The Journal of Open Source Software

# Phonemizer: text to phones conversion for multiple languages in Python

**Mathieu Bernard**[1]

**1** LSCP/ENS/CNRS/EHESS/Inria/PSL Research University, Paris, France

**DOI:** DOIunavailable

**Software**

- Review
- Repository
- Archive

**Editor:** Pending Editor

**Reviewers:**

- @Pending Reviewers

**Submitted:** N/A
**Published:** N/A



## Summary

The `phonemizer` software is used to turn an input text into phonetic alphabet. A wrapper on four different backends:

- espeak
- espeak-mbrola
- Festival (Black et al., 2014)
- Segments (Forkel et al., 2019)

## Statement of Need

Text phonemization is a preprocessing step required in different fields of natural language processing and speech processing. The `phonemizer` is used for word segmentation in the wordseg toolbox (Bernard et al., 2020). It is also in use in the preprocessing pipeline of deep learning text-to-speech systems (Ideas Engineering, 2021; Mozilla, 2021; Zhang et al., 2020).

## Acknowledgements

We are thankful to Alex Cristia who initiated this project and to Emmanuel Dupoux for his support and advices. This work is funded by the European Research Council (ERC-2011-AdG-295810 BOOTPHON), the Agence Nationale pour la Recherche (ANR-17-EURE-0017 Frontcog, ANR-10-IDEX-0001-02 PSL, ANR-19-P3IA-0001 PRAIRIE 3IA Institute) and grants from CIFAR (Learning in Machines and Brains), Facebook AI Research (Research Grant), Google (Faculty Research Award), Microsoft Research (Azure Credits and Grant), and Amazon Web Service (AWS Research Credits).

## References

Bernard, M., Thiolliere, R., Saksida, A., Loukatou, G. R., Larsen, E., Johnson, M., Fibla, L., Dupoux, E., Daland, R., Cao, X. N., & others. (2020). WordSeg: Standardizing unsupervised word form segmentation from text. *Behavior Research Methods*, *52*(1), 264–278.

Black, A. W., Clark, R., Richmond, K., Yamagishi, J., Oura, K., & King, S. (2014). *The festival speech synthesis system* (Version 2.4) [Computer software]. CSTR, University of Edinburgh. https://www.cstr.ed.ac.uk/projects/festival

Forkel, R., Moran, S., List, J.-M., Greenhill, S. J., Ashby, L., Gorman, K., & Kaiping, G. (2019). *Cldf/segments: Unicode standard tokenization* (Version v2.1.3). Zenodo. https://doi.org/10.5281/zenodo.3549784

Ideas Engineering. (2021). Non-autoregressive transformer based neural network for text-to-speech. In *GitHub repository*. GitHub. https://github.com/as-ideas/TransformerTTS

Mozilla. (2021). Deep learning for text to speech. In *GitHub repository*. GitHub. https://github.com/mozilla/TTS

Zhang, J.-X., Ling, Z.-H., & Dai, L.-R. (2020). Non-parallel sequence-to-sequence voice conversion with disentangled linguistic and speaker representations. In *GitHub repository*. GitHub. https://github.com/jxzhanggg/nonparaSeq2seqVC_code