) when patients were reading words. S/T: ratio between electrodes showing suprathreshold correlation in each ECoG frequency band and the total number of electrodes in all ECoG frequency bands. **Fig. 4b.** Same as above but values referred to patients reading sentences.

were not was that electrodes that had a correlation value above background when those in the speech-arrest had maximal correlation values were significantly different (chi-squared 746,429, DF 529, $p < 0.0001$) from those that had maximal correlation values when the electrodes not located in speech arrest areas showed maximal correlation values (Fig. 3d).

The average anticipation of the ECoG over the sound envelope precedes phonological processing

Language related neural activity in Broca's area as shown by large field potentials (LFP) starts approximately 400 ms before sound is produced and is characterized by 3 further components at 280 ms, 150 ms and 50 ms¹². In our experiment, the average anticipation of the ECoG over the sound envelope was 245 ms (S.D. + 11.27). This value that maximized the correlation between the two signals is consistent with the second component of LFPs that is correlated with grammatical processing and precedes later components related to the phonological and the phonetic processing and the articulatory motor commands (12, 13). This value was similar when patients were reading aloud as well as when they were reading the same sentences and words mentally. Further indication that the correlation between sound envelope and ECoG has biological significance come from the observation that the length of the segments that significantly start to reduce correlation after shuffling falls into the same range of length of the segments of an utterance that, when locally time reversed, make it partially unintelligible (14).

Differences in spectral distributions of the correlation amplitude distinguish words vs. sentences

Utterances may be distinguished by their phonological and grammatical contents. We found that values above background of the cross-correlation between the sound envelope and the ECoG are differentially distributed in the frequency bands according to the presence of specific morphosyntactic structure (i.e. words vs. sentences) contained in the text read by the patients (Fig. 4). Differences in spectral distributions are statistically significant and allow a distinction between words pronounced in isolation vs. sentences (chi-squared 1112.469, DF 7, $p < 0.0001$). This difference remains significant even when sentences and words of similar time duration are compared (chi-squared 23.717, DF 7, $p < 0.0013$). Interestingly, most differences are in the high gamma range of the ECoG (Fig. 4), where individual suprasegmental features including prosody effects should be less relevant (1).

Discussion

Our experiment shows that neural activity in Broca's area, a prototypical high level language area governing morphosyntactic structures across languages (15-19), and in the high level language areas of the dominant temporal lobe, is informed by the envelope of the sound of the linguistic expressions that that same activity is encoding. This is true whether the encoded linguistic items will then be uttered or not (as when we think). Our findings concerning the correlation of cortical activity in high level language areas to the sound envelope of the encoded linguistic items are symmetrical to those emerging by studying cortical activity during speech perception (3, 5, 6). Even if we assume that subvocalization with subthreshold activation of the phonatory apparatus is always active when we think (20), the correlation between ECoG and the envelope of the sound of the linguistic items involved is not limited to the late phonological and the phonetic processing and the encoding of the articulatory motor commands that immediately precede sound production. The value of 245 ms (S.D. + 11.27) we found in order to maximize the correlation between the sound envelope and ECoG signals precedes of about 90ms the start of those later activities and is consistent with earlier activities in high level language areas that are correlated with more abstract linguistic feature like grammatical processing (12, 13). Our ability to find significant differences in the spectral distributions of the correlation amplitude between words and sentences of equal duration also support the notion that sound encoding is involved in early step of the linguistic production and, at least in normal subjects is not limited to speech production. Prelingually deaf schizophrenic that depended on the same left hemispheric areas that control language production in hearing people for sign language (21) report of hearing inner voices with frequencies comparable to normally hearing people (22). Although the current interpretation of hearing voices in congenitally deaf people is controversial (23) it is possible that some rudimentary representation of the auditory consequence of articulation is present in some deaf people (23).

Our results suggest that, in normal hearing people, sound representation is at the heart of language and not simply a vehicle for expressing some otherwise mysterious symbolic activity of our brain.

