# A BIO-MECHANISM FOR THE ENHANCEMENT OF PRESENCE AND DIRECTIONALITY CUES IN SOUND PERCEPTION

**PACS:** Stereophonic reproduction (43.38 Vk)

Diego Gonzalez, Davide Bonsi, Domenico Stanzial
The Musical and Architectural Acoustics Laboratory, FSSG-CNR
Fondazione G. Cini, Isola di san Giorgio Maggiore,
I-30124-Venezia, Italy
Tel: ++39 041 5207757
Fax: ++39 041 5208135; E-mail:
diego.gonzalez@cini.ve.cnr.it, davide.bonsi@cini.ve.cnr.it, domenico.stanzial@cini.ve.cnr.it

## ABSTRACT

The two main detection mechanisms for azimuthal sound localisation are the interaural time difference (ITD) and the interaural level difference (ILD). The action mechanisms of these cues are supposed to be independent each from the other and acting at different frequency regions. However, recent research on insect hearing physiology suggest that interaural time differences in the arrival of the acoustic wave are also responsible for interaural level differences thanks to a sort of mixed mechanism. On this basis we propose a scheme in which interaural time differences produce also an interaural level difference adjusting itself dynamically. This technique can be used for enhancing stereo recordings and also for artificially controlling the azimuthal localisation of arbitrary sound sources.

## INTRODUCTION

Interaural time differences (ITDs) and interaural level differences (ILDs) are the two main cues for spatial localisation of sounds in the horizontal plane (azimuthal localisation) [1].
ITD and ILD are supposed to be acted by two evaluative mechanisms that, to a certain extent, act independently each from the other. The first mechanism is able to evaluate time shifts in the fine structure of the incoming stimuli and represents the main mechanism for localisation of sounds containing frequency information in the band below 1.6 kHz. The second mechanism evaluates azimuth position by interpreting sound pressure differences and probably also envelope time shifts. Its frequency operation range is mainly above 1.6 kHz being complementary to the first mechanism.
The precedence effect, which consists in the ability to emphasise the leading sound in making a localisation decision [2] puts into evidence the presence of inhibitory phenomena in binaural processing. However, the precedence effect is activated with a relatively long latency, its effect being directed mainly to the cancellation of delayed echoes due to ambience reflections. In fact, typical activation times are above 1 ms, which is about the upper limit for lateral localization by means of ITD cues (defined by the travel time of a sound wave through a distance equal to the ears separation).
Several models of binaural time-amplitude interaction have been studied which are extensions of the cross-correlation approach first proposed by Jeffres [3]. These are the “multiplicative intensity weighting pulse” of Stern & Colburn [4], the “separate cross-correlation weighting” of Sayers and Cherry [5] and the “lateral inhibition model” of Lindemann [6-7] (see ref. [8] for a review on models of binaural interaction). In particular, the Lindeman’s extension can describe several observations on the precedence effect as, summing localisation, the law of the first wave front, and echo suppression. However, it has not yet been possible to determine the physiological plausibility of the inhibitory mechanism proposed, which is supposed to be of central origin.

On the other hand, recent studies on insect hearing physiology have suggested the existence of peripheral inhibitory mechanisms with very short activation times if compared with those of the precedence effect [9]. In this time scale, interaction between ITDs and ILDs has been pointed out. In fact, the experiments of Miles et al. [10] have shown that in the ears of the parasitic fly *Ormia ochracea*, ITDs can be enhanced allowing the detection of time intervals of the order of a few microseconds which cannot be detected with usual neural circuitry. Moreover, arrival time differences of transients are converted in level differences. This last effect can be viewed as a proportional precedence effect for times when localisation mechanisms are still active.
A similar periphery inhibitory mechanism is at least theoretically possible in mammals by electromotive activation of controlateral outer hear cells through efferent nervous pathways. On this basis, we propose an audio enhancement scheme which converts ITDs between envelopes on ILDs which adjust itself dynamically.

