COMPREHENSIVE ANNOTATED BIBLIOGRAPHY: DEEP LEARNING ERA PHONOLOGICAL FEATURES

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 Core Theoretical Foundations

Chomsky & Halle (1968)
The Sound Pattern of English
Foundational work establishing generative phonology theoretical basis with approximately 20 binary distinctive features for systematic world language phonological phenomena description. Introduces SPE-style rules (AB / [context_context]) enabling formal phonological derivation. Provides theoretical foundation for modern computational phonology and deep learning phonological feature representation, serving as starting point for contemporary computational phonological research.

Prince & Smolensky (2004)
Optimality Theory: Constraint Interaction in Generative Grammar
Revolutionary constraint-based phonological theory proposing phonological grammars through constraint interaction. Three-component framework: Generator (Gen), Constraints (Con), Evaluator (Eval) explaining typological variation through universal constraint language-specific ranking. Provides theoretical foundation for constraint-based deep learning models and neural architecture constraint integration guidelines through weighted constraint systems.

Goldsmith (1976)
Autosegmental Phonology
Revolutionary multi-tier phonological representation theory where different features exist on independent parallel tiers connected by association lines, explaining suprasegmental phenomena. Introduces autosegmental and association line concepts demonstrating phonological representation complexity beyond linear sequences. Provides theoretical foundation for hierarchical representation learning in multi-layer neural architectures and parallel information processing streams.

Clements (1985)
The Geometry of Phonological Features
Establishes hierarchical organization theory of distinctive features through feature geometry, organizing features under class nodes sharing dependencies. Groups features under organizing nodes (Place, Laryngeal, Manner) defining natural classes and constraining possible processes. Provides theoretical foundation for deep learning models hierarchical feature representation organization and modern neural phonological modeling fundamental principles.

2 Self-Supervised Learning in Speech

Baevski et al. (2020)
wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations
Landmark work introducing contrastive learning and masked language modeling fusion demonstrating self-supervised speech representation learning superiority with minimal labeled data. LibriSpeech experiments show 1-hour labeled data achieving equivalent performance to traditional 100-hour approaches. Revolutionizes speech recognition paradigm with 4,500+ citations driving research acceleration and practical low-resource language applications through quantization and masking innovations.

Hsu et al. (2021)
HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units
Hidden-Unit BERT addressing speech-specific SSL challenges including multiple acoustic units, vocabulary-free pretraining, variable-length units through offline clustering aligned target provision methodology. Achieves wav2vec 2.0 equivalent or superior performance across LibriSpeech benchmarks. 1B parameter model achieves maximum 19% and 13% relative WER reduction demonstrating methodological advancement in SSL.

S. Chen et al. (2022)
WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing
Framework extension through 94,000-hour training achieving universal representation for full stack speech processing. Masked speech prediction and denoising joint learning enables multi-faceted information modeling including speaker, paralinguistic, and content information. Achieves state-of-the-art performance on SUPERB benchmark serving as Microsoft speech processing systems foundation model with practical deployment.

Mohamed et al. (2022)
Self-Supervised Speech Representation Learning: A Review
Comprehensive field integration providing systematic taxonomy of generative, contrastive, predictive methods with multimodal extensions and evaluation methodologies for rapid field development integration. Covers architectural choices, training objectives, evaluation protocols across phonetic discrimination, word segmentation, downstream tasks. Identifies key challenges including efficiency, multilingual learning, interpretability establishing methodological research foundations.

3 Vector Quantization and Discrete Representations

van den Oord et al. (2017)
Neural Discrete Representation Learning
Seminal Vector Quantized-VAE introduction enabling discrete latent models achieving continuous equivalent performance. Speech experiments demonstrate 64x compression achieving 49.3% phoneme classification accuracy versus 7.2% random baseline. Establishes discrete speech representation foundation demonstrating phonological feature computational implementation feasibility through straight-through estimator and commitment loss technical innovations.

Zhang et al. (2024)
SpeechTokenizer: Unified Speech Tokenizer for Speech Language Models
Unified semantic-acoustic token framework through hierarchical information disentanglement within single architecture. First layer captures content with HuBERT guidance while subsequent layers encode paralinguistic details via RVQ implementation. Achieves superior content preservation (WER 5.04 vs EnCodec 5.11) and perceptual quality (MUSHRA 90.55 vs 79.86) with zero-shot TTS surpassing VALL-E performance.

