SINGLE-ENDED SPATIAL ENHANCEMENT USING A CROSS-COUPLED LATTICE EQUALIZER

Robert C. Maher

EuPhonics, Inc.
4840 Pearl East Circle, Suite 201E
Boulder, CO 80301-6115
rmaher@euphonics.com

ABSTRACT

A stereophonic enhancement system is described which expands the perceived width of the stereo sound image. The system accepts electrical audio signals comprising a two-channel (left and right) stereo pair and produces enhanced left and right stereo signals for use with conventional two-channel audio recording and playback systems. The system includes control circuitry which monitors the dissimilarity of the left and right input signals and optionally the dissimilarity of the left and right processed output signals. An all-pass decorrelation subsystem can also be included for mono-to-stereo conversion of monophonic input signals prior to the spatial enhancement system.

1. INTRODUCTION

Spatial enhancement processes are intended to produce the aural impression that the stereo sound field has become larger, and optionally to simulate a stereo effect from a monophonic recording. These features are useful in order to simulate a more spacious and natural sonic impression than can ordinarily be obtained via the small and closely spaced pair of speakers frequently found, for example, in multimedia personal computing systems.

The introduction of two-channel stereo reproduction systems in the consumer marketplace in the 1950's caused a growing interest in mono-to-stereo conversion schemes that would create pseudo-stereo signals from a pre-existing monophonic recording. Several well-known methods include [1] (a) sending the mono signal directly to one output channel while sending a slightly delayed or phase shifted version to the other channel, (b) sending a low pass filtered version of the mono signal to one channel and a high pass filtered version to the other channel, (c) sending a comb filtered version of the mono signal to one channel and a version processed by a complementary comb filter to the other channel, and (d) creating an incoherent pair of output signals by passing the mono input signal through separate channels of a stereo reverberation system.

Spatial enhancement systems may either be *single-ended* or *complementary* [2]. Single-ended systems act on a standard stereo audio signal pair without the cooperation of the audio producer, while complementary systems are designed to utilize information that was specifically encoded into the audio stream for enhancement purposes, e.g., the surround sound technology licensed by Dolby Laboratories. Some audio enthusiasts include a single-ended spatial enhancement device (typically a reverb unit) as part of the home audio reproduction system.

In addition to the standard delay and reverb techniques, a variety of spatial enhancement methods for two-channel (L and R) stereo recordings have been proposed that are similar to the sum and difference (L+R, L-R) methods of Blumlein [3]. Blumlein discovered that altering the level of the difference signal in a stereo recording was useful in changing the perceived stereo image. This technique is also widely used in the middle-side (MS) microphone arrangement.

Single-ended spatial enhancement systems typically introduce sonic artifacts that are characteristic of the particular processing techniques that are employed. For example, sum and difference systems often produce an imbalance between the bass and treble frequency ranges: too much bass if L+R dominates, too much treble if L-R dominates. Furthermore, attempts to obtain an expanded sound field may produce noticeable "phasiness" in the aural response. These spectral effects are undesirable and often result in dissatisfied users.

Therefore, a new single-ended spatial enhancement technique is proposed in this paper. This new technique is designed to avoid the strong spectral coloration that is frequently noted as a shortcoming of existing enhancement systems. The system can operate automatically to follow changes in the input signal characteristics. It is also computationally efficient and suitable for implementation in a multimedia personal computer system.

2. ENHANCEMENT STRATEGY

There are a variety of approaches for performing single-ended spatial enhancement. As mentioned, many of the published methods operate on the stereo sum and difference signals. Although the sum and difference approaches can be effective, they often suffer from noticeable spectral coloration and the creation of unnaturally strong phantom images due to the explicit presence of inverted components in the opposite channel (-L in right output and -R in left output). The inverted components are helpful in producing a diffuse sound field, but their level needs to be controlled carefully. The investigation leading to the proposed

spatial enhancement method was intended to avoid the shortcomings of the sum and difference approach.

2.1 Characteristics of Broadened Sound Source Reproduction

It is expected that a small or compact sound source recorded in a non-reverberant environment will be reproduced in a stereo recording at a particular perceived location in space. For small sound sources located in front, the listener is typically able to at least identify whether the source is to the left or to the right of center. The resulting stereo sound field has a relatively precise relationship between the phase of the signals in the left and right channel. If the sound source is large or if several spaced sound sources are present, the level and phase of the corresponding signals becomes spread over some appropriate range. Therefore, one desirable characteristic in a spatial enhancement system is the ability to exaggerate the differences between the left and right channel signals so as to simulate the differences that would occur in a real environment in which the sound sources were spread apart and/or made larger.

