Using IIDs to Estimate Sound Source Direction

Leslie S. Smith
Department of Computing Science and Mathematics
University of Stirling, Stirling FK9 4LA, UK
lss@cs.stir.ac.uk

Abstract

Poster. Sounds recorded using a binaural head are analysed to find the azimuthal direction of a sound source. Two techniques for inter-aural intensity difference (IID) estimation are compared. In both, signals were filtered into a number of wideband logarithmically spaced frequency bands. In method 1, an estimate of the IID in each frequency band was made every 20ms, and in method 2, estimates were made only when a cluster of onsets had been found. Onsets were detected using a biologically plausible spike-based technique. IID vectors were converted to directions using estimates of the impulse response of the binaural head. The onset-based technique provides better results, particularly in reverberant environments.

1 Introduction

Locating the source of a sound is an important task for an animal. It is normally achieved using a mixture of IIDs (resulting from ears having a non-uniform omindirectional response) and inter-aural temporal differences (ITDs) (resulting from differing path lengths from the sound source to each ear) (**?**). However, sound generally reaches the ears not only from the direct path, but also from paths which include reflections. The onset of a sound always comes from the shortest, direct path. Human IIDs tend to be small below about 1000Hz, due to diffraction round the head. At high frequencies ITDs tend to become ambiguous because each ITD estimate e is actually $e \pm np$ (p is the signal period). Using the same model head (**?**), we showed that ITDs at low frequencies can be used to estimate the azimuth of sound source. Here we use IIDs measured at frequencies above 1000Hz.

In method 1, we compute the IID in each frequency band every 20ms. One problem in using IIDs for azimuthal source determination is that the instantaneous IID depends on all the paths sound has taken to reach the ears. In method 2, we examine IIDs only at onsets. Below, we describe how we measure IID vectors, how these are turned into directions, and we compare the effectiveness of methods 1 and 2. The most similar approach is in (**?**), but he uses brief pulses, whereas we are using normal utterances.

2 Methods

Sounds are played using a small amplifier and loudspeaker to the binaural recording system. This consists of two matched omni-directional microphones, placed at the outer ends of the auditory canals of a model head. This head consists of a realistic model skull with a latex covering modelling real flesh and skin with latex pinnae. The head was mounted on a simple model torso. Signals were played to the binaural recording system at the same height as the pinnae, at a distance of 1.5m. They were played at 10° intervals. After digitisation at 96000 samples/second, 16 bit, the signal was filtered into 32 overlapping bands using a gammatone filterbank with centre frequencies and bandwidths based on (**?**).

For both methods 1 and 2, an instantaneous IID value for each filtered band was calculated as follows. First the sum of the squares of the sample values inside an interval of length 5ms was computed for both the left and the right channels. Then the sum from the left channel was divided by the sum from the right channel. For method 1, IID vectors were calculated every 20 ms throughout the sound. For method 2, these values were calculated only at onsets, using only the channels in which onsets had occurred. Onset times need to be estimated accurately, with low latency, and across a wide range of signal levels. Firstly each channel signal was coded as a set of 20 spike trains, with spike probability in each spike train depending on signal strength but saturating, and with spikes generated in phase with the signal. These spike trains were fed into a set of depressing synapses on an integrate-and-fire neuron with a high leakage. This results in a low latency phase locked onset detector similar to that in (**?**). We considered as onsets only those occasions in which at least a certain number of onset spikes had occurred within a brief period of time, in both left and right channels. One problem with this approach is that at high source angles, the signal strength in the contralateral ear is low, and some onsets went undetected.

Both methods produce IID vectors. Computing azimuths requires estimation of the difference in response between the ears for the frequencies of the filterbank at different angles. The impulse response of each ear was computed using the MLS technique (**?**) at a range of azimuthal angles. This was truncated to 2 ms, as we were interested only in the effect of the head and torso, not environmental reflections. The difference between the left and right ear response was computed by taking the ratio of the power spectra of the Fourier transforms of the truncated impulse responses. The IID vector was compared with these, and the closest used to estimate the source angle.