Chang et al. (2024)
The Interspeech 2024 Challenge on Speech Processing Using Discrete Units
Comprehensive evaluation framework establishment for discrete speech units across ASR, TTS, SVS tasks with 40+ submission systematic comparison. Introduces standardized evaluation protocols including bitrate calculation and multi-task effectiveness measurement. Finds SSL-based units outperform codec-based units for linguistic tasks while codecs preserve acoustic details, providing critical guidance for discretization strategy selection.

Higy et al. (2021)
Discrete Representations in Neural Models of Spoken Language
Critical analysis of discrete speech representation evaluation through four-indicator systematic comparison emphasizing evaluation methodology selection potential bias identification. Demonstrates systematic evaluation approach importance for discrete representation quality assessment revealing indicator selection influences conclusions. Provides

methodological guidelines for evaluation precision enhancement and bias mitigation in representation comparison studies.

4 Computational Phonology

Hayes & Wilson (2008)
A Maximum Entropy Model of Phonotactics and Phonotactic Learning
Revolutionary Maximum Entropy framework for phonotactic constraint learning from positive data achieving high correlation (r=.946) with human acceptability judgments. Introduces automatic constraint induction algorithm discovering relevant generalizations without manual specification, successfully capturing gradient well-formedness intuitions. Provides principled probabilistic interpretation enabling neural parameterization while maintaining interpretability through statistical-theoretical integration.

B. Tesar & Smolensky (1998)
Learnability and the Optimality Hierarchy
Foundational computational learning theory establishing Optimality Theory grammar constraint ranking learnability from positive data with polynomial-time convergence proofs. Introduces Recursive Constraint Demotion algorithm iteratively demoting violated constraints below satisfied ones addressing Credit Problem through Minimal Violation principle. Provides crucial baselines for comparing symbolic and neural learning approaches in phonological acquisition.

Mayer (2021)
Capturing Gradience in Phonology through Corpus-based Evidence
Formal extension through probabilistic Tier-based Strictly Local (pTSL) grammars enabling stepwise phenomena partial regular language expansion. Comprehensive corpus-based evidence demonstrating long-distance phonological dependencies are gradient rather than categorical with continuous probability distributions. Bridges symbolic phonology with statistical learning approaches through information-theoretic dependency strength measurement.

Jarosz (2019)
Computational Models of Learning and Processing in Phonology
Comprehensive computational phonology learning research integration spanning decades of development forming current research directions. Reviews symbolic and statistical approaches discussing hidden structure problems, ambiguous data challenges, bias-variance tradeoffs comparing different algorithms and representations. Emphasizes developmental trajectories and cognitive constraints providing unified framework for understanding diverse computational approaches.

5 Neuro-Symbolic Integration

Begu (2020)
Generative Adversarial Phonology: Modeling Unsupervised Phonetic and Phonological Learning with Neural Networks
Revolutionary GAN application to phonological learning demonstrating unsupervised acoustic data learning without explicit symbolic supervision. Successfully learns allophone distribution patterns including pre-vocalic aspiration of voiceless stops suggesting statistical-theoretical integration possibility. Generator develops context-appropriate production through adversarial training arguing phonological knowledge emerges from production-perception tension naturally occurring in neural architectures.

d'Avila Garcez et al. (2009)
Neural-Symbolic Cognitive Reasoning
Comprehensive theoretical and practical foundations for neural-symbolic integration providing mathematical and theoretical basis for symbolic reasoning and neural computation unification. Presents multiple paradigms including knowledge compilation, extraction, hybrid dynamic interaction covering differentiable logic, graph neural networks, neurosymbolic program synthesis. Provides architectural principles directly informing phonological modeling applications.

Panchendrarajan & Zubiaga (2024)
Synergizing Machine Learning and Symbolic Methods: A Comprehensive Survey
Systematic survey examining NLP hybrid method success patterns through 200+ paper analysis providing comprehensive integration framework. Finds integrated approaches with differentiably embedded logical constraints most

promising while identifying common failure modes and scalability issues. Guides integration strategy for phonological knowledge and neural learning systematic combination through differentiation-maintaining constraint compilation.

