3D Sensors with 5.1 audio system and microphones

Robin Lobel, December 2007

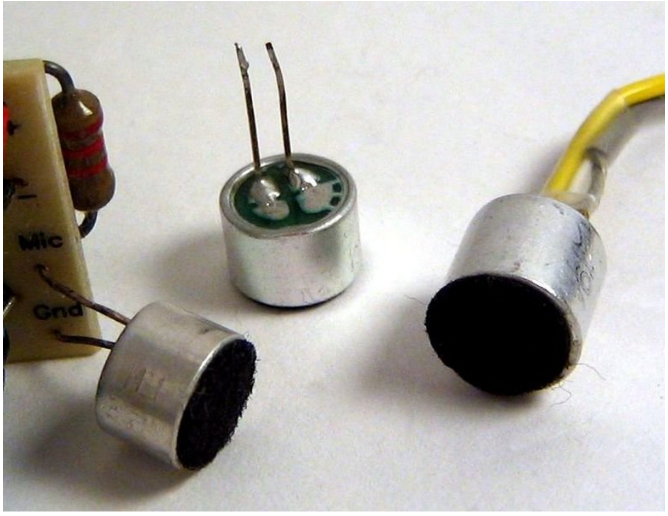

Fig 1. Electret microphones

Abstract

This paper present a way to turn standard, low-cost electret microphones into 3d sensors using a computer with 5.1 audio system.

1. The idea behind audio 3d sensors

Since sonar use sound to measure distances, a computer audio system may as well be used to measure distances with appropriate set of mics, speakers and signal coding.

1.1 Hardware

For the experiment we will use a low cost set of electret microphones (Fig 1.) and small speakers (such as thoses used in ear speakers, Fig 2).

Fig 2. Microphone and speaker

For a dual 3d sensors configuration, we will use 3 speakers and 2 microphones.
Theses microphones/speakers must be have good response at 18500Hz and more.
Using a 5.1 audio system we can connect the 3 speakers, and the 2 microphones are plugged to the stereo microphone input of the soundcard.

The speakers are disposed in an equally spaced triangle whose position is known, while the microphones will be the mobile parts.

1.2 Signal

The frequencies we chose to use are in a very narrow band: high enough not to be heard by humans (experiences shows that any frequency starting at 18500Hz and up is not heard), and low enough to be supported by the electret microphone, speakers, and computer audio card.
While most sound cards now support 96000Hz sampling, it's very unlikely our hardware has good response around theses frequencies. Plus, if we want to maintain backward compatibility with older soundcards, we should stick at 44100Hz sampling.
In consequence, we will use frequencies between 18500Hz and 20000Hz.

To detect the exact distance between speaker and microphone, both impulse response and phase response will be used. To do so, we create our signal starting with a pure frequency (18500Hz for instance) and we modulate its amplitude by a serie of Blackmann-Harris windows.
This window has good properties for short time Fourier analysis, as it does not alter much the original signal. So the signal will be near-pure 18500Hz with a slight margin due to impulse modulation (Fig 3).

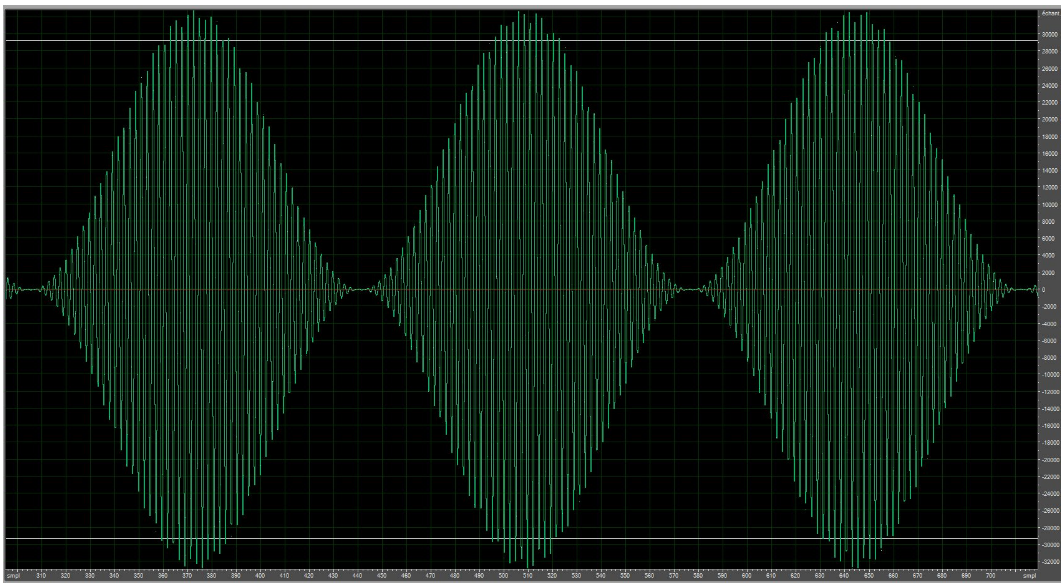

Fig 3. Blackmann-Harris modulated 18500Hz tone

While we send this infinite signal speakers (with slightly different frequencies so they don't mix), we record at the same time from the microphone (Fig 4).

