

rsif.royalsocietypublishing.org

CrossMark

[click for updates](#)

Research

Cite this article: Luque J, Luque B, Lacasa L.

2015 Scaling and universality in the human voice.

J. R. Soc. Interface **12**: 20141344.<http://dx.doi.org/10.1098/rsif.2014.1344>

Received: 9 December 2014

Accepted: 29 January 2015

Subject Areas:

biophysics

Keywords:

criticality, speech, scaling

Author for correspondence:

Lucas Lacasa

e-mail: l.lacasa@qmul.ac.ukElectronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2014.1344> or via <http://rsif.royalsocietypublishing.org>.Jordi Luque¹, Bartolo Luque² and Lucas Lacasa³¹Telefonica Research, Edificio Telefonica-Diagonal 00, Barcelona, Spain²Departamento de Matemática Aplicada y Estadística, EIAE, Universidad Politécnica de Madrid, Madrid, Spain³School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E14NS, UK

Speech is a distinctive complex feature of human capabilities. In order to understand the physics underlying speech production, in this work, we empirically analyse the statistics of large human speech datasets ranging several languages. We first show that during speech, the energy is unevenly released and power-law distributed, reporting a universal robust Gutenberg–Richter-like law in speech. We further show that such ‘earthquakes in speech’ show temporal correlations, as the interevent statistics are again power-law distributed. As this feature takes place in the intraphoneme range, we conjecture that the process responsible for this complex phenomenon is not cognitive, but it resides in the physiological (mechanical) mechanisms of speech production. Moreover, we show that these waiting time distributions are scale invariant under a renormalization group transformation, suggesting that the process of speech generation is indeed operating close to a critical point. These results are put in contrast with current paradigms in speech processing, which point towards low dimensional deterministic chaos as the origin of nonlinear traits in speech fluctuations. As these latter fluctuations are indeed the aspects that humanize synthetic speech, these findings may have an impact in future speech synthesis technologies. Results are robust and independent of the communication language or the number of speakers, pointing towards a universal pattern and yet another hint of complexity in human speech.

1. Introduction

The description, understanding and modelling of speech is an interdisciplinary topic of current interest for physics [1], social and cognitive sciences [2,3], data mining as well as engineering [4–7]. Classical speech synthesis technologies and algorithms were firstly based on linear stochastic models and linear prediction [6–8], and their underlying theory, the so-called source-filter theory of speech production, initially relied on several key assumptions including uncoupled vocal tract and speech source, laminar airflow propagating linearly, periodic fold vibration and homogeneous tract conditions [6,8,9]. Despite the successes of this benchmark theory, the synthetic speech generated by linear models fails to be ‘natural’. As a matter of fact, current speech synthesizers usually require the incorporation of pieces of real speech, e.g. concatenation of smaller speech units (phoneme and diphone-based synthesis [6,7]) to improve their synthetic output.

With the popularization of nonlinear dynamics, fractals and chaos theory, the modelling paradigm slightly shifted and nonlinear speech processing emerged [10]. Accordingly, a number of authors pointed towards low dimensional, chaotic phenomena as the underlying mechanism governing the fluctuations in speech. This paradigm shift has fostered a number of inspiring theoretical vocal-fold and tract models [11] of increasing complexity, displaying a range of nonlinear phenomena such as bifurcations and chaos [12] or irregular oscillations [13] to cite some. Glottal airflow induced by fluid–tissue interaction has also been modelled using Navier–Stokes equations [14]. Nevertheless, the empirical justification for low dimensional chaos is, up to a certain extent, based on preliminary evidence [15,16], and this modelling approach is only theoretically justified through an analogy between turbulent states, chaos and fractals [10]. As a matter of fact, it has been recently acknowledged in the nonlinear dynamics community that caution should be applied when empirically