**THE MODEL**

Fig.1 reports a block diagram of the processing scheme implemented in Symulink®. Two envelope detectors monitor left and right channels. Let us suppose that a transient arrives first to the right channel, thus, the right envelope detector exits a positive value. This positive value enters a first order dynamical subsystem whose output reaches asymptotically the input value following an exponential law

$$x_{out}(t) = [F_0 - x_{in}(t_0)].e^{-kt}$$

where $F_0$ is the initial value at the input of the subsystem and K its time constant which depends on the feedback gain (block Gain 1 in Fig. 1-b).

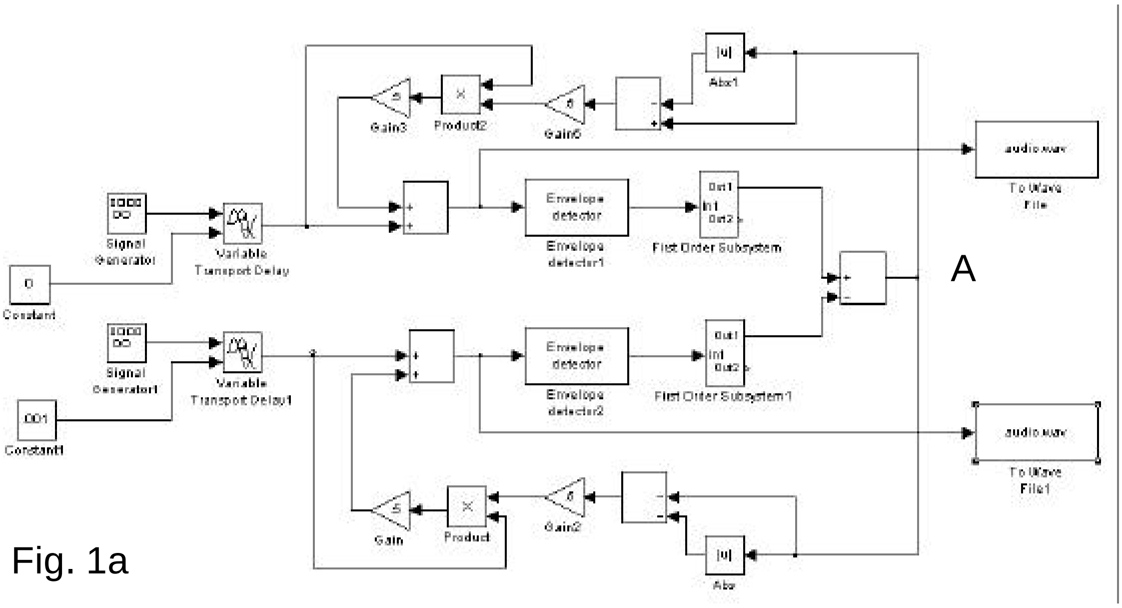


Fig. 1a

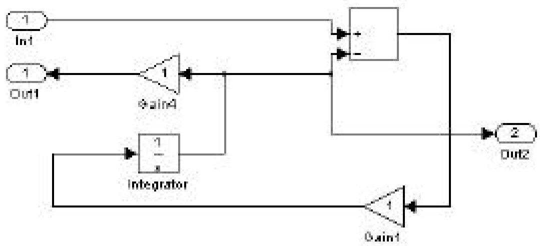


Fig. 1b

Fig. 1 a,b – a) Block diagram of the envelope ITD to ILD converter implemented by means of Symulink blocks; b) block diagram of the first order dynamical subsystem.

The same processing is applied to the left channel. Consequently, the output of the left dynamical subsystem remains zero until the arrival of the signal to the left channel. At this moment, a positive value exits the difference block (point A in Fig. 1-a). The arithmetical net converts this value, which depends on the time delay between channel signals, in a zero for the right channel and minus the value of the difference for the left channel. This difference is then multiplied by the corresponding input, scaled with an adjustable gain, and added to the same input. In this way, for the conditions before described at the arrival of the transient at the left channel, the right channel output remains equal to the input and the left channel is attenuated by an initial factor depending of the delay between channel signals. After this time, if the signal at the left channel has an amplitude equal to that of the right channel, a positive value exits the left dynamical subsystem and the difference at the point A (Fig. 1-a) starts to decrease. Consequently, also the attenuation of the left channel starts to decrease reaching asymptotically a zero value.