Hamilton et al. (2024)
Is Neuro-Symbolic AI Meeting Its Promise in Natural Language Processing? A Structured Review
Critical systematic evaluation examining neuro-symbolic integration success across 200+ NLP papers categorizing approaches and analyzing promise-achievement gaps. Finds integrated methods with differentiable logical constraints most promising while identifying common failure modes including optimization difficulties and scalability issues. Provides realistic assessment guiding integration strategy with evidence-based recommendations for theoretical claims versus practical achievements.

6 Phonological Representation Analysis

Mikolov et al. (2013)
Distributed Representations of Words and Phrases and their Compositionality
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013)

Improved Skip-gram model learning high-quality distributed vector representations capturing precise syntactic and semantic relationships. Introduces frequent word subsampling and negative sampling achieving speedup and quality enhancement. Demonstrates novel phrasal representation methods and additive compositionality with examples like vec(“French”) + vec(“actress”) vec(“Juliette Binoche”). Foundation for distributed representation revolution in modern NLP and deep learning with profound impact on phonological embedding approaches.

Silfverberg et al. (2018)
Sound Analogies with Phoneme Embeddings
Demonstrates distributed phoneme embeddings learned from text encode systematic phonological relationships enabling analogical reasoning without explicit supervision. Vector arithmetic captures phonological alternations with voicing and place relationships emerging systematically. Successfully predicts held-out alternations and enables cross-linguistic transfer showing phonological structure emergence more clearly from phoneme than orthographic sequences.

Kolachina & Magyar (2019)
What Do Phone Embeddings Learn about Phonology?
Systematic analysis revealing neural phonological learning differential capabilities through vowel harmony learning success versus consonant constraint learning failure discovery. Probing study investigates phonological knowledge encoded in phone embeddings across tasks and architectures finding contextual embeddings capture allophonic variation while static embeddings capture phonemic categories with feature-based distance correlations.

Venkateswaran et al. (2025)
Probing for Phonology in Self-Supervised Speech Representations: A Case Study on Accent Perception
Sociolinguistic application demonstrating neural representation capability in capturing phonological variation and accent cognition relationships through systematic accent-conditioned process modeling. Tests social linguistic variation encoding revealing models capture systematic pronunciation differences across dialect groups. Provides benchmarks for evaluating sociolinguistic phonological knowledge in contemporary speech processing systems.

Astrach & Pinter (2025)
Probing Subphonemes in Morphology Models
Architectural insights revealing local phonological features (embedding layer) versus long-distance dependencies (encoder layer) differential representation through hierarchical phonological information encoding investigation. Probing methodology tests articulatory and acoustic features in neural representations finding models capture fine-grained phonetic detail relevant for morphophonological processes with visualization techniques for representation analysis.

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

Introduces t-SNE technique visualizing high-dimensional data by assigning each datapoint a location in two or three-dimensional maps. Improves upon Stochastic Neighbor Embedding (SNE) reducing tendency to crowd points together in center while revealing structure at multiple scales. Functions as crucial visualization tool for phonological representation analysis and neural network internal state interpretation enabling researchers to understand learned feature organization and clustering patterns.

7 Evaluation Methodologies and Benchmarks

Ardila et al. (2020)
Common Voice: A Massively-Multilingual Speech Corpus
Ardila, R. et al. (2020)

Massive multilingual speech dataset aggregating 2,500 hours of audio from 50,000+ contributors through crowdsourcing data collection and validation. Establishes largest public domain speech recognition corpus with 29 languages demonstrating Mozilla DeepSpeech average 5.99% Character Error Rate improvement across 12 languages. Functions as foundational database for multilingual speech technology research and low-resource language development with demographic metadata enhancing recognition system accuracy.

Panayotov et al. (2015)
LibriSpeech: An ASR Corpus Based on Public Domain Audio Books
Large-scale ASR corpus establishing de facto standard for English speech recognition evaluation with 4,770+ citations providing consistent performance comparison foundation. Contains 1000 hours read English audiobooks with carefully designed train/dev/test splits avoiding speaker overlap. Provides clean and other conditions with varying acoustic quality enabling fair comparison across representation learning approaches.

McAuliffe et al. (2017)
Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi
Essential infrastructure enabling large-scale corpus phonological research through trainable text-speech alignment system using Kaldi toolkit with speaker adaptation. Provides pretrained models for 50+ languages achieving accuracy within 20ms of manual annotations. Enables theory-empirical bridging through automated phonetically annotated corpora creation for detailed phonetic evaluation and research applications.

