# Stereo and Surround Imaging Within the Automotive Environment

Robert Reams, Chief Scientist
Neural Audio

Recent advances in digital radio make it possible to address the age old problem of stereophonic imaging within the automotive environment. By proper spatial environment management it is possible to expand the listener's "sweet spot" preserving the dialog clarity and sense of envelopment for all occupants in the vehicle.

## History

Past research has pointed out weaknesses in the presentation of stereophonic or spatial content within the automotive environment. Of greatest import is the dependency of the stereo or