We propose a realtime multimodal system that combines face pose tracking and speaker diarization, for analyzing group meetings. System aims to estimate "who is looking at whom" (=visual focus of attention) and "who is speaking now", using a new omnidirectional camera–microphone system. It can also create a 3-D visualization of the meeting scene in realtime. To the best of our knowledge, this is the first system that can realize both face "pose" tracking (not only position) and speaker diarization, in realtime. This system will open the door to realtime applications such as teleconferencing systems.

**References**