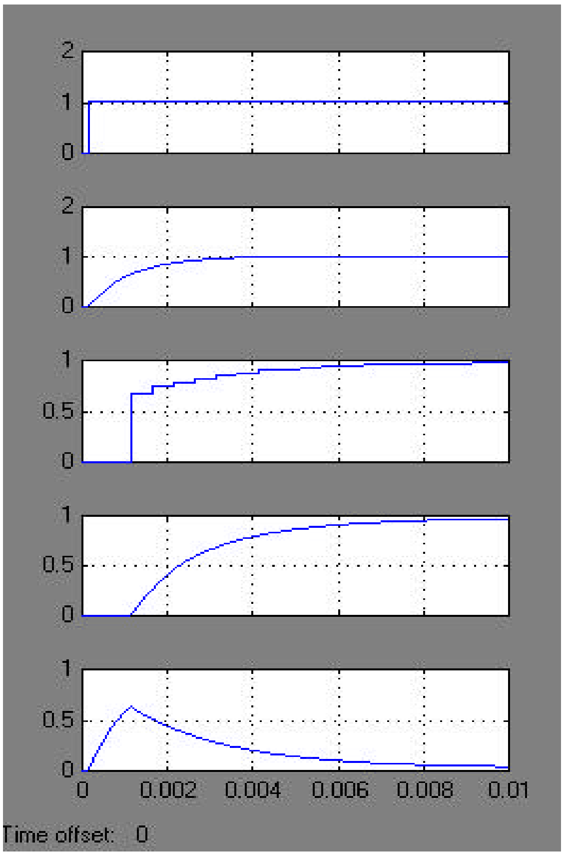


Fig. 2a

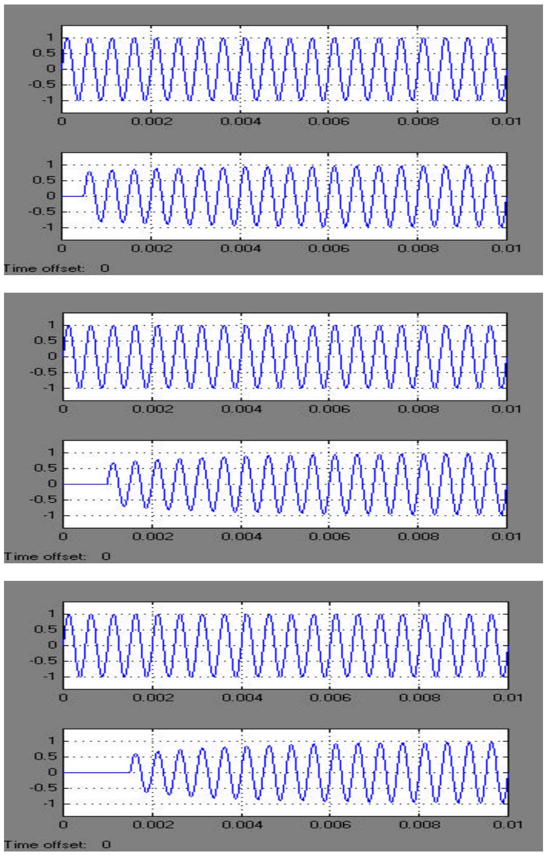


Fig. 2b

Fig. 2 a,b – a) From top to bottom, input and output of the right first order dynamical subsystem, input and output of the second subsystem and difference at point A (see block diagram, Fig. 1a); b) processing of a sinusoid of 2 kHz with 0.5, 1 and 1.5 ms of delay between channels.

