

{tag}

{/tag}

IJCA Proceedings on National Conference on
Innovative Paradigms in Engineering and Technology (NCIPET 2012)

© 2012 by IJCA Journal

ncipet - Number 1

Year of Publication: 2012

Authors:

N. P. Jawarkar

R. S. Holambe

T. K. Basu

{bibtex}ncipet1001.bib{/bibtex}

Abstract

This paper describes a technique for unsupervised audio segmentation. Main objective of the work presented in this paper is to study the performance of audio segmentation system using metric-based method. The system first classifies the audio signal into speech and nonspeech signal using variance of zero crossing rate. The feature Line spectral pair is used for automatically detecting the speaker change point. Hotelling T2 distance metric is used in the

first stage for coarse speaker change detection. The Bayesian information criterion (BIC) is used in the second stage to validate the potential speaker change point detected by the coarse segmentation procedure to reduce the false alarm rate. Database of four files containing the speech recorded from different combinations of male and female speakers mixed with nonspeech signal such as music and environmental sound are used for segmentation. The database-file with one male and one female gives the best performance with F1 measure of 0.9474.

References

ences

- A. Solomonoff, A. Mielke, M. Schmidt, H. Gish, "Clustering speakers by their voices," in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, Seattle, USA, pp. 757–760, May 1998.
- S. E. Tranter and D. A. Reynolds, "An overview of automatic speaker diarization systems," IEEE Trans. Audio, Speech and language process., vol. 14, no. 5, pp. 1557-1565, Sept. 2006.
- M. Kotti, V. Moschou, and C. Kotropoulos, "Speaker segmentation and clustering," Signal Processing, 88(2008), pp. 1091-1124.
- L. Lu, H. Jiang, and H. J. Zhang, "A robust audio classification and segmentation method," in Proc. 9th ACM Int. Conf. Multimedia, 2001, pp. 203–211.
- R. Huang, J. H. L. Hansen, "Advances in unsupervised audio classification and segmentation for the broadcast news and NGSW corpora," IEEE Trans. Audio, speech, Language Process., vol. 1, no. 3, pp. 07919, May 2006.
- S. Chen and P. Gopalakrishnan, "Speaker, environment and channel change detection and clustering via the Bayesian information criterion," in Proc. Broadcast News Transcr. Under. Workshop, Lansdowne, VA, 1998, pp. 127–132.
- B. Zhou and J. H. L. Hansen, "Unsupervised audio stream segmentation and clustering via the Bayesian information criterion," in Proc. ICSLP 2000, vol. 1, Beijing, China, Oct. 2000, pp. 714–717.
- M. Siegler, U. Jain, B. Raj, and R. M. Stern, "Automatic segmentation, classification and clustering of broadcast news audio," in Proc. DARPA Speech Recognition Workshop, Chantilly, VA, 1997, pp. 97–99.
- C. Barras, X. Zhu, S. Meigner, J. L. Gauvain, "Multistage Speaker diarization of broadcast news," IEEE Trans. Audio Speech Language Process., vol. 14, no. 5, pp. 1557-1565, Sept. 2006.
- M. Cettolo and M. Federico, "Model selection criteria for acoustic segmentation," in Proc. ISCA ITRWASR 2000 Workshop, Paris, France, Sep. 2000, pp. 221–227. National Conference on Innovative Paradigms in Engineering & Technology (NCIPET-2012) Proceedings published by International Journal of Computer Applications® (IJCA) 5
- S. Wegmann, P. Zhan, and L. Gillick, "Progress in broadcast news transcription at dragon systems," Proc. ICASSP, vol. 1, pp. 33–36, Mar. 1999.
- T. Zhang and C.-C. J. Kuo, "Audio content analysis for online audiovisual data segmentation and classification," IEEE Trans. Speech Audio Process., vol. 9, no. 4, pp. 441–457, Jul. 2001.

- L. Lu, H. Zhang, and H. Jiang, "Content analysis for audio classification and segmentation," IEEE Trans. Speech Audio Process., vol. 10, no. 7, pp. 504–516, Oct. 2002.
- S. Z. Li, "Content-based audio classification and retrieval using the nearest feature line method," IEEE Trans. Speech Audio Process., vol. 8, no. 5, pp. 619-625, Sept. 2000.
-] L. Lu and H. Zhang, "Speaker change detection and tracking in real-time news broadcasting analysis," in Proc. ACM Multimedia, Juan-les-Pins, France, Dec. 2002, pp. 602–610.
- S.E. Tranter, K. Yu, G. Evermann, P.C. Woodland, "Generating and valuating segmentations for automatic speech recognition of conversational telephone speech," in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada, pp. 433–477, May 2004.
- A. Adami, S. Kajarekar, and H. Hermansky, "A new speaker change detection method for two-speaker segmentation," in Proc. ICASSP, vol. 4, Orlando, FL, 2002, pp. 13–17.
- T. K. Truong, C. Lin, S. Chen, " Segmentation of specific speech signals from multi-dialog environment using SVM and wavelet," Pattern Recognition Letters, 28, pp. 1307-1313, 2007.
- L. R. Rabiner and R. W. Schafer, Digital signal processing of speech signals, Englewood, NJ: Prentice-Hall, 1978.
- Itakura F., "Line spectrum representation of linear predictive coefficients of speech signals," J. Acoust. Soc Am., 57, 537(A), 1975.
- J. Ajmera, I. McCowan, H. Bourlard, "Robust speaker change detection," IEEE Signal Process. Lett. 11 (8), pp. 649–651, August 2004.

Index Terms

Computer Science

Engineering and Technology

Keywords

Speaker segmentation LSP audio segmentation VZCR