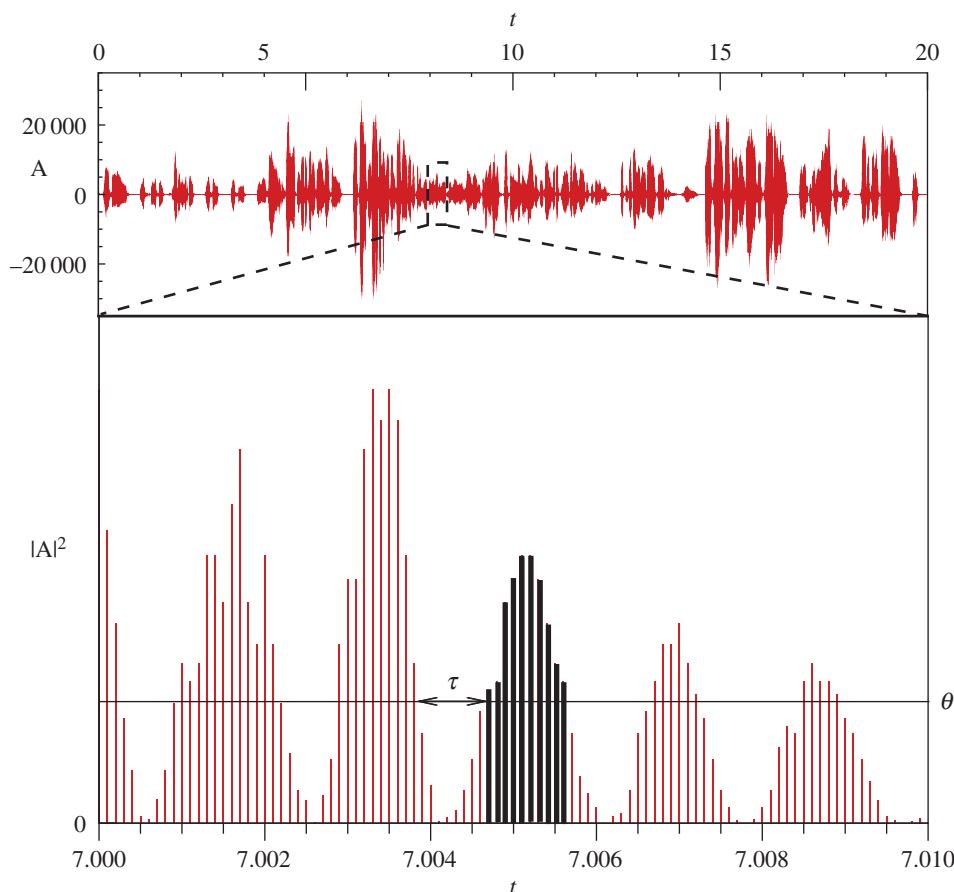


Figure 1. (Top panel) Sample speech waveform $A(t)$ of a single-speaker recording (Spanish language). (Bottom panel) Instantaneous energy per unit time $\varepsilon(t) = |A(t)|^2$ in an excerpt from the top panel. The energy threshold Θ , defined as the instantaneous energy level for which a fixed percentage of the entire data remains above that level, helps us to unambiguously define, for a given threshold, what we call ‘speech events’ (events for which $\varepsilon(t) > \Theta$) from silence or speech interevents of duration τ . The energy released E in a speech event is computed from the integration of the instantaneous energy over the duration of that event $E \sim \sum_{\text{event}} \varepsilon(t)$ (dark area in the figure denotes the energy released in a given speech event). (Online version in colour.)

analysing the low dimensional nature of short, noisy and non-stationary empirical signals, as not only the computation of dynamical invariants such as fractal dimensions or Lyapunov exponents is difficult in those cases [17], but furthermore correlated stochastic noise can be misleadingly described as having a low dimensional attractor [18]. It is therefore reasonably unclear what modelling paradigm we should follow to pinpoint the nonlinear nature of the fine grained details of speech.

In this work, we make no *a priori* assumptions about the underlying adequate dynamical model, and we follow a data-driven approach. We thoroughly analyse the statistics of speech waveforms in extensive real datasets [19] that extend all the way into the intraphoneme range ($t < 10^{-2}$ s) [20], enabling us to dissect the purely physiological (i.e. mechanical) aspects that play a role in the production of speech, from other such as cognitive aspects [3,21]. Our main results are the following: (i) energy releases in speech are power-law distributed with a language-independent universal exponent, and accordingly a Gutenberg–Richter-like law [22] is proposed within speech. (ii) In the intraphoneme range ($t < 10^{-2}$ s), the interevent times (silences of duration τ) between energy releases are also power-law distributed, suggesting long-range correlations in the time fluctuations of the amplitude signal. (iii) Furthermore, these distributions are invariant under a time renormalization group (RG) transformation [23–26]. On the basis of these results, we should

conclude that the physiological mechanism of speech production self-organizes close to a critical point.

