

New Csound Opcodes for Binaural Processing

Brian Carty

LAC '08

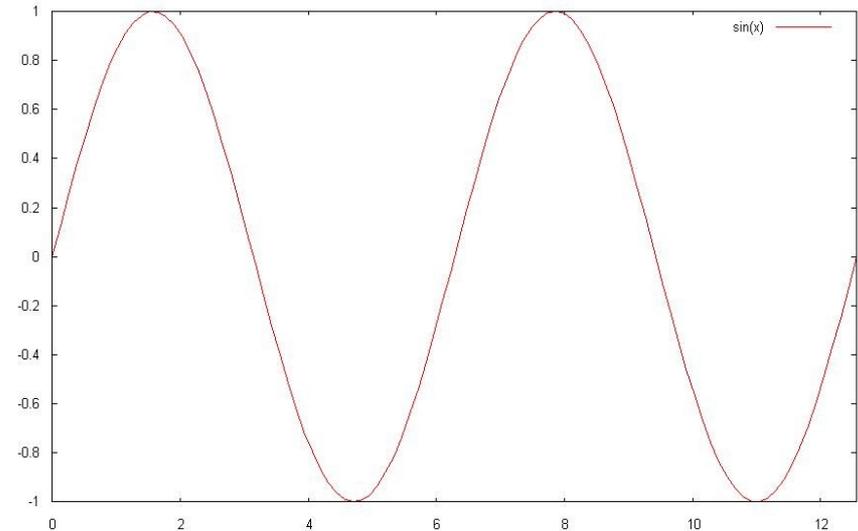
Cologne

Goal and Definitions

- Goal: A comprehensive, generic, accurate and efficient toolset for binaural artificial recreation of audio spatialisation.
- A number of Csound opcodes, using a Head Related Transfer Function based approach, are presented.
- Novel approaches (specifically methods using phase truncation and functionally derived itd respectively), as well as a method based on established digital signal processing methods (minimum phase plus delay) are implemented.
- *Sound localisation* deals with how and why we can locate sound sources in our spatial environment.
- *Sound spatialisation* defines how sound is distributed in this environment.

Background

- Sound travels in waves created when a vibrating source disturbs the air.
- Parameters of simple periodic sinusoidal wave: magnitude, frequency and phase
- ‘Real world’ sounds are made up of combinations of simple periodic sounds, with different frequencies, magnitudes and phases.
- Upon reaching the ears, these component frequencies trigger different areas of the Basilar Membrane which consequently transmits electric signals to the brain to be perceived as sound.
- The *frequency domain* represents the frequencies of the simple sine waves present in a sound, their relative amplitudes/magnitudes and phases.



Sound Localisation: An Introduction

- *Binaural hearing* is the term given to listening with two ears rather than one, and is the main factor involved in sound localisation
- One such binaural indication of a sound's spatial characteristics is *Interaural Time Difference (ITD)*: the name given to the time it takes a sound to reach one ear after it has first reached the other.

Interaural Time Difference

Interaural Intensity Difference

- Interaural *Intensity* Difference (IID) uses varying respective intensities of a signal at each ear to locate source sounds.
- It is generally accepted that interaural time and intensity differences work together to provide a well-defined spatial image, with ITD working best for low frequencies and IID for high.
- Monaural information (independent information from one ear) also plays an important role in sound localisation.
- The pinna and concha both have a non-linear frequency response over the audible spectrum, altering incoming sounds.

HRTFs

- *Head Related Transfer Functions* (HRTFs) are functions that describe how a sound from a specific location is altered from source to inner ear.
- The frequency domain process of simulating an auditory location using HRTFs can be summarised thus:
 - Record the impulse response of the left and right ear for the desired point in space.
 - Analyse the frequency content of the sound you wish to spatialise.
 - Impose the HRTF for the left and right ears on the sound (boost or attenuate and delay the frequencies contained in the input in accordance with how the ear treats the appropriate frequencies), using the process of convolution.
 - Finally, play the signals derived to the left and right ears respectively, **on headphones**.
- In summary: find out how the ears treat all frequencies for your desired location, and treat the frequencies contained in your input sound in the same way.

