

DAFTAR PUSTAKA

- Anguera, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G. dan Vinyals, O., 2012, Speaker Diarization: A Review of Recent Research, *IEEE Transactions on Audio, Speech, and Language Processing*, [Online] 20 (2), 356–370, tersedia di DOI:10.1109/TASL.2011.2125954.
- Anidjar, O.H., Estève, Y., Hajaj, C., Dvir, A. dan Lapidot, I., 2023, Speech and multilingual natural language framework for speaker change detection and diarization, *Expert Systems with Applications*, [Online] 213119238, tersedia di DOI:<https://doi.org/10.1016/j.eswa.2022.119238>.
- Aperdannier, R., Schacht, S. dan Piazza, A., 2024, *A Review of Common Online Speaker Diarization Methods*.
- Appiahene, P., Ussiph, N. dan Missah, Y., 2018, Information Technology Impact on Productivity: A Systematic Review and Meta-Analysis of the Literature, *International Journal of Information Communication Technologies and Human Development*, [Online] 1039–61, tersedia di DOI:10.4018/IJICTHD.2018070104.
- Bahdanau, D., Cho, K. dan Bengio, Y., 2014, Neural Machine Translation by Jointly Learning to Align and Translate, *CoRR*, [Online] abs/1409.0473, tersedia di <https://api.semanticscholar.org/CorpusID:11212020>.
- Bapat, R.B., 2014, *Graphs and matrices*, [Online]. tersedia di DOI:10.1007/978-1-4471-6569-9.
- Boersma, P. dan Weenink, D., 2025, *Praat: doing phonetics by computer [Computer program]*. [Online]. tersedia di <https://archive.mpi.nl/tla/elan>.
- Brynjolfsson, E. dan McAfee, A., 2014, *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.*, W W Norton & Co, New York, NY, US.
- Bullock, L., Bredin, H. dan Garcia-Perera, L.P., 2020, Overlap-Aware Diarization: Resegmentation Using Neural End-to-End Overlapped Speech Detection, *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, [Online], 2020 hlm. 7114–7118, tersedia di DOI:10.1109/ICASSP40776.2020.9053096.
- Chen, W., Anh, T.T., Zhong, X. dan Chng, E.S., 2024, Enhancing Low-Latency Speaker Diarization with Spatial Dictionary Learning, *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, [Online], 2024 hlm. 11371–11375, tersedia di DOI:10.1109/ICASSP48485.2024.10446666.
- Chilkuri, N. dan Eliasmith, C., 2021, Parallelizing Legendre Memory Unit Training, *ArXiv*, [Online] abs/2102.11417, tersedia di <https://api.semanticscholar.org/CorpusID:232014544>.

- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. dan Bengio, Y., 2014, Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, Alessandro Moschitti, Bo Pang, dan Walter Daelemans (ed.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, [Online], Oktober 2014 Association for Computational Linguistics, Doha, Qatar., hlm. 1724–1734, tersedia di DOI:10.3115/v1/D14-1179.
- Davenport, T. dan Harris, J., 2007, *Competing on Analytics: The New Science of Winning*.
- European Broadcasting Union, 2023, *R 128: Loudness Normalisation and Permitted Maximum Level of Audio Signals*. [Online]. (R 128). tersedia di <https://tech.ebu.ch/docs/r/r128.pdf>.
- Fayek, H.M., 2016, *Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between*. [Online]. tersedia di <https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>.
- Fujita, Y., Kanda, N., Horiguchi, S., Nagamatsu, K. dan Watanabe, S., 2019, End-to-End Neural Speaker Diarization with Permutation-Free Objectives, *Interspeech 2019*, [Online], 2019 hlm. 4300–4304, tersedia di DOI:10.21437/Interspeech.2019-2899.
- Fujita, Y., Watanabe, S., Horiguchi, S., Xue, Y. dan Nagamatsu, K., 2020, End-to-End Neural Diarization: Reformulating Speaker Diarization as Simple Multi-label Classification, *ArXiv*, [Online] abs/2003.02966, tersedia di <https://api.semanticscholar.org/CorpusID:212628366>.
- Gan, Z. dan Ma, F., 2024, An Improved Multi-Speaker Tibetan Speaker Diarization Based on EDA-EEND, *2024 7th International Conference on Computer Information Science and Application Technology (CISAT)*, [Online], 2024 hlm. 322–325, tersedia di DOI:10.1109/CISAT62382.2024.10695270.
- Gaurav, R., Stewart, T.C. dan Yi, Y., 2023, Reservoir Based Spiking Models for Univariate Time Series Classification. *Frontiers in Computational Neuroscience*. [Online]. tersedia di DOI:10.3389/fncom.2023.1148284.
- Géron, A., 2023, *Recurrent Neural Networks*. [Online]. tersedia di <https://www.oreilly.com/library/view/neural-networks-and/9781492037354/ch04.html#idm139624956152448-marker>.
- Gold, B., Morgan, N. dan Ellis, D., 2011, *Speech and audio signal processing*, John Wiley & Sons.
- Gupta, M. dan George, J.F., 2016, Toward the development of a big data analytics capability, *Information & Management*, [Online] 53 (8), 1049–1064, tersedia di DOI:<https://doi.org/10.1016/j.im.2016.07.004>.
- Hao, F., Li, X. dan Zheng, C., 2023, End-to-end neural speaker diarization with an iterative adaptive attractor estimation, *Neural Networks*, [Online] 166566–578, tersedia di DOI:<https://doi.org/10.1016/j.neunet.2023.07.043>.