2. Results

A TV broadcast speech database named KALAKA-2 [19] is utilized to analyse language-dependent speech (see the online supplementary material for empirical analysis on a different database gathering a total of 12 languages and additional explanations of the experimental methodology.) It was originally designed for language recognition evaluation purposes and consists of wide-band TV broadcast speech recordings (roughly 4 h per language) featuring six different languages: Basque, Catalan, Galician, Spanish, Portuguese and English. TV broadcast shows were recorded and sampled using 2 bytes at a rate of $16\,000 \text{ samples s}^{-1}$, taking care of including as much diversity as possible regarding speakers and speech modalities. It includes both planned and spontaneous speech throughout diverse environmental conditions, such as studio or outside journalist reports but excluding telephonic channels. Therefore, audio excerpts may contain voices from several speakers but only a single language. For illustrative purposes in figure 1, we depict a sample speech waveform amplitude $A(t)$ and its squared, semi-definite positive instantaneous energy $\varepsilon(t) = |A(t)|^2$, respectively. Without loss of generality, dropping the irrelevant constants, $\varepsilon(t)$ has units

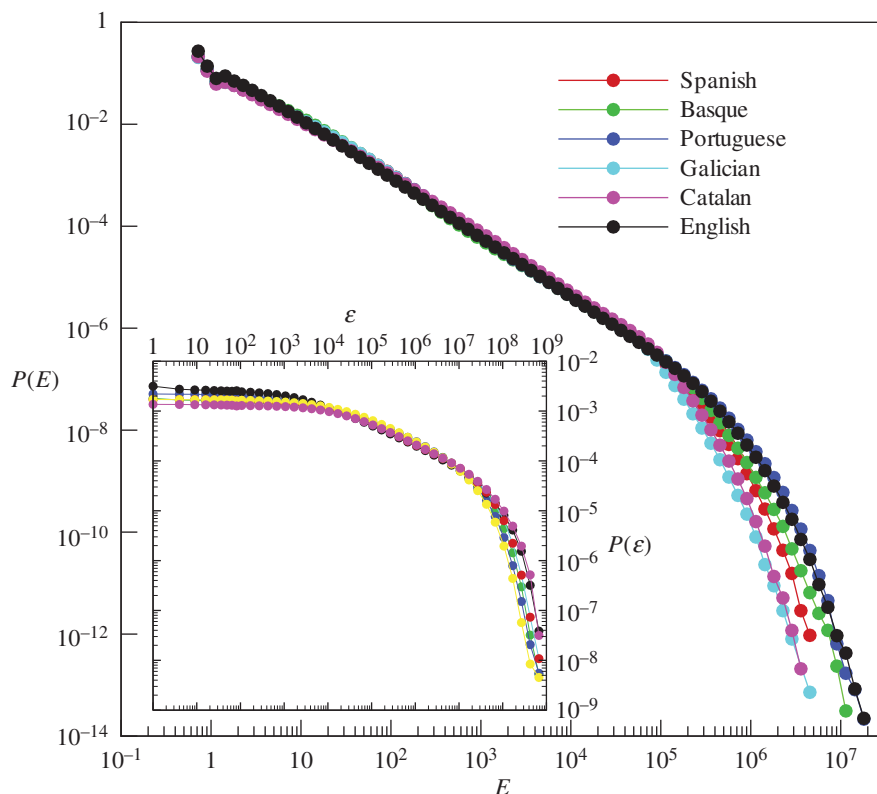


Figure 2. Log–log plot of the integrated energy distribution $P(E)$ released in the statistics of all languages considered in this study, after a logarithmic binning. The figure shows a power law $P(E) \sim E^{-\alpha}$, that holds for five decades, truncated for large (rare) events by a finite-size cutoff. (Inset panel) Log–log plot of the associated instantaneous energy histogram $P(\varepsilon)$ of each language, which we include for completeness. We note that this non-scaling shape of instantaneous energy is similar to the one found for precipitation rates [28]. (Online version in colour.)

