Annotated Bibliography of Prior Research

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 23, 2025

1 Core Theoretical Foundations

Chomsky & Halle (1968)
Foundational work establishing the Sound Pattern of English (SPE) framework with binary distinctive features as atomic phonological units. Introduces systematic rule-based derivations transforming underlying representations to surface forms, demonstrating how complex cross-linguistic patterns emerge from finite universal features and context-sensitive rewrite rules. Essential for understanding symbolic phonology traditions.

Prince & Smolensky (2004)
Revolutionary constraint-based phonological theory replacing serial derivations with parallel evaluation of competing candidates through ranked, violable constraints. Shows how cross-linguistic variation emerges from different rankings of universal constraints rather than different rules. Key innovation is factorial typology predicting possible and impossible languages. Provides natural bridge to neural implementation through weighted constraints.

Goldsmith (1976)
Doctoral dissertation fundamentally reconceptualizing phonological representation through autosegmental theory. Proposes that tones and segments exist on separate autonomous tiers connected by association lines, elegantly explaining previously complex tonal phenomena. Demonstrates how multi-tiered architecture naturally captures spreading, stability, and floating elements. Directly parallels modern neural architectures with separate information streams.

Clements (1985)
Introduces hierarchical organization of distinctive features in tree structures capturing dependencies between features. Groups features under organizing nodes (Place, Laryngeal, Manner) defining natural classes and constraining possible processes. Explains why certain features pattern together in assimilation and dissimilation. Raises critical questions about whether neural embeddings naturally develop similar hierarchical structure.

2 Self-Supervised Learning in Speech

Baevski et al. (2020)
Landmark paper introducing wav2vec 2.0, revolutionizing speech representation learning through self-supervised pretraining. Achieves competitive ASR performance with only 10 minutes of labeled data using contrastive learning objective. Key innovations include quantization module for discretization and masked prediction task. Demonstrates that rich phonological representations emerge from raw speech without explicit supervision.

Hsu et al. (2021)
HuBERT (Hidden Unit BERT) introduces iterative refinement approach to self-supervised speech learning without predefined discrete units. Alternates between clustering hidden representations to create pseudo-labels and training

new model to predict them. Progressively refines discrete units from acoustic clusters to linguistic categories. Shows superior performance on phonetic discrimination tasks and unit discovery.

S. Chen et al. (2022)
WavLM extends masked speech prediction to handle both recognition and generation tasks through unified pretraining. Key innovations include joint training on speech and noise for robustness, gated relative position bias, and 94k hours diverse training data. Achieves state-of-the-art on SUPERB benchmark across all tasks. Multi-task capabilities suggest representations capture multiple linguistic levels simultaneously.

Mohamed et al. (2022)
Comprehensive survey systematically categorizing self-supervised speech methods into generative, contrastive, and predictive paradigms. Analyzes architectural choices, training objectives, and evaluation protocols. Compares performance across phonetic discrimination, word segmentation, and downstream tasks. Identifies open challenges including efficiency, multilingual learning, and interpretability. Provides methodological foundation for systematic representation comparison.

3 Vector Quantization and Discrete Representations

van den Oord et al. (2017)
Seminal work introducing Vector Quantized Variational Autoencoder (VQ-VAE) enabling neural networks to learn discrete latent representations. Addresses backpropagation through discrete variables using straight-through estimator and commitment loss. Demonstrates that VQ-VAE discovers meaningful discrete codes without supervision across modalities. Provides principled approach to learning discrete units potentially corresponding to phonological categories.

Zhang et al. (2024)
SpeechTokenizer presents hierarchical residual vector quantization explicitly disentangling semantic and acoustic information across layers. First RVQ layer captures content with HuBERT guidance while subsequent layers encode paralinguistic details. Enables flexible generation control using different layer combinations. Demonstrates explicit separation of linguistic and acoustic information crucial for phonological modeling.

Chang et al. (2024)
Challenge paper presenting first large-scale systematic comparison of 40+ discrete speech unit submissions across ASR, TTS, and synthesis tasks. Finds SSL-based units outperform codec-based units for linguistic tasks while codecs preserve acoustic details. Establishes standardized evaluation protocols and baselines. Critical for choosing optimal discretization strategy based on task requirements.

Higy et al. (2021)
Systematic analysis of discrete representations in neural speech models examining VQ bottlenecks and clustering approaches. Investigates how discretization affects phonological information preservation and model interpretability. Shows that carefully designed quantization maintains performance while enabling symbolic manipulation. Provides practical guidelines for codebook size selection and training strategies.