Belinkov & Glass (2019)
Analysis Methods in Neural Language Processing: A Survey
Comprehensive systematization of neural language processing analysis methods providing interpretability research foundation establishment. Categorizes approaches into visualization, diagnostic probing, adversarial evaluation, behavioral analysis discussing probe complexity, significance testing, correlation versus causation problems. Establishes best practices for analyzing phonological representations in neural models with implementation guidelines and limitation discussions.

Conneau et al. (2020)
Unsupervised Cross-lingual Representation Learning for Speech Recognition
Introduces XLSR scaling wav2vec 2.0 to 53 languages covering diverse phonological systems demonstrating 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 providing framework and baselines for cross-linguistic evaluation enabling phonological universals investigation.

8 Recent Advances and Applications

Yang et al. (2025)
k2SSL: A Faster and Better Framework for Self-Supervised Speech Representation Learning
Revolutionary efficiency innovation introducing highly optimized SSL framework achieving 34.8% WER reduction with 3.5x training acceleration through Zipformer architecture U-Net style downsampling. Additional optimizations include ScaledAdam, progressive training, improved augmentation demonstrating efficiency-performance compatibility. Shows architectural innovations improve phonological learning while reducing computational requirements challenging traditional efficiency-quality tradeoffs.

Cho et al. (2025)
Sylber: Syllabic Embedding Representation of Speech from Raw Audio
Revolutionary tokenization innovation introducing syllable granularity dynamic tokenization achieving 4.27 tokens/second performance representing 6-7x improvement over traditional approaches. Demonstrates linguistically motivated tokenization effectiveness through intermediate granularity optimization between phones and words. Shows syllable-level discretization advantages for downstream tasks requiring prosodic information with efficiency gains.

9 Cross-linguistic and Typological Studies

Moran & McCloy (2019)
PHOIBLE: A Large-Scale Database of Phonological Inventories
Comprehensive phonological database aggregating 2,186 language 3,020 phonological inventories representing world's largest systematic collection. Includes segment lists with IPA transcriptions, distinctive features, prosodic information covering all language families enabling typological generalizations about universals and variation. Reveals statistical tendencies and implicational relationships for testing whether learned representations capture cross-linguistic phonological universals.

Mortensen et al. (2016)
PanPhon: A Resource for Mapping IPA Segments to Articulatory Feature Vectors
Comprehensive database mapping 5,000+ IPA segments to 21 articulatory feature vectors combining binary features with gradient phonetic properties. Enables categorical and continuous representations including algorithms for feature inference and segment distance calculation with high linguist judgment agreement validation. Provides ground-truth for evaluating whether neural models discover articulatory structure naturally.

Dunbar et al. (2019)
The Zero Resource Speech Challenge 2019: TTS without T
Zero Resource Challenge evaluating unsupervised linguistic unit discovery from raw speech without text using TTS as downstream task assessing representation quality. Includes typologically diverse languages testing universality combining objective metrics with subjective naturalness assessment. Shows different systems discover different granularities providing evidence about units emerging from unsupervised learning across language families.

Parcollet et al. (2024)
LeBenchmark 2.0: A Standardized, Replicable and Enhanced Framework for Self-Supervised Representations of French Speech
Comprehensive evaluation framework for French speech processing including detailed phonetic tasks providing standardized benchmarks, pretrained models, and evaluation protocols. Covers diverse tasks from phoneme recognition to semantic understanding enabling systematic comparison across architectures and training approaches. Demonstrates language-specific evaluation importance beyond English establishing evaluation standards for non-English speech representation learning.

10 Language Acquisition and Cognitive Modeling

Matusevych et al. (2020)
Evaluating Computational Models of Infant Phonetic Learning across Languages
Evaluates five computational models of infant phonetic learning across three cross-linguistic phonetic contrasts (English []-[l], Mandarin [t]-[t], Catalan [e]-[]) using ABX discrimination tasks. Two models—DPGMM (unsupervised frame-level clustering) and CAE-RNN (weakly supervised sequence learning)—successfully predicted infant-like discrimination patterns for English and Mandarin contrasts, demonstrating that unsupervised learning from natural speech can capture early phonetic development patterns.