of energy per time. Then, a threshold Θ is defined as the instantaneous energy level for which a fixed percentage of data is larger than the threshold. For instance, $\Theta = 70\%$ is the threshold for which 30% of the data falls under this energy level (this allows data to be compared across different empirical signals [27]). Θ not only works as a threshold of ‘zero energy’ that filters out background (environmental) noise, but help us to unambiguously define what we might call a speech event, a sequence of *consecutive* measurements $\varepsilon(t) > \Theta$, from a silence event, whose duration is τ (see figure 1). Note at this point that, as Θ can take arbitrary values, speech events are not necessarily the events that gather true speech (also, according to psychoacoustic theory of perception, speech might not be an adequate word for very short timescales). From now on, we adopt the term speech event as a working definition. Accordingly, speech can now be seen as a dynamical process of energy releases or ‘speech earthquakes’ separated by silence events, with in principle different statistical properties given different thresholds Θ . In what follows, we address all these properties.

2.1. Energy release: a Gutenberg–Richter-like scaling law in speech

The energy of a speech event is computed from the integration of the instantaneous energy over the duration of that event $E = \int_{\text{event}} \varepsilon(t) dt \approx \sum_{\text{event}} \varepsilon(t) \Delta t$, where $\Delta t = 1/16\,000$ s is the inverse of the sampling frequency (and therefore E has arbitrary units of energy). In order to get rid of environmental noise, we set a fixed threshold $\Theta = 80\%$ and for each language, we compute its histogram $P(E)$. In figure 2,

we draw, in log–log scales, this histogram for all languages considered (note that a logarithmic binning was used to smooth out the data). We find out a robust power-law scaling $P(E) \sim E^{-\alpha}$ over five decades saturated by the standard finite-size cutoff, where the fitted exponents are all consistent with a language-independent universal behaviour: $\alpha = 1.13 \pm 0.06$ for Spanish, $\alpha = 1.16 \pm 0.05$ for Basque, $\alpha = 1.16 \pm 0.07$ for Portuguese, $\alpha = 1.13 \pm 0.05$ for Galician, $\alpha = 1.10 \pm 0.05$ for Catalan and $\alpha = 1.15 \pm 0.02$ for English, all having a correlation coefficient $r^2 > 0.999$ (for completeness, in the inset of figure 2, we also depict the binned histogram of instantaneous energy $P(\varepsilon)$ for all languages).

As long as the magnitude in seismicity [22] is related to the logarithm of the energy release, under this definition $P(E)$ can be identified as a new ‘Gutenberg–Richter-like law’ in speech. This may be seen as related to other scaling laws in cognitive sciences [21], although at this stage it is still unclear what particular contributions come from both the mechanical (vocal folds and resonating cavity) and the cognitive systems.

2.2. Scaling and universality of waiting time distributions

In a second part, we study the temporal orchestration of fluctuations, that is, the arrangement of silences or speech interevents of duration τ . We will pay special attention to the intraphoneme range (timescales $t < 3 \times 10^{-2}$ s [6,20]), where we assume no cognitive effects are present, in order to focus on the physiological aspects of speech. At this point, we introduce a RG transformation to explore the