The insertion gain (see Symulink® block gain for the left channel in Fig. 2-a) control the initial value of the attenuation and also the recovering time constant $K_{eff}$ because of the in-the-loop placement of envelope detectors. In Fig. 2-a we can see the time evolution of all the control signals (input-output of the subsystems and their difference for a stereo signal consisting of a 2.000 Hz. sinusoid at the right channel and a delayed version of it at the left one. In Fig. 2b we can see also the effect of the processing for three different values of the delay: .5, 1 and 1.5 ms (this stimulus is also simulated in the block diagram of Fig.1-a). The rise time constant is 1 ms (determined by k=1000 in both subsystems) and the insertion gain is fixed at .5 . For these parameter values and a delay of 1 ms the initial attenuation is about 3 dB. The value of the insertion gain must be less than unity in order to ensure the stability of the loop. In Fig. 3 a, the effect of an increment of the insertion gain to a value of 0.9, is shown.

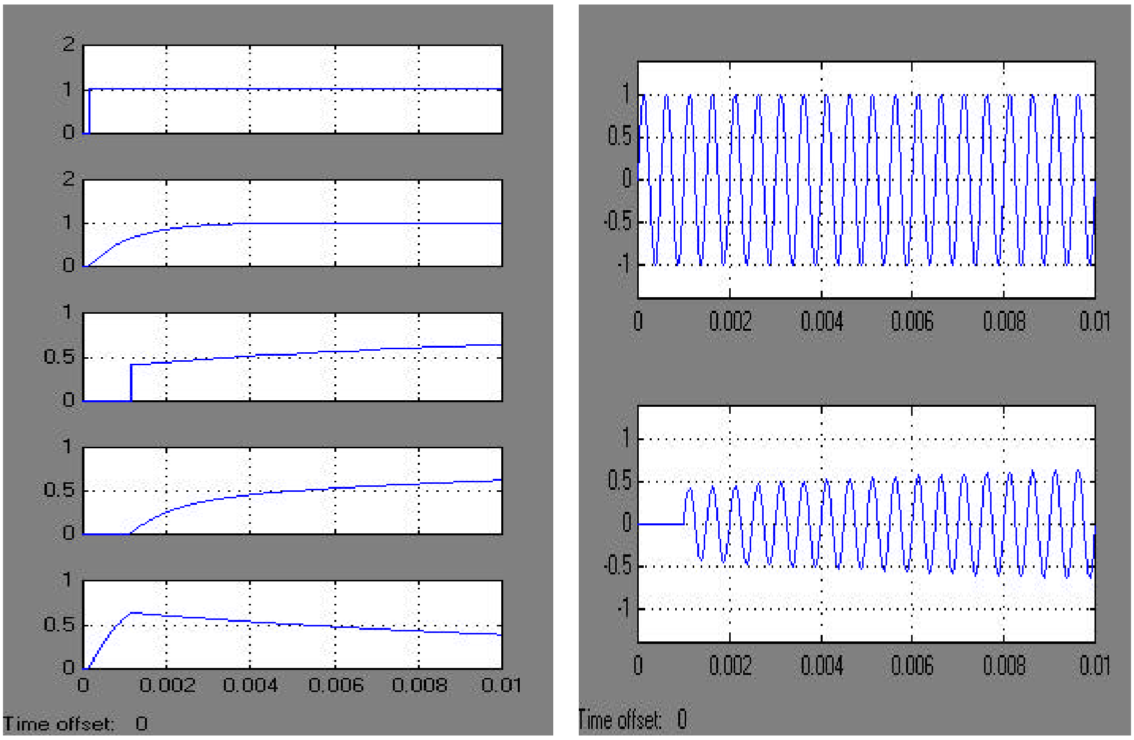


Fig. 3 – Evolution of control signals and result of the processing for an insertion gain of 0.9 and a delay between channels of 1 ms.

**Stereo Enhancement**

Stereo recording techniques are mainly of two types, those utilising coincident or near coincident microphonic arrays (which rely mainly on ILD information due to directionality properties of the array) and those using spaced-microphone arrays (which record mainly ITD information) [11]. For the last technique there is a strong tradition of using spaced omnidirectional microphones. The acoustical images so generated do not elicit the high definition localisation of those produced by a coincident pair of microphones due mainly to the absence of ILD cues. A method for enhancing the stereo effect in this kind of recording consists in placing the microphones at greater distances than those characterising the human ear separation. The typical separation between microphones is 0.6 meters, greater distances producing the effect of “hole in the middle”. This enhancement technique can be explained using the diagrams of Franssen [12] where (see Fig. 4) the position of the phantom source can be roughly determined by the combined effect of ITD and ILD cues.

