Springer Link

Search 🔍　　　🛒　　　Log in

Analysis of influencing features with spectral feature extraction and multi-class classification using deep neural network for speech recognition system

Dinesh Kumar Anguraj, J. Anitha, S. John Justin Thangaraj, L. Ramesh, Seetha Rama Krishna & D. Mythrayee ✉

Metrics

Abstract

There is a drastic need for extracting information from non-linguistic features of the audio sources. It leads to the eminent rise of speech technology over the past few decades. It is termed computational para-linguistics. This research concentrates on extracting and providing a robust feature that examines the characteristics of speech data. The factors are analysed in a spectral way which stimulates the auditory elements. The speech enhancement technological process is being initiated with pre-processing, feature extraction, and classification. Initially, the input data conversion is done with ADC of 16 kHz sampling frequency. The spectral features are extracted with minimal Mean Square Error to enhance the re-

construction ability and eliminate the redundancy characteristics. Finally, the deep neural network is adopted for multi-class classification. The simulation is performed in MATLAB 2020a environment, and the empirical outcomes are evaluated with existing approaches. Here, metrics like Mean Square Error, accuracy, Signal-to-Noise ratio (SNR) and features retained are computed efficiently. The anticipated model shows a trade-off in contrast to prevailing approaches. The outcomes demonstrate a better recognition rate and offer significant characteristics in selecting the most influencing features.

This is a preview of subscription content, <u>access via your institution</u>.

Access options

Buy article PDF

34,95 €

Price includes VAT (India)
Tax calculation will be finalised during checkout.

Instant access to the full article PDF.

<u>Learn more about Institutional subscriptions</u>

References

Deng, J., Xu, X., Zhang, Z., Frühholz, S., &

Schuller, B. (2016). Exploitation of phase-based features for whispered speech emotion recognition. *IEEE Access, 4*, 4299–4309.

El-Maghraby, E. E., Gody, A., & Farouk, M. (2020). Noise-robust speech recognition system based on multimodal audio-visual approach using different deep learning classification techniques. *Egyptian Journal of English Language, 7*(1), 27–42.

Fan, Y., Jiang, Q.-Y., Yu, Y.-Q., & Li, W.-J. (2019). Deep hashing for speaker identification and retrieval. In *Proceedings of Interspeech* Graz, pp. 2908–2912.

He, N., & Zhao, H. (2017). A retrieval algorithm of encrypted speech based on syllable-level perceptual hashing". *Computer Science and Information Systems, 14*(3), 703–718.

Ioffe, S., & Szegedy, C. (2015). Batch normalisation: Accelerating deep network training by reducing internal covariate shift. In *The Proceedings of Machine Learning Research.*, pp. 1–9.

Khonglah, B.K. & Prasanna, S. R. M. (2016). Low frequency region of vocal tract information for speech/music classification. In *Proceedings of IEEE Region 10 Conference*, pp. 2593–2597.

Kim, B. & Pardo, B.A. (2019). Improving content-based audio retrieval by vocal imitation feedback.' In *Proceedings of the IEEE International Conference on Acoustics Speech Signal Process. (ICASSP)*, Brighton, pp. 4100–4104.

Kruspe, A., Zapf, D., & Lukashevich, H. (2017) Automatic speech/music discrimination for broadcast signals. In Proceeding of INFORMATIK *Gesellschaft für Informatik*, Bonn, pp. 151–162.

Li, L. W., Fu, T., & Hu, W. (2019). Piecewise supervised deep hashing for image retrieval. *Multimedia Tools and Applications, 78*(17), 24431–24451.

Li, Y., Kong, X., & Fu, H. (2019). Exploring geometric information in CNN for image retrieval". *Multimedia Tools and Applications, 78*(21), 30585–30598.

Li, Y., Tao, J., Chao, L., Bao, W., & Liu, Y. (2017). CHEAVD: A Chinese natural emotional audio_visual database. *Journal of Ambient Intelligence and Humanized Computing, 8*(6), 913–924.