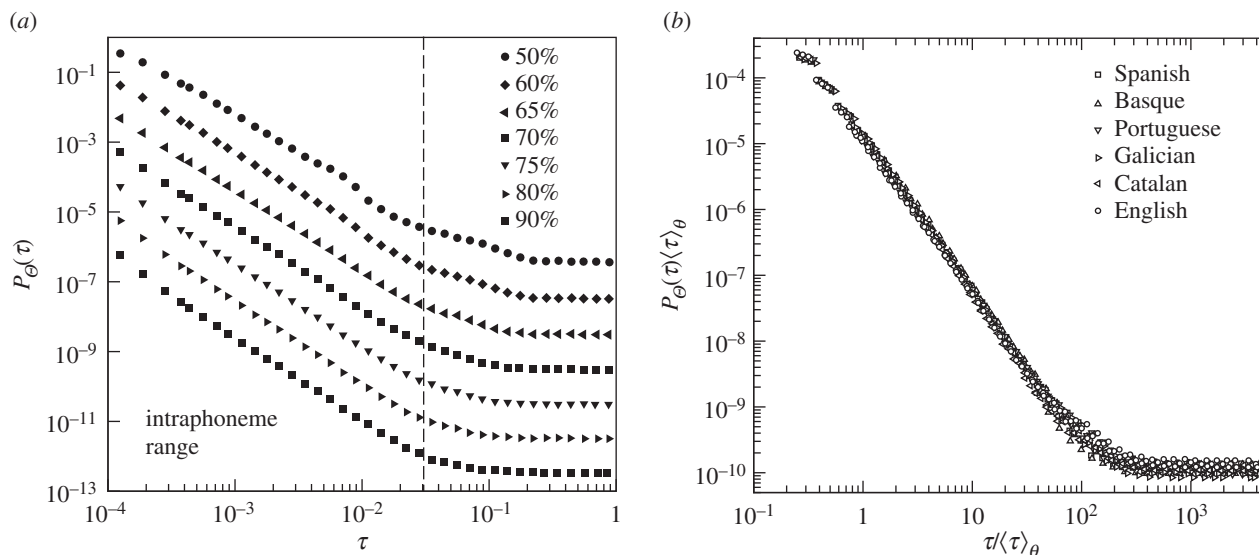


Figure 3. (a) Log–log plot of the speech interevent distribution $P_\Theta(\tau)$ computed for the English database, for different energy thresholds Θ (there is a vertical translation performed for illustrative purposes, and data have been logarithmically binned). In the intraphoneme range (timescales $t < 3 \times 10^{-2}$ s, described by all times smaller than the vertical dotted line), we find power-law distributed interevent times fairly independently from the threshold, which suggests the onset of long-range correlations in this regime. (b) Rescaled interevent distributions $P(\tau)$ according to the decimation method (see the text), for several languages and different thresholds (additional languages can be found in the electronic supplementary material). In the intraphoneme range, all curves collapse into a universal function and are invariant under the RG transformation, suggesting that the underlying physiological mechanism of speech synthesis is critical.

origin of temporal correlations. This particular technique originates from the statistical physics community and has been previously used in the context of earthquakes [23,24,27] and tropical-cyclone statistics [25]. The first part of the transformation consists of a decimation: we raise the threshold Θ . This in general leads to different interevent distributions $P_\Theta(\tau)$. The second part of the transformation consists of a scale transformation in time, such that renormalized systems become comparable: $\tau \rightarrow \tau/\langle \tau \rangle_\Theta$, $P(\tau) \rightarrow P_\Theta(\tau)\langle \tau \rangle_\Theta$, where $\langle \tau \rangle_\Theta$ is the mean interevent time of the system for a particular Θ . Invariant distributions under this RG transformation collapse into a threshold-independent universal curve: an adimensional waiting time distribution. While the complete fixed point structure of this RG is not well understood yet, recent advances [24,26] rigorously found that stable (attractive) fixed points include the exponential distribution and a somewhat exotic double power-law distribution, which are attractors for both memoryless and short-range correlated stochastic point processes under the RG flow. Invariant distributions other than the previous fixed points are likely to be unstable solutions of the RG flow, therefore encompassing criticality in the underlying dynamics [26].

In figure 3a, we plot in log–log, for different thresholds, the interevent histogram $P_\Theta(\tau)$ associated with the English language. In the intraphoneme range, interevents are power-law distributed in every case $P_\Theta(\tau) \sim \tau^{-\beta}$. In figure 3b, we plot the rescaled histograms $P(\tau)$ ($\Theta = 60\%$, 65% , 70% for every language, yielding a total of 18 curves), collapsing under a single curve (see the online supplementary material for empirical analysis on a different database gathering a total of 12 languages and additional explanations on the experimental methodology). Note that the collapse is quite good for those timescales that belong to the intraphoneme range, where only physiological mechanisms are in place, and such collapse is lost for larger timescales. This suggests that for every language, the statistics are invariant under this RG transformation, and the pattern is robust across languages.

