摘要

我們這個研究是嘗試探究離散單元與音位之間的關係。

首先原因是因為語音本身捕捉的是連續的訊號變化，因此我們有了離散單元。然而，人類本身語音學就已經有了離散的符號 — 文字與音位。於是，我們可以試圖去比較現今語音離散表徵與人類對語音音位歸類的差異來理解這些離散表徵是否有捕捉到類似於人類發音的特性。並且，藉由語音對音位標註之間的分組，我們可以觀察這些離散表徵是否也有相似的分組特性。

最後，因為人類對音位的感知往往多於單一的語音表徵音框，因此我們可以嘗試借鑑文字處理中的次詞單位，重新編碼語音訊號再次確認這些次詞單位是否類似於人類的發音特性。

關鍵字：語音基石模型、離散單元、語音表徵、語音學

目錄

第 五 章　結論與展望

5.1　研究貢獻與討論

本篇研究論文旨在細部探討和比較語音基石模型得到的離散表徵與人們理解語音的最小單位 — 音位之間的關係，藉助語音學知識為語音標註提供的分組方式，拓展純度與相互資訊給予的意義，比較共同或條件機率分佈各自的熵與純度等資訊，細部觀察機器學習到的離散表徵與音位標註之間的相似性與差異。其中，藉由分群演算法所獲得的單一離散單元，以及引入文字處理中分詞演算法重新編碼出的次詞單位 — 聲學片段，兩者都是將語音訊號離散化的方式。我們比較了離散單元與聲學片段在音位標註之間共通性的變化，探討用不同方式對語音訊號取得符記造成的影響。

首先，論文第三章介紹了與無文字架構以及語音表徵相關的分析研究，隨後簡介語音學知識中，對於不同音位之間如何按照發音特性分門別類。接著，透過純度與相互資訊定義中統計的共同機率分佈，將機率分佈從離散單元與音位兩個角度切入、比較各自的條件機率分佈特性，觀察不同模型、不同分群參數或不同音位之間是否有特定的集中或分散關係，以及不同離散表徵模型對語音訊號特性歸類的能力。結果可發現，HuBERT 作為目前無文字架構最常用的語音離散表徵模型的理由，很可能來自於它們的音位純度與相互資訊都相對較高，因而更能捕捉到語音中與內容相關的重要資訊，且同樣的趨勢在語音學分類的標註也可以被觀察到。(((寫一下細部的三方向觀察結果))) 從

其後在論文第四章，我們將離散單元以文字處理中的單一詞演算法重新分組編碼為次詞單位序列，以使得不同的離散單元之間可以重新分組成新的符記，並與第三章的結果對照比較，觀察是否在對發音特性的捕捉效果上有所變化。(((寫

一下沒什麼好處的結果)))

5.2 未來展望

希望這些對離散單元與分詞方法應用的嘗試，能幫助我們在訓練任務之前，決定哪種語音基石模型更適合作為離散編碼語音訊號的基礎。接下來，我們期望能針對常見的語音任務，特別是語音辨識和語音翻譯等內容處理相關的任務，比對離散單元促成的實際成效和分析數據之間的關係，並對這些任務中的錯誤案例進行統計和個案探討。

另外，對於如何結合語音離散單元，除了將其視為文字進行分詞演算法外，我們還可以使用其他方式對離散單元序列進行分組，以達成壓縮序列長度並使其與音位等語音內容更加一致的目標。例如，將此目標形塑為語音分段（Speech Segmentation）任務等，也是未來可以嘗試的離散單元分組方式。

最後，利用語音學分組的切入點，或許可以在未來分析離散單元或連續語音表徵時，不再僅限於參考音位或文字，還可以從語音學知識提供的相似性資訊出發，為錯誤發音修正等任務提供衡量的依據。

參 考 文 獻

[1] T. Kudo, "Subword regularization: Improving neural network translation models with multiple subword candidates," 2018.

[2] ——, "Subword regularization: Improving neural network translation models with multiple subword candidates," 2018.

[3] I. P. Association, *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press, 1999.

[4] A. Klautau, "Arpabet and the timit alphabet," *an archived file. https://web. archive. org/web/20160603180727/http://www. laps. ufpa. br/aldebaro/papers/ak_arpabet01. pdf (Accessed Mar. 12, 2020)*, 2001.

[5] "The CMU Pronouncing Dictionary." [Online]. Available: http://www.speech.cs.cmu.edu/cgi-bin/cmudict?stress=-s&in=CITE

[6] S. Watanabe, T. Hori, S. Karita, T. Hayashi, J. Nishitoba, Y. Unno, N. Enrique Yalta Soplin, J. Heymann, M. Wiesner, N. Chen, A. Renduchintala, and T. Ochiai, "ESPnet: End-to-end speech processing toolkit," in *Proceedings of Interspeech*, 2018, pp. 2207–2211. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2018-1456

[7] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, 2016.

