

Context Sensitive Audio Analysis for Interactive Multimedia Indexing

What if?	Applications
<p>Cell phones could know when their owner is speaking, and ring accordingly?</p> <p>Room acoustics could cue a PDA to remind its user of location-specific tasks?</p> <p>Mobile devices could index and log data relevant to the user's daily activities, for example time spent on the phone, in the car, in meetings?</p>	<ul style="list-style-type: none"> • Automatic generation of an indexed script of a meeting or other event • Identification of speaker and creation of metrics for researching group dynamics such as interruption patterns, dominance, etc. • Identification of active speakers and extraction of their individual voices from a video or audio feed from a room full of people • Enhancement of the active speakers in a meeting or teleconference • Speaker tracking for automatic video steering • Context aware mobile devices for meetings.
Research	People and Partners
<p>This project sets out to study and create a family of techniques and algorithms to facilitate synchronous and asynchronous group meetings by analyzing ambient audio information.</p> <p>--> More below</p>	<p>Julius Smith, CCRMA</p> <p>Mark Bolas, Mechanical Engineering</p> <p>Larry Leifer, Mechanical Engineering</p> <p>Roy Pea, Education</p> <p>Pamornpol Jinachitra, Electrical Engineering</p> <p>Ian McDowall, Polaris Road, Inc.</p>

Research In a room full of events like people talking or music playing, it is useful to be able to locate and identify the active sources of sounds. Each source sound can then be enhanced, extracted, analyzed for its content, tracked for its movement, or even get its pattern extracted in relation to other sources for social study purposes. A growing need exists for multimedia content indexing. The automatic creation of an index enables the fast searching of multimedia recordings. For example, a target sound or a particularly insightful comment could quickly be found in a video recording of a meeting which may contain hours of irrelevant material. It enables the user to have a convenient access to desired parts of the multimedia clip more quickly, as well as having a better idea of what the clip is about through automatic summarization. This will increase efficiency and productivity in the workplace and home.

The project aims to explore the use of a microphone array and digital signal processing technology that can locate and track the active sound sources through time. The sources can subsequently be identified according to our mapping, given labels or information from synchronous video, and possibly extracted and analyzed for contents. While current automatic summarization and indexing mostly analyze only the content, this project will investigate the use of spatial information in indexing. While we anticipate most sources to be speech, non-speech sounds (e.g., music, or non speech cues such as clapping) will also be accommodated. Special attention will be placed on cases with multiple active sources in real room environments. This is a challenging scenario where most existing techniques do not perform well, and is partly responsible for the relatively inferior performance of speech recognition systems. The appropriate output of the analysis will anticipate integration with other multimedia content platforms.

Contact colas@well.com