Macwhinney (2000)
The CHILDES Project: Tools for Analyzing Talk
Comprehensive Child Language Data Exchange System containing 50+ million transcribed conversation words across 30+ languages. Details CHAT transcription format encoding phonetic details, gestures, context with CLAN analysis tools for automated phonological development analysis. Captures gradual development from babbling through complex phonology providing ecologically valid training data for computational acquisition model evaluation.

Dupoux (2018)
Cognitive Science in the Era of Artificial Intelligence: A Roadmap for Reverse-Engineering the Infant Language-Learner
Articulates research program using AI systems as human cognitive development models focusing on language acquisition with evaluation based on learning trajectories, error patterns, critical periods. Outlines key phenomena including categorical perception emergence and perceptual narrowing introducing cognitive benchmark approach assessing human-like behavior across developmental stages. Essential framework for evaluating cognitive plausibility in phonological learning models.

T. A. Nguyen et al. (2020)
The Zero Resource Speech Benchmark 2021: Metrics and Baselines for Unsupervised Spoken Language Mod-

eling
Comprehensive evaluation metrics and baselines for unsupervised speech learning without transcriptions including ABX discrimination measuring phonemic contrast distinction, word segmentation boundary detection, syntactic/semantic probing. Provides dynamic time warping implementation and statistical methods for acoustic confound control revealing phonetic discrimination emerges early while word representations require extensive data.

Cruz Blandón et al. (2023)
Introducing Meta-Analysis in the Evaluation of Computational Models of Infant Language Development
Meta-analysis framework for evaluating computational infant language development models synthesizing findings across multiple studies identifying consistent patterns and contradictions. Proposes standardized evaluation protocols for model comparison emphasizing ecological validity and developmental trajectories. Provides statistical methods for aggregating results across heterogeneous studies identifying robust findings in computational language acquisition research.

Benders & Blom (2023)
Computational Modelling of Language Acquisition: An Introduction
Introduction bridging theoretical language acquisition frameworks with implementation details covering different modeling paradigms from symbolic to neural approaches. Discusses challenges in modeling realistic input and developmental constraints reviewing evaluation methods comparing models with child data. Emphasizes cognitive plausibility importance beyond task performance for authentic developmental modeling applications.

McMurray (2023)
The Acquisition of Speech Categories: Beyond Perceptual Narrowing, beyond Unsupervised Learning and beyond Infancy
Critical analysis challenging traditional categorical perception views presenting evidence for gradient representations and continuous processing dynamics. Discusses computational model implications for speech perception and acquisition arguing against strict categorical boundaries favoring probabilistic representations. Provides behavioral benchmarks for evaluating model predictions about perceptual categorization in developmental phonological systems.

11 Additional Foundational References

B. B. Tesar (1995)
Computational Optimality Theory
Doctoral dissertation providing computational Optimality Theory implementation demonstrating constraint ranking learnability through Robust Interpretive Parsing handling hidden structure in learning. Develops Error-Driven Constraint Demotion algorithm with convergence proofs showing symbolic grammar learnability from data. Foundational for understanding computational OT approaches and establishing symbolic learning baselines.

Silverman (2012)
Neutralization
Comprehensive theoretical treatment examining neutralization phenomena phonological and phonetic aspects discussing complete versus incomplete neutralization with phonological theory implications. Provides cross-linguistic neutralization pattern typology addressing controversies about underlying representations and abstractness. Important for understanding categorical versus gradient phenomena in phonological representation systems.

Staples & Graves (2020)
Neural Components of Reading Revealed by Distributed and Symbolic Computational Models
Acoustic and articulatory evidence for gradient allophonic variation challenging strict categorical views using ultrasound and acoustic analysis showing continuous variation in supposedly categorical processes. Demonstrates speaker-specific and context-dependent gradience arguing for probabilistic representations capturing variation. Supports representations spanning categorical to gradient spectrums in phonological processing.

D. Nguyen et al. (2016)
Computational Sociolinguistics: A Survey
Analyzes phonological variation, optionality, probability learning in acquisition and diachrony presenting computational variation learning models from ambiguous input. Shows how probabilistic patterns emerge from competing constraints discussing implications for phonological theory and acquisition. Bridges categorical grammar with probabilistic implementation through computational modeling approaches.