A more careful statistical analysis is required here as the range of the power law, restricted to the intraphoneme regime, is smaller. Accordingly, exponent β is estimated now following Clauset *et al.*'s method [29] which employs maximum-likelihood estimation (MLE) of a power-law model, where goodness-of-fit test and confidence interval are based on Kolmogorov–Smirnov (KS) tests (comparing the actual distribution with 100 synthetic power-law models whose exponent is the one found in the MLE to obtain p -values). This method yields a fairly universal exponent $\beta = 2.06$ within the 95% confidence interval [1.84, 2.20], KS p -value of 0.99 and the statistical support for the power-law hypothesis given by a p -value of 0.95 (see the online supplementary material for empirical analysis on a different database gathering a total of 12 languages and additional explanations on the experimental methodology).

3. Discussion

It is well known that the dynamics of speech generation are complex, nonlinear and certainly poorly explained by the benchmark source-filter theory. The fact that a Gutenberg–Richter law for the energy release probability distribution during speech emerges opens the possibility of understanding speech production in terms of crackling noise, a highly nonlinear behaviour that was first described in condensed matter physics [30] and later found in a variety of natural hazards systems including earthquakes [27], rain [28] or solar flares [31] to cite some [26]. The underlying theory might then describe energy releases as the resonating response function of the system under airflow perturbation (the so-called susceptibility in the statistical physics jargon), and the fact that it is a power-law distributed quantity is the first evidence of criticality in these systems.

The fact that the waiting time distributions are invariant under decimation (and different from the well-known stable

exponential laws or random-walk statistics) suggests that the system might be operating close to a critical point. Note that the currently established low dimensional chaotic hypothesis suffers from being a process with short-range correlations and therefore chaotic speech interevents should typically renormalize into the exponential law. The most plausible conclusion of this work is that the physiological process of speech production evidences long-range correlations and criticality. This argument can be put in the same grounds as equilibrium critical phenomena, where only long-range correlations allow the system to escape from the basin of attraction of the trivial RG (high or low temperature) fixed points. As long as the critical solution is an *unstable* fixed point of the RG flow, the fact that the underlying dynamics seems to be poised near this point is somewhat remarkable. If we assume that the dynamics of speech generation at the level of the vocal folds are influenced by the properties of the glottal airflow, then our results would suggest that a similar critical behaviour might take place in other physiological processes involving such airflow. Interestingly, this is something that has been found empirically in the mechanism of lung inflation [32], what suggests a possible generative model based on lung structure.

Although there is no general relation between interevent statistics and power spectra, note that whereas a simple stochastic processes such as a Brownian motion cannot explain these results, under the more general paradigm of fractional Brownian motion with first return distribution $P(T) \sim T^{H-2}$ [33], through a hand-waving analogy, our current findings would be consistent with a fBm model with $H \approx 0$, which leads to the so-called $1/f$ noise. This resulting phenomenological description is in agreement with previous evidence [3,34], this latter being yet another trait of long-range correlations in human communication [34,35]. This gives further credit to our results since, whereas the mechanism of criticality does not necessarily imply the presence of $1/f$ noise or vice versa, they are usually found together [22].

Although we acknowledge that weakly chaotic systems such as intermittent ones can operate close to criticality and that a rather convoluted combination of basic stochastic processes might also reproduce some of the features, our results suggest that the paradigm of self-organized criticality (SOC) [22]—out of equilibrium dissipative systems that self-organize towards a critical state—might be a more straightforward and adequate modelling scenario than standard low dimensional chaos or phenomenological combinations of stochastic processes. If this was to be the case, threshold dynamics [22] would appear as the essential ingredient that encompasses the nonlinear properties in speech at the

physiological level. These new approaches could further contribute to the development of both (i) microscopic (generative) models of speech production and (ii) complementary methods of speech synthesis, that would profit from the self-similar properties of speech at the intraphoneme range to make refined speech interpolation without the need to incorporate pieces of real speech in the synthesis model.

From a speech analysis perspective, it is not clear at this stage what is the particular contribution of sounds articulated as voiced fricatives, e.g. /z/ in zip or /v/ in vine, which are produced by continuous friction of turbulent and noisy airflow at the place of articulation, generating small irregularities in jitter and shimmer measurements, which capture deviation from periodicity of the glottal source (in both time and amplitude, respectively). Although this may contribute in some extent to the observed phenomenology (and further analysis is required), we should highlight that interevents of short-range correlated dynamics tend to renormalize towards the exponentially distributed trivial fixed point, at odds with the phenomenology that is presented here. Also it is worth mentioning the wide variability within the technical set-up and hence the possibility of the presence of sub-harmonics, which are related to the fundamental frequency by ratios and usually generated by signal amplification.

