Speech as a complex signal

Ramón Guevara (✉ guevara.erra@gmail.com)

University of Padua

Lukas Bogge

University of Paris

Research Article

Keywords:

Posted Date: February 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1365372/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Speech as a complex signal

Ramón Guevara^{1*} and Lukas Bögge²

¹University of Padua, Department of Physics and Astronomy, Padua, 35131, Italy

²University of Paris, Memory, Brain and Cognition Laboratory (MC2Lab), Boulogne-Billancourt, 92774, France

*guevara.erra@gmail.com

ABSTRACT

Speech is an auditory signal that needs to robustly transmit complex messages under noisy conditions in an efficient way. How can this be achieved? To answer this question we developed a method for the analysis of time series complexity and applied it to a collection of artificial and natural auditory signals, including speech in different languages. Our complexity measure is the Shannon entropy of the recurrence networks of time series. We found that speech has higher complexity than animal vocalizations, pink noise and a quasiperiodic synthetically generated signal and its complexity is of the order of environmental sounds. We have also found that the global connectivity of recurrence networks correlates with speech rhythm, the temporal modulation of the speech envelope, as well as with a spectral measure of the scaling properties of efficient speech codes. We observe that signals with too large or too small phase autocorrelation have relatively small complexity values, and can thus transmit little information. Speech is situated at intermediate values of phase autocorrelation, in the region of large complexity, in agreement with the idea that complex systems lie at the frontier between order and disorder and that speech has evolved to match the human mind in complexity.

Introduction

Humans are highly social primates and use different tools for communication, among which speech is by far the most powerful one. How does the speech signal reflect the infinite expressive power of language itself? Speech is the result of the co-evolution of brain, language, the auditory system and the articulators, by means of natural selection, to fulfill the necessity of externalizing the growing inner richness of the human mind. Of course, it is possible to communicate complex thought using simple codes, but efficient communication must be fast. For example, a conversation in Morse code would take much longer than its equivalent in a natural language.

What are the general constraints on signals used for biological communication in natural conditions, and do such constraints affect speech complexity? Such signals are subject to a trade-off between redundancy in order to be robust against degradation and efficiency in order to operate fast and effortlessly¹.

A redundant message cannot convey much information^{1,2}. The least redundant message possible, a random signal, contains maximal information. In neuroscience, the efficient neural coding principle argues that sensory systems encode environmental signals in an optimal way, representing the greatest amount of information at the lowest possible cost^{2,3}, which is achieved by minimizing redundancy in the neural encoding. The efficient coding principle has been successfully applied to the visual system,⁴⁻⁶ and more recently to the auditory system⁷⁻¹⁰. Yet, signals often show some level of redundancy. A redundant signal is easy to interpret (decode) and redundancy confers robustness against noise in communication: if part of the message is not correctly encoded or received, e.g. due to environmental noise, it can be retrieved by another intact part of the message.

If a signal is to be used as a vehicle for communication in a noisy environment and under limited encoding and decoding capabilities, but has to convey as much information as possible, it should not be either too redundant, or too random. Signals that achieve this balance are often called "complex". In statistical physics, complexity is defined as a property of systems composed of interactive parts, poised between order and disorder, in the sense that the system contains a mixture of ordered states and disordered states, none of which dominates over the other¹¹. For instance, when ice melts under a phase transition, we can observe a mix of ice and liquid water. In other words, a system is complex if it exhibits a mixture of order and disorder, or regularity and randomness¹².

Consider the two extremes of this trade-off. The first extreme is a signal that consists of a strict alternation of two symbols, 0 and 1, in a periodic fashion: 01010101.... Even if part of the message is contaminated by noise (e. g. 0101XX01) the message can easily be recovered from the rest of the signal. In other words, this is a robust message. In this case, however, there is essentially no information content. The same is true for a purely sinusoidal signal, which is fully redundant. One way to convey information using this type of signals is to add some kind of modulation to the bare signal. For instance, in radio-communication information is encoded by the frequency or amplitude modulation of a sinusoidal carrier signal, and not by the carrier itself. In the process of conveying information by adding modulation, the signal becomes more complex.

