:

PHONOLOGICAL FEATURES IN THE ERA OF DEEP LEARNING: A MULTI-DIMENSIONAL INVESTIGATION INTO OPTIMAL REPRESENTATIONAL UNITS FOR LANGUAGE MODELING

DOCTORAL THESIS WRITING QUALIFICATION REVIEW

Sora Nagano
Graduate School of Arts and Sciences
The University of Tokyo

s-oswld-n@g.ecc.u-tokyo.ac.jp

August 30, 2025

1

Chomsky & Halle (1968)
Chomsky, N. & Halle, M. (1968)

20SPEAB / [_]

Prince & Smolensky (2004)
: Prince, A. & Smolensky, P. (2004)

GenConEval

Goldsmith (1976)
Goldsmith, J. (1976)

Clements (1985)
Clements, G. N. (1985)

2

Baevski et al. (2020)
wav2vec 2.0: Baevski, A., Zhou, Y., Mohamed, A. & Auli, M. (2020)

LibriSpeech11004,500

Hsu et al. (2021)
HuBERT: Hsu, W.-N., Bolte, B., Tsai, Y.-H. H., Lakhotia, K., Salakhutdinov, R. & Mohamed, A. (2021)

SSLHidden-Unit BERTLibriSpeechwav2vec 2.01B19%13%WERSSL

S. Chen et al. (2022)
WavLM: Chen, S. et al. (2022)

94,000SUPERBMicrosoft

Mohamed et al. (2022)
: Mohamed, A. et al. (2022)

3

van den Oord et al. (2017)
van den Oord, A., Vinyals, O. & Kavukcuoglu, K. (2017)

Vector Quantized-VAE6449.3%7.2%

Zhang et al. (2024)
SpeechTokenizer: Zhang, Z. et al. (2024)

-HuBERTRVQEnCodecWER 5.04 vs 5.11MUSHRA 90.55 vs 79.86TTSVALL-E

Chang et al. (2024)
Interspeech 2024 Chang, H.-J. et al. (2024)

40ASRTTSSVSSSL

Higy et al. (2021)
Higy, B., Millet, J. & Dunbar, E. (2021)

4

Hayes & Wilson (2008)
Hayes, B. & Wilson, C. (2008)

r=.946-

B. Tesar & Smolensky (1998)
Tesar, B. & Smolensky, P. (1998)

Mayer (2021)
Mayer, C. (2021)

pTSL

Jarosz (2019)
Jarosz, G. (2019)

-

5

Begu (2020)
: Begu, G. (2020)

GAN--

d'Avila Garcez et al. (2009)
d'Avila Garcez, A. S., Lamb, L. C. & Gabbay, D. M. (2009)

Panchendrarajan & Zubiaga (2024)
: Panchendrarajan, R. & Zubiaga, A. (2024)

200NLP

Hamilton et al. (2024)
AI Hamilton, K. et al. (2024)

-200NLP

6

Mikolov et al. (2013)
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013)

Skip-gramvec(“”) + vec(“”)vec(“”)NLP

Silfverberg et al. (2018)
Silfverberg, M., Mao, L. & Hulden, M. (2018)

Kolachina & Magyar (2019)
Kolachina, S. & Magyar, B. (2019)

Venkateswaran et al. (2025)
: Venkateswaran, A. et al. (2025)

Astrach & Pinter (2025)
Astrach, N. & Pinter, Y. (2025)

Maaten & Hinton (2008)
t-SNE van der Maaten, L. & Hinton, G. (2008)

t-SNESNE

7

Ardila et al. (2020)
Common Voice: Ardila, R. et al. (2020)

50,0002,50029Mozilla DeepSpeech125.99%

Panayotov et al. (2015)
LibriSpeech: ASR Panayotov, V., Chen, G., Povey, D. & Khudanpur, S. (2015)

4,770ASRtrain/dev/test1000cleanother

McAuliffe et al. (2017)
Montreal Forced Aligner: Kaldi- McAuliffe, M. et al. (2017)

Kaldi-5020ms-

Belinkov & Glass (2019)
: Belinkov, Y. & Glass, J. (2019)

Conneau et al. (2020)
Conneau, A. et al. (2020)

53wav2vec 2.0XLSR

8

Yang et al. (2025)
k2SSL: Yang, S. et al. (2025)

ZipformerU-Net3.534.8% WERSSLScaledAdamprogressive--

Cho et al. (2025)
Sylber: Cho, H. et al. (2025)

6-74.27/

9

Moran & McCloy (2019)
PHOIBLE: Moran, S. & McCloy, D. (2019)

2,1863,020IPA

Mortensen et al. (2016)
PanPhon: IPA Mortensen, D. R. et al. (2016)