Finally, the combined fact that (i) the results are robust for different human languages and that (ii) the timescales involved in this analysis require that the process is purely physiological, lead us to conclude that this mechanism is a universal trait of human beings. It is another open question whether the onset of SOC in this system is the result of an evolutionary process [22], where human speech waveforms would have evolved to be independent of speaker and receiver distance and perception thresholds. In such context, further work should be done to investigate whether similar patterns originate in the communication of other species, and up to which extent other variables, such as ageing, may play a role.

Data accessibility. Datasets used in this work can be downloaded from <https://147.83.50.223/RSIF20141344>.

Acknowledgements. The authors would like to thank Luis Javier Rodríguez-Fuentes and Mikel Peñagarikano for recording and hand-labelling the speech corpus, Alvaro Corral for helpful discussions and anonymous referees for interesting comments. After submission of this manuscript we learned of a recent publication [36] where the authors explore scaling and complexity matching in conversational speech, finding similar power-law distribution for interevent statistic, although considering different timescales where cognitive and behavioural factors are present.

Funding statement. B.L. acknowledges funding from Spanish Education Council under project FIS2013–41057-P.

References

- Hegger R, Kantz H, Matassini L. 2000 Denoising human speech signals using chaoslike features. *Phys. Rev. Lett.* **84**, 3197–3200. (doi:10.1103/PhysRevLett.84.3197)
- Hickok G, Poepper D. 2007 The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402. (doi:10.1038/nrn2113)
- Kello CT, Anderson GG, Holden JG, Van Orden GC. 2008 The pervasiveness of $1/f$ scaling in speech reflects the metastable basis of cognition. *Cogn. Sci.* **32**, 1217–1231. (doi:10.1080/03640210801944898)
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. 1991 Better speech recognition with cochlear implants. *Nature* **352**, 236–238. (doi:10.1038/352236a0)
- Zeng F-G, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargava A, Wei C, Cao K. 2005 Speech recognition with amplitude and frequency modulations. *Proc. Natl Acad. Sci. USA* **102**, 2293–2298. (doi:10.1073/pnas.0406460102)
- Flanagan JL, Allen JB, Mark A. 2008 *Hasegawa-Johnson, speech analysis synthesis and perception*, 3rd edn. New York, NY: Springer.
- Taylor P. 2009 *Text-to-speech synthesis*. Cambridge, UK: Cambridge University Press.
- Rabiner LR, Schafer RW. 1978 *Digital processing of speech signals*. Englewood Cliffs, NJ: Prentice-Hall.

