Power Laws and Scaling in the Waiting Time Distribution of Speech

J. Luque[1,2*], L. Lacasa[2], B. Luque[2]
[1] *Dept. de Teoria del Senyal i Comunicacions, TALP Research Center, Universitat Politècnica de Catalunya (UPC), Barcelona, (Spain)*
[2] *Dept. de Matemática Aplicada y Estadística, ETSI Aeronáuticos, Universidad Politécnica de Madrid (UPM), Madrid, (Spain)*

Speech is a distinctive feature of human capabilities. Knowledge of the speech amplitude-squared density, which constitutes an estimation of the released energy, provides basic data concerning statistics of the continuous speech waveform and understanding of the physics of speech production. In this work we present and investigate the statistics of intra-phoneme waiting time distributions in several human speech databases with sufficient size to describe the longtime speech energy density. We find that the waiting time statistics evidence a robust power-law shape, which suggests long-range temporal correlations in the energy of speech. Whereas long-range correlations have been previously reported in several human activities[1], we emphasize that statistics addressed in this work restrict to the intra-phoneme range (characteristic time $t < 10$ ms). This suggests that the responsible of this complex phenomenon is not cognitive[2], but resides on the physiological speech production mechanism by itself.
Furthermore, we show that these waiting time distributions follow a scaling law in the sense that they are invariant under a renormalization group transformation[3,4]. Such a result suggests that the speech production process is indeed critical, at odds with standard approaches in speech processing which rely on linear stochastic approaches[5,6].
Results are robust and independent of the communication language, the number of speakers or the speech style, pointing towards an universal pattern and a hint of the complexity of human speech that evokes an explanation in terms of a critical phenomenon.

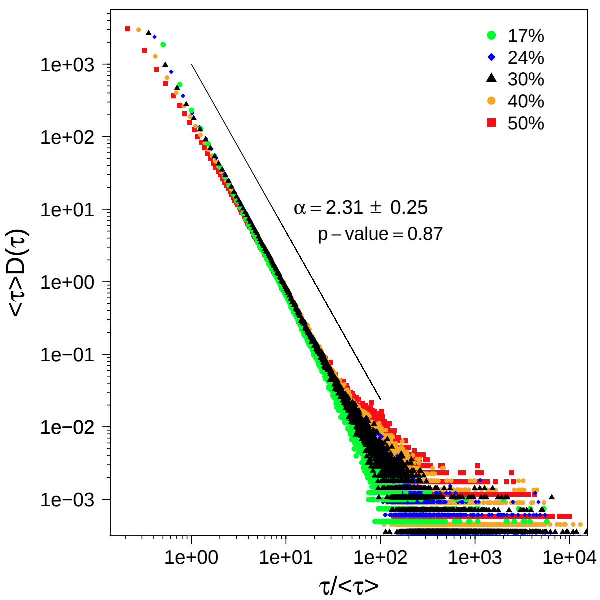

FIG. 1. **Rescaled power-law distributions of speech waiting times for different energy thresholds.** The speech waveform corresponds to an excerpt, from Catalonian broadcast news television, lasting four hours and sampled at 16 KHz. Waiting times distributions – $D(\tau)$ – are calculated for different energy thresholds, ranging in $17\% - 50\%$ of the maximum of the speech amplitude-squared. It is worth to mention that energy speech values below that threshold are discarded from the waiting time computation. In all cases the data can be fitted by a power law distribution with plausible evidence. The maximum likelihood power-law fit obtained setting the energy threshold to 30% is also depicted. It yields $\alpha = 2.31 \pm 0.25$ with a p-value around 87%.

---

* jorge.luque@upc.edu

[1] Barabasi, A. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207 (2005)
[2] Kello, C.T., Brown, G.D.A., Ferrer-i-Cancho, R., Holden, J.G., Linkenkaer-Hansen, K., Rhodes, T. and Van Orden, G.C., Scaling laws in cognitive sciences, *Trends in Cognitive Sciences* **14**, 5 (2010).
[3] Corral, Álvaro. Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes. *Phys. Rev. Lett.* **92**, 108501 (2004).
[4] Corral, A., Ossó, A. and Llebot, J. E. Scaling of Tropical Cyclone Dissipation. *Nature Phys.* **6**, 693-696 (2010).
[5] Taylor, P., Text-to-Speech Synthesis, (Cambridge University Press, 2009).
[6] Flanagan, J. L., Allen, J. B. and Hasegawa-Johnson, M. A., Speech analysis synthesis and perception, (Springer-Verlag, 3rd edition, 2008).