The opposite extreme is pure noise. White noise could in principle be ideal for communication because a signal with a Gaussian probability distribution (as is the case for white noise) contains the maximum amount of information among all possible signals with a given variance. However, white noise is difficult to decode, extremely non-robust and very easily confused with environmental noise. In other words, noise is not complex enough for the purposes of communication in a noisy environment.

For organisms communicating through acoustic signals, the problems of encoding/decoding messages and distinguishing the signal from the background have led to the evolution of communication systems that rely on different strategies across species to maximize information transfer while making communication robust. The natural acoustic environment is rich in sounds having a short duration and a broad-band spectrum (breaking branches, objects hitting each other, etc⁷; we refer to these sounds here as "environmental sounds"). One strategy to communicate in such environments is illustrated by mammalian vocalizations. They are typically of long duration and have a narrow-band spectrum, which makes them easily distinguishable from environmental sounds, and therefore ideal for communication in a noisy background. But for the same reason, they are limited in their information transfer capabilities (as in the example above of a periodic signal).

What about human speech? Speech is sufficiently different from environmental sounds so as not to be confused with background noise, but being a broad-band signal, it can convey more information than animal vocalizations. Here, we seek to formally model this intuition, comparing the complexity of the speech signal to that of other acoustic signals.

In the current work, we therefore propose to study speech and other acoustic signals in terms of information theory and complexity theory. As explained above, a signal is complex if it is sufficiently redundant while maximizing information transfer. We investigate the complexity of speech signals in comparison with other acoustic signals. We propose that speech complexity could be the result of an optimization process, the interplay between two competing factors, code robustness and maximization of information transfer. To this end we develop a new measure of complexity of time series, based on information theory.

The paper is structured as follows. First, we describe the complexity measure we have developed, and use it to study acoustic time series, analyzing speech in several natural languages, animal vocalizations, environmental sounds of non-animal origin and synthetic signals having different levels of autocorrelation (Study 1). The goal is to compare speech with other types of acoustic signals in terms of complexity, in order to test the hypothesis that speech has evolved to solve the trade-off between robustness and information transfer in an optimal way.

We then test (Study 2) whether artificially reducing the linguistic contents of the speech signal leads to a corresponding decrease in its complexity. This allows us to directly test the casual relationship between linguistic information content and complexity. We test speech signals that were transformed using a resynthesis technique that selectively degrades certain linguistic features (such as phoneme inventory), while preserving others¹³. Two types of transformed speech signals were used, known in the literature as "saltanaj" and "sasasa" transformations^{9,13}. These transformations reduce the consonant and vowel inventory while preserving all other properties (such as sentence structure, prosody, etc). Our prediction is that the complexity of re-synthesized speech is lower than that of natural speech and the decrease is proportional to the amount of linguistic information suppressed.

Finally, we compare the complexity measure derived here with existing metrics of speech rhythmicity (Study 3). Our prediction is that if complexity captures a linguistically relevant aspect of the speech signal, it should correlate with existing phonological metrics of the acoustic properties of speech.

Results

Study 1

In Study 1 we compare the complexity of speech with that of animal vocalizations, environmental sounds and synthetic signals. To this effect we have designed a new measure to quantify complexity. We use this complexity metrics, and not others already existing in the literature (e.g. algorithmic complexity), because those metrics do not take into account the spectro-temporal properties of time series. Indeed, our measure is specifically designed and adapted for the study of speech and other acoustic signals used for communication, because it takes into account the auto-correlation of the signals, and auto-correlation is directly linked to spectral properties and with rhythmicity, as we discuss later.

Our complexity measure is based on information theory, but it is not simply the information content of the signal. Indeed, the information content of a signal is not a measure of complexity. Let us see why. The information content of a signal is given by its Shannon entropy¹, which for a continuous signal $x(t)$ is

$$S = - \int \ln \rho(x) \rho(x) dx, \quad (1)$$

and only depends on the probability distribution of the series ($\rho(x)$, the histogram of x), not on its temporal structure. Shuffling (randomizing) a time series does not change its entropy, but the time dimension is highly relevant for speech. Therefore we

expect the complexity measure to be sensitive to randomization^{14–16}. For an easy illustration consider the time series of strictly alternating 0's and 1's discussed in the Introduction (01010101....). Shuffling the positions of the 0's and 1's in this time series at random produces another time series with the same amount of information as the original one (so having the same entropy). This is because shuffling does not change the number of 0's and 1's (the histogram of x), and therefore does not change the information content, but does change where the symbols are located. But it is intuitively clear that the original and the shuffled signals are different. In other words, time matters. Therefore, the information contained in a signal is not sufficient to characterize the complexity of time series. Our proposed complexity measure is able to capture the difference between the original and the shuffled signal.

