

# References

## DATASET

Çano, Erion; Morisio, Maurizio (2017). Music Mood Dataset Creation Based on Last.fm Tags. In: Fourth International Conference on Artificial Intelligence and Applications (AIAP 2017), Vienna, Austria, 27-28 May 2017. pp. 15-26.

## EMOTION DETECTION

Yi-Hsuan Yang and Homer H. Chen. 2011. Music Emotion Recognition (1st. ed.). CRC Press, Inc., USA.

Çano, Erion. (2018). Text-based Sentiment Analysis and Music Emotion Recognition.

Sailunaz, K., Dhaliwal, M., Rokne, J. *et al.* Emotion detection from text and speech: a survey. *Soc. Netw. Anal. Min.* 8, 28 (2018). <https://doi.org/10.1007/s13278-018-0505-2>.

Parisi, Loreto & Francia, Simone & Olivastri, Silvio & Tavella, Maria. (2019). Exploiting Synchronized Lyrics And Vocal Features For Music Emotion Detection.

Martín-Gómez, Lucía & Cáceres, María. (2018). Applying Data Mining for Sentiment Analysis in Music. 198-205. 10.1007/978-3-319-61578-3\_20.

Laurier, C., & Herrera, P. (2009). Automatic Detection of Emotion in Music: Interaction with Emotionally Sensitive Machines.

Hızlısoy, Serhat & Yildirim, Serdar & Tüfekci, Zekeriya. (2020). Music emotion recognition using convolutional long short term memory deep neural networks. *Engineering Science and Technology an International Journal*. 24. 760-767. 10.1016/j.jestch.2020.10.009.

Laurier, Cyril & Herrera, Perfecto. (2007). Audio music mood classification using support vector machine. *MIREX Task on Audio Mood Classification*.

Han, Byeong-jun & Rho, Seungmin & Dannenberg, Roger & Hwang, Eenjun. (2009). SMERS: Music Emotion Recognition Using Support Vector Regression.. *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*. 651-656.

Wagner, Johannes & Kim, Jonghwa & Andre, Elisabeth. (2005). From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification.. 940-943. 10.1109/ICME.2005.1521579.

## MUSIC INFORMATION RETRIEVAL

Meinard Miller. 2015. Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications (1st. ed.). Springer Publishing Company, Incorporated.

Lartillot O., Toivainen P., Eerola T. (2008) A Matlab Toolbox for Music Information Retrieval. In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-78246-9\\_31](https://doi.org/10.1007/978-3-540-78246-9_31).

Deepali. Y. Loni, & Dr. Shaila Subbaraman. Extracting Acoustic Features of Singing Voice for Various Applications Related to MIR: A Review.

Fujihara, Hiromasa & Goto, Masataka. (2007). A Music Information Retrieval System Based on Singing Voice Timbre.. 467-470.

Devopedia. 2021. "Audio Feature Extraction." Version 8, May 23. Accessed 2021-09-09. <https://devopedia.org/audio-feature-extraction>

F. Eyben et al., "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," in *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190-202, 1 April-June 2016, doi: 10.1109/TAFFC.2015.2457417.

Gupta, U., Chatterjee, A., Srikanth, R., & Agrawal, P. (2017). A sentiment-and-semantics-based approach for emotion detection in textual conversations. *arXiv preprint arXiv:1707.06996*.

Tang, Duyu & Wei, Furu & Yang, Nan & Zhou, Ming & Liu, Ting & Qin, Bing. (2014). Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*. 1. 1555-1565. 10.3115/v1/P14-1146.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

## DATA FUSION

Smolinska, A., Engel, J., Szymanska, E., Buydens, L., & Blanchet, L. (2019). General Framing of Low-, Mid-, and High-Level Data Fusion With Examples in the Life Sciences. In *Data Fusion Methodology and Applications* (pp. 51-79). (Data Handling in Science and Technology; Vol. 31). Elsevier Ltd, Academic Press. <https://doi.org/10.1016/B978-0-444-63984-4.00003-X>

Castanedo, Federico. (2013). A Review of Data Fusion Techniques. *TheScientificWorldJournal*. 2013. 704504. 10.1155/2013/704504.