5,000IPA21

Dunbar et al. (2019)
Zero Resource2019: T Dunbar, E. et al. (2019)

TTSZero Resource

Parcollet et al. (2024)
LeBenchmark 2.0: Parcollet, T. et al. (2024)

10

(**schatz2021?**)
: Schatz, T., Feldman, N. H., Goldwater, S., Cao, X.-N. & Dupoux, E. (2021)

Macwhinney (2000)
CHILDES: MacWhinney, B. (2000)

305000CLANCHAT

Dupoux (2018)
: Dupoux, E. (2018)

AI

T. A. Nguyen et al. (2020)
Zero Resource2021: Nguyen, T. A. et al. (2020)

ABX

Cruz Blandón et al. (2023)
Cruz-Blandón, M. A. et al. (2023)

Benders & Blom (2023)
: Benders, T. et al. (2023)

McMurray (2023)
: McMurray, B. (2023)

11

B. B. Tesar (1995)
Tesar, B. (1995)

Robust Interpretive ParsingError-Driven Constraint DemotionOT

Silverman (2012)
Silverman, D. (2012)

Staples & Graves (2020)
Staples, R. et al. (2020)

D. Nguyen et al. (2016)
: Nguyen, J. et al. (2016)

Reubold et al. (2010)
F0: Reubold, U. & Harrington, J. (2010)

ASR

Kazanina et al. (2018)
: Kazanina, N. (2018)

Tsvilodub et al. (2025)
Tsvilodub, P. et al. (2025)

-

Medin et al. (2024)
: Medin, R. et al. (2024)

SSL

Pouw et al. (2024)
Pouw, W. et al. (2024)

Gosztolya et al. (2024)
Wav2vec 2.0— Gosztolya, G. et al. (2024)

SSLSSLSSL

12 PDF

J. Chen & Elsner (2023)
Chen, J. & Elsner, M. (2023)

ciwGANGANGAN

Guriel et al. (2023)
Guriel, D., Goldman, O. & Tsarfaty, R. (2023)

LSTMtransducer

Pasad et al. (2024)
Pasad, A., Chien, C.-M., Settle, S. & Livescu, K. (2024)

Pandian (2025)
- Pandian, S. M. (2025)

-

Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2020). *Common voice: A massively-multilingual speech corpus* (arXiv:1912.06670). arXiv. https://doi.org/10.48550/arXiv.1912.06670

Astrach, G., & Pinter, Y. (2025). *Probing subphonemes in morphology models* (arXiv:2505.11297). arXiv. https://doi.org/10.48550/arXiv.2505.11297

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). Wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, *33*, 12449–12460.

Begu, G. (2020). Generative adversarial phonology: Modeling unsupervised phonetic and phonological learning with neural networks. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.00044

Belinkov, Y., & Glass, J. (2019). Analysis methods in neural language processing: A survey. *Transactions of the Association for Computational Linguistics*, *7*, 49–72. https://doi.org/10.1162/tacl_a_00254

Benders, T., & Blom, E. (2023). Computational modelling of language acquisition: An introduction. *Journal of Child Language*, *50*(6), 1287–1293. https://doi.org/10.1017/S0305000923000429

Chang, X., Shi, J., Tian, J., Wu, Y., Tang, Y., Wu, Y., Watanabe, S., Adi, Y., Chen, X., & Jin, Q. (2024). *The interspeech 2024 challenge on speech processing using discrete units* (arXiv:2406.07725). arXiv. https://doi.org/10.48550/arXiv.2406.07725

Chen, J., & Elsner, M. (2023). *Exploring how generative adversarial networks learn phonological representations* (arXiv:2305.12501). arXiv. https://doi.org/10.48550/arXiv.2305.12501

Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., Li, J., Kanda, N., Yoshioka, T., Xiao, X., Wu, J., Zhou, L., Ren, S., Qian, Y., Qian, Y., Wu, J., Zeng, M., Yu, X., & Wei, F. (2022). WavLM: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*, *16*(6), 1505–1518. https://doi.org/10.1109/JSTSP.2022.3188113

Cho, C. J., Lee, N., Gupta, A., Agarwal, D., Chen, E., Black, A. W., & Anumanchipalli, G. K. (2025). *Sylber: Syllabic embedding representation of speech from raw audio* (arXiv:2410.07168). arXiv. https://doi.org/10.48550/arXiv.2410.07168

Chomsky, N., & Halle, M. (1968). *The sound pattern of english.*

Clements, G. N. (1985). The geometry of phonological features. *Phonology*, *2*, 225–252.

Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2020). *Unsupervised cross-lingual representation learning for speech recognition* (arXiv:2006.13979). arXiv. https://doi.org/10.48550/arXiv.2006.13979

Cruz Blandón, M. A., Cristia, A., & Räsänen, O. (2023). Introducing meta-analysis in the evaluation of computational models of infant language development. *Cognitive Science*, *47*(7), e13307. https://doi.org/10.1111/cogs.13307

d'Avila Garcez, A. S., Lamb, L. C., & Gabbay, D. M. (2009). *Neural-symbolic cognitive reasoning*. Springer.

Dunbar, E., Algayres, R., Karadayi, J., Bernard, M., Benjumea, J., Cao, X.-N., Miskic, L., Dugrain, C., Ondel, L., Black, A. W., Besacier, L., Sakti, S., & Dupoux, E. (2019). *The zero resource speech challenge 2019: TTS without t* (arXiv:1904.11469). arXiv. https://doi.org/10.48550/arXiv.1904.11469

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, *173*, 43–59. https://doi.org/10.1016/j.cognition.2017.11.008

Goldsmith, J. A. (1976). *Autosegmental phonology* [PhD thesis]. Massachusetts Institute of Technology.

Gosztolya, G., Kiss-Vetráb, M., Svindt, V., Bóna, J., & Hoffmann, I. (2024). *Wav2vec 2.0 embeddings are no swiss army knife-a case study for multiple sclerosis*.

Guriel, D., Goldman, O., & Tsarfaty, R. (2023). *Morphological inflection with phonological features* (arXiv:2306.12581). arXiv. https://doi.org/10.48550/arXiv.2306.12581

Hamilton, K., Nayak, A., Boi, B., & Longo, L. (2024). Is neuro-symbolic AI meeting its promise in natural language processing? A structured review. *Semantic Web*, *15*(4), 1265–1306. https://doi.org/10.3233/SW-223228

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379–440. https://doi.org/10.1162/ling.2008.39.3.379

Higy, B., Gelderloos, L., Alishahi, A., & Chrupaa, G. (2021). Discrete representations in neural models of spoken language. In J. Bastings, Y. Belinkov, E. Dupoux, M. Giulianelli, D. Hupkes, Y. Pinter, & H. Sajjad (Eds.), *Proceedings of the fourth BlackboxNLP workshop on analyzing and interpreting neural networks for NLP* (pp. 163–176). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.blackboxnlp-1.11

Hsu, W.-N., Bolte, B., Tsai, Y.-H. H., Lakhotia, K., Salakhutdinov, R., & Mohamed, A. (2021). *HuBERT: Self-supervised speech representation learning by masked prediction of hidden units* (arXiv:2106.07447). arXiv. https://doi.org/10.48550/arXiv.2106.07447

Jarosz, G. (2019). Computational modeling of phonological learning. *Annual Review of Linguistics*, *5*(1), 67–90. https://doi.org/10.1146/annurev-linguistics-011718-011832

Kazanina, N., Bowers, J. S., & Idsardi, W. (2018). Phonemes: Lexical access and beyond. *Psychonomic Bulletin & Review*, *25*(2), 560–585. https://doi.org/10.3758/s13423-017-1362-0

Kolachina, S., & Magyar, L. (2019). What do phone embeddings learn about phonology? In G. Nicolai & R. Cotterell (Eds.), *Proceedings of the 16th workshop on computational research in phonetics, phonology, and morphology* (pp. 160–169). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-4219

Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605.

Macwhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Transcription format and programs*. Lawrence Erlbaum Associates Publishers.

Mayer, C. (2021). Capturing gradience in long-distance phonology using probabilistic tier-based strictly local grammars. *Proceedings of the Society for Computation in Linguistics 2021*, 39–50.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. *Interspeech*, *2017*, 498–502.

McMurray, B. (2023). The acquisition of speech categories: Beyond perceptual narrowing, beyond unsupervised learning and beyond infancy. *Language, Cognition and Neuroscience*, *38*(4), 419–445. https://doi.org/10.1080/23273798.2022.2105367

Medin, L. B., Pellegrini, T., & Gelin, L. (2024). Self-supervised models for phoneme recognition: Applications in children's speech for reading learning. *Interspeech 2024*, 5168–5172. https://doi.org/10.21437/Interspeech.2024-1095

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*.

Mohamed, A., Lee, H., Borgholt, L., Havtorn, J. D., Edin, J., Igel, C., Kirchhoff, K., Li, S.-W., Livescu, K., Maaløe, L., Sainath, T. N., & Watanabe, S. (2022). Self-supervised speech representation learning: A review. *IEEE Journal of Selected Topics in Signal Processing*, *16*(6), 1179–1210. https://doi.org/10.1109/JSTSP.2022.3207050

Moran, S., & McCloy, D. (Eds.). (2019). *Phoible 2.0*. Max Planck Institute for the Science of Human History.

