THE IMPORTANCE OF THE FUNDAMENTAL FREQUENCY IN CHIMPANZEES VOCAL COMMUNICATION

Marion Laporte[*1,2], Louis-Jean Boë [2,3] Nicolas Audebert [4]

[*]Corresponding Author: marion.laporte@mnhn.com

[1] UMR 7194, PaleoFED, Muséum National d'Histoire Naturelle, Paris, France

[2] Institut des Sciences du Calcul et des Données, Sorbonne Université, Paris, France

[3] GIPSA Lab, CNRS, Grenoble, France

[4] CEDRIC (EA4629), Conservatoire National des Arts et Métiers, HESAM Université, Paris, France

In the quest for understanding the evolution of human language, primates have often been taken as an approximation of the human ancestral condition despite their equally long evolutionary path. Recent studies have shown that several primates, including chimpanzees, have the capacity to produce vowel-like vocalizations, despite a high larynx (Boë et al., 2017; Fitch et al., 2016; Grawunder et al., 2022). However, many of these primates live in complex social systems in environments where the visibility is low and where they need to communicate over long distances. In these cases, the vocal signals produced involve high intensity signals associated to high fundamental frequencies (*fo*) that are better suited for long-distance propagation as they maximize the sound power radiation (Titze & Palaparthi, 2018). This is especially the case for chimpanzees, who live in complex fission-fusion societies in forests or savannah habitats (Goodall, 1986).

Here, we argue that these long-distance vocalizations which constitute two-thirds of all of male chimpanzees' vocal communication (Arcadi, 2000) cannot always provide complete formantic information or contrast, as is the case for human vowels which are produced with low fundamental frequencies. One way around this problem is to look at the fundamental frequency modulation, which is equally well conserved over long distances and generally robust against surrounding noise. While the chimpanzee vocal repertoire has been extensively studied over

the years, no study has yet looked at the *fo* modulation per se and its importance in the vocal repertoire of chimpanzees. We aim to evaluate this by determining a vocal space of the entire vocal repertoire through different yet complementary machine learning approaches using the *fo* modulation and Mel-frequency cepstral coefficients (MFCCs) that are commonly used in human speech recognition and take into account all the spectral information. We manually extracted the temporal evolution of the *fo* for N=+6000 vocal units recorded from wild Ugandan chimpanzees (11 adult males and 12 adult females). This allowed us to treat a vocalization as a univariate time series, making them suitable for further analysis. We tested both *supervised* classification methods (Support Vector Machines (SVM) (Schölkopf & Smola, 2001) using the GAK kernel (Cuturi, 2011)) and trained an auto-encoder neural network on the frequential time series to perform a non-linear dimension reduction for *unsupervised* classification. The interesting aspect of working without supervision is that it does not rely on an a priori manual classification of vocalizations into different groups and helps to identify patterns in the data without human bias (t-SNE). We compared the results obtained on the *fo* profiles with equivalent classification and dimension reduction techniques applied to the spectral information *(MFCCs)*, similarly to how previous works performed data analysis on primate vocalisations (Wadewitz et al., 2015) and CNN 2D typically used for image classification but with good performances on spectrogram classification (Hershey et al., 2017).
Our results show that the modulation of the fundamental frequency is an important factor in differentiating the vocalisations of the chimpanzee repertoire, despite an important gradation between calls. An analysis of the *fo* modulation alone obtains good results. The classification does not improve with the MFCC approach, but gains a finer-grained categorization with the CNN 2D approach. We present clustering provided through t-SNE visualization which permits to better understand the links between the different vocalisations and the importance of their gradation.

We examine these results from an evolutionary perspective and discuss the importance of the *fo* modulation in chimpanzee vocal communication compared to human speech production.

Acknowledgements

This work is part of the Apesvoice project, funded by Sorbonne University (Emergence 2021-2022) and the ISCD as part of the *Origins of Speech* team. We would like to thank Amélie Vialet and Pascal Perrier for their comments.

References

Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., Becker, Y., Rey, A., & Fagot, J. (2017). Evidence of a Vocalic Proto-System in the Baboon (Papio papio) Suggests Pre-Hominin Speech Precursors. PLOS ONE, 12(1), e0169321. https://doi.org/10.1371/journal.pone.0169321

Cuturi, M. (2011). Fast Global Alignment Kernels. In Proceedings of the 28th international conference on machine learning (ICML-11), pp. 929–936.

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. Science Advances, 2(12), e1600723. https://doi.org/10.1126/sciadv.1600723

Goodall, J. (1986). The chimpanzees of Gombe: Patterns of behavior. Cambridge: Harvard University Press.

Grawunder, S., Uomini, N., Samuni, L., Bortolato, T., Girard-Buttoz, C., Wittig, R. M., & Crockford, C. (2022). Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1841), 20200455. https://doi.org/10.1098/rstb.2020.0455

Hershey, S. et al. (2017) « CNN architectures for large-scale audio classification ». In : 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). p. 131-135.

Schölkopf, B., & Smola, A. J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

Titze, I., & Palaparthi, A. (2018). Radiation efficiency for long-range vocal communication in mammals and birds. The Journal of the Acoustical Society of America, 143(5), 2813–2824. https://doi.org/10.1121/1.5034768

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing Vocal Repertoires—Hard vs. Soft Classification Approaches. PLOS ONE, 10(4), e0125785. https://doi.org/10.1371/journal.pone.0125785