[8] T. Kudo and J. Richardson, "SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, E. Blanco and W. Lu, Eds. Brussels, Belgium: Association for Computational Linguistics, Jan. 2018, pp. 66–71. [Online]. Available: https://aclanthology.org/D18-2012

[9] P. Gage, "A new algorithm for data compression," *C Users J.*, vol. 12, no. 2, p. 23–38, feb 1994.

[10] R. Sennrich, B. Haddow, and A. Birch, "Neural Machine Translation of Rare Words with Subword Units," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, K. Erk and N. A. Smith, Eds. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 1715–1725. [Online]. Available: https://aclanthology.org/P16-1162

[11] Z. Ma, Z. Zheng, G. Yang, Y. Wang, C. Zhang, and X. Chen, "Pushing the Limits of Unsupervised Unit Discovery for SSL Speech Representation," in *Proc. INTERSPEECH 2023*, 2023, pp. 1269–1273.

[12] Y. Yang, F. Shen, C. Du, Z. Ma, K. Yu, D. Povey, and X. Chen, "Towards universal speech discrete tokens: A case study for asr and tts," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 10 401–10 405.

[13] H.-J. Chang and J. Glass, "R-spin: Efficient speaker and noise-invariant representation learning with acoustic pieces," *arXiv preprint arXiv:2311.09117*, 2023.

[14] A. Elkahky, W.-N. Hsu, P. Tomasello, T.-A. Nguyen, R. Algayres, Y. Adi, J. Copet, E. Dupoux, and A. Mohamed, "Do coarser units benefit cluster prediction-based speech pre-training?" in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, June 2023, pp. 1–5.

[15] F. Shen, Y. Guo, C. Du, X. Chen, and K. Yu, "Acoustic bpe for speech generation with discrete tokens," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 11 746–11 750.

[16] B. Shi, W.-N. Hsu, K. Lakhotia, and A. Mohamed, "Learning audio-visual speech representation by masked multimodal cluster prediction," in *International Conference on Learning Representations*, 2021.

[17] M. de Seyssel, M. Lavechin, Y. Adi, E. Dupoux, and G. Wisniewski, "Probing phoneme, language and speaker information in unsupervised speech representations," in *Proc. Interspeech 2022*, 2022, pp. 1402–1406.

[18] B. M. Abdullah, M. M. Shaik, B. Möbius, and D. Klakow, "An Information-Theoretic Analysis of Self-supervised Discrete Representations of Speech," in *Proc. INTERSPEECH 2023*, 2023, pp. 2883–2887.

[19] A. H. Liu, H.-J. Chang, M. Auli, W.-N. Hsu, and J. Glass, "Dinosr: Self-distillation and online clustering for self-supervised speech representation learning," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[20] Z. Huang, C. Meng, and T. Ko, "Repcodec: A speech representation codec for speech tokenization," *arXiv preprint arXiv:2309.00169*, 2023.

[21] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 5206–5210, iSSN: 2379-190X. [Online]. Available: https://ieeexplore.ieee.org/document/7178964

[22] A. Sicherman and Y. Adi, "Analysing discrete self supervised speech representation for spoken language modeling," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, June 2023, pp. 1–5.

[23] L. Strgar and D. Harwath, "Phoneme Segmentation Using Self-Supervised Speech Models," in *2022 IEEE Spoken Language Technology Workshop (SLT)*, Jan. 2023, pp. 1067–1073. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10022827

[24] H. Tan and M. Bansal, "Vokenization: Improving language understanding with contextualized, visual-grounded supervision," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 2066–2080. [Online]. Available: https://aclanthology.org/2020.emnlp-main.162

[25] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in neural information processing systems*, vol. 33, pp. 12 449–12 460, 2020.

[26] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[28] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[29] A. Van Den Oord, O. Vinyals *et al.*, "Neural discrete representation learning," *Advances in neural information processing systems*, vol. 30, 2017.

[30] T. Maekaku, X. Chang, Y. Fujita, L.-W. Chen, S. Watanabe, and A. Rudnicky, "Speech representation learning combining conformer cpc with deep cluster for the zerospeech challenge 2021," 2022.

[31] X. Zhang, D. Zhang, S. Li, Y. Zhou, and X. Qiu, "Speechtokenizer: Unified speech tokenizer for speech large language models," 2024.