In addition to the interaural phase differences, a variety of other spectral and temporal characteristics are expected for sound sources located at different points in space. Some of these differences are due to the geometry and acoustical properties of the real or simulated recording environment, while other differences are attributable to the geometry and acoustical properties of the listener, e.g., the appropriate head-related transfer functions (HRTFs). For single-ended spatial enhancement it is desired to alter the existing spectral and temporal behavior of the signal in such as way as to exaggerate the size and separation of the reproduced virtual sound sources.

Through the examination of a variety of theoretical and empirical models for HRTFs corresponding to zero azimuth (straight ahead) and other directions, a set of salient spectral features can be determined. By applying this equalization to the signal from a source located near zero azimuth, it is possible to give the aural impression that the source has moved out from the center, as suggested by Schroeder [4, 5]. A low order approximation to these features indicates the spectral magnitude equalization depicted in Figure 1. The corresponding phase response for a sound source at azimuth near 90° compared to a zero azimuth source can be simulated by adding a controlled amount of inverted signal to the opposite channel.
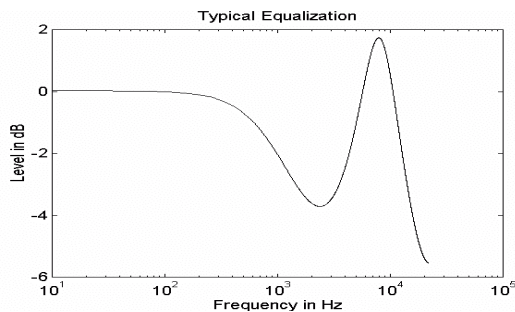
Figure 1: Equalization magnitude for side/front comparison

2.2 Mono-to-Stereo by Decorrelation

For reasons of convenience, economy, or source availability, it is sometimes desirable to convert a monophonic recording into a pseudo-stereo signal. As noted above, it is possible to perform mono-to-stereo conversion using a variety of delay and equalization procedures. However, these standard techniques are often unsatisfactory due to the spectral alterations that they create.

Decorrelation is a term that can be defined as a procedure to increase the perceived dissimilarity between audio signals without introducing audible spectral or temporal artifacts [6]. In the proposed mono-to-stereo system the decorrelation process is accomplished by *phase dispersion*, i.e., the introduction of different frequency-dependent delays to a pair of signal paths.

In order to obtain a largely flat response in the frequency domain and a dense, aperiodic impulse response in the time domain, the proposed mono-to-stereo procedure involves two cascades (left and right) of inverting digital all pass filters of the form $H(z)= -(a+z^{-N})/(1+az^{-N})$. Each cascade comprises several (typically five) single-coefficient all pass stages of different delay lengths. The design choice of delays and coefficients is made to result objectively in a low value of cross correlation and subjectively in an uncolored response. An example set of delays and coefficients determined empirically are given in Table 1. These cascades result in impulse responses of approximately 50 msec with a 44.1kHz sample rate.

|		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Left	N:	169	51	18	13	5
	a:	0.684	0.678	-0.673	0.692	0.686
Right	N:	150	69	21	9	7
	a:	-0.694	-0.689	0.683	0.677	-0.672

Table 1: Example all pass delay lengths, N, and gain coefficients, a, for left and right 5 stage decorrelator cascades

2.3 Spatial Enhancement Implementation

The fundamental equalization characteristic for the proposed spatial enhancement procedure is chosen to reduce the mid-band signal level (reduction of ~3dB near 2kHz), and to provide a high frequency boost (~2dB near 8kHz) to simulate the natural increase in brightness of sound sources located to the side of the head due to the azimuth-dependent response of the external ear (pinna) and the orientation of the ear canal. The required equalization can be accomplished with a parallel combination of a first-order low pass filter, a second-order band pass filter, and a direct path, in order to approximate the equalization of Figure 1.

The current procedure has been developed to identify and boost dissimilar components in the L and R signals and to insert controlled (attenuated and inverted) versions of the dissimilar components into the opposite output channel. This controlled image inversion procedure correlates strongly with a perceptual impression of sound field expansion by introducing phase and amplitude effects that would occur naturally for large and widely separated sound sources.

There are several ways to implement a system with these characteristics. A lattice (cross-coupled) structure is chosen for its simplicity of construction and analysis. The basic structure is shown in Figure 2.
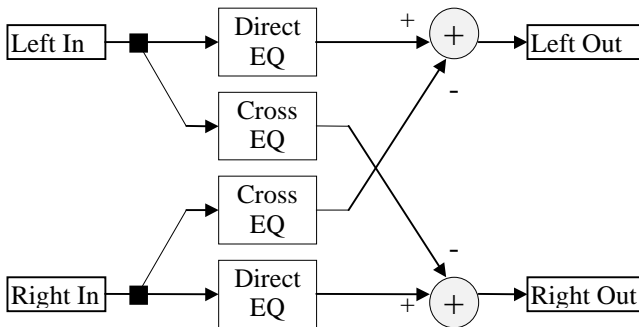

Figure 2: Lattice arrangement for spatial enhancement procedure

In the most basic embodiment, the direct and cross equalizer (EQ) blocks perform spectral shaping similar to the characteristic shown in Figure 1.

