On-Line Speaker Indexing
Soonil Kwon
Shrikanth Narayanan

Research Goal
• Sequentially detect points where a speaker identity changes in a multi-speaker audio stream.
• Categorize each speaker segment without any prior knowledge about the target speakers.

Role in IMSC
• Speaker indexing, the process of determining who is talking when, is an integral element of speech data monitoring and content-based data mining applications.
• Example: multimedia meeting/teleconference monitors and browsers can be useful for conveniently obtaining meeting information, such as who is saying what and when, remotely through on-line or off-line systems.

Research Approach
• This research addresses two challenges: The first relates to sequential speaker change detection. The second relates to speaker modeling in light of the fact that the number/identity of the speakers is unknown.
• To address these issues, a predetermined generic speaker-independent model set, called the Sample Speaker Models (SSM), is proposed.

Accomplishments
• About 17% accuracy improvement through SSM based speaker indexing.
• Publications

Uniqueness & Related Work
• Methods based on speaker verification using speaker subspace for speaker indexing (by Nishida and Ariki).
• Iterative speaker segmentation using the Generalized Likelihood Ratio (GLR) Test (Rosenberg et al).
• Our work is for on-line speaker segmentation and clustering without prior knowledge of speakers and speaker models with higher accuracy.

5-Year Plan
• The optimal number of sample speaker models and positions in the feature space to use for unsupervised speaker indexing.
• Higher level linguistic information and multi-modal features can be integrated.