We derive a complexity measure that takes into account the temporal dynamics of time series. Our approach is based on the idea that the dynamics of a time series can be captured as a recurrence network. A recurrence network is described by a connectivity matrix, quantifying the autocorrelation of the time series for all possible time shifts (Fig. 1). In other words, it is a measure of how much a time series is autocorrelated with itself at increasing time shifts, therefore describing recurrences in phase-space¹⁷. In a recurrence network, strong links between nodes imply that the corresponding phase states are similar. For example, an oscillatory time series has high values of autocorrelation and a densely connected recurrence network, whereas a very noisy signal has low autocorrelation and a disconnected recurrence network. Using this method, dynamical patterns in the time series can be revealed, which would be difficult to observe with more traditional methods, but can be readily described by the formal tools derived from complex network theory.

The obtained complexity of speech, animal vocalizations, environmental sounds, and synthetic sounds is shown in Fig. 2. Data points $(C, S) = (\text{connectivity}, \text{entropy})$ for all the time series used are represented in the figure.

Speech signals and environmental sounds have large complexity and intermediate values of connectivity, so they are near the top of the inverted U curve. Synthetically generated pink noise and animal vocalizations have lower complexity values and larger connectivity. A synthetic signal composed of two harmonics has even larger connectivity and lower complexity than the rest of the signals.

Study 2

In this study, we degrade speech by reducing its linguistic information content and as a consequence its acoustic variability, and investigate if such degradation impacts the complexity of the resultant acoustic signal. The type of transformations used were the "sasasa" and "saltanaj" transformations¹³. In the "saltanaj" transformation all vowels are replaced by /a/ and all consonants within a consonant class by a representative of that class (fricatives by /s/, liquids by /l/, occlusives by /t/, nasals by /n/, and glides by /j/; thus, the word "pen" becomes "tan", for instance). For the "sasasa" transformation all vowels are substituted by /a/ and all consonants by /s/. In other words, the "saltanaj" transformation preserves consonant class identity, whereas the "sasasa" transformation does not. Both transformations thus reduce the information content of the speech signal as well as its acoustic variability, as many phoneme contrasts get neutralized. The "sasasa" transformation creates a more radical degradation, reducing phonemes to a single consonant and a single vowel, not unlike the maximally redundant time series 010101.... If our complexity measure is sensitive to information content, then we predict that a decrease in complexity should be observed for both transformations, and that this decrease should be larger for the "sasasa" than for the "saltanaj" transformation.

As predicted, we found that the "sasasa" and "saltanaj" transformations alter the complexity of speech signals (Fig. 3). Natural speech signals have the highest complexity, followed by the "saltanaj" transformed signals and ending with the "sasasa" transformed signals ($S_{\text{normal}} > S_{\text{saltanaj}} > S_{\text{sasasa}}$). Sounds belonging to the three different categories (normal, "saltanaj", "sasasa") form clearly non-overlapping classes. This reduction was observed for all the languages tested.

Study 3

The results of Study 1 and Study 2 suggest that speech is indeed a complex signal representing a trade-off between regularity and efficiency (a complex signal is a signal with high entropy, posed near the center of the inverted U curve, that is, having intermediate values of network connectivity). Different lines of research in linguistics and neuroscience have investigated one or the other aspect of the speech signal separately. In Study 3 we compare the network connectivity to these existing metrics of the regularity and efficiency of speech. As network connectivity captures both of these (a very regular signal has a highly connected network, and vice-versa), we predict that it will be correlated with the existing measures, but this correlation will not be perfect, as the existing measures cannot capture both of the aspects at the same time.