9. Kent RD, Read C. 1992 *The acoustic analysis of speech*. London, UK: Whurr Publishers, Singular Publishing Group.
10. Marshall S, Sicuranza G (eds). 2006 *Advances in nonlinear signal and image processing*. New York, NY: Hindawi Publishing Corporation.
11. Steinecke I, Herzel H. 1995 Bifurcations in an asymmetric vocal-fold model. *J. Acoust. Soc. Am.* **97**, 1874–1884. (doi:10.1121/1.412061)
12. Jiang JJ, Zhang Y, Stern J. 2001 Modeling of chaotic vibrations in symmetric vocal folds. *J. Acoust. Soc. Am.* **110**, 2120–2128. (doi:10.1121/1.1395596)
13. Neubauer J, Mergell P, Eysholdt U, Herzel H. 2001 Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes. *J. Acoust. Soc. Am.* **110**, 3179–3192. (doi:10.1121/1.1406498)
14. Tao C, Jiang JJ. 2008 Chaotic component obscured by strong periodicity in voice production system. *Phys. Rev. E* **77**, 061922. (doi:10.1103/PhysRevE.77.061922)
15. Kumar A, Mullick SK. 1996 Nonlinear dynamical analysis of speech. *J. Acoust. Soc. Am.* **100**, 615. (doi:10.1121/1.415886)
16. Maragos P, Potamianos A. 1999 Fractal dimensions of speech sounds: computation and application to automatic speech recognition. *J. Acoust. Soc. Am.* **105**, 1925. (doi:10.1121/1.426738)
17. Kantz H, Schreiber T. 2005 *Nonlinear time series analysis*, 2nd edn. New York, NY: Cambridge University Press.
18. Osborne AR, Provenzale A. 1989 Finite correlation dimension for stochastic systems with power-law spectra. *Phys. D* **35**, 357–381. (doi:10.1016/0167-2789(89)90075-4)
19. Rodríguez-Fuentes LJ, Peñagarikano M, Varona A, Dez M, Bordel G. 2012 KALAKA-2: a TV broadcast speech database for the recognition of Iberian languages in clean and noisy environments. In *Proc. of the 8th Int. Conf. on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey, 21–27 May*. See http://www.lrec-conf.org/proceedings/lrec2012/pdf/486_Paper.pdf.
20. Crystal TH, House AS. 1988 Segmental durations in connected-speech signals: syllabic stress. *J. Acoust. Soc. Am.* **83**, 1574–1585. (doi:10.1121/1.395912)
21. Kello CT, Brown GDA, Ferrer-i-Cancho R, Holden J, Linkenkaer-Hansen K, Rhoded T, Van Orden GC. 2010 Scaling laws in cognitive sciences. *Trends Cogn. Sci.* **14**, 223–232. (doi:10.1016/j.tics.2010.02.005)
22. Christensen K, Moloney NR. 2005 *Complexity and criticality*. London, UK: Imperial College Press.
23. Corral A. 2004 Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Phys. Rev. Lett.* **92**, 108501. (doi:10.1103/PhysRevLett.92.108501)
24. Corral A. 2005 Renormalization-group transformations and correlations of seismicity. *Phys. Rev. Lett.* **95**, 028501. (doi:10.1103/PhysRevLett.95.028501)
25. Corral A, Osso A, Llebot JE. 2010 Scaling of tropical-cyclone dissipation. *Nat. Phys.* **6**, 693–696. (doi:10.1038/nphys1725)
26. Corral A. 2009 Point-occurrence self-similarity in crackling-noise systems and in other complex systems. *J. Stat. Mech.* **2009**, P01022. (doi:10.1088/1742-5468/2009/01/P01022)
27. Bak P, Danon L, Scanlon T. 2002 Unified law for earthquakes. *Phys. Rev. Lett.* **88**, 178501. (doi:10.1103/PhysRevLett.88.178501)
28. Peters O, Deluca A, Corral A, Neelin JD, Holloway CE. 2010 Universality of rain event size distributions. *J. Stat. Mech.* **2010**, P11030. (doi:10.1088/1742-5468/2010/11/P11030)
29. Clauset A, Shalizi CR, Newman MEJ. 2009 Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703. (doi:10.1137/070710111)
30. Sethna JP, Dahmen KA, Myers CR. 2001 Cracking noise. *Nature* **418**, 242–250. (doi:10.1038/35065675)
31. de Arcangelis L, Godano C, Lippiello E, Nicodemi M. 2006 Universality in solar flare and earthquake occurrence. *Phys. Rev. Lett.* **96**, 051102. (doi:10.1103/PhysRevLett.96.051102)
32. Suki B, Barabasi A-L, Hantos Z, Petak F, Stanley HE. 1994 Avalanches and power-law behaviour in lung inflation. *Nature* **368**, 615–618. (doi:10.1038/368615a0)
33. Ding M, Yang W. 1995 Distribution of the first return time in fractional Brownian motion and its application to the study of on–off intermittency. *Phys. Rev. E* **52**, 207–213. (doi:10.1103/PhysRevE.52.207)
34. Voss RF, Clarke J. 1975 1/f noise in music and speech. *Nature* **258**, 317–318. (doi:10.1038/258317a0)
35. Altmann EG, Pierrehumbert JB, Motter AE. 2009 Beyond word frequency: bursts, lulls, and scaling in the temporal distributions of words. *PLoS ONE* **4**, e7678. (doi:10.1371/journal.pone.0007678)
36. Abney DH, Paxton A, Dale R, Kello CT. 2014 Complexity matching in dyadic conversations. *J. Exp. Psychol. Gen.* **143**, 2304–2315. (doi:10.1037/xge0000021)