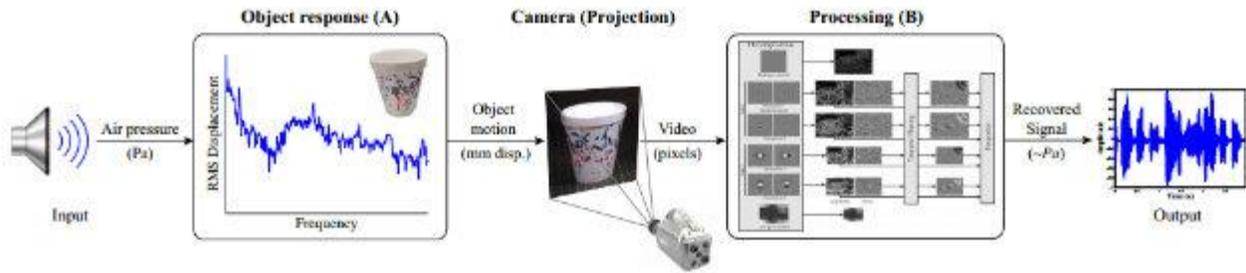# Recovering Audio from Video: James Bond Technology for Future Broadcasters?

Researchers at MIT have recently found a way to extract audio from a real scene using only the visual signal captured by a video camera. When a sound wave impinges on an object, it causes small vibrations on the surface of the object. If these vibrations are captured by a video camera with sufficiently high resolution and frame rate, advanced signal processing algorithms can partially recreate the sound wave. A group of MIT researchers have set up a system and conducted this sound extraction experiment with different kinds of sound (music, speech, etc.) on different kinds of objects (a bag of potato chips, foam cup, plants, etc.), and they have published their results along with clips of the original and recovered sounds at http://people.csail.mit.edu/mrub/VisualMic. The recovered sound in this experiment shows strong correlation with the original input.

The fundamental concept of detecting sound-induced vibration on an object's surface and then reproducing the sound from that vibration is not new. Infrared devices based on this principle were used by the eavesdroppers during the cold war era. Laser microphones, which sometimes use an additional laser Doppler vibrometer to resolve the phase ambiguity due to motion, operate on the same principle. These devices, however, require that a beam be pointed at the target object and that a receiver be precisely positioned to receive the reflected beam. The approach taken by the MIT researchers is different: Their vision is to extract the sound from the video recorded by a commercially available camera.

The following figure describes the concept. Sound wave generated by a speaker on the left is the input to the system. A foam cup in the picture serves as a visual microphone that vibrates with the sound wave. The camera records the subtle movements of the cup and thus transforms sound induced physical displacement into pixel motions of video. The video is then processed by a signal processing block, and the subtle motions of the cup is finally translated into sound that is correlated with the input signal.

Object response (A) | **Camera (Projection)** | **Processing (B)**

Input — Air pressure (Pa) — RMS Displacement / Frequency — Object motion (mm disp.) — Video (pixels) — Recovered Signal (~Pa) — Output

In addition to using high frame rate cameras for this experiment, the MIT researchers also experimented with a normal video camera using a rolling shutter. A rolling shutter imaging technique, common in CMOS sensor based cameras, means that the sensor pixels are read out row-by-row sequentially at different times. Since each row of the video frame is captured at different times, the sampled information is also different. Thus the audio signal corresponding to a row of video signal is effectively sampled at a much higher rate than the video frame rate in rolling shutter cameras. With the knowledge of frame period, exposure time, frame delay, and line delay, the desired audio signal can be partially reconstructed using sophisticated algorithms.

If this technology matures, the real time sound extraction capability from video could potentially be integrated into a TV camera. Such cameras, in addition to actively recording high quality audio as they do today, may also passively record many different bits of sounds. This could be a useful tool for news gathering, targeted advertising, and social media-based applications. For example, while capturing video of a crowd around an event, the news crew may be able to find out what the individuals in the crowd are discussing. A video of spectators at a ball park with cups of beer in their hands may also reveal, in real time, what they are chatting about. The potential business opportunities and applications of this technology are many, and broadcasters have a natural advantage in this game as they own the cameras with the highest resolution and frame rates. (Privacy issues may also have to be considered, however.) Although still in its infancy, this technology may someday give broadcasters an upper hand in social media and targeted advertising in the information age.

Important Dates and Upcoming Events

NAB/SBE Satellite Uplink Operators Training Workshop

September 29 - October 2, 2014

Washington, DC

137th International AES Convention

October 9 - October 12, 2014

Los Angeles, CA

2014 IEEE Broadcast Symposium

October 15 - October 17, 2014

San Antonio, TX

SMPTE Annual Technical Conference and Exhibition

October 20 - October 23, 2014

Hollywood, CA