3 Results

The figure shows the results of processing one second of male speech in a slightly reverberant environment. Method 1 underestimates the angle, and has a larger standard deviation than method 2. Method 2 overestimates the angle and discriminates poorly at large angles, partly due to the low energy of the contralateral signal. The instantaneous results at the bottom of the figure show that even single onsets can provide reasonable estimates of source azimuth. At large angles, some onsets are undetected due to the low energy of the contralateral signal. Other results (not shown) demonstrate that in highly reverberant environments, estimating IIDs only at onsets is even more crucial.

4 Discussion and further work

We have shown that using IIDs at onsets can permit sound source azimuthal estimation for sounds of short duration, even in a reverberant environment. This IID based technique should be combined with the ITD based technique in (**?**) which works at lower frequencies. Together, they should be able to substantially improve sound direction estimation (including elevation, given impulse responses at different elevations) for wideband sounds such as speech. The use of onsets makes the system emulate the precedence effect (**?**) since a secondary onset after the first one (but without any intermediate offset) will be ignored because the depressing synapses will not have time to recover. In addition, clustering onsets should allow the location of more than one simultaneous (but not simultaneously onsetting) sound source.

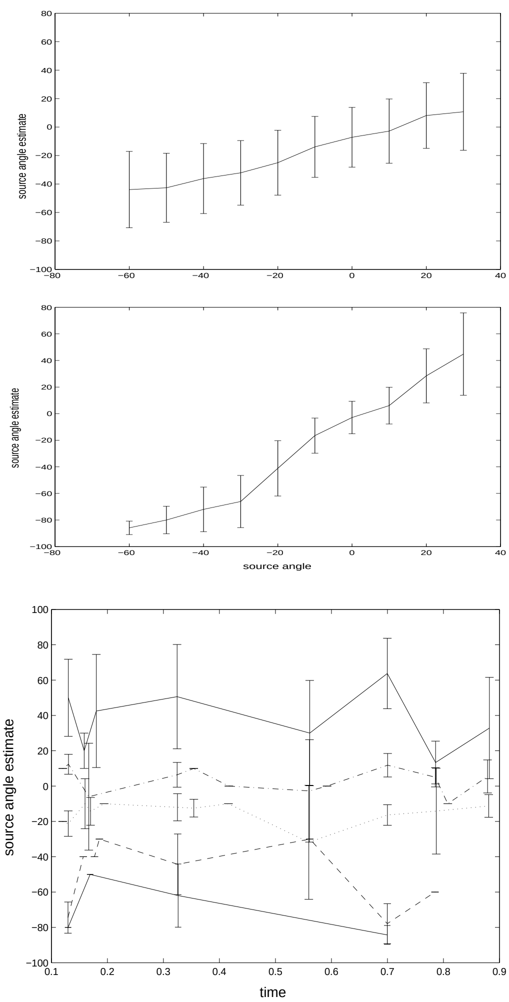

Figure 1: Result of processing a 1 second male speech signal. Top, middle: IID angles computed for varying source angles using method 1 (top) and method 2 (middle). Bottom: IID angle computed at each onset, for angles from -50 to + 30 degrees for 1 second of speech. line styles: line for -50, +30 degrees, dashed for -30 degrees, dotted line for -10 degrees and dash-dot for +10 degrees.

References

Blauert, J. (1996). *Spatial Hearing*. MIT Press, revised edition.

Macpherson, E. (1991). A computer model of binaural localization for stereo imaging measurement. *Journal of the Audio Engineering Society*, 39:604–622.

Moore, B. and Glasberg, B. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74(3):750–753.

Rife, D. and Vanderkooy, J. (1989). Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society*, 37:419–444.

Smith, L. (2001). Using depressing synapses for phase locked auditory onset detection. In Dorffner, G., Bischof, H., and Hornik, K., (Eds.), *Artificial Neural Networks: ICANN 2001*, volume 2130 of *LNCS*, pages 1103–1108. Springer.