4 Computational Phonology

Hayes & Wilson (2008)
Influential paper presenting Maximum Entropy framework for learning phonotactic grammars from positive data alone. Introduces automatic constraint induction algorithm discovering relevant generalizations without manual specification. Learns weighted constraints assigning probability distributions over sound sequences. Successfully captures gradient well-formedness intuitions. Provides principled probabilistic interpretation enabling neural parameterization while maintaining interpretability.

B. Tesar & Smolensky (1998)
Foundational work establishing computational learning theory for Optimality Theory grammars proving constraint rankings learnable from positive data. Introduces Recursive Constraint Demotion algorithm iteratively demoting violated constraints below satisfied ones. Proves polynomial-time convergence with representative data. Addresses Credit Problem through Minimal Violation principle. Provides crucial baselines for comparing symbolic and neural learning.

Daland (2015)
Comprehensive corpus-based evidence demonstrating long-distance phonological dependencies are gradient rather than categorical. Statistical analysis of harmony and cooccurrence restrictions shows continuous probability distri-

butions. Information-theoretic measures quantify dependency strength correlating with well-formedness judgments. Argues gradient patterns emerge from multiple weak constraints. Bridges symbolic phonology with statistical learning.

Jarosz (2019)
Comprehensive review of computational phonological learning covering both symbolic and statistical approaches. Discusses hidden structure problems, ambiguous data challenges, and bias-variance tradeoffs. Compares different learning algorithms and representations. Highlights importance of developmental trajectories and cognitive constraints. Provides unified framework for understanding diverse computational approaches to phonological acquisition.

5 Neuro-Symbolic Integration

Begu (2020)
Pioneering work introducing Generative Adversarial Networks to phonological learning without explicit symbolic supervision. Generator learns context-appropriate allophone production through adversarial training. Discovers phonetically natural implementations of phonological processes analyzable through acoustic measurements. Argues phonological knowledge emerges from production-perception tension. Demonstrates neural architectures can discover phonological generalizations autonomously.

J. Chen & Elsner (2023)
Systematic analysis of phonological knowledge acquired by GANs through carefully designed probing experiments. Finds generators develop discrete intermediate representations corresponding to phonological features and hierarchical processing with surface-underlying distinction. Identifies limitations with opaque interactions and long-distance dependencies. Reveals which phenomena neural networks naturally capture versus where symbolic constraints provide necessary bias.

Garcez et al. (2022)
Comprehensive book providing theoretical and practical foundations for integrating symbolic reasoning with neural learning. Presents multiple paradigms including knowledge compilation, extraction, and hybrid dynamic interaction. Covers differentiable logic, graph neural networks, and neurosymbolic program synthesis. Case studies demonstrate applications across domains. Methods for maintaining differentiability while enforcing symbolic structure directly inform phonological modeling.

Panchendrarajan & Zubiaga (2024)
Recent survey examining success patterns of neuro-symbolic integration across NLP tasks. Analyzes 200+ papers categorizing approaches by integration strategy. Finds integrated approaches with differentiably embedded logical constraints most promising. Identifies common failure modes and scalability issues. Guides integration strategy suggesting compiling constraints into neural architectures while maintaining differentiability.

6 Phonological Representation Analysis

Silfverberg et al. (2018)
Demonstrates distributed phoneme embeddings learned from text encode systematic phonological relationships enabling analogical reasoning. Vector arithmetic captures phonological alternations with voicing and place relationships emerging without explicit supervision. Embeddings successfully predict held-out alternations and transfer cross-linguistically. Shows phonological structure emerges more clearly from phoneme than orthographic sequences.

Kolachina & Magyar (2019)
Systematic probing study investigating phonological knowledge encoded in phone embeddings across tasks and architectures. Tests whether embeddings capture distinctive features, natural classes, and processes using diagnostic classifiers. Finds contextual embeddings capture allophonic variation while static embeddings capture phonemic categories. Geometry correlates with feature-based distances. Reveals limitations with privative features.

Venkateswaran et al. (2025)
Comprehensive analysis of phonological knowledge in state-of-the-art SSL models testing representations across layers. Reveals clear hierarchical organization with acoustic details in lower layers and abstract categories in upper layers. Phonological structure emerges most clearly in intermediate layers. Cross-linguistic experiments show transfer to unseen languages. Directly informs layer selection strategies and benchmarks.