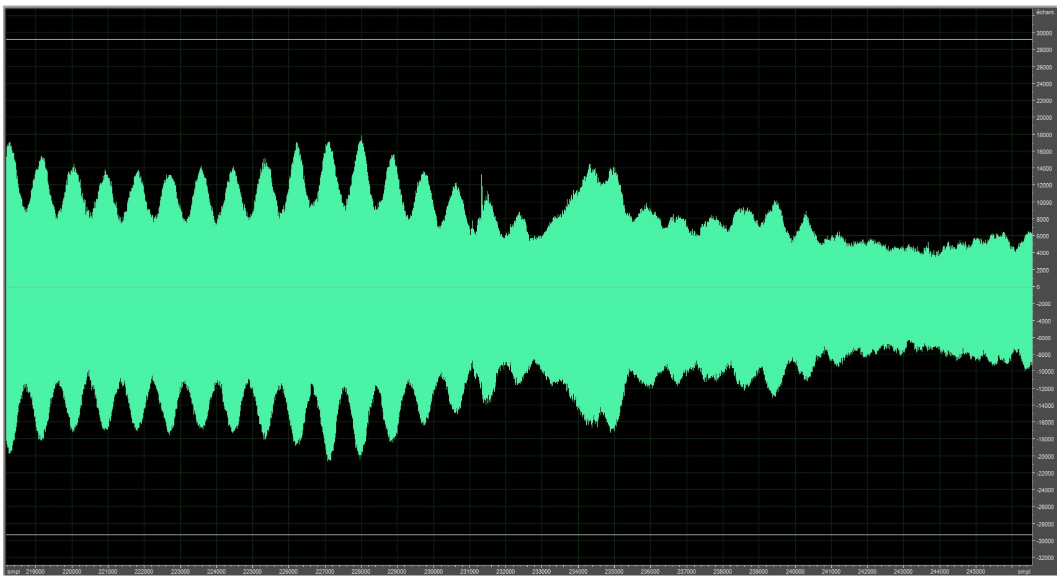

Fig 4. Recorded signal (we can clearly see impulses here).

Then all we have to do is apply a FFT on the recorded signal (using, one again, a Blackmann-Harris window for instance) so we can separate the 3 speakers, compare amplitude with the signal we send to get a rough idea of time difference, then fine tune the time difference by comparing the phase.

As we have 3 speakers, we have 3 distances, we can triangulate and recover the 3d position of each microphone in realtime.

3. Results

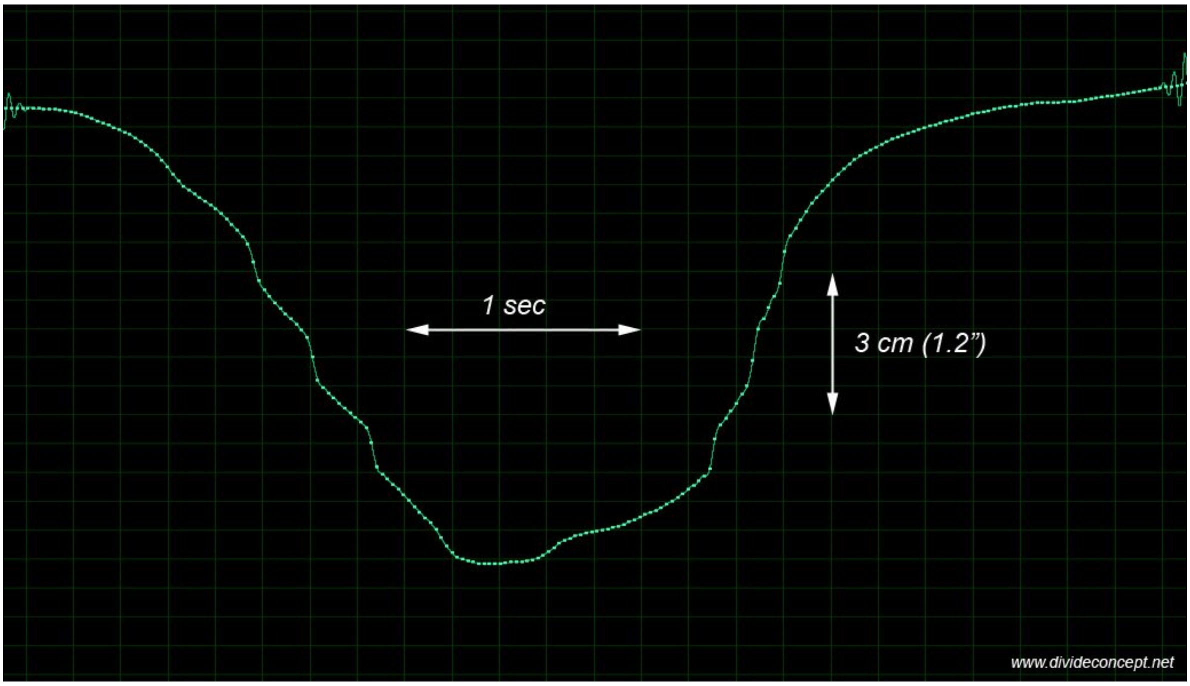

Fig 5. Recorded distance with a microphone getting closer to one speaker, then moving away

The system shows high accuracy with sub-millimeter precision, and we can check the position 50 times per second. Delay between action and response depends on the audio card, on the test system it was around one quarter second.

A video can be seen here: http://www.vimeo.com/8983229

4. Limitations and future development

As the system use air, which is highly compressible and subject to subtle changes all the time, experiences shows some spring effect when quick moves are made. However it doesn't long more than a second in most cases.