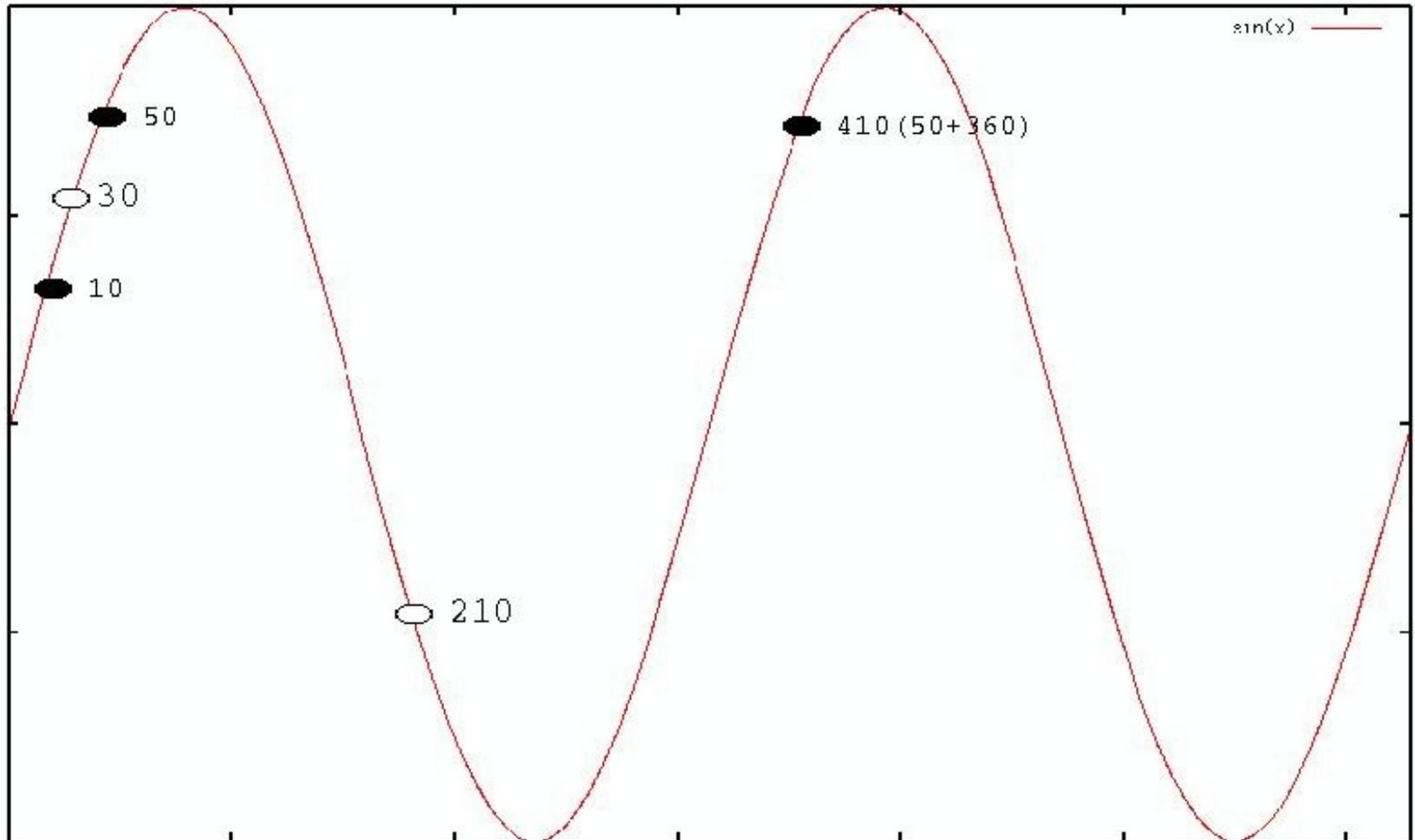
HRTFs Continued

- HRTF data sets are typically measured at discrete, equidistant points around a listener or dummy head.
- As the physiology of everyone's ears is different, HRTFs vary considerably from subject to subject, but a generalised data set gives good results.
- If a location is required that has not been measured, or if a sound is required to move smoothly from one location to another, some kind of averaging or interpolation must be done.

