Real-Time Sound Source Separation and Localisation

Summary:

Multiple sound sources in a noisy environment can be separated, emphasized, suppressed, modified and then recombined in any 3D spatial configuration to give clean sound signals. All processing is done in real-time and no prior knowledge of the sources or environment is needed.

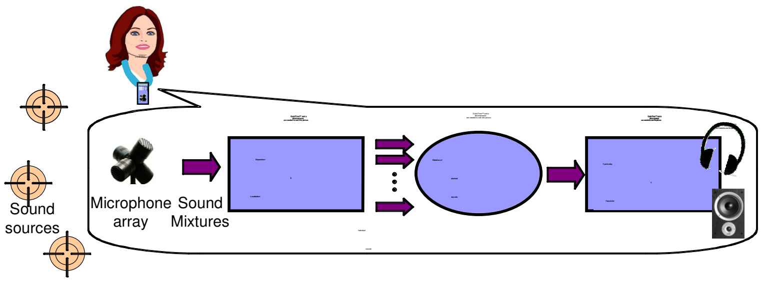

Fig 1. System block diagram

Main Features:

- User can optimize the system in real-time to achieve the best sound quality
- Small, lightweight microphone array; even suitable for MEMS microphones in a very small compact array
- Real-time separation <1/40th sec delay
- Isolate and lock-on to a moving sound source.
- Compatible with frequency manipulation to compensate for hearing loss
- Only 2 to 4 audio channels; easy to interface and process
- Fully 3D binaural presentation of separated sounds to improve intelligibility
- Up to 26dB improvement in SNR, on average 14.4dB
- Localise sound in any horizontal or vertical axis with no confusion of spatial orientation e.g. mixing up front and back
- Full audio range from 20 Hz to 22 kHz.

Development status:

Fully functional portable demonstrator

A generic (application / market independent), portable technology demonstrator is fully functional. It comprises special data capture hardware and uses off-the-shelf components, including small conventional microphones, and software. Use of even smaller MEMS microphones is possible. The system runs on a PC or MAC laptop and separation and localisation parameters can be changed in real-time using the specially developed Graphical User Interface (GUI) provided with the software. The system is suitable for demonstrating the capabilities of the source separation algorithm and carrying out performance testing.

Availability: Available for licensing.

IP status: Patent filed by University of Surrey

Contact: Martyn Buxton-Hoare
Assistant Director - Technology Transfer
T +44 (0) 1483 683 670
M +44 (0) 7714 604 823
m.buxton-hoare@surrey.ac.uk

Potential Applications:

Hearing-aids, enhanced hearing

Listening to selected sounds/conversations and improved speech intelligibility for hearing-impaired.

Mobile phones

Environmental noise and interference suppression

Human-computer interfaces – speech recognition

Pre-processing to improve signal-to-noise ratio

Broadcasting, content production

Real-time audio capturing and synthesis for 3D TV productions, ensuring spatial synchronicity of sound and picture, including multi-view rendering.

Immersive remote collaboration

Selective transmission of multiple speech sounds and their processing for 3D reproduction.

In-car and hands-free communication

Noise and acoustic echo cancellation.

Surveillance and security CCTV systems

automatically select and point to sounds and listen;

automatic keyword / threat detection in noisy, multi-speaker environments such as airports.

Biometric speaker identification

Pre-processing to improve signal-to-noise ratio

Technical:

Block-based processing enables fast, near real-time implementation even at high sampling rates such as 44.1 kHz. Closed-form separation does not require adaptation or iteration. Performance tests in semi-reverberant and reverberant environments with two or three sound sources provide good signal-to-interference (SIR) ratios (Fig 2).

When used with automatic keyword recognition systems as a front end processor the technology significantly improves the recognition rate.