Specifically, we correlate network connectivity with (i) percent vowel (%V), the relative proportion of vowels in the full speech signal, proposed as a measure of the rhythmicity of speech¹⁸, and (ii) an information-theoretically derived measure of the non-redundancy (efficiency) of speech^{7,9}. These two measures were chosen because they represent the two opposing forces relevant for the complexity of a signal discussed in the Introduction: redundancy and maximal information coding.

The first measure, percent vowel (%V), conceives of the speech signal as essentially rhythmic and captures the regular alternation of vowels and consonants¹⁸. This operationalization thus focuses on the redundancy and regularity of the signal, related to the property of robustness discussed earlier. In languages with a rhythmic alternation of vowels and consonants

(e.g. Japanese where each consonant is accompanied by a vowel, forming a syllable), %V is large. Network connectivity is accordingly large for these languages, because a periodic string of vowels and consonants (VCVCVC...) has large autocorrelation. On the contrary, in languages like English, where more complex patterns can be formed, %V and connectivity are larger.

The other measure, by contrast, assesses to what extent the speech signal obeys the information theoretical principle of non-redundancy and optimal encoding, as suggested by the theory of efficient neural coding². The efficient encoding of acoustic signals is equivalent to a decomposition of signals into independent components at the cochlea⁷, which is hypothesized to act as a bank of filters having output signals that minimize information redundancy across neural representations. The statistical structure of the filters is encapsulated by their spectral parameters, bandwidth (Δf) and central frequency (f_c). These parameters are related by a scaling relationship, $\Delta f = A f_c^{1-k}$, with A and k constants for a given language⁹. The set of filters for a given language can be characterized by their sharpness, defined as $Q = f_c / \Delta f$. Broadband filters are less sharp than narrow-band filter for a given central frequency. In terms of sharpness, the spectral scaling relationship for a filter population reads $Q = m f_c + q$, with m and q being constants. This scaling relationship can be represented in a log-log plot of sharpness versus central frequency. The slope (m) of the linear regression of sharpness by central frequency can be used as a metric to characterize the set of filters for a given language⁹.

In Study 3, we compare network connectivity with %V as a measure of rhythmicity and with the slope m as a measure of efficiency.

Figure 4 shows the correlation between network connectivity and %V. Connectivity did not correlate with %V ($R^2 = 0.142$; $p = 0.405$). Figure 5 shows the correlation between network connectivity and m . Connectivity correlates with m ($R^2 = 0.576$; $p = 0.0479$).

Discussion

Speech is our species-specific communication signal. Throughout its evolution, it had to meet a challenge: develop robustness against noise while efficiently encoding the maximal amount of information with the least effort possible. Biological systems often meet this challenge by converging on a complex state, in between order and disorder. Here we asked whether speech could be formally characterized as being such a complex system. We have, therefore, developed a novel method for the estimation of time-series complexity and applied it, in three studies, to speech with the aim of assessing whether it represents an optimal trade-off between efficient information encoding and redundancy.

Our results suggest that speech can be found at the maximum of the complexity curve together with environmental signals (Figs.2), exceeding animal vocalizations, pink noise and simple synthetic signals in complexity. Importantly, speech in different languages has similar complexity. To ascertain that complexity is causally related to the amount of linguistic information encoded, we manipulated this latter, reducing the phoneme repertoire of speech in different languages, either preserving consonant class information ("saltanaj" transform) or suppressing it ("sasasa"). This manipulation decreased complexity as predicted and did so proportionally to the amount of linguistic information suppressed. When relating the complexity measure to existing metrics of either the rhythmicity, i.e. order, or the efficiency of speech, we have obtained moderate correlations, as expected, since complexity captures both aspects simultaneously.

Several of these results have important implications for our understanding of speech and language processing. In particular, they suggest a unifying framework for two influential, but seemingly opposing theories of speech perception, the theory of efficient neural coding^{2,3,5} and the theory of embedded neural oscillations^{19,20}.