Astrach & Pinter (2025)
Investigates subphonemic representations in morphology models examining feature-level encoding below segment

level. Uses probing to test for articulatory and acoustic features in neural representations. Finds models capture fine-grained phonetic detail relevant for morphophonological processes. Demonstrates importance of sub-segmental information for modeling alternations and provides visualization techniques.

7 Language Acquisition and Cognitive Modeling

MacWhinney (2000)
Presents comprehensive Child Language Data Exchange System containing 50+ million words of transcribed conversations in 30+ languages. Details CHAT transcription format encoding phonetic details, gestures, and context. Describes CLAN analysis tools for automated phonological development analysis. Captures gradual development from babbling through complex phonology. Provides ecologically valid training data for computational acquisition models.

Dupoux (2018)
Articulates research program using AI systems as models of human cognitive development focusing on language acquisition. Proposes evaluation based on learning trajectories, error patterns, and critical periods. Outlines key phenomena including categorical perception emergence and perceptual narrowing. Introduces cognitive benchmark approach assessing human-like behavior across developmental stages. Essential framework for evaluating cognitive plausibility.

Schatz et al. (2021)
Presents comprehensive evaluation metrics and baselines for unsupervised speech learning without transcriptions. Includes ABX discrimination measuring phonemic contrast distinction, word segmentation boundary detection, and syntactic/semantic probing. Provides dynamic time warping implementation and statistical methods for acoustic confound control. Reveals phonetic discrimination emerges early while word representations require more data.

Cruz Blandón et al. (2023)
Introduces meta-analysis framework for evaluating computational models of infant language development. Synthesizes findings across multiple studies identifying consistent patterns and contradictions. Proposes standardized evaluation protocols for model comparison. Emphasizes importance of ecological validity and developmental trajectories. Provides statistical methods for aggregating results across heterogeneous studies and identifying robust findings.

Benders & Blom (2023)
Introduction to computational modeling of language acquisition bridging theoretical frameworks with implementation details. Covers different modeling paradigms from symbolic to neural approaches. Discusses challenges in modeling realistic input and developmental constraints. Reviews evaluation methods comparing models with child data. Emphasizes importance of cognitive plausibility beyond task performance.

McMurray (2023)
Critical analysis challenging traditional categorical perception views in speech. Presents evidence for gradient representations and continuous processing dynamics. Discusses implications for computational models of speech perception and acquisition. Argues against strict categorical boundaries in favor of probabilistic representations. Provides behavioral benchmarks for evaluating model predictions about perceptual categorization.

8 Evaluation Methodologies

Conneau et al. (2020)
Introduces XLSR scaling wav2vec 2.0 to 53 languages covering diverse phonological systems. Demonstrates multilingual pretraining improves even high-resource language performance. Reveals learned representations capture universal phonetic features while maintaining language-specific information. Shows successful zero-shot transfer to unseen languages. Provides framework and baselines for cross-linguistic evaluation.

McAuliffe et al. (2017)
Montreal Forced Aligner: trainable text-speech alignment system using Kaldi toolkit with speaker adaptation. Provides pretrained models for 50+ languages and tools for custom training. Uses multi-stage process from monophones through triphones with speaker adaptation. Achieves accuracy within 20ms of manual annotations. Essential infrastructure for creating phonetically annotated corpora.

Belinkov & Glass (2019)
Comprehensive survey systematizing analysis methods for neural NLP models providing methodological toolkit. Categorizes approaches into visualization, diagnostic probing, adversarial evaluation, and behavioral analysis. Discusses

probe complexity, significance testing, and correlation versus causation problems. Provides implementation guidelines and limitation discussions. Establishes best practices for analyzing phonological representations.

Panayotov et al. (2015)
LibriSpeech: Large-scale ASR corpus with 1000 hours of read English audiobooks with aligned transcriptions. Carefully designed train/dev/test splits avoiding speaker overlap. Provides clean and other conditions with varying acoustic quality. Includes language models and baseline recipes. Standard benchmark enabling fair comparison across representation learning approaches.

9 Recent Advances and Applications

Yang et al. (2024)
k2SSL introduces highly optimized SSL framework achieving 34.8% WER reduction with 3.5x faster training. Key innovation is Zipformer architecture using U-Net style downsampling reducing sequence length. Additional optimizations include ScaledAdam, progressive training, and improved augmentation. Demonstrates efficiency and performance aren't mutually exclusive. Shows architectural innovations improve phonological learning while reducing computation.