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L. (2016). PanPhon: A resource for mapping IPA segments to articulatory feature vectors. In Y. Matsumoto & R. Prasad (Eds.), *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 3475–3484). The COLING 2016 Organizing Committee.

Nguyen, D., Doruöz, A. S., Rosé, C. P., & de Jong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*, *42*(3), 537–593. https://doi.org/10.1162/COLI_a_00258

Nguyen, T. A., Seyssel, M. de, Rozé, P., Rivière, M., Kharitonov, E., Baevski, A., Dunbar, E., & Dupoux, E. (2020). *The zero resource speech benchmark 2021: Metrics and baselines for unsupervised spoken language modeling* (arXiv:2011.11588). arXiv. https://doi.org/10.48550/arXiv.2011.11588

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). Librispeech: An ASR corpus based on public domain audio books. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5206–5210. https://doi.org/10.1109/ICASSP.2015.7178964

Panchendrarajan, R., & Zubiaga, A. (2024). *Synergizing machine learning & symbolic methods: A survey on hybrid approaches to natural language processing* (arXiv:2401.11972). arXiv. https://doi.org/10.48550/arXiv.2401.11972

Pandian, S. M. (2025). *Hybrid symbolic-neural architectures for explainable artificial intelligence in decision-critical domains*.

Parcollet, T., Nguyen, H., Evain, S., Boito, M. Z., Pupier, A., Mdhaffar, S., Le, H., Alisamir, S., Tomashenko, N., Dinarelli, M., Zhang, S., Allauzen, A., Coavoux, M., Esteve, Y., Rouvier, M., Goulian, J., Lecouteux, B., Portet, F., Rossato, S., … Besacier, L. (2024). *LeBenchmark 2.0: A standardized, replicable and enhanced framework for self-supervised representations of french speech* (arXiv:2309.05472). arXiv. https://doi.org/10.48550/arXiv.2309.05472

Pasad, A., Chien, C.-M., Settle, S., & Livescu, K. (2024). What do self-supervised speech models know about words? *Transactions of the Association for Computational Linguistics*, *12*, 372–391. https://doi.org/10.1162/tacl_a_00656

Pouw, C., Kloots, M. de H., Alishahi, A., & Zuidema, W. (2024). Perception of phonological assimilation by neural speech recognition models. *Computational Linguistics*, *50*(3), 1557–1585. https://doi.org/10.1162/coli_a_00526

Prince, A., & Smolensky, P. (2004). Optimality theory: Constraint interaction in generative grammar. In *Optimality theory in phonology* (pp. 1–71). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756171.ch1

Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on *F*0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication*, *52*(7), 638–651. https://doi.org/10.1016/j.specom.2010.02.012

Silfverberg, M. P., Mao, L., & Hulden, M. (2018). Sound Analogies with Phoneme Embeddings. *Society for Computation in Linguistics*, *1*(1). https://doi.org/10.7275/R5NZ85VD

Silverman, D. (2012). *Neutralization*. Cambridge University Press.

Staples, R., & Graves, W. W. (2020). Neural components of reading revealed by distributed and symbolic computational models. *Neurobiology of Language (Cambridge, Mass.)*, *1*(4), 381–401. https://doi.org/10.1162/nol_a_00018

Tesar, B. B. (1995). *Computational optimality theory* [PhD thesis]. University of Colorado at Boulder.

Tesar, B., & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry*, *29*(2), 229–268. https://doi.org/10.1162/002438998553734

Tsvilodub, P., Hawkins, R. D., & Franke, M. (2025). *Integrating neural and symbolic components in a model of pragmatic question-answering* (arXiv:2506.01474). arXiv. https://doi.org/10.48550/arXiv.2506.01474

van den Oord, A., Vinyals, O., & kavukcuoglu, koray. (2017). Neural discrete representation learning. *Advances in Neural Information Processing Systems*, *30*.

Venkateswaran, N., Tang, K., & Wayland, R. (2025). *Probing for phonology in self-supervised speech representations: A case study on accent perception* (arXiv:2506.17542). arXiv. https://doi.org/10.48550/arXiv.2506.17542

Yang, Y., Zhuo, J., Jin, Z., Ma, Z., Yang, X., Yao, Z., Guo, L., Kang, W., Kuang, F., Lin, L., Povey, D., & Chen, X. (2025). *k2SSL: A faster and better framework for self-supervised speech representation learning* (arXiv:2411.17100). arXiv. https://doi.org/10.48550/arXiv.2411.17100

Zhang, X., Zhang, D., Li, S., Zhou, Y., & Qiu, X. (2024). *SpeechTokenizer: Unified speech tokenizer for speech large language models* (arXiv:2308.16692). arXiv. https://doi.org/10.48550/arXiv.2308.16692