Efficient neural coding theory claims that sensory systems capture the statistical structure of environmental signals as a means to encode them in the most efficient way possible, capturing the greatest amount of information at the lowest possible cost^{2,3,5}. This has been well established in the visual system²¹, and has more recently been extended to the auditory system⁷. This theory focuses on the optimality and non-redundancy of the speech signal. From this perspective, the 1/f, broad-band power spectrum of speech could be seen as a means to maximize information without redundancy. Such a system, obeying a power law, is scale-free, having no characteristic time scale.

Seemingly opposed to this is the theory of embedded neural oscillations in speech processing^{19,20}, which emphasizes the existence of neural oscillations at a small number of frequency bands, specifically delta (1-3 Hz), theta (4-8 Hz) and gamma (25-35 Hz), involved in the simultaneous processing of prosodic units, syllables and phonemes, respectively, building on the rhythmic organization of the speech signal itself²². Such a model thus operates over privileged time scales. It thus appears to be in contradiction with the efficient coding model, since from the point of view of information theory, a quasi-periodic signal is not ideal. Rhythmic oscillations do not minimize redundancy.

How are we to reconcile the idea that neural oscillations are crucial for auditory processing with the efficient encoding of speech? Our results suggest a possible solution. According to our approach, speech is fine-tuned to have the right amount of rhythmicity and randomness in such a way as to have maximal complexity. The neural signals encoding speech are similarly

a mixture of oscillations and noise. Complexity is thus the unifying framework that allows us to capture the order and the disorder of speech as a communication signal.

Methods

Study 1

Data

The data was composed of speech, animal vocalizations, environmental sounds and synthetic signals.

The speech samples were taken from nine typologically different languages obtained from previous studies [Basque, Spanish²³; Dutch, English, Japanese, Polish¹⁸; Marathi, Turkish^{9,24}; Tashlhiyt Berber²⁵]. Each dataset consisted of sentences recorded by four native female speakers in their respective languages, at a sampling rate of 16 kHz. They were simple, declarative news-like sentences containing between 15 to 21 syllables, roughly matched in meaning across languages. These languages were chosen, because they are geographically as well as historically unrelated and represent typologically different structures⁹. Furthermore, they were analyzed in previous studies for speech rhythm^{18,9,23–25} and we will use the metrics obtained in these previous studies to compare to our complexity measure in Study 3. The 1476 samples, which constituted the input for the first analysis, had a duration of around 2 – 5 seconds each. Since the recordings differed in size, the time length of each file was shortened to the length of the shortest sample in order to have equal-size input signals. The shortening was done in such a way as to preserve the center and discard the edges of the signals.

Environmental sounds and animal vocalizations were obtained from a variety of sources, following⁷. Environmental sounds comprise sounds which are typically broadband and non-harmonic, such as the cracking of ice and branches or the dripping of water. Animal vocalizations were selected from a set of mammalian vocalizations²⁶, following⁷. Periods of silence were removed from the data set. To reduce the amount of computation required, the environmental and animal vocalization stereo sounds were converted to mono and down-sampled from 44.1 kHz to 14.7 kHz⁷.

Synthetic signals were generated as the sum of two sinusoidal signals (at frequencies of 35 and 500 Hz). Their length was adjusted to those of the shortened speech files. Pink noise was generated as a superposition of relaxation processes, a method commonly used in the literature (see e.g.,²⁷). White and pink noise signals had the same length as the other signals.

All signals were normalized so as to have the same root mean square power.

Recurrence network

We calculated the recurrence network for each data set in the following way (see Fig. 1 for a schematic representation of the method). Each signal was divided into $N_s = 80$ equally long segments which did not overlap with each other. We denote the original signal as $x(t)$, and the segments as $y_i(t)$ with $i = 1 \dots 80$. The number of samples in each segment was 417 (of an approximate duration of 26 ms). Each of the segments constitute a node of the recurrence network. The edges of the network represent the temporal correlation between segments. The network is described by an adjacency matrix¹⁷. We quantified the edges in terms of phase-autocorrelation (phase synchronization between time-shifted versions of the time series), because phase synchronization is directly related to physical coupling²⁸ and has been extensively used to characterize nonlinear relationships between time series.