Materials and Methods

For an extended description of the techniques employed see the supplemental Materials and Methods section.

Participants and cortical mapping

The present study (named LIP51) was approved by the Fondazione I.R.C.C.S. Policlinico S. Matteo institutional review board and ethic committee on Human Research. All analyses, whose results are presented in our study, were done off-line and did not interfere with the clinical management of the patients. Acute intraoperative recordings obtained from 16 Italian native speaker patients (12 males and 4 females) affected by primary or secondary malignant tumors growing in the dominant frontal, temporal or parietal lobes were the subject of our study. Stimulation for finding areas of speech arrest were performed according to standard neurosurgical techniques for awake neurosurgical operations (10, 11). The intensity of the current used for cortical stimulation mapping varied from patient to patient being the maximum intensity that did not produce after-discharges in the simultaneous ECoG traces. The position of the stimulating and recording electrodes was determined visually and recorded on a neuronavigator (Vectorvision Brain Lab, Munich, Germany). This allowed the unambiguous classification of all electrodes in the dominant frontal lobes as electrodes corresponding to the anatomical Broca's area (24) or as electrodes that were outside it.

Neural and audio Recordings

Electroencephalographic (ECoG) recordings were obtained using a multi-channel electroencephalographer (System Plus, Micromed, Mogliano, Italy) with sampling rate = 2048 Hz and ADC resolution of 16bit. One or two HDMs grids with an inter-electrode distance of 5 mm measured from center -to-center were simultaneously employed for ECoG recordings. Sound tracks and neural activities were simultaneously recorded during testing; sound was acquired by H1 X/Y stereo microphone (Zoom H1, Zoom corp. Tokyo, Japan) placed at the basis of the neck on the same side of the operated hemisphere at sampling rates of 96KHz.

Neuropsychological testing

545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612

613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680

681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748

Linguistic items: linguistic expressions were based on standard Italian words and sentences aloud pronouncing all the utterances as they are used to while trying to keep the sound intensity as uniform as possible between different reading session. This resulted in a uniform mean absolute amplitude of the sound envelopes obtained from all patients (mean 2.48 mV, s.d. + 0.30 mV). Reading mentally: after successfully completing the reading aloud testing, if the patient was still comfortable and attentive, we asked her/him to read mentally the same words and sentences she/he just read aloud without changing their reading pace. They were also explicitly instructed to avoid lips movement or other voluntary movement mimicking sound emission. Pushing the button: 3 patients were also recorded while they were pushing a button with the hand contralateral to the operated dominant hemisphere. The button was pushed every time was displayed on a computer screen the drawing of a hand with the index finger pushing a red button over a black background. The image of the pushing finger was randomly alternated with black screens at the same pace rate used for the patient when she/he was tested for reading.

Signal analysis

The frequency components of the high-resolution audio signal were separated into 5 contiguous bands (audio band 1: 750-1250, audio band 2 1250-1750, audio band 3 1750-2250, audio band 4 2250-2750, audio band 5 2750-3250 Hz). For each audio band the sound envelope was extracted from the audio track by interpolation followed by low-pass filtering. The ECoG and the envelope were separated into 8 contiguous frequency bands in octave ratio, ranging from 0.04 to 128 Hz (ECoG band 1: 0.04-1 Hz, ECoG band 2: 1-2 Hz, ECoG band 3: 2-4 Hz, ECoG band 4: 4-8Hz, ECoG band 5: 8-16, ECoG band 6: 16-32 Hz, ECoG band 7: 32-64 Hz, ECoG band 8: 64-128 Hz) by using standard numerical Gaussian filters. After normalization, the cross-correlation between the sound envelope and the corresponding ECoG traces of the different electrodes was separately computed for all frequency bands. We highlighted periodicities in the obtained cross correlations by computing

1. Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neurosci.* 15, 511-517.
2. Pulvermüller, F. et al. (2006) Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. USA* 103, 7865-7870.
3. Nourski K-V. et al. (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29, 15564-15574.
4. Bouchard K-E, Mesgarani N, Johnson K, Chang E-F (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495, 327-332. Erratum in: *Nature* 498, 526 (2013).
5. Mesgarani N, Cheung C, Johnson K, Chang E-F (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006-1010.
6. Pasley B-N et al. (2012) Reconstructing speech from human auditory cortex. *PLoS Biol.* 10, e1001251. doi: 10.1371/journal.pbio.1001251.
7. Monson B-B, Han S, Purves D (2013) Are auditory percepts determined by experience? *PLoS One* 8, e63728. doi: 10.1371/journal.pone.0063728.
8. Kubanek J, Brunner P, Gunduz A, Poeppel D, Schalk, G. (2013) The Tracking of Speech Envelope in the Human Cortex. *PLoS One* 8, e53398. doi: 10.1371/journal.pone.0053398.
9. Egidí G, Caramazza A (2013) Cortical systems for local and global integration in discourse comprehension. *Neuroimage* 71, 59-74.
10. Ojemann G, Ojemann J, Lettich E, Berger M (1989) Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. *J. Neurosurg.* 71, 316-326.
11. Sanai N, Mirzadeh Z, Berger M-S (2008) Functional outcome after language mapping for glioma resection. *N. Engl. J. Med.* 358, 18-27.
12. Sahin N-T, Pinker S, Cash S-S, Schomer D, Halgren E (2009) Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326, 445-449.
13. Hanna J et al. (2014) Early activation of Broca's area in grammar processing as revealed by the syntactic mismatch negativity and distributed source analysis. *Cogn. Neurosci.* 5, 66-76.

the periodogram of the signal by a standard fast Fourier transform algorithm (26, 27).

The ECoG signals averaged before evaluating the cross-correlation with the audio envelope started 500ms before the audio track.

Shuffling and statistical analysis

Shuffling was implemented by slicing the electrode signal, corresponding to a given sentence or word, into time subintervals of increasing length, starting from 0.01s with 0.01s step up to 0.4s, and with 0.05 step from 0.4s up to the full time span. Time subintervals of equal duration were then randomly mixed before computing the cross-correlation of the shuffled signal with the unshuffled positive envelope of the audio track. Statistical analysis was performed on absolute values (modules) of cross-correlation. Statistical significance was assessed by parametric (unpaired Student t test) and nonparametric (Mann-Whitney test) methods when amplitude of correlations and delay of envelope of the audio over the ECoG over the envelope of the audio trace were compared. Distribution of electrodes simultaneously active were compared by chi-square statistic. All statistical analyses were calculated by using MedCalc for Windows, version 14.8.1 (MedCalc Software, Ostend, Belgium).

Acknowledgments.

We thank Franco Bottoni MSc. for help with sound recording tools, Mauro Tachimiri together with all EEG technicians for help with intraoperative recordings; Dr. Mauro Benedetti for preliminary discussions; The Neurocognition and Theoretical Syntax (NeTS) Research Center at IUSS for endowing Dr. G. Aromataris with a fellowship; The Department of Electrical, Computer, and Biomedical Engineering at the University of Pavia for endowing Dr. G. Aromataris with a fellowship; prof. Stefano Cappa, prof. Noam Chomsky and prof. Ian Tattersall for useful comments on the manuscript. **Author Contributions** L.M. and A.M. conceived and designed the research; L.M. performed the surgery and collected the data; A.M. designed the linguistic stimuli. G.A. A.C. V.A. designed and implemented the codes, L.M. G.A. A.C. V. A. analyzed the data; L.M. A.M. V.A. G.A. A.C. discussed the results and corrected the drafts, L.M. and A.M. wrote the manuscript that was approved in its final version by all the authors.

