Binaural Source Localization and Filtering

Functional Description

By:
Daniel Hauer

Project Advisor:
Dr. Brian D Huggins

Date:
October 11, 2005

This project will use binaural sound concepts to spatially filter two binaural or stereo source signals based on the perceived source azimuth angle. The source material will be prerecorded either in commercial form, as in a music CD, or a recording made specifically for the system, as in testing situations.

Stereo music encoding is a system that is familiar to anyone who has every listened to a CD, radio, or virtually any other commercial music medium. Stereo recording are based on the separation of sound information into two discreet channels, namely left and right channels. While the two channels are stored and played back separately in a stereo system, nearly every sound source (an instrument, voice, or other auditory source) is played back both in the left and right channels with varying levels between the two channels. This produces the psychoacoustic (a subjective experience of sound as interpreted by the brain) effect of the sound source emanating from a point somewhere between the two extremes of pure left and right. Stereo recordings can be accurately reproduced using loudspeakers or headphones. They are not, however accurate reproductions of the original (source) sound environment.

Binaural recording is a type of sound recording that provides a remarkably accurate reproduction of the original sound source and soundstage. This is accomplished with the use of specialized recording techniques usually involving a binaural head. A binaural head is an artificial human head with microphones imbedded in the ears. (See figure 1.) Binaural recordings aim to capture the sound received by each human ear as accurately as possible. When played back through headphones, the original soundstage is recreated almost exactly thanks to the precise coupling from headphone transducer to ear present in headphone playback systems. While binaural recording signals are completely compatible with stereo mediums, the correct binaural effect cannot be reproduced by conventional stereo loudspeakers and thus the reproduction is restricted to the use of headphones. Theoretically, an infinite number of loudspeakers placed in a spherical manner is capable of reproducing binaural sounds. While an infinite number of loudspeakers is not possible, a goal of this project is to spatially filter a binaural signal so that a large amount of loudspeakers can be used to reproduce binaural recordings with a high degree of accuracy.

The system will rely on detecting three specific relationships between the two source channels. These relationships are known as the inter-channel time difference (ICTD), the inter-channel level difference (ICLD). As the names suggest, the ICTD is simply the difference in time in the arrival of a particular auditory event between the two binaural or stereo channels. Physically, this is equivalent to sound waves arriving at a listener's ears at slightly different times due to the finite speed of sound traveling in air. Similarly, the ICLD is the difference in level for an auditory event between the two channels. Intuitively, a sound originating from the side of a listener's head will be louder in the nearest ear as compared to the away-facing ear. The ICLD is a measurement of this level difference. Also used by the system is the inter-channel coherence (ICC). Simply put, this is a measure of how alike the two signal channels are i.e. how much common auditory material they share. The system will simulate the binaural source localization abilities of the brain when the source material is presented via headphones.

Thus, in this case, the ICTD, ICLD, and ICC can be considered equivalent to the interaural time difference (ITD), the interaural level difference (ILD), and the interaural coherence (IC), respectively. This connection can be made because of the aforementioned very precise coupling of the audio signal to the ear. The method used will be based on a paper by Harald Viste and Gianpaolo Evangelista titled "Binaural Source Localization" as presented to the 7th International Conference on Digital Audio Effects. (http://dafx04.na.infn.it/WebProc/Proc/P_145.pdf)

Figure 2 shows the overall block diagram of the system. As shown, the inputs to the system include the two stereo or binaural channels. Other inputs will include user inputs to the system to specify system parameters and performance. These inputs will be described subsequently. The inputs are fed into a processing unit, which may be a PC or a DSP chip. An output of the system is labeled "Visual Display" in figure 2 and will be implemented on a PC screen in the form of a polar graph of sound intensities. This output may be included in the final system or just used as a debugging and proof of concept tool. The main outputs of the system will be the spatially filtered signals. These signals will correspond to separate azimuth angle ranges in relation to the simulated listener. Figure 3 illustrates this spatial filtering. The two channel audio presented to the listener via headphones is spatially filtered into the depicted regions. These regions are polar symmetric and correspond to the perceived direction of the source material. The number of regions will be dictated by the user input. In every case though, the azimuth separation between region boundaries is (360/N) º for a binaural source, or (180/N) º for a stereo source, where N is the dictated number of output channels. That is, a binaural recording will be filtered in based on a circular soundstage around the listener while a stereo recording will be filtered with a semicircular soundstage oriented directly in front of the listener as in figure 3.

The aforementioned user inputs to the system will include the type of recording (binaural or stereo) to be processed and the number of channels to be decoded from the source signals. Other user stipulations may include a so-called spatial Q factor that is similar to Q factors in frequency based filers. This parameter will dictate how sharply the transition from juxtaposed regions occurs. In this sense, the parameter may be considered a spatial slope.

Ultimately, the filtered output channels may be fed to separate loudspeakers placed in accordance to their input signal's spatial assignment. In this way, a more accurate soundstage can be recreated using conventional transducers.