Phase synchronization is defined as²⁹,

$$R_{jk} = | \langle e^{i\Delta\theta_{jk}} \rangle |, \quad (2)$$

for any two signals $y_j(t)$ and $y_k(t)$. In this equation $\Delta\theta_{jk}$ is the instantaneous phase difference between the two given signals. The R_{jk} 's are the entries of the connectivity matrix $\mathbf{R}$, quantifying the phase autocorrelation of the signal $x(t)$. The instantaneous phases $\theta(t)$ were calculated in the standard way by means of the analytic signal approach⁽³⁰⁾. Briefly, for each time series $y(t)$, an analytic (complex) signal $z(t)$ is derived as,

$$z(t) = y(t) + iH[y(t)], \quad (3)$$

where $H[y(t)]$ is the Hilbert transform of the signal $y(t)$, calculated as the principal value of the integral,

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{y(\tau)}{t - \tau} d\tau. \quad (4)$$

The connectivity matrix has values in the range $0 < R_{jk} < 1$. From the connectivity matrix $\mathbf{R}$, the adjacency matrix (a Boolean matrix) $\mathbf{B}$ was calculated by using a threshold v , according to the rule

$$B_{jk} = \begin{cases} 1 & \text{if } R_{jk} \geq v \\ 0 & \text{if } R_{jk} < v \end{cases} \quad (5)$$

The adjacency matrix **B** gives a binary representation of the connectivity matrix **R**, with the intuitive meaning that if there is strong phase correlation between the time series x_j and x_k , then the corresponding entry in the matrix is 1, and otherwise it is 0. The threshold was fixed at the value $v = 0.16$, at which the white noise sample in our data set has zero network entropy. With this threshold choice, white noise was taken to be the baseline (zero complexity) compared to which the other times series complexities were calculated.

Complexity measure

The complexity measure developed in the current paper is inspired by a series of recent studies on the complexity of electrophysiological recordings in altered states of consciousness^{31–33}. In particular,³¹ have shown that conscious and functional states of brain activity are associated with high complexity and intermediate values of network synchronization, and the complexity measure can discriminate between conscious and unconscious states. Complexity in that work is the Shannon entropy of brain connectivity networks, therefore it is the information content of relevant functional networks. That is, complexity is understood as the entropy of a spatial network. In the current work, complexity is defined as the entropy of a temporal network, the recurrence network described above.

The temporal dynamics of the sound time series investigated here are taken into account by establishing recurrence networks for the sound samples as described above. The Shannon entropy of the recurrence networks is then calculated to quantify the number of possible configurations that can exist for a given network connectivity. By definition, this is the information content of the recurrence network, and therefore it measures the information contained in the temporal dynamics of the time series.

Entropy was obtained for each adjacency matrix by calculating the number of possible patterns of configurations of binary strings with a given number of connected pairs (that is, the number of possible strings of 1's and 0's with a fixed length and a fixed number of 1's). This is a standard combinatorial problem. For an $N_s \times N_s$ Boolean matrix **B** there are $N = N_s(N_s - 1)/2$ different entries (this is because the connectivity matrix is symmetric; entries on the diagonal are not counted as they represent the synchronization of a given signal with itself). The length of the binary strings is N . If there are C connected pairs (so the number of 1's in the binary strings is C) we have $P = N!/C!(N - C)!$ possible patterns for each value of C . Therefore C gives the connectivity of the recurrence network. In this binary representation of the connectivity matrix, entropy can be calculated as $S = \ln(P)$. However, direct calculation of S from this equation is precluded by numerical overflow. In practice, we used the Stirling approximation for the natural logarithm, $\ln(n!) = n \ln(n) - n$ (with n a large integer), from which we obtain, $S = N \ln(N/(N - C)) - C \ln(C/(N - C))$. This results in the inverted U curve depicted in Figs.2 and 3.

The X-axis of these curves is the degree of connectivity (C) of the recurrence network of the time series. It represents how much the time series is correlated with itself in phase. In other words, it is a measure of phase autocorrelation of the time series for all possible time shifts. It gives low values for completely uncorrelated noise and high values for periodic signals. The Y-axis is the complexity of the time series (S), calculated as the amount of information in its recurrence network. The inverted U curve is the functional dependence of S with C , that is, the complexity of time series as a function of phase autocorrelation. Noisy signals with low values of phase autocorrelation and periodic signals with high values of phase autocorrelation both have low complexity. Signals with intermediate values of phase autocorrelation have large complexity.