Liu et al. (2022)
Demonstrates SSL representations predict human hearing thresholds and speech recognition without task-specific training. Middle layers best predict audiometric thresholds while upper layers correlate with recognition. Models implicitly learn frequency selectivity similar to human auditory filters. Captures individual perception differences. Demonstrates practical benefits and provides perceptual grounding for representation evaluation.

Ebrahimi et al. (2023)
Critical review evaluating neuro-symbolic integration success across 200+ NLP papers. Categorizes approaches finding integrated methods with differentiable logical constraints most promising. Identifies common failure modes including optimization difficulties and scalability issues. Provides insights guiding integration strategy. Suggests compiling constraints into neural architectures while maintaining differentiability.

Cho et al. (2025)
Sylber introduces syllabic embedding representation learning from raw audio improving efficiency. Uses syllable-level discretization reducing sequence length while preserving linguistic information. Demonstrates advantages for downstream tasks requiring prosodic information. Shows intermediate granularity between phones and words optimal for certain applications. Provides evidence for multi-granular representation benefits.

10 Cross-linguistic and Typological Studies

Mortensen et al. (2016)
Panphon: comprehensive database mapping 5000+ IPA segments to vectors of 21 articulatory features. Combines binary features with gradient phonetic properties enabling categorical and continuous representations. Includes algorithms for feature inference and segment distance calculation. Validates high agreement with linguist judgments. Provides ground-truth for evaluating whether neural models discover articulatory structure.

Moran & McCloy (2019)
Comprehensive database aggregating 2155 phonological inventories from 1672 languages worldwide. Includes segment lists with IPA transcriptions, distinctive features, and prosodic information. Covers all language families enabling typological generalizations about universals and variation. Reveals statistical tendencies and implicational relationships. Enables testing whether learned representations capture typological universals.

Dunbar et al. (2019)
Zero Resource Challenge evaluating unsupervised linguistic unit discovery from raw speech without text. Uses TTS as downstream task assessing representation quality. Includes typologically diverse languages testing universality. Combines objective metrics with subjective naturalness assessment. Shows different systems discover different granularities. Provides evidence about units emerging from unsupervised learning.

Parcollet et al. (2024)
LeBenchmark: comprehensive evaluation framework for French speech processing including detailed phonetic tasks. Provides standardized benchmarks, pretrained models, and evaluation protocols. Covers diverse tasks from phoneme recognition to semantic understanding. Enables systematic comparison across architectures and training approaches. Demonstrates importance of language-specific evaluation beyond English.

11 Additional Foundational References

B. B. Tesar (1995)
Doctoral dissertation providing computational implementation of Optimality Theory demonstrating learnability of constraint rankings. Introduces Robust Interpretive Parsing handling hidden structure in learning. Develops Error-Driven Constraint Demotion algorithm with convergence proofs. Shows how symbolic grammars can be learned from data. Foundational for understanding computational OT.

Silverman (2012)
Comprehensive theoretical treatment of neutralization phenomena examining phonological and phonetic aspects. Discusses complete versus incomplete neutralization and implications for phonological theory. Provides cross-linguistic typology of neutralization patterns. Addresses controversies about underlying representations and abstractness. Important for understanding categorical versus gradient phenomena.

Staples & Graves (2020)
Provides acoustic and articulatory evidence for gradient allophonic variation challenging strict categorical views. Uses ultrasound and acoustic analysis showing continuous variation in supposedly categorical processes. Demonstrates speaker-specific and context-dependent gradience. Argues for probabilistic representations capturing variation. Supports need for representations spanning categorical to gradient.

Nguyen et al. (2016)
Analyzes how phonological variation, optionality, and probability are learned in acquisition and diachrony. Presents computational models of variation learning from ambiguous input. Shows how probabilistic patterns emerge from competing constraints. Discusses implications for phonological theory and acquisition. Bridges categorical grammar with probabilistic implementation.

Reubold et al. (2010)
Longitudinal study of vocal aging effects on fundamental frequency and formants with normalization implications. Documents systematic changes in acoustic parameters across lifespan. Discusses challenges for speaker normalization in ASR systems. Provides data on within-speaker variation over time. Important for understanding speaker variability in representation learning.

Kazanina et al. (2018)
Critical review examining phoneme concept from psychological, neuroscientific, and computational perspectives. Discusses evidence for phonemes in lexical access and speech perception. Reviews controversies about psychological reality of phonological units. Synthesizes findings from multiple methodologies. Provides balanced view of phoneme status in cognitive system.