Figure 1 - The Neumann KU 100 dummy head binaural stereo microphone.

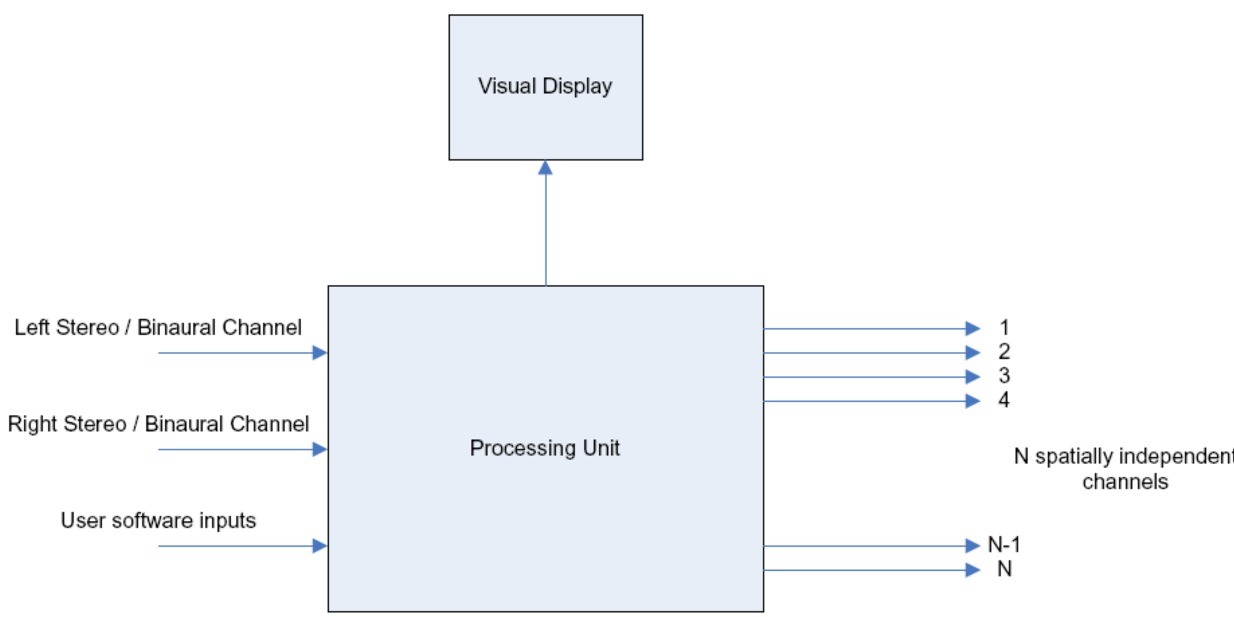

Figure 2 – Overall system block diagram.

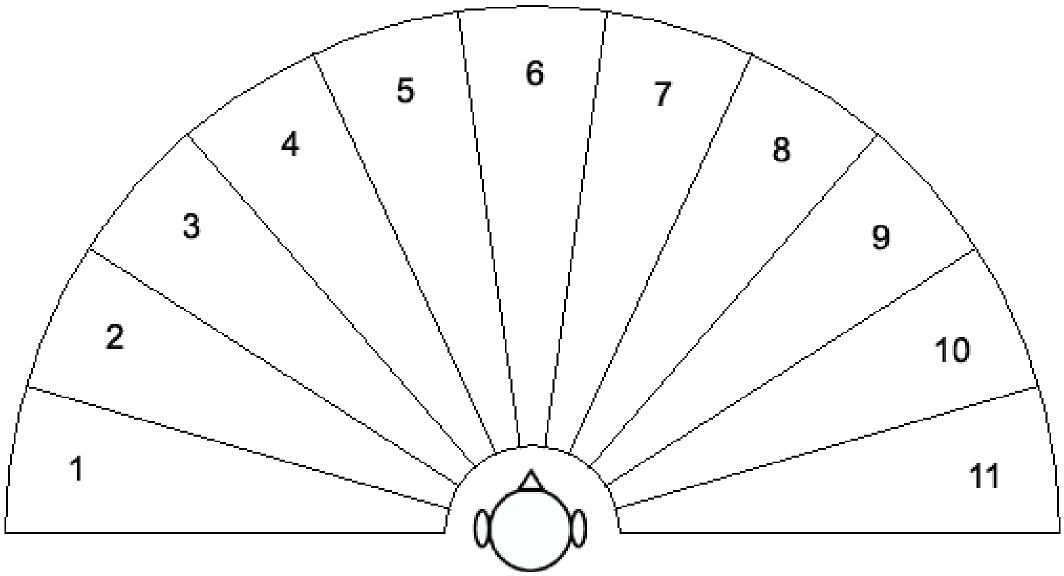

Figure 3 – An example of spatial filtering. Here, a stereo soundstage is filtered into 11 distinct regions or zones.