However, to achieve suitable spatial enhancement it is important to provide signal-dependent equalization. In cases where the left and right input signals are very similar it is necessary to apply the most aggressive processing in order to extract and exaggerate any small interchannel differences that may be present. On the other hand, if the left and right input signals are already substantially different, the spatial enhancement processing can be applied more conservatively.

The required signal-dependent processing is achieved by detecting the difference between the left and right input signals. The "difference" can be simply a smoothed version of |L-R|, an estimate of the cross-correlation between L and R, or some other comparison function. The difference function can then be used to alter the direct and cross EQ gains to achieve the proper level of spatial enhancement.

An example of this simple structure is shown in Figure 3. The smoothed difference signal, labeled PFACTOR, is used to adjust the gains of the direct and cross paths. The direct and cross gains will, in general, be different functions of PFACTOR.
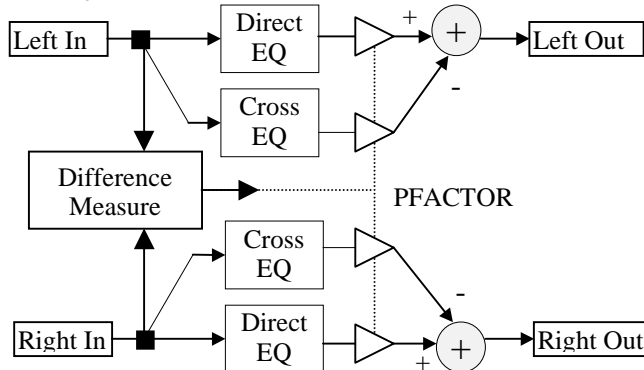
Figure 3: Equalizer with signal-dependent gain control

One possible implementation of this structure is depicted in Figure 4. In this form, the equalization and gain adjustments are accomplished using a combination of shared low pass and band pass filters with different signal-dependent gains for the direct and cross EQ paths. The resulting configuration offers a low-complexity implementation that is convenient for use on a programmable DSP platform.
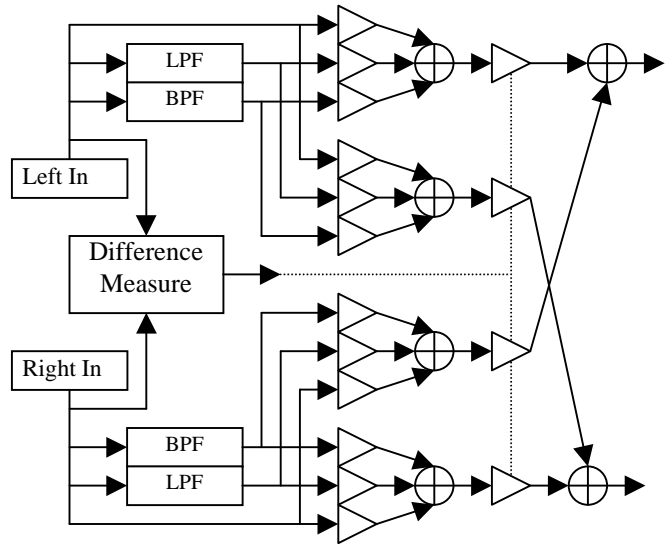
Figure 4: Low-complexity implementation of the signal-dependent spatial enhancement procedure

The performance of this configuration can be improved still further through the use of a second difference detection block monitoring the left and right *output* signals. The difference signal at the output is used in a feedback manner to adjust the LPF, BPF, and direct gains separately. The feedback arrangement enables a change in the shape of the spectral equalization that exaggerates the differences between the left and right output signals.

4. PERFORMANCE EXAMPLES

The behavior of the mono-to-stereo and spatial enhancement systems described in the previous section can be demonstrated in the time and frequency domains.

4.1 Decorrelation Behavior

Because all pass filters are used, the decorrelator cascades described in Section 2.2 have a flat magnitude response but differing phase responses. The effectiveness of the decorrelation can be observed by determining the output cross correlation function, as shown in Figure 5. For the specified set of coefficients the cross correlation never exceeds 0.1 in magnitude. The audible effectiveness of the decorrelation has been verified through informal listening tests with a wide variety of signal sources.
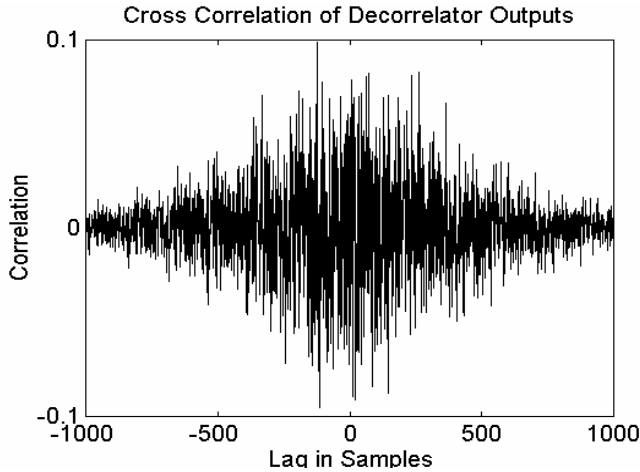
Figure 5: Cross correlation function between left and right decorrelator output signals