## MULTIMODAL SYSTEMS

Mohammad Soleymani, David Garcia, Brendan Jou, Bjo'rn Schuller, Shih-Fu Chang, Maja Pantic, A Survey of Multimodal Sentiment Analysis, *Image and Vision Computing* (2017), doi: 10.1016/j.imavis.2017.08.003.

R. A. Calvo and S. D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," in *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18-37, Jan. 2010, doi: 10.1109/T-AFFC.2010.1.

## LOW-LEVEL FUSION

Louis-Philippe Morency, Rada Mihalcea, and Payal Doshi. 2011. Towards multimodal sentiment analysis: harvesting opinions from the web. In *Proceedings of the 13th international conference on multimodal interfaces (ICMI '11)*. Association for Computing Machinery, New York, NY, USA, 169–176. DOI:<https://doi.org/10.1145/2070481.2070509>.

Poria, Soujanya & Cambria, Erik & Howard, Newton & Huang, Guang-Bin & Hussain, Amir. (2015). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*. 174. 10.1016/j.neucom.2015.01.095.

V. Pérez Rosas, R. Mihalcea and L. Morency, "Multimodal Sentiment Analysis of Spanish Online Videos," in *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 38-45, May-June 2013, doi: 10.1109/MIS.2013.9.

Behjat Siddiquie, Dave Chisholm, and Ajay Divakaran. 2015. Exploiting Multimodal Affect and Semantics to Identify Politically Persuasive Web Videos. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 203–210. DOI:<https://doi.org/10.1145/2818346.2820732>.

Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces (ICMI '04)*. Association for Computing Machinery, New York, NY, USA, 205–211. DOI:<https://doi.org/10.1145/1027933.1027968>

## MID-LEVEL FUSION

Jamdar, Adit & Abraham, Jessica & Khanna, Karishma & Dubey, Rahul. (2015). Emotion Analysis of Songs Based on Lyrical and Audio Features. *International Journal of Artificial Intelligence & Applications*. 6. 10.5121/ijaia.2015.6304.

Poria, Soujanya & Cambria, Erik & Gelbukh, Alexander. (2015). Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-Level Multimodal Sentiment Analysis.

Zhihong Zeng, Yuxiao Hu, Yun Fu, Thomas S. Huang, Glenn I. Roisman, and Zhen Wen. 2006. Audio-visual emotion recognition in adult attachment interview. In *Proceedings of the 8th international conference on Multimodal interfaces (ICMI '06)*. Association for Computing Machinery, New York, NY, USA, 139–145. DOI:<https://doi.org/10.1145/1180995.1181028>

## HIGH-LEVEL FUSION

F. H. Rachman, R. Samo and C. Fatichah, "Song Emotion Detection Based on Arousal-Valence from Audio and Lyrics Using Rule Based Method," 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), 2019, pp. 1-5, doi: 10.1109/ICICoS48119.2019.8982519.

Poria, Soujanya & Cambria, Erik & Gelbukh, Alexander. (2015). Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-Level Multimodal Sentiment Analysis.

M. Wöllmer et al., "YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context," in *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 46-53, May-June 2013, doi: 10.1109/MIS.2013.34.

Poria, Soujanya & Cambria, Erik & Howard, Newton & Huang, Guang-Bin & Hussain, Amir. (2015). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*. 174. 10.1016/j.neucom.2015.01.095.

Behjat Siddiquie, Dave Chisholm, and Ajay Divakaran. 2015. Exploiting Multimodal Affect and Semantics to Identify Politically Persuasive Web Videos. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 203–210. DOI:<https://doi.org/10.1145/2818346.2820732>.

Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces (ICMI '04)*. Association for Computing Machinery, New York, NY, USA, 205–211. DOI:<https://doi.org/10.1145/1027933.1027968>