HRTF Interpolation

- Interpolation in the frequency domain gives better results.
- Magnitude can be interpolated linearly.
- Phase interpolation is more complex, as phase is a periodic quantity.
- Uncertainty arises when trying to interpolate phases, as a phase value can be +/- any amount of full cycles.

Phase Interpolation Problem

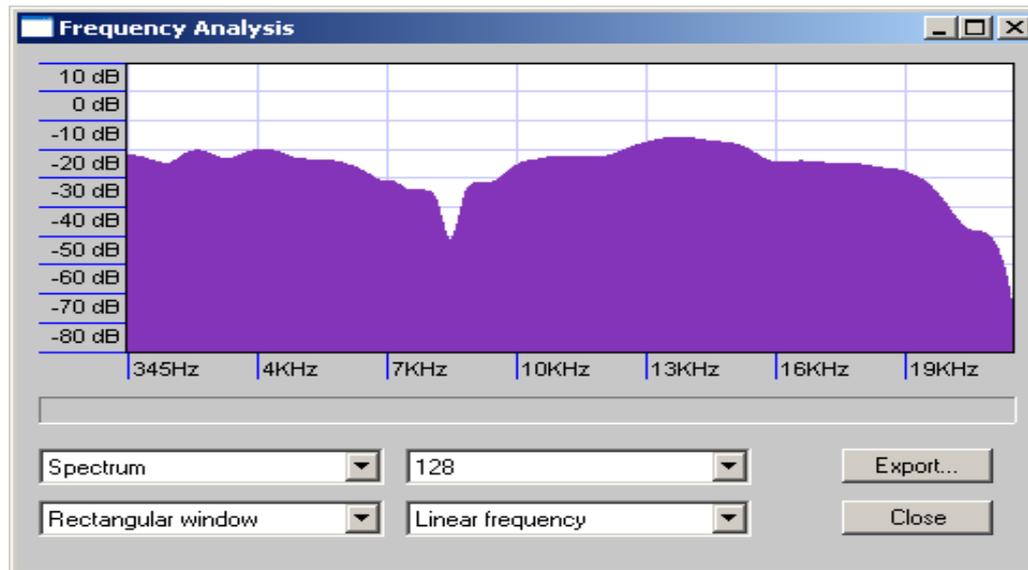
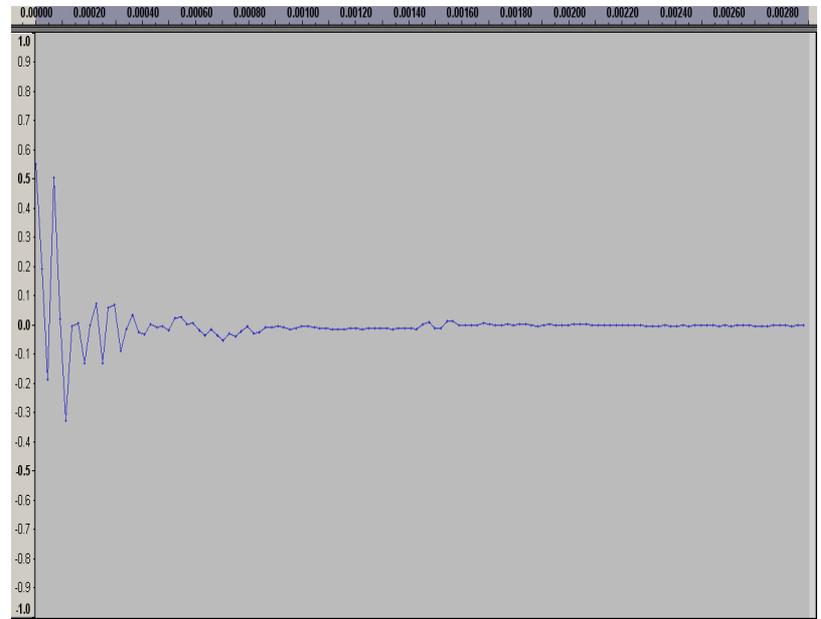
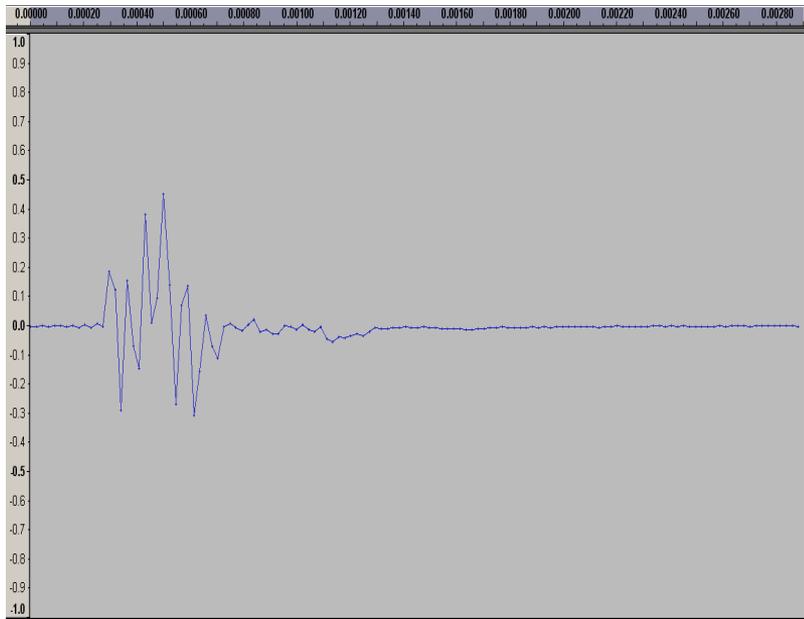


