On-Line Speaker Indexing

1. Research Team

Project Leader: Prof. Shrikanth Narayanan, *Electrical Engineering*

Post Doc(s): Panayiotis Georgiou

Graduate Students: Soonil Kwon

2. Statement of Project Goals

In this project, we propose an on-line method that picks out the speech segments from an audio stream, and classifies them by speakers. On-line speaker indexing method should be only sequentially executed. In other words, we make any decision of indexing with only current and previous speech data. Moreover it gets more difficult if there is no prior knowledge about speakers in speech data. Since the models of speakers are not available for indexing, we need to build them on the fly with only previous speech data. We try to solve this critical problem using an alternative method.

3. Project Role in Support of IMSC Strategic Plan

Automatic segmentation and classification of multi-speaker audio data have been given considerable attention as multimedia communication/information systems are getting an integral part of our daily life. For example, multimedia meetings or teleconferences are one of the most common but important events. It is impossible to attend all relevant meetings that are held all over the world. Multimedia meeting or teleconference browsers can be used as on-line or off-line systems to overcome spatiotemporal problems [1,2]. Off-line speaker indexing cannot be used for real-time meeting or teleconferencing systems. Our on-line speaker indexing algorithm can give speaker index information to real-time or on-line multimedia systems.

4. Discussion of Methodology Used

The audio samples come into the system, and are classified into different speech and background noise types. Only speech data are used for the next stage (speaker change detection). In the next step, the system detects the end of speech data of a current speaker. When it finds the boundary, the whole data between front and rear ends are used for clustering. After clustering, a selected model for the current speaker is adapted into the current speaker dependent model. The adapted model is kept in the reference model set, the original model before adaptation is deleted. Next audio samples after the boundary of the current speaker come into the system, and the system repeats the previous steps until all data are consumed.

5. Short Description of Achievements in Previous Years

Developed algorithms for speaker change detection and speaker clustering.

5a. Detail of Accomplishments During the Past Year

Our new technique gave 96.5\% indexing accuracy for a telephone conversation data source and 84.3\% accuracy for a broadcast news source.

6. Other Relevant Work Being Conducted and How this Project is Different

On-line speaker indexing has not been actively studied yet. A few works have been reported about speaker indexing. One of these works is about self-organized modeling by Nishida and Ariki [3]. Their method can perform the speaker indexing in on-line mode. This means that speaker indexing and model construction can be performed sequentially without storing all the testing data in advance. But the problem is that sequentially constructed models cannot represent speakers well due to the initial small amount of data. This problem also makes some error propagation. We try to solve this critical drawback using an alternative method. The idea is to build generic models of speakers, which are independent of speakers for a test with the assumption that some speakers of the reference set are acoustically close to the test speaker. Although we do not know the exact number of speakers, we assume that the number is finite. With this assumption, some models are built through training with data irrelevant to the test. This reference model set can make it possible for an on-line system to run well without training.

7. Plan for the Next Year

Improve the speaker change detection and model adaptation algorithm.

8. Expected Milestones and Deliverables

Create a robust and elaborate on-line multi-speaker indexing algorithm.

9. Member Company Benefits

N/A

10. References

[1] Kwon, S. and Narayanan, S., ``Speaker Change Detection Using a New Weighted Distance Measure'', International Conference on Spoken Language Processing, Volume Two, p.2537-2540, 2002.
[2] Yang, J., Zhu, X., Gross, R., Kominek, J., Pan, Y., and Waibel, A., "Multimodal People ID for a Multimedia Meeting Browser", Proceedings of 7th ACM International Conference on Multimedia, Part One, p.159-168, 1999.
[3] Nishida, M. and Ariki, Y., "Speaker Indexing for News Articles, Debates, and Drama in Broadcasted TV Programs", IEEE International Conference on Multimedia Computing and systems, Volume Two, p.466-471, 1999.