Study 2

Data

The speech samples consisted of "sasasa" and "saltanaj" versions of the speech samples used in study 1 for Dutch, English, Japanese, Polish and Spanish. The samples were taken from^{13,18,34}. The sampling rate was 16 kHz. The transformations were not available for the other languages.

Recurrence network

Same as in Study 1.

Complexity

Same as in Study 1.

Study 3

We obtained %V for English, Spanish, Polish, Japanese, Dutch, Turkish and Marathi from^{18,23,24}, and the slope m from⁹. We conducted a linear regression between the connectivity value and each of these measures. To fit the linear regression model we employed a robust regression method, iteratively reweighted least squares, that is less sensitive to outliers than standard linear regression^{35,36}.

Author contributions statement

R.G. conceived the ideas, developed the methods and wrote the algorithms, L.B. analysed the data, R.G. and L.B. analysed the result, R.G. wrote the manuscript. All authors reviewed the manuscript.

Additional information

Competing interests The authors declare no competing interests.

References

1. Shannon, C. E. A mathematical theory of communication. *Bell system technical journal* **27**, 379–423 (1948).
2. Barlow, H. B. Possible principles underlying the transformations of sensory messages. (1961).
3. Attneave, F. Some informational aspects of visual perception. *Psychol. review* **61**, 183 (1954).
4. Field, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *Josa a* **4**, 2379–2394 (1987).
5. Olshausen, B. A. & Field, D. J. Sparse coding of sensory inputs. *Curr. opinion neurobiology* **14**, 481–487 (2004).
6. Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural computation* **7**, 1129–1159 (1995).
7. Lewicki, M. S. Efficient coding of natural sounds. *Nat. neuroscience* **5**, 356–363 (2002).
8. Smith, E. C. & Lewicki, M. S. Efficient auditory coding. *Nature* **439**, 978–982 (2006).
9. Guevara Erra, R. G. & Gervain, J. The Efficient Coding of Speech: Cross-Linguistic Differences. *PloS one* **11**, e0148861 (2016).
10. Stilp, C. E. & Lewicki, M. S. Statistical structure of speech sound classes is congruent with cochlear nucleus response properties. In *Proceedings of Meetings on Acoustics 166ASA*, vol. 20, 050001 (ASA, 2013).
11. Kauffman, S. A. *The origins of order: Self-organization and selection in evolution* (OUP USA, 1993).
12. Huberman, B. A. & Hogg, T. Complexity and adaptation. *Phys. D: Nonlinear Phenom.* **22**, 376–384 (1986).
13. Ramus, F. & Mehler, J. Language identification with suprasegmental cues: A study based on speech resynthesis. *The J. Acoust. Soc. Am.* **105**, 512–521 (1999).
14. De Boer, E. & Dreschler, W. Auditory psychophysics: spectrotemporal representation of signals. *Annu. review psychology* **38**, 181–202 (1987).
15. Hirsh, I. J. & Watson, C. S. Auditory psychophysics and perception. *Annu. review psychology* **47**, 461–484 (1996).
16. Kluender, K. R., Coady, J. A. & Kiefte, M. Sensitivity to change in perception of speech. *Speech Commun.* **41**, 59–69 (2003).
17. Donner, R. V., Zou, Y., Donges, J. F., Marwan, N. & Kurths, J. Recurrence networks—a novel paradigm for nonlinear time series analysis. *New J. Phys.* **12**, 033025 (2010).
18. Ramus, F., Nespors, M. & Mehler, J. Correlates of linguistic rhythm in the speech signal. *Cognition* **73**, 265–292 (1999).
19. Giraud, A.-L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. neuroscience* **15**, 511–517 (2012).
20. Poeppel, D. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech communication* **41**, 245–255 (2003).
21. Bell, A. J. & Sejnowski, T. J. The “independent components” of natural scenes are edge filters. *Vis. research* **37**, 3327–3338 (1997).
22. Langus, A., Mehler, J. & Nespors, M. Rhythm in language acquisition. *Neurosci. & Biobehav. Rev.* **81**, 158–166 (2017).
23. Molnar, M., Carreiras, M. & Gervain, J. Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition* **152**, 150–159 (2016).
24. Mehler, J., Sebastian-Galles, N. & Nespors, M. Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. *The new cognitive neuroscience* **3**, 825–836 (2004).
25. Ridouane, R. Leading issues in tashlhiyt phonology. *Lang. Linguist. Compass* **10**, 644–660 (2016).