Minimum Phase

- Any rational system function can be broken into a minimum phase and an all pass system.
- The magnitude of the minimum phase all pass decomposition is represented solely by the minimum phase system and the phase is reconstituted by both the allpass and minimum phase representations.
- A unique and, in this case, extremely useful property of minimum phase systems is that phase values for each component frequency can be derived from the corresponding magnitude values.

Minimum Phase HRTFs

- The auditory system approaches minimum phase.
- Therefore HRTFs can be thought of as minimum phase filters, with a linear allpass component, which can be implemented using a simple delay line.
- A pair of HRTFs (for the left and right ears) can consequently be broken down into 3 parts: a minimum phase representation of each empirical HRTF pair (left and right ear), and an interaural delay.

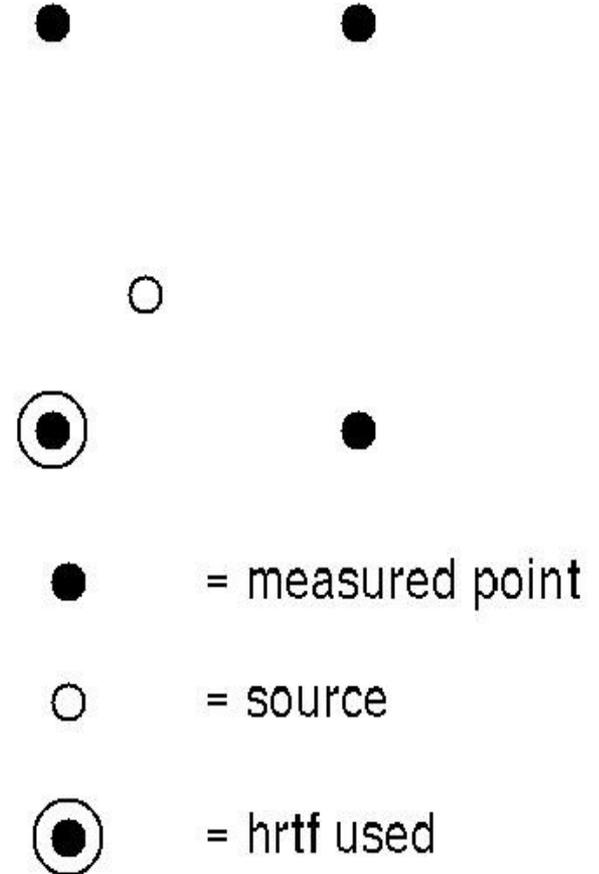


Minimum Phase Implementation

- The process of HRTF interpolation thus involves analysing each HRTF pair to find the relevant interaural delay and reducing them to minimum phase representations.
- The minimum phase magnitude values and extracted delay can then be linearly interpolated. Interpolated minimum phase phase spectra can be derived from interpolated magnitude spectra.
- Overall studies suggest that minimum phase plus delay models are adequate for most source locations, although the approximations involved in assuming the system is minimum phase have been noted.
- The minimum phase method employs complex digital signal processing of the HRTF data, and is quite computationally expensive.
- Therefore, novel alternatives are suggested, using data more directly and avoiding the minimum phase assumption.

HRTFer

- HRTFer, the current Csound opcode for HRTF based binaural localisation, provides accurate spatialisation for static locations which correspond exactly to HRTF measured points.
- However, if a static point is required that has not been measured, the system simply chooses the nearest point.
- A dynamic, rather than static source will skip from one nearest measured point to the next along the user defined trajectory. This staggered movement causes irregularities in the output.
- Crossfades are suggested by the authors.



Magnitude Interpolation, Phase Truncation

- The first new opcode introduces an interpolation algorithm which works by storing the four nearest HRTF values to the desired location, left and right below and above.
- Linear interpolation of the magnitude values is performed.
- The nearest measured phase value is used for intermediate filters.
- Crossfades are performed when new, nearer *phase* values are available.