- Harris, F.J., 1978, On the use of windows for harmonic analysis with the discrete Fourier transform, *Proceedings of the IEEE*, [Online] 66 (1), 51–83, tersedia di DOI:10.1109/PROC.1978.10837.
- Hochreiter, S. dan Schmidhuber, J., 1997, Long Short-Term Memory, *Neural Computation*, [Online] 9 (8), 1735–1780, tersedia di DOI:10.1162/neco.1997.9.8.1735.
- Horiguchi, S., Watanabe, S., Garcia, P., Takashima, Y. dan Kawaguchi, Y., 2022, *Online Neural Diarization of Unlimited Numbers of Speakers Using Global and Local Attractors*, [Online] tersedia di DOI:10.1109/TASLP.2022.3233237.
- India, M., Hernando, J. dan Fonollosa, J.A.R., 2023, Language modelling for speaker diarization in telephonic interviews, *Computer Speech & Language*, [Online] 78101441, tersedia di DOI:<https://doi.org/10.1016/j.csl.2022.101441>.
- Jurafsky, D. dan Martin, J.H., 2024, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd edisi, [Online]. tersedia di <https://web.stanford.edu/~jurafsky/slp3/>.
- Kangala, V.K. dan Ramisetty, R.R., 2023, Optimized speaker change detection approach for speaker segmentation towards speaker diarization based on deep learning, *Data & Knowledge Engineering*, [Online] 144102121, tersedia di DOI:<https://doi.org/10.1016/j.datak.2022.102121>.
- Khambra, G., 2023, IMPLICATIONS OF SPEECH RECOGNITION THROUGH NLP: AN OVERVIEW, *BSSS Journal of Computer*, [Online] tersedia di <https://api.semanticscholar.org/CorpusID:261598185>.
- Kinnunen, T. dan Li, H., 2010, An overview of text-independent speaker recognition: From features to supervectors, *Speech Communication*, [Online] 52 (1), 12–40, tersedia di DOI:<https://doi.org/10.1016/j.specom.2009.08.009>.
- Ko, T., Peddinti, V., Povey, D., Seltzer, M.L. dan Khudanpur, S., 2017, A study on data augmentation of reverberant speech for robust speech recognition, *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, [Online], 2017 hlm. 5220–5224, tersedia di DOI:10.1109/ICASSP.2017.7953152.
- Loshchilov, I. dan Hutter, F., 2017, Decoupled Weight Decay Regularization, *International Conference on Learning Representations*, [Online], 2017 hlm. tersedia di <https://api.semanticscholar.org/CorpusID:53592270>.
- Lu, H., Cheng, G. dan Yan, Y., 2024, Conversational Short-Phrase Speaker Diarization via Self-Adjusting Speech Segmentation and Embedding Extraction, *IEEE Signal Processing Letters*, [Online] 312340–2344, tersedia di DOI:10.1109/LSP.2024.3453772.
- Maas, A.L., 2013, *Rectifier Nonlinearities Improve Neural Network Acoustic Models*, [Online], 2013 hlm. tersedia di <https://api.semanticscholar.org/CorpusID:16489696>.