[32] G.-T. Lin, Y.-S. Chuang, H.-L. Chung, S.-w. Yang, H.-J. Chen, S. Dong, S.-W. Li, A. Mohamed, H.-y. Lee, and L.-s. Lee, "Dual: Discrete spoken unit adaptive learning for textless spoken question answering," *arXiv preprint arXiv:2203.04911*, 2022.

[33] M. Rivière, A. Joulin, P.-E. Mazaré, and E. Dupoux, "Unsupervised pretraining transfers well across languages," 2020.

[34] A. Baevski, S. Schneider, and M. Auli, "vq-wav2vec: Self-supervised learning of discrete speech representations," *arXiv preprint arXiv:1910.05453*, 2019.

[35] S. Schneider, A. Baevski, R. Collobert, and M. Auli, "wav2vec: Unsupervised Pre-Training for Speech Recognition," in *Proc. Interspeech 2019*, 2019, pp. 3465–3469.

[36] W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed, "Hubert: Self-supervised speech representation learning by masked prediction of hidden units," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3451–3460, 2021, publisher: IEEE.

[37] M. Ravanelli, J. Zhong, S. Pascual, P. Swietojanski, J. Monteiro, J. Trmal, and Y. Bengio, "Multi-Task Self-Supervised Learning for Robust Speech Recognition," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 6989–6993, iSSN: 2379-190X.

[38] Y.-A. Chung and J. Glass, "Generative Pre-Training for Speech with Autoregressive Predictive Coding," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 3497–3501, iSSN: 2379-190X.

[39] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems*, vol. 33.

Curran Associates, Inc., 2020, pp. 1877–1901. [Online]. Available: https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

[40] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," 2019.

[41] L. T, LiShang-Wen, and LeeHung-yi, "TERA: Self-Supervised Learning of Transformer Encoder Representation for Speech," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, publisher: IEEE. [Online]. Available: https://dl.acm.org/doi/10.1109/TASLP.2021.3095662

[42] A. T. Liu, S.-w. Yang, P.-H. Chi, P.-c. Hsu, and H.-y. Lee, "Mockingjay: Unsupervised Speech Representation Learning with Deep Bidirectional Transformer Encoders," Oct. 2019. [Online]. Available: https://arxiv.org/abs/1910.12638v2

[43] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[44] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep Contextualized Word Representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long*

Papers), M. Walker, H. Ji, and A. Stent, Eds. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 2227–2237. [Online]. Available: https://aclanthology.org/N18-1202

[45] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Sep. 2013, arXiv:1301.3781 [cs]. [Online]. Available: http://arxiv.org/abs/1301.3781

[46] J. Chorowski, R. J. Weiss, S. Bengio, and A. Van Den Oord, "Unsupervised speech representation learning using wavenet autoencoders," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 27, no. 12, pp. 2041–2053, 2019, publisher: IEEE.

[47] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[48] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, D. Wu, M. Carpuat, X. Carreras, and E. M. Vecchi, Eds. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 103–111. [Online]. Available: https://aclanthology.org/W14-4012

[49] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," 2014.

[50] D. M. Eberhard, G. F. Simons, and C. D. Fennig, *Ethnologue: Languages of the World*, 27th ed. Dallas, Texas: SIL International, 2024. [Online]. Available: https://www.ethnologue.com

[51] K. Lakhotia, E. Kharitonov, W.-N. Hsu, Y. Adi, A. Polyak, B. Bolte, T.-A. Nguyen, J. Copet, A. Baevski, A. Mohamed, and E. Dupoux, "Generative Spoken Language Modeling from Raw Audio," Sep. 2021, arXiv:2102.01192 [cs]. [Online]. Available: http://arxiv.org/abs/2102.01192

[52] "Textless NLP: Generating expressive speech from raw audio," Sep. 2021. [Online]. Available: https://ai.meta.com/blog/textless-nlp-generating-expressive-speech-from-raw-audio/

[53] S. Ren, S. Liu, Y. Wu, L. Zhou, and F. Wei, "Speech Pre-training with Acoustic Piece," in *Interspeech 2022*. ISCA, Sep. 2022, pp. 2648–2652. [Online]. Available: https://www.isca-archive.org/interspeech_2022/ren22_interspeech.html

[54] F. Wu, K. Kim, S. Watanabe, K. J. Han, R. McDonald, K. Q. Weinberger, and Y. Artzi, "Wav2Seq: Pre-Training Speech-to-Text Encoder-Decoder Models Using Pseudo Languages," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island, Greece: IEEE, Jun. 2023, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/10096988/

[55] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," in *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, pp.