26. Emmons, L. H., Ross, D. L. & Whitney, B. M. *Sounds of Neotropical Rainforest Mammals: An Audio Fieldguide* (Conservation International, 1997).
27. Schroeder, M. *Fractals, chaos, power laws: Minutes from an infinite paradise* (Courier Corporation, 2009).
28. Pikovsky, A., Rosenblum, M. & Kurths, J. *Synchronization: a universal concept in nonlinear sciences*, vol. 12 (Cambridge university press, 2003).
29. Mormann, F., Lehnertz, K., David, P. & Elger, C. E. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Phys. D: Nonlinear Phenom.* **144**, 358–369 (2000).
30. Gabor, D. Theory of communication. Part 1: The analysis of information. *Electr. Eng. III: Radio Commun. Eng. J. Inst.* **93**, 429–441 (1946).
31. Guevara Erra, R., Mateos, D., Wennberg, R. & Perez Velazquez, J. Statistical mechanics of consciousness: Maximization of information content of network is associated with conscious awareness. *Phys. Rev. E* **94** (2016).
32. Mateos, D., Wennberg, R., Guevara Erra, R. & Velazquez, J. P. Consciousness as a global property of brain dynamic activity. *Phys. Rev. E* **96**, 062410 (2017).
33. Mateos, D., Guevara Erra, R., Wennberg, R. & Velazquez, J. P. Measures of entropy and complexity in altered states of consciousness. *Cogn. neurodynamics* **12**, 73–84 (2018).
34. Molnar, M., Gervain, J. & Carreiras, M. Within-rhythm class native language discrimination abilities of basque-spanish monolingual and bilingual infants at 3.5 months of age. *Infancy* **19**, 326–337 (2014).
35. Holland, P. W. & Welsch, R. E. Robust regression using iteratively reweighted least-squares. *Commun. Stat. Methods* **6**, 813–827 (1977).
36. Huber, P. J. *Robust statistics*, vol. 523 (John Wiley & Sons, 2004).

Figure 1. Recurrence network from time series: a schematic illustration of the method. Left: the signal under investigation (speech) is divided into non-overlapping blocks of equal length (for simplicity of illustration, only 4 blocks are depicted). The arrows connecting blocks represent phase synchronization between the corresponding signals within each block. Each arrow was labeled with a capital letter (A...F). Right: recurrence network. Each of the blocks constitute a node of the network (blue circles numbered according to their corresponding block). The lines connecting the nodes are the edges of the network, and correspond to the arrows depicted on the left part of the figure, both in color and label

Figure 2. Entropy of recurrence networks (S) as a function of network connectivity (C). The black solid line is the inverted U curve, the theoretically derived relationship between S and C . Individual data points are the (C, S) values for each type of signal used. In the case of natural languages, environmental sounds and animal vocalizations, they were averaged across all the samples in the set. Tas. Berber: Tashlhiyt Berber; synth. sounds: synthetic sounds; env. sounds: environmental (non animal) sounds; vocalizations: animal vocalizations. The synchronization threshold used was $\nu = 0.16$. The inset is a magnification of the tip of the curve, to show in detail the complexity of speech in different languages

Figure 3. Entropy of recurrence networks (S) as a function of network connectivity (C) for five natural languages and their corresponding "sasasa" and "saltanaj" transforms. For English and Polish, the "sasasa" transforms coincide, as indicated. As in the previous figures, the black solid line is the theoretically derived relation between S and C . The synchronization threshold used was $\nu = 0.16$.

Figure 4. Correlation between network connectivity and percent vowels ($\%V$) for speech in natural languages.

Figure 5. Correlation between network connectivity and the slope of the linear regression of sharpness by central frequency (m), for speech in natural languages.