

DAFTAR PUSTAKA

- Atmaja, B. T., Akhmad, F., Arifianto, D., & Akagi, M. (2018). *Speech recognition on Indonesian language by using time delay neural network*.
- Bailey, G. (2016). Automatic detection of sociolinguistic variation using forced alignment. *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAV 44*, 22(2), 10–20. <https://doi.org/10.1017/9781107279865>
- Bigi, B., & Meunier, C. (2018). Automatic Segmentation of Spontaneous Speech / Segmentação automática da fala espontânea. *Revista De Estudos Da Linguagem*, 26(4), 1489–1489. <https://doi.org/10.17851/2237-2083.26.4.1489-1530>
- Boersma, P., & van Heuven, V. (2001). *Speak and unSpeak with PRAAT*. 5(9), 7.
- Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. (2004). Fisher English Training Speech Part 1 Transcripts. In *Philadelphia: Linguistic Data Consortium*. <https://catalog.ldc.upenn.edu/LDC2004T19>
- Dahl, G. E., Dong Yu, Li Deng, & Acero, A. (2012). Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1), 30–42. <https://doi.org/10.1109/TASL.2011.2134090>
- Dave, N. (2013). *Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition*. 1, 5.
- Fromont, R., & Watson, K. (2016). Factors influencing automatic segmental alignment of sociophonetic corpora. *Corpora*, 11(3), 401–431. <https://doi.org/10.3366/cor.2016.0101>
- Georgescu, A.-L., & Cucu, H. (2018). Automatic Annotation of Speech Corpora Using Complementary GMM and DNN Acoustic Models. *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, 1–4. <https://doi.org/10.1109/TSP.2018.8441374>
- Goldman, J. P. (2011). EasyAlign: An automatic phonetic alignment tool under Praat. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 3233–3236.
- Gonzalez, S., Grama, J., & Travis, C. E. (2019). *Comparing the performance of major forced aligners used in sociophonetic research*.
- Gonzalez, S., Travis, C. E., Grama, J., Barth, D., Ananthanarayan, S., Gonzalez, S., Travis, C. E., Grama, J., Barth, D., & Ananthanarayan, S. (2018). *Recursive forced alignment: A test on a minority language ARC Centre of Excellence for the Dynamics of Language , Australian National University University of Texas at Austin*. 2–5.
- Gorman, K., Howell, J., & Wagner, M. (2011). *Prosodylab-Aligner: A Tool for Forced Alignment of Laboratory Speech*. 39(3), 192–193.
- Gunawan, M., Nurfadhilah, E., Aini, L. R., Uliniansyah, M. T., & Junde, J. (2018). *Uji Coba Korpus Data Wicara BPPT sebagai Data Latih Sistem Pengenalan Wicara Bahasa Indonesia*. 1(2), 6.
- Heriyanto, H., Hartati, S., & Putra, A. E. (2018). EKSTRAKSI CIRI MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC) DAN RERATA COEFFICIENT UNTUK PENGECEKAN BACAAN AL-QUR'AN. *Telematika*, 15(2), 99. <https://doi.org/10.31315/telematika.v15i2.3123>
- Johnson, L. M., Di Paolo, M., & Bell, A. (2018). Forced alignment for understudied language varieties: Testing Prosodylab-Aligner with Tongan data. *Language Documentation and Conservation*, 12, 80–123.

- Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall.
- Li, H., Wu, J., & Narasimhalu, A. D. (2002). *Method and Apparatus for Voice Annotation and Retrieval of Multimedia Data*.
- Liu, F.-H., Stern, R. M., Huang, X., & Acero, A. (1993). Efficient cepstral normalization for robust speech recognition. *Proceedings of the Workshop on Human Language Technology - HLT '93*, 69. <https://doi.org/10.3115/1075671.1075688>
- Lotus Quality Assurance. (2016). *Lotus Quality Assurance*. <https://www.lotus-qa.com>
- MacKenzie, L., & Turton, D. (2019). *Assessing the accuracy of existing forced alignment software on varieties of British English*.
- Matoušek, J., & Tihelka, D. (2013). Annotation errors detection in TTS corpora. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, August*, 1511–1515.
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, 498–502. <https://doi.org/10.21437/Interspeech.2017-1386>
- Misbullah, A., Nazaruddin, N., Marzuki, M., & Zulfan, Z. (2020). Penerapan Time Delay Neural Network pada Model Akustik untuk Sistem Voice-to-Text Berbahasa Sunda. *Journal of Data Analysis*, 2(2), 61–70. <https://doi.org/10.24815/jda.v2i2.15235>
- Mohri, M., Pereira, F., & Riley, M. (2002). Weighted finite-state transducers in speech recognition. *Computer Speech & Language*, 16(1), 69–88. <https://doi.org/10.1006/csla.2001.0184>
- Mohri, M., Pereira, F., & Riley, M. (2008). Speech Recognition with Weighted Finite-State Transducers. In J. Benesty, M. M. Sondhi, & Y. A. Huang (Eds.), *Springer Handbook of Speech Processing* (pp. 559–584). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-49127-9_28
- Ochshorn, R. M., & Hawkins, M. (2016). *Gentle: A robust yet lenient forced aligner built on Kaldi*. <https://lowerquality.com/gentle/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., & Glembek, O. (2011). *Kaldi: Decoding graph construction in Kaldi*. <https://kaldi-asr.org/doc/graph.html>
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., & Vesely, K. (2011). *The Kaldi Speech Recognition Toolkit*. <http://infoscience.epfl.ch/record/192584>
- Rao, K. S., & K E, M. (2017). *Speech Recognition Using Articulatory and Excitation Source Features*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-49220-9>
- Reddy, S., & Stanford, J. N. (2015). Toward completely automated vowel extraction: Introducing DARLA. *Linguistics Vanguard*, 1(1), 15–28. <https://doi.org/10.1515/lingvan-2015-0002>
- Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). FAVE (forced alignment and vowel extraction) program suite. URL <Http://Fave.Ling.Upenn.Edu>.
- Saon, G., Soltau, H., Nahamoo, D., & Picheny, M. (2013). Speaker adaptation of neural network acoustic models using i-vectors. *2013 IEEE Workshop on Automatic*

- Speech Recognition and Understanding*, 55–59.
<https://doi.org/10.1109/ASRU.2013.6707705>
- Weide, R. L. (1998). *The CMU pronouncing dictionary*.
<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>
- Young, S. (1994). *The HTK Hidden Markov Model Toolkit: Design and Philosophy*
Cambridge University Engineering Department Trumpington Street , Cambridge ,
CB2 1PZ. May, 1–29.
- Yu, D., & Deng, L. (2015). *Automatic Speech Recognition*. Springer London.
<https://doi.org/10.1007/978-1-4471-5779-3>
- Zhang, X., Trmal, J., Povey, D., & Khudanpur, S. (2014). Improving deep neural network
acoustic models using generalized maxout networks. *2014 IEEE International
Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 215–219.
<https://doi.org/10.1109/ICASSP.2014.6853589>