12 449–12 460. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/92d1e1eb1cd6f9fba3227870bb6d7f07-Abstract.html

[56] X. Chang, B. Yan, K. Choi, J.-W. Jung, Y. Lu, S. Maiti, R. Sharma, J. Shi, J. Tian, S. Watanabe, Y. Fujita, T. Maekaku, P. Guo, Y.-F. Cheng, P. Denisov, K. Saijo, and H.-H. Wang, "Exploring Speech Recognition, Translation, and Understanding with Discrete Speech Units: A Comparative Study," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2024, pp. 11 481–11 485, iSSN: 2379-190X. [Online]. Available: https://ieeexplore.ieee.org/document/10447929

[57] P.-J. Chen, K. Tran, Y. Yang, J. Du, J. Kao, Y.-A. Chung, P. Tomasello, P.-A. Duquenne, H. Schwenk, H. Gong, H. Inaguma, S. Popuri, C. Wang, J. Pino, W.-N. Hsu, and A. Lee, "Speech-to-Speech Translation for a Real-world Unwritten Language," in *Findings of the Association for Computational Linguistics: ACL 2023*, A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 4969–4983. [Online]. Available: https://aclanthology.org/2023.findings-acl.307

[58] H.-y. Lee, A. Mohamed, S. Watanabe, T. Sainath, K. Livescu, S.-W. Li, S.-w. Yang, and K. Kirchhoff, "Self-supervised Representation Learning for Speech Processing," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, M. Ballesteros, Y. Tsvetkov, and C. O. Alm, Eds. Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 8–13. [Online]. Available: https://aclanthology.org/2022.naacl-tutorials.2

[59] K. Lakhotia, E. Kharitonov, W.-N. Hsu, Y. Adi, A. Polyak, B. Bolte, T.-A. Nguyen, J. Copet, A. Baevski, A. Mohamed, and E. Dupoux, "On Generative Spoken Language Modeling from Raw Audio," *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 1336–1354, 2021, place: Cambridge, MA Publisher: MIT Press. [Online]. Available: https://aclanthology.org/2021.tacl-1.79

[60] W.-N. Hsu, Y.-H. H. Tsai, B. Bolte, R. Salakhutdinov, and A. Mohamed, "Hubert: How Much Can a Bad Teacher Benefit ASR Pre-Training?" in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, ON, Canada: IEEE, Jun. 2021, pp. 6533–6537. [Online]. Available: https://ieeexplore.ieee.org/document/9414460/

[61] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation Learning with Contrastive Predictive Coding," Jan. 2019, arXiv:1807.03748 [cs, stat]. [Online]. Available: http://arxiv.org/abs/1807.03748

[62] X. Zhao, Q. Zhu, J. Zhang, Y. Zhou, and P. Liu, "Speech Enhancement with Multi-granularity Vector Quantization," in *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Oct. 2023, pp. 1937–1942, iSSN: 2640-0103. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10317485

[63] L.-W. Chen, S. Watanabe, and A. Rudnicky, "A Vector Quantized Approach for Text to Speech Synthesis on Real-World Spontaneous Speech," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, pp. 12 644–12 652, Jun. 2023,

number: 11. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/26488

[64] X. Chang, B. Yan, Y. Fujita, T. Maekaku, and S. Watanabe, "Exploration of Efficient End-to-End ASR using Discretized Input from Self-Supervised Learning," in *INTERSPEECH 2023*. ISCA, Aug. 2023, pp. 1399–1403. [Online]. Available: https://www.isca-archive.org/interspeech_2023/chang23b_interspeech.html

[65] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Jan. 1998, conference Name: Proceedings of the IEEE. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/726791

[66] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of Physiology*, vol. 148, no. 3, pp. 574–591, Oct. 1959. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1363130/

[67] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943, publisher: Springer. [Online]. Available: https://doi.org/10.1007/BF02478259

[68] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958, publisher: American Psychological Association. [Online]. Available: https://doi.apa.org/doi/10.1037/h0042519

[69] D. Wells, H. Tang, and K. Richmond, "Phonetic Analysis of Self-supervised Representations of English Speech," 2022, pp. 3583–3587. [Online]. Available: https://www.isca-archive.org/interspeech_2022/wells22_interspeech.html

[70] K.-I. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, no. 3, pp. 183–192, Jan. 1989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0893608089900038

[71] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/323533a0

[72] D. E. Rumelhart and J. L. McClelland, "Learning Internal Representations by Error Propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. MIT Press, 1987, pp. 318–362. [Online]. Available: https://ieeexplore.ieee.org/document/6302929