Liu, Z. T., Xie, Q., Wu, M., Cao, W.-H., Mei, Y., & Mao, J.-W. (2018). Speech emotion recognition

based on an improved brain emotion learning model. *Neurocomputing, 309*, 145–156.

Mao, Q., Dong, M., Huang, Z., & Zhan, Y. (2014). Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE Transactions on Multimedia, 16*(8), 2203–2213.

Masoumeh, S., & Mohammad, M. B. (2014). A review on speech-music discrimination methods. *International Journal of Computer Science and Network Solutions, 2*, 67–78.

Mirsamadi, S, E. Barsoum, & C. Zhang, (2017). Automatic speech emotion recognition using recurrent neural networks with local attention. In *Proceedings of the IEEE International Conference on Acoustics, (ICASSP)*, pp. 2227–2231.

Papakostas, M., & Giannakopoulos, T. (2018). Speech-music discrimination using deep visual feature extractors. *Expert Systems with Applications, 114*, 334–344.

Rahmani, A.M. & F. Razzazi. (2019). An LSTM auto-encoder for single-channel speaker attention system. In *Proceeding of the 9th International Conference Computer Knowledge. Engineering (ICCKE)*, Mashhad, pp. 110–115.

Song, P., Zheng, W., Ou, S., Zhang, X., Jin, Y., Liu, J., & Yu, Y. (2016). Crosscorpus speech emotion recognition based on transfer non-negative matrix factorisation. *Speech Communication, 83*, 34–41.

Srinivas, D., Roy, D., & Mohan, C.K. (2014). Learning sparse dictionaries for music and speech classification. In *Proceeding of the 19th International Conference on Signal Processing,* pp. 673–675.

Sun, W., Zou, B., Fu, S., Chen, J., & Wang, F. (2019). Speech emotion recognition based on DNN-decision tree SVM model. *Speech Communication, 115*, 29–37.

Swain, M., Routray, A., & Kabisatpathy, P. (2018). Databases, features and classi_ers for speech emotion recognition: A review. *International Journal of Speech Technology, 21*(1), 93–120.

Wang, Z.Q. & Tashev, I. (2017) Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks. In *Proceedings of the IEEE International Conference on Acoustics, (ICASSP),* pp. 5150–5154.

Wen, G., Li, H., Huang, J., Li, D. & Xun, E (2017). Random deep belief networks for recognising

emotions from speech signals. *Computational Intelligence and Neuroscience*

Xu, Y., Kong, Q., Wang, W. & Plumbley, M.D. (2018). Large-scale weakly supervised audio classification using gated convolutional neural network. In *Proceedings of the IEEE International Conference on Acoustics, Speech Signal Process. (ICASSP)*, Calgary, pp. 121–125.

Yu, J., Markov, K., & Matsui, T. (2019). Articulatory and spectrum information fusion based on deep recurrent neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27*(4), 742–752.

Zhang, Z. (2016). Mechanics of human voice production and control". *Journal of the Acoustical Society of America, 140*(4), 2614–2635.

Zhang, H., Li, Y., Hu, Y., & Zhao, X. (2020). An encrypted speech retrieval method based on deep perceptual hashing and CNN-BiLSTM. *IEEE Access, 8*, 148556–148569.

Zhang, Z., Pi, Z., & Liu, B. (2015). Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on Biomedical Engineering, 62*(2), 522–531.

Zhang, N. & Lin, J. (2018). An efficient content based music retrieval algorithm. In *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Big Data Smart City (ICITBS)*, Xiamen, pp. 617–620.

Author information

Affiliations

Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

Dinesh Kumar Anguraj

Department of CSE, Malla Reddy Engineering College (A), Dhulapally, Secunderabad, Telangana, India

J. Anitha

Computer Science and Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai, Tamil Nadu, India

S. John Justin Thangaraj

Dr MGR Educational and Research Institute, Maduravoyal, Chennai, Tamil Nadu, India

L. Ramesh

CreovSys Solutions India Private Limited, Tiruppur, Tamil Nadu, India

Seetha Rama Krishna & D. Mythrayee

Corresponding author

Correspondence to [D. Mythrayee](#).

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article