In Fig. 5 we see the result of processing a stereo recording of a clarinet played in an anechoic chamber obtained with two B&K omnidirectional microphones characterised by a flat frequency response and spaced 0.67 meters. For this distance we can implement an ILD enhancement reaching a maximal attenuation level of 6 dB with impulsive transients.

The application of the ITD to ILD conversion proposed for stereo recordings can enhance the localisation process giving a more realistic feeling. This is due mainly to the generation of ILDs in the high frequency range and thus to the recreation of the head shadow effect which is present in a normal hearing localisation task (also in a live musical performance). Moreover, the exponential recovering of the original level can enhance the perceived quality of the sound without noticeable degradation of the localisation performance. In fact, once the azimuth of the sound source has been localised at the beginning of the transient, the binaural “sluggishness” ensures its perceptual continuity when new information is added to the controlateral channel;

this added information, however, can contribute positively to the overall quality of the perceived sound.

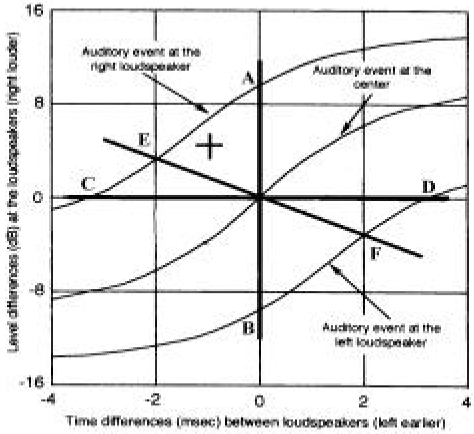


Fig. 4 – Frasser's diagram showing the tradeoffs between amplitude and delay times for localisation on the stereo soundstage. Extreme curves represent perception completely on one side (left or right ) and the central one represents centred auditory events. The vertical A-B line represents the typical excursus of auditory events as recorded by a coincident microphone array. The horizontal GD line instead, represents approximately such of the spaced-pair with omnidirectional microphones (there is in fact a level difference with omnidirectional microphones due to the distance to the different distances to the sound source). The E-F line represents one of the simulations performed with a separation of 67 cm (around 2 ms of ITD) and an introduced maximal attenuation of around 3dB (modified from ref. [11]).

This wealth of information on both channels is perhaps the clue for the preference of stereo recordings with spaced arrays of omnidirectional microphones expressed by many musicians and sound engineers. In this sense, in order to take profit of the enhancement, the best recording configuration seems to be the one with near coincident omnidirectional microphones with a spatial separation close to the normal human inter-ear distance. In this way, natural ITD cues are created, and the two microphones record nearly identical amplitude signals with the same frequency content. The extra ILDs necessary for enhancing the localisation range (see the Frasser's diagram of Fig. 4) can be artificially introduced by the described processor.

**CONCLUSIONS**

We have proposed a method for the enhancement of stereo signals, which convert envelope ITD cues in self-adjusting ILD changes, which follow transient attacks. The method, as showed, can be utilised for enhancing stereo signals recorded with omnidirectional spaced-microphones.
The method can be modified in different ways, for example, including appropriate filters in the negative feedback path, so limiting the band of frequencies for which the enhancement is produced. A tracking comb filter may consent the enhancement of a single musical instrument (which, for example, needs a localisation enhancement due to wrong recording).
Moreover, the processing can be acted separately at several frequency bands (for example by means of a cochlear filter bank able to separate the input signal in critical bands), which can be ultimately added to obtain the processed signal. Another different kind of application may be the enhancement of spatial cues for hearing impaired subjects. The original strategy of the *Ormia ochracea* fly consents, in addition to the ILD enhancement, an ITD enhancement. In this way, the fly can localise sounds produced by the cricket host in the frequency range around 5 kHz which produce ITDs on the order of about only 1 microsecond. The mechanical model proposed in reference [10] consists of a pair of coupled differential equations of the second order. Our scheme is equivalent to a pair of switched unidirectional coupled first order differential equations (exponential solutions) and may be very interesting to explore the consequences for human perception of a processing based on a similar scheme but with a model of the second order (in this case stability problems can be significant due to the possibility of self-excited oscillations