- The user can define the length of these crossfades.

Functional Model

- A second approach interpolates magnitude as before, and attempts to model phase assuming the head is a sphere.
- Mathematically, the ITD for a particular source location, assuming a spherical head can be defined as:

$$\frac{r(\theta + \sin(\theta))}{c} \cos(\phi)$$

where r is the head (/sphere) radius, θ is the angle and ϕ the elevation of the source.

- A low frequency, frequency dependent scaling factor is introduced as a more complete solution.

Functional Model Continued

- Essentially, frequency dependent ITD is extracted from the empirical HRTFs for the low frequency bands of interest.
- These new values are then used as frequency dependent scaling factors in the synthesis of the phase spectrum for the desired HRTF.
- This model provides an accurate average low frequency ITD for this particular dataset, and a steady Woodworth based ITD for higher frequencies.
- This provides an accurate interpolation algorithm for static sources.
- An STFT based process is used for dynamic sources.

Csound Implementations

- Three new opcodes have been designed.
- The first allows phase truncation (with user definable crossfades) or minimum phase binaural processing.
- The second and third are based on the functional model, one for the more efficient static spatialisation and one for dynamic sources. These opcodes allow choice of spherical head radius for ITD calculation, and STFT overlap for dynamic trajectories.
- All models allow sampling rates of 44.1, 48 and 96 kilohertz.
- Data files containing the HRTF data at the appropriate sampling rate, as well as minimum phase delay data are also required.
- Despite the addition of magnitude interpolation, and algorithms for appropriate phase representation, the new, optimised opcodes perform favourably in comparison to the HRTFer opcode.

Conclusions

- Minimum phase requires data preparation and knowledge of complex digital signal processing.
- Casual listening tests show there is often an audible discrepancy between minimum phase and empirical data convolution for musical and test sources, although localisation is good for both.
- Phase truncation output appears to give a result more consistent with the empirical dataset as a whole.
- Adding non linear low frequency scaling factors to a spherical head model will reintroduce some of the finer phase detail involved in the non uniformly spherical shape of the head, the pinnae and the torso.
- The binaural processing capabilities of Csound have thus been updated and improved, using existing and novel approaches. Smooth, artefact free dynamic and static binaural processing is now realisable using the various novel techniques described above.