- Majeed, S.A., Husain, H., Samad, S.A. dan Idbeaa, T.F., 2015, Mel frequency cepstral coefficients (Mfcc) feature extraction enhancement in the application of speech recognition: A comparison study, *Journal of theoretical and applied information technology*, [Online] 7938–56, tersedia di <https://api.semanticscholar.org/CorpusID:145817537>.
- Maseri, M. dan Mamat, M., 2020, *Performance Analysis of Implemented MFCC and HMM-based Speech Recognition System*, [Online]. tersedia di DOI:10.1109/IICAIET49801.2020.9257823.
- Mayer-Schönberger, V. dan Cukier, K., 2013, *Big data: A revolution that will transform how we live, work, and think.*, Houghton Mifflin Harcourt, Boston, MA.
- Morrone, G., Cornell, S., Serafini, L., Zovato, E., Brutti, A. dan Squartini, S., 2024, End-to-end integration of speech separation and voice activity detection for low-latency diarization of telephone conversations, *Speech Communication*, [Online] 161103081, tersedia di DOI:<https://doi.org/10.1016/j.specom.2024.103081>.
- Oppenheim, A. V dan Schafer, R.W., 2010, *Discrete-time signal processing*, Prentice Hall.
- Oppenheim, A. V, Schafer, R.W. dan Buck, J.R., 1999, *Discrete-time signal processing*.
- Pang, B. dan Lee, L., 2008, Opinion Mining and Sentiment Analysis, *Foundations and Trends® in Information Retrieval*, [Online] 2 (1–2), 1–135, tersedia di DOI:10.1561/1500000011.
- Papala, G., Ransing, A. dan Jain, P., 2023, Sentiment Analysis and Speaker Diarization in Hindi and Marathi Using using Finetuned Whisper: Sentiment Analysis in Hindi and Marathi, *Scalable Comput. Pract. Exp.*, [Online] 24835–846, tersedia di <https://api.semanticscholar.org/CorpusID:265310061>.
- Park, T.J., Kanda, N., Dimitriadis, D., Han, K.J., Watanabe, S. dan Narayanan, S., 2022, A review of speaker diarization: Recent advances with deep learning, *Computer Speech & Language*, [Online] 72101317, tersedia di DOI:<https://doi.org/10.1016/j.csl.2021.101317>.
- Peterson, K., Tong, A. dan Yu, Y., 2022, OpenASR21: The Second Open Challenge for Automatic Speech Recognition of Low-Resource Languages, *Interspeech*, [Online], 2022 hlm. tersedia di <https://api.semanticscholar.org/CorpusID:252348636>.
- Proakis, J.G. dan Manolakis, D.G., 2007, *Digital signal Processing*, Pearson.
- Rabiner, L.R. dan Juang, B.-H., 1993, Fundamentals of speech recognition, *Prentice Hall signal processing series*, [Online], 1993 hlm. tersedia di <https://api.semanticscholar.org/CorpusID:268095015>.
- Raj, D., Huang, Z. dan Khudanpur, S., 2021, Multi-Class Spectral Clustering with Overlaps for Speaker Diarization, *2021 IEEE Spoken Language Technology*

Workshop (SLT), [Online], 2021 hlm. 582–589, tersedia di DOI:10.1109/SLT48900.2021.9383602.

Ravanelli, M. dkk., 2021, *SpeechBrain*. *GitHub repository*.

Ryant, N., Church, K.W., Cieri, C., Cristia, A., Du, J., Ganapathy, S. dan Liberman, M.Y., 2019, The Second DIHARD Diarization Challenge: Dataset, task, and baselines, *Interspeech*, [Online], 2019 hlm. tersedia di <https://api.semanticscholar.org/CorpusID:195069497>.

Rybicka, M., Villalba, J., Thebaud, T., Dehak, N. dan Kowalczyk, K., 2024, End-to-End Neural Speaker Diarization With Non-Autoregressive Attractors, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, [Online] 323960–3973, tersedia di <https://api.semanticscholar.org/CorpusID:271776874>.

Sakti, S., Kelana, E., Riza, H., Sakai, S., Markov, K. dan Nakamura, S., 2008, Development of Indonesian Large Vocabulary Continuous Speech Recognition System within A-STAR Project, *Proceedings of the Workshop on Technologies and Corpora for Asia-Pacific Speech Translation (TCAST)*, [Online], 2008 hlm. tersedia di <https://aclanthology.org/I08-8004/>.

Sanchez-Lengeling, B., Reif, E., Pearce, A. dan Wiltschko, A.B., 2021, A Gentle Introduction to Graph Neural Networks, *Distill*, [Online] tersedia di DOI:10.23915/distill.00033.

Satriawan, C.H. dan Lestari, D.P., 2018, Average Window Smoothing for an Indonesian Language Online Speaker Identification System, *International Journal on Electrical Engineering and Informatics*, [Online] tersedia di <https://api.semanticscholar.org/CorpusID:86551696>.

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. dan Monfardini, G., 2009, The Graph Neural Network Model, *IEEE Transactions on Neural Networks*, [Online] 20 (1), 61–80, tersedia di DOI:10.1109/TNN.2008.2005605.

Sejdić, E., Djurović, I. dan Jiang, J., 2009, Time-frequency feature representation using energy concentration: An overview of recent advances, *Digital Signal Processing: A Review Journal*, [Online] 19 (1), tersedia di DOI:10.1016/j.dsp.2007.12.004.

Sell, G. dan Garcia-Romero, D., 2014, Speaker diarization with plda i-vector scoring and unsupervised calibration, *2014 IEEE Spoken Language Technology Workshop (SLT)*, [Online], 2014 hlm. 413–417, tersedia di DOI:10.1109/SLT.2014.7078610.