Reubold et al. (2010)
Vocal Aging Effects on F0 and the First Formant: A Longitudinal Analysis in Adult Speakers
Longitudinal study documenting vocal aging effects on fundamental frequency and formants with normalization impli-

cations for speech processing systems. Documents systematic acoustic parameter changes across lifespan discussing ASR speaker normalization challenges. Provides within-speaker variation over time data important for understanding speaker variability in representation learning and acoustic model development.

Kazanina et al. (2018)
Phonemes: Lexical Access and Beyond
Critical review examining phoneme concept from psychological, neuroscientific, computational perspectives discussing phoneme evidence in lexical access and speech perception. Reviews controversies about phonological unit psychological reality synthesizing findings from multiple methodologies. Provides balanced phoneme status view in cognitive systems informing computational phonological unit representation approaches.

Tsvilodub et al. (2025)
Integrating Neural and Symbolic Components in a Model of Pragmatic Question-Answering
Recent neural-symbolic integration advances for linguistic structure learning combining pattern recognition with reasoning through modular architectures separating perception from symbolic computation. Shows structured representation benefits for compositional generalization demonstrated in question-answering systems. Provides architectural blueprints for phonological applications requiring symbolic-neural integration.

Medin et al. (2024)
Self-Supervised Models for Phoneme Recognition: Applications in Children's Speech for Reading Learning
Explores educational phonetic analysis applications for language learning and pronunciation training developing systems providing explicit pronunciation error feedback. Uses SSL models for detailed phonetic assessment showing interpretable representation benefits for pedagogical applications. Demonstrates phonologically-informed model practical value in educational technology and language instruction contexts.

Pouw et al. (2024)
Perception of Phonological Assimilation by Neural Speech Recognition Models
Analyzes allophonic variation patterns in spontaneous speech using large corpora and neural models testing whether models capture context-dependent variation similar to human productions. Examines gradience in supposedly categorical processes providing benchmarks for evaluating allophonic knowledge. Shows naturalistic data importance for realistic phonological variation modeling.

Gosztolya et al. (2024)
Wav2vec 2.0 Embeddings Are No Swiss Army Knife—A Case Study for Multiple Sclerosis
Analysis revealing SSL embedding limitations for specialized speech tasks like pathological speech assessment showing domain-specific features sometimes outperform general SSL representations. Identifies conditions where specialized representations necessary providing cautionary evidence about universal SSL applicability. Important for understanding representation limitations in specialized phonological analysis applications.

12 PDF-Based Additional References

J. Chen & Elsner (2023)
Exploring How Generative Adversarial Networks Learn Phonological Representations
Investigates ciwGAN phonological representation learning through nasality feature analysis in French and English vowels finding interactive effects between latent variables rather than one-to-one phonological feature correspondence. Shows GANs distinguish contrastive versus non-contrastive features across languages but learned representations differ from traditional linguistic phonological representations challenging claimed GAN advantages over other neural approaches.

Guriel et al. (2023)
Morphological Inflection with Phonological Features
Explores incorporating phonological features into morphological reinflection tasks using two methods: data manipulation replacing characters with phonological features and model manipulation adding self-attention layer for feature-aware representations. Tests eight shallow-orthography languages with LSTM and transducer models showing comparable performance to graphemic baselines suggesting character-level models already capture phonological information implicitly.

Pasad et al. (2024)
What Do Self-Supervised Speech Models Know About Words?
Comprehensive analysis of word-level linguistic properties in ten self-supervised speech models using canonical correlation analysis and task-based evaluations. Finds word-identifying information concentrates near segment cen-

ters, pre-training objectives influence layer-wise information distribution, and visually grounded models outperform speech-only counterparts on word discrimination, segmentation, and semantic similarity tasks.

Pandian (2025)
Hybrid Symbolic-Neural Architectures for Explainable Artificial Intelligence in Decision-Critical Domains
Proposes hybrid symbolic-neural architectures combining transparent symbolic reasoning with neural network learning capabilities for decision-critical applications including healthcare, legal compliance, and finance. Explores integration strategies including loosely coupled approaches (neural outputs feeding symbolic rules) and tightly coupled approaches (joint training) where interpretability and explainability are paramount for human trust and regulatory approval.

References

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.

Matusevych, Y., Schatz, T., Kamper, H., Feldman, N. H., & Goldwater, S. (2020). *Evaluating computational models of infant phonetic learning across languages* (arXiv:2008.02888). arXiv. https://doi.org/10.48550/arXiv.2008.02888

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