14. Saberi K, Perrott D-R (1999) Cognitive restoration of reversed speech. *Nature* 398, 760.
15. Moro A et al. (2001) Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage* 13, 110-118.
16. Friederici A-D (2002) Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78-84.
17. Musso M et al. (2003) Broca's area and the language instinct. *Nature Neurosci.* 6, 774-781.
18. Pallier C, Devauchelle A-D, Dehaene S (2011) Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. USA*, 108, 2522-2527
19. Moro, A. (2008) The boundaries of Babel. The brain and the enigma of impossible languages. The MIT Press, Cambridge MA, USA.
20. Smith J-D, Wilson M, Reisberg D (1995) The role of subvocalization in auditory imagery. *Neuropsychologia.* 33(11), 1433-1454.
21. MacSweeney M, Capek C-M, Campbell R, Woll B (2008) The signing brain: the neurobiology of sign language. *Trends Cogn. Sci.* 12(11), 432-440.
22. du Feu M, McKenna P (1999) Prelingually profoundly deaf schizophrenic patients who hear voices: a phenomenological analysis. *Acta Psychiatr. Scand.* 99, 453-459.
23. Atkinson J-R. (2006) The perceptual characteristics of voice-hallucinations in deaf people: insights into the nature of subvocal thought and sensory feedback loops. *Schizophr Bull.* 32, 701-708.
24. Keller S-S, Crow T, Foundas A, Amunts K, Roberts N (2009) Broca's area: nomenclature, anatomy, typology and asymmetry. *Brain Lang.* 109, 29-48.
25. Hickok G. (1993) Parallel parsing: Evidence from reactivation in garden-path sentences. *J Psycholinguist Res.* 22, 239-250.
26. Oppenheim A-V, Schaffer R-W (1989) Discrete Time Signal Processing. Prentice Hall, Englewood Cliffs, NJ.
27. Orfanidis, S-J. (1995) Introduction to Signal Processing. Englewood Cliffs, NJ: Prentice Hall, Englewood Cliffs, NJ.

Supplementary Materials and Methods Section

Participants and cortical mapping

The present study (named LIPS1) was approved by the Fondazione I.R.C.C.S. Policlinico S. Matteo institutional review board and ethic committee on Human Research. Patient testing, cortical mapping and high-density surface multielectrode arrays (HDM) placement were determined entirely by clinical criteria. All analyses, whose results are presented in our study, were done off-line and did not interfere with the clinical management of the patients. Acute intraoperative recordings obtained from 16 Italian native speaker patients (12 males and 4 females) affected by primary or secondary malignant tumors growing in the dominant frontal, temporal or parietal lobes were the subject of our study. All participants' clinical history was shorter than 3 months and none of the participants had a long-term history of epilepsy (only 4 of them had sporadic epileptic fits before surgery), making it unlikely that any long-term changes in connectivity and brain excitability resulting from chronic epilepsy could limit the relevance of our results for human physiology. Stimulation for finding areas of speech arrest were performed according to standard neurosurgical techniques for awake neurosurgical operations (10, 11). Biphasic square-wave pulses at 60Hz delivered by hand held bipolar electrodes separated by a distance of 5mm (Ojemann cortical stimulator, Radionics, Burlington, MA, USA) were used. The intensity of the current used for cortical stimulation mapping varied from patient to patient being the maximum intensity that did not produce after-discharges in the simultaneous ECoG traces. The position of the stimulating and recording electrodes was determined visually and recorded on a neuronavigator (Vectorvision Brain Lab, Munich, Germany). This allowed the unambiguous classification of all electrodes in the dominant frontal lobes as electrodes corresponding to the anatomical Broca's area (24) or as electrodes that were outside it.