4.2 Behavior for L and R Very Similar

If the left and right input signals are very similar, it is possible to describe the input signals as $L=M+L_o$ and $R=M+R_o$, where M represents the common (mono) components shared between left and right, and L_o and R_o are the left-only and right-only components, respectively. In this situation it is desirable to attempt to boost whatever small differences between L and R that are present. In this situation the input difference signal (L_o-R_o) is near minimum. Using a control function which results in a slightly greater boost of the direct path compared to the cross path, the response at the output for the "mono" portion (M) of the L and R input pair receives a lower gain than the direct and cross components which differ between the inputs. The effective response functions are shown in Figure 6.
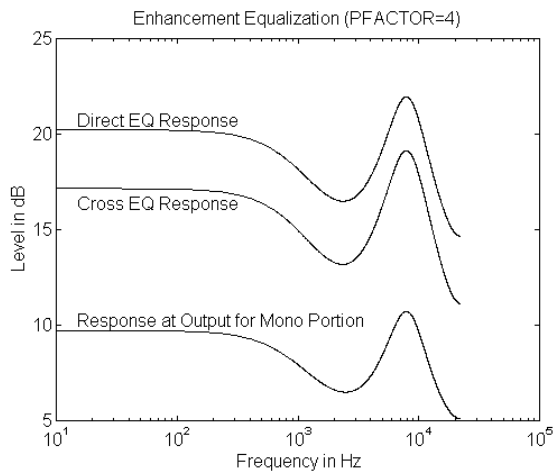

Figure 6: Frequency response magnitudes for input L and R very similar

4.3 Behavior for L and R Very Dissimilar

If L and R are already very dissimilar at the input, it is not necessary to alter the residual mono signal very much. The L, R difference is near maximum, and the direct path gain is automatically adjusted to be much greater than the gain for the cross path. In this configuration the major audible effect of the enhancement procedure is the slight spectral shaping between 2 and 8 kHz. The response functions are shown in Figure 7.
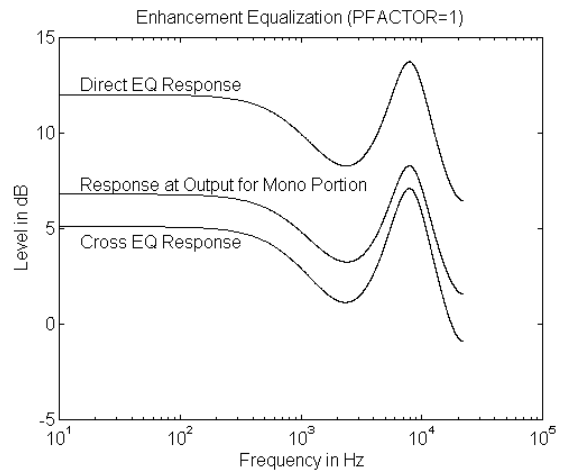

Figure 7: Frequency response magnitudes for L and R very dissimilar

5. CONCLUSIONS

In this paper a description has been given of a stereo spatial enahancement system. A mono-to-stereo all pass decorrelation structure has also been described. The proposed structures are effective and of sufficiently low complexity to be suitable for use in multimedia personal computer systems.

REFERENCES

1. Eargle, J., *Handbook of Recording Engineering*, New York: Van Nostrand Reinhold, 1986.

2. Maher, R. C., "Old and new techniques for artificial stereophonic image enhancement," *Proc. 101st Audio Eng. Soc. Convention*, Preprint 4371, November, 1996.

3. Blumlein, A. D., "Improvements in and relating to sound-transmission, sound-recording, and sound-reproducing systems," British Patent 394,325. Reprinted in *J. Audio Eng. Soc.*, vol. 6, 1958.

4. Schroeder, M., and Atal, B., "Computer simulation of sound transmission in rooms," *IEEE Conv. Record*, Part 7, March, 1963.

5. Schroeder, M., "Digital simulation of sound transmission in reverberant spaces," *J. Acoust. Soc. Am.*, vol. 47, 1970.

6. Kendall, G., "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music J.*, vol. 19, 1995.