Maestro: Conductor of Multimedia Analysis Technologies

SRI International
High-level Goal

Develop techniques to provide direct and natural access to a large database of multimedia data through multiple modalities, including video, audio, and text.
The Maestro Goal

- Research and demonstration testbed.
- To discover, develop, and test synergistic strategies of combining multimedia analysis technologies.
- To achieve archiving performance exceeding that possible with any of the technologies on their own.
- Maestro provides integrating architecture and visual interface (The Maestro score).
SRI’s Multimedia Analysis Technologies

- **Speech processing**
  - Automatic speech recognition (ASR)
  - Speaker identification
  - Speaker tracking/grouping
  - Sentence boundary/disfluency detection

- **Video analysis**
  - Scene segmentation
  - Scene tracking/grouping
  - Camera flash detection

- **Optical character recognition (OCR)**
  - Video caption
  - Scene text (light or dark)
  - Person identification

- **Information extraction (IE)**
  - Names of people, places, organizations
  - Temporal terms
  - Story segmentation/classification
Broadcast News Application
Key Characteristics

- Geographic location
- Topic of the news segment
- Names of news-making people
- Who or what is in the picture
- Who is speaking
Augmented Lexicon Improves Recognition Results

Without lexicon:  TONY  BLAKJB  WNITEE SIATEE
With lexicon:     TONY  BLAIR     UNITED STATES
Example Combination

Text Recog

“BLAIR” +

add to lexicon

ASR

blair

NL

“Blair”