Neural and audio Recordings

Electrocorticographic (ECoG) recordings were obtained using a multichannel electroencephalographer (System Plus, Micromed, Mogliano, Italy) with sampling rate = 2048 Hz and ADC resolution of 16bit. One or two HDMs grids with an inter-electrode distance of 5 mm measured from center -to-center were simultaneously employed for ECoG recordings. One HDM was made of 20 electrodes arranged in a rectangular array (1.5 cm X 2 cm) and a second was made of 4 electrodes arranged in a 2cm long row. The larger HDM was either centered over the main speech arrest area identified by direct cortical stimulation in the dominant frontal lobe of the hemisphere undergoing surgery (left 13 patients, right 2 patient) or the left superior parietal gyrus (2 patients used as controls). The smaller HDM consisted of 4 electrodes arranged in a row and was positioned over the dominant temporal lobe in 10 patients that also had the large HDM positioned over the frontal lobe. The specific positions of all electrodes in the left hemisphere of all patients are shown in fig. 1B. All ECoG signals were visually inspected by a standard viewer (EDF Browser) and traces, and parts of traces, with obvious artifacts (including excessive electromagnetic noise from operating room equipment and poor contact with the cortical surface) were removed. The 50 Hz AC interference was suppressed along with its harmonics (up to 150 Hz) by means of a digital comb filter.

Sound tracks and neural activities were simultaneously recorded during testing; sound was acquired by H1 X/Y stereo microphone (Zoom H1, Zoom corp. Tokyo, Japan) placed at the basis of the neck on the same side of the operated hemisphere at sampling rates of 96KHz. In parallel a low-resolution sound trace with 2048 Hz sampling rate was directly recorded in one channel of the ECoG and later exploited to synchronize the high-resolution sound trace and the electrode signals. To this end, a suitable triggering audio signal was used at the beginning of each recording session.

Neuropsychological testing

Linguistic items: linguistic expressions were based on standard Italian taken from the "Lessico di Frequenza dell'Italiano Parlato" (LIP) included either simple words or sentences. Words included 6

singular nouns depicting common objects either manipulable or not, 2 deverbal nouns with same suffixation expressing actions, 1 complex name of number, 1 verb; 2 words where ambiguous as they could be interpreted either as objects or verbs. The number of syllables varied from 2 to a maximum of 10 (average 3.5) and included all vocalic phonemes and all basic types of consonantic phonemes represented alphabetically. Sentences were all affirmative active present tensed clauses; they included 6 single and 5 complex clauses including a declarative sentence. The number of syllables varied from 7 to 13 (average 10) and included all vocalic phonemes represented and all basic types of consonantic phonemes represented alphabetically: 7 sentences included a non-expressed subject as well-known common property of Italian syntax. As for their semantics: all nouns and verbs were taken from high frequency lexemes; they were all describing simple conditions or states; only one sentence contained a semantic clash. All sentences were strictly unambiguous and were expressly designed not involve “garden-path” effects to avoid rereading (25).

All linguistic items were randomly repeated at least 5 times (min. 5, max. 12) during every testing section (each lasting about 5 minutes) and each testing section was also repeated at least 2 times up to a maximum of 40 minutes.

All words and sentences were written in white characters on a black background shown on a computer screen which was placed at a distance determined by the patient for effortless and comfortable reading.

Reading aloud: The patients were asked to read standard Italian words and sentences aloud pronouncing all the utterances as they are used to while trying to keep the sound intensity as uniform as possible between different reading session. This resulted in a uniform mean absolute amplitude of the sound envelopes obtained from all patients (mean 2.48 mV, s.d. \pm 0.30 mV). The dimension of the fonts and the pace of presentation were determined by the patient for her/his maximal comfort in preliminary trials. These were administered before surgery in order to familiarize the patients with the procedure and the texts.

Reading mentally: after successfully completing the reading aloud testing, if the patient was still comfortable and attentive, we asked her/him to read mentally the same words and sentences she/he just read aloud without changing their reading pace. They were also explicitly instructed to avoid lips movement or other voluntary movement mimicking sound emission. During silent reading an audio trace coming from the microphone on the neck, was continuously recorded and trials from patients that documented sound emission by the patient during mental reading were discarded.