and chaos). ITD enhancement can also be applied for electronically assisted resolution of very near sound sources and also for the recording of ILD cues in near-coincident stereo microphones arrays. Further work is being performed regarding the psychoacoustical characterisation as a function of parameters setting and some of the possible extensions related to the enhanced localisation of isolated musical sounds are being developed.

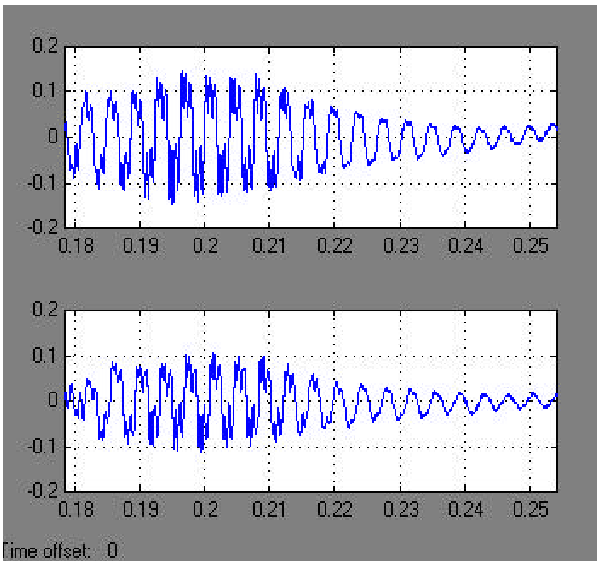


Fig. 5 – Enhancement of a clarinet note emitted from an azimuthal angle of 22 degrees from the right. The parameters of the processor are the same as in Fig. 2.

**BIBLIOGRAPHICAL REFERENCES**


[1] Blauert J., *Spatial Hearing, the Psychophysics of Human Sound Localization,* MIT Press, Cambridge, (1999).
[2] Grantham D.W., Spatial hearing and related phenomena, in *Hearing*, *Handbook of Perception and Cognition,* pp. 297-339, edited by B.J.C. Moore, Academic Press, London (1995).
[3] Jeffres L.A., A place theory of sound localization, *Journal of Comparative and Physiological Psychology,* 41, pp 35-39 (1948)
[4] Stern R.M. jr. and Colburn H.S., Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position, *Journal of the Acoustical Society of America*, 64, pp. 127-140, (1978).
[5] Sayers B.M. and Cherry E.C., Mechanism of binaural fusion in the hearing of speech, *Journal of the Acoustical Society of America*, 29, pp. 973-987, (1957).
[6] Lindemann W., Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals, *Journal of the Acoustical Society of America*, 80, pp. 1608-1622, (1986).
[7] Lindemann W., Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wavefront, *Journal of the Acoustical Society of America*, 80, pp. 1623-1630, (1986).
[8] Stern R.M. and Trahiotis C., Models of Binaural Interaction, in *Hearing*, *Handbook of Perception and Cognition,* pp. 347-386, edited by B.J.C. Moore, Academic Press, London (1995).
[9] Whitlow W.L.Au, Some hot topics in animal bioacoustics, *Journal of the Acoustical Society of America*, 101, pp. 2433-2441, (1996).
[10] Miles R.N., Robert D. and Hoy R.R., Mechanically coupled ears for directional hearing in the parasitoid fly *Ormia ochracea*, *Journal of the Acoustical Society of America*, 98, pp. 3059-3070, (1995).
[11] Eargle J., *The Microphone Book*, Focal Press, Boston, (2001).
[12] Franssen, *Stereophony*, Philips Technical Bibliography, Eindhoven, The Netherlands, (1964).