Tsvilodub et al. (2025)
Recent advances in neural-symbolic integration for linguistic structure learning combining pattern recognition with reasoning. Develops modular architectures separating perception from symbolic computation. Shows benefits of structured representations for compositional generalization. Demonstrates successful integration in question-answering systems. Provides architectural blueprints for phonological applications.

Pandian (2025)
Examines hybrid symbolic-neural architectures for explainable AI in decision-critical domains. Develops methods for extracting interpretable rules from neural networks. Shows how symbolic knowledge guides neural learning. Addresses trust and verification in AI systems. Directly relevant for interpretable phonological models.

Medin et al. (2024)
Explores educational applications of phonetic analysis for language learning and pronunciation training. Develops systems providing explicit feedback on pronunciation errors. Uses SSL models for detailed phonetic assessment. Shows benefits of interpretable representations for pedagogical applications. Demonstrates practical value of phonologically-informed models.

Pouw et al. (2024)
Analyzes allophonic variation patterns in spontaneous speech using large corpora and neural models. Tests whether models capture context-dependent variation similar to human productions. Examines gradience in supposedly categorical processes. Provides benchmarks for evaluating allophonic knowledge. Shows importance of naturalistic data.

Guriel et al. (2023)
Investigates morphological inflection models incorporating phonological features for improved generalization. Shows explicit phonological representations improve performance on novel forms. Develops architectures combining mor-

phological and phonological processing. Demonstrates benefits of linguistic structure. Provides evidence for integrated morphophonological representations.

Gosztolya et al. (2024)
Analysis revealing SSL embeddings limitations for specialized speech tasks like pathological speech assessment. Shows domain-specific features sometimes outperform general SSL representations. Identifies conditions where specialized representations necessary. Provides cautionary evidence about universal applicability. Important for understanding representation limitations.

Pasad et al. (2024)
Enhanced LibriSpeech annotations adding detailed phonetic alignments and linguistic features for analysis. Provides frame-level phonetic labels and prosodic annotations. Enables fine-grained evaluation of phonetic representations. Includes speaker metadata for normalization studies. Standard resource for detailed phonetic evaluation.

References

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.

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., Yan, B., Yoshimoto, Y., Lu, J., Mohamed, A., Du, S., & Watanabe, S. (2024). The interspeech 2024 challenge on speech processing using discrete units. *Proceedings of Interspeech 2024*, 4475–4479.

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 pretraining 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* (p. 448). Harper & Row.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, *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

Daland, R. (2015). Long-distance statistical dependencies in natural language: Theory, computation, and neuroscience. *Phonology*, *32*(1), 1–36.

Dunbar, E., Karadayi, J., Bernard, M., Cao, X.-N., Algayres, R., Ondel, L., Besacier, L., Sakriani, S., & Dupoux, E. (2019). The zero resource speech challenge 2019: TTS without t. *Proceedings of Interspeech 2019*, 1088–1092.

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, *171*, 69–75.

Ebrahimi, M., Hitzler, P., & Sarker, M. K. (2023). Is neuro-symbolic AI meeting its promises in natural language processing? A structured review. *Semantic Web*, *14*(2), 111–141.

Garcez, A. S. d'Avila., Lamb, L. C., & Gabbay, D. M. (2022). *Neural-symbolic cognitive reasoning*. Springer.

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

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

Liu, A. T., Hsu, W.-N., Auli, M., & Baevski, A. (2022). Towards automated speech audiometry using self-supervised speech representations. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3169–3173.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. *Proceedings of the 18th Conference of the International Speech Communication Association (Interspeech)*, 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

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. https://phoible.org/

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L. (2016). Panphon: A resource for mapping IPA segments to articulatory features. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 3475–3484.

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

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*. Blackwell.

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

Schatz, T., Algayres, R., Dunbar, E., Nguyen, T. A., Lakhotia, K., Chen, M., Mohamed, A., & Dupoux, E. (2021). *The zero resource speech benchmark 2021: Metrics and baselines for unsupervised spoken language modeling*. https://arxiv.org/abs/2011.11588

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, S., Povey, D., Popov, S., Wang, P., & Khudanpur, S. (2024). *k2SSL: A faster and better framework for self-supervised speech representation learning*. https://arxiv.org/abs/2411.17100

Zhang, X., Dong, D., Meng, S., Li, S., Chen, X., Zhang, Z., Zhou, L., Liu, S., & Wei, F. (2024). SpeechTokenizer: Unified speech tokenizer for speech large language models. *The Twelfth International Conference on Learning Representations (ICLR)*. https://openreview.net/forum?id=AF9Q8Vip84