Serafini, L., Cornell, S., Morrone, G., Zovato, E., Brutti, A. dan Squartini, S., 2023, An experimental review of speaker diarization methods with application to two-speaker conversational telephone speech recordings, *Computer Speech & Language*, [Online] 82101534, tersedia di DOI:<https://doi.org/10.1016/j.csl.2023.101534>.

Shamsi, M., Larcher, A., Barrault, L., Meignier, S., Prokopalo, Y., Tahon, M., Mehrish, A., Petitrenaud, S., Galibert, O., Gaist, S., Anjos, A., Marcel, S. dan

- Costa-jussà, M.R., 2023, Towards lifelong human assisted speaker diarization, *Computer Speech & Language*, [Online] 77101437, tersedia di DOI:<https://doi.org/10.1016/j.csl.2022.101437>.
- Sloetjes, H. dan Wittenburg, P., 2008, Annotation by Category: ELAN and ISO DCR, Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, dan Daniel Tapias (ed.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, [Online], Mei 2008 European Language Resources Association (ELRA), Marrakech, Morocco., hlm. tersedia di <https://aclanthology.org/L08-1034/>.
- Smith, J.O., 2011, *Spectral Audio Signal Processing*.
- Snyder, D., Chen, G. dan Povey, D., 2015, *MUSAN: A Music, Speech, and Noise Corpus*. [Online]. tersedia di <https://arxiv.org/abs/1510.08484>.
- Snyder, D., Garcia-Romero, D., Povey, D. dan Khudanpur, S., 2017, Deep Neural Network Embeddings for Text-Independent Speaker Verification, *Interspeech 2017*, [Online], 2017 hlm. 999–1003, tersedia di DOI:10.21437/Interspeech.2017-620.
- Snyder, D., Garcia-Romero, D., Sell, G., Povey, D. dan Khudanpur, S., 2018, X-Vectors: Robust DNN Embeddings for Speaker Recognition, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, [Online], 2018 hlm. 5329–5333, tersedia di DOI:10.1109/ICASSP.2018.8461375.
- Sohn, J., Kim, N.S. dan Sung, W., 1999, A statistical model-based voice activity detection, *IEEE Signal Processing Letters*, [Online] 61–3, tersedia di <https://api.semanticscholar.org/CorpusID:9808701>.
- Sun, J., 2019, Research on vocal sounding based on spectrum image analysis, *EURASIP Journal on Image and Video Processing*, [Online] 2019, tersedia di DOI:10.1186/s13640-018-0397-0.
- The Language Archive, 2024, *ELAN (Version 6.9) [Computer software]*. [Online]. tersedia di <https://archive.mpi.nl/tla/elan>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. ukasz dan Polosukhin, I., 2017, Attention is All you Need, I Guyon, U Von Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, dan R Garnett (ed.), *Advances in Neural Information Processing Systems*, [Online], 2017 Curran Associates, Inc., hlm. tersedia di https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. dan Bengio, Y., 2017, Graph Attention Networks, *6th International Conference on Learning Representations*,
- Voelker, A., Kajić, I. dan Eliasmith, C., 2019, Legendre Memory Units: Continuous-Time Representation in Recurrent Neural Networks, H Wallach, H Larochelle, A Beygelzimer, F d Alché-Buc, E Fox, dan R Garnett (ed.),

Advances in Neural Information Processing Systems, [Online], 2019 Curran Associates, Inc., hlm. tersedia di https://proceedings.neurips.cc/paper_files/paper/2019/file/952285b9b7e7a1be5aa7849f32ffff05-Paper.pdf.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K. dan Lang, K.J., 1989, Phoneme recognition using time-delay neural networks, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, [Online] 37 (3), 328–339, tersedia di DOI:10.1109/29.21701.

Wang, Z., Jiang, P., Wang, Z., Han, B., Liang, H., Ai, Y. dan Pan, W., 2024, Enhancing Air Traffic Control Communication Systems with Integrated Automatic Speech Recognition: Models, Applications and Performance Evaluation, *Sensors*, [Online] 24 (14), tersedia di DOI:10.3390/s24144715.

Wei, Y., Guo, H., Ge, Z. dan Yang, Z., 2023, Graph attention-based deep embedded clustering for speaker diarization, *Speech Communication*, [Online] 155102991, tersedia di DOI:<https://doi.org/10.1016/j.specom.2023.102991>.

Zewoudie, A.W., Luque, J. dan Pericás, F.J.H., 2014, *Jitter and Shimmer Measurements for Speaker Diarization*, [Online], 2014 hlm. tersedia di <https://api.semanticscholar.org/CorpusID:59409231>.

Zhou, J., 1990, *Discrete cosine transform*, [Online]. tersedia di DOI:10.1016/c2009-0-22279-3.