Pushing the button: 3 patients were also recorded while they were pushing a button with the hand contralateral to the operated dominant hemisphere. The button was pushed every time was displayed on a computer screen the drawing of a hand with the index finger pushing a red button over a black background. The image of the pushing finger was randomly alternated with black screens at the same pace rate used for the patient when she/he was tested for reading.

Signal analysis

The frequency components of the high-resolution audio signal were separated into 5 contiguous bands (audio band 1:750-1250, audio band 2 1250-1750, audio band 3 1750-2250, audio band 4 2250-2750, audio band 5 2750-3250 Hz). For each audio band the sound envelope was extracted from the audio track by interpolation followed by low-pass filtering. The ECoG and the envelope were separated into 8 contiguous frequency bands in octave ratio, ranging from 0.04 to 128 Hz (ECoG band 1: 0.04-1 Hz, ECoG band 2: 1-2 Hz, ECoG band 3: 2-4 Hz, ECoG band 4: 4-8Hz, ECoG band 5: 8-16, ECoG band 6: 16-32 Hz, ECoG band 7: 32-64 Hz, ECoG band 8: 64-128 Hz) by using standard numerical Gaussian filters. The amplitudes of the ensuing signals were then normalized imposing that their autocorrelation at zero lag was unit. After normalization the cross-correlation between the sound envelope and the corresponding ECoG traces of the different electrodes was separately computed for all frequency bands.

The cross-correlations were computed over time intervals of several minutes inclusive of many sentences and words, by standard algorithms (26, 27) on ECoG and audio envelope signals that

were not averaged. This allowed a robust analysis of the spectral contributions of the cross-correlation. We then highlighted periodicities in the obtained cross correlations by computing the periodogram of the signal by a standard fast Fourier transform algorithm (26, 27).

In order to increase the temporal resolution in the study of the correlation between sound envelope and ECoG during production of the linguistic items, the periods of the ECoG and audio traces containing the same linguistic item in a single trial were put in register and averaged before calculating the cross-correlation. Utterances were recognized by direct listening aided by suitable software (Audacity). The ECoG signals averaged before evaluating the cross-correlation with the audio envelope started 500ms before the audio track. We extended of 500 ms the ECoG trace because the starting of Broca's area activity anticipates speech production of about 400 ms (Sahin 2009), and 100 ms were added as compensation of the patients' differences in speech rates.

Shuffling and statistical analysis

Shuffling was implemented by slicing the electrode signal, corresponding to a given sentence or word, into time subintervals of increasing length, starting from 0.01s with 0.01s step up to 0.4s, and with 0.05 step from 0.4s up to the full time span. Time subintervals of equal duration were then randomly mixed before computing the cross-correlation of the shuffled signal with the unshuffled positive envelope of the audio track. Depending on the number of the available subintervals we calculated the average of the correlation values of 50 independent shuffled version of the original utterance or the maximum number of the possible combinations of subintervals. We determined the maximum number of independent shuffled version after we verified on selected words and sentences that the obtained correlation mean did not significantly change by averaging over 20, 30, 50, 100 shuffled cases. The minimum time interval of the slices (0.01s) was selected because it is typically considered an appropriate sampling frame for human speech in artificial speech recognition technology and it is below the 50 ms frame whose reversion does not affect speech comprehension (14).

Statistical analysis was performed on absolute values (modules) of cross-correlation. Only values of cross correlation within a window of 500ms delay of the spoken utterances with respect to the ECoG trace were considered in the analyses. Statistical significance was assessed by parametric (unpaired Student t test) and nonparametric (Mann–Whitney test) methods when amplitude of correlations and delay of envelope of the audio over the ECoG over the envelope of the audio trace were compared. Distribution of electrodes simultaneously active were compared by chi-square statistic. All statistical analyses were calculated by using MedCalc for Windows, version 14.8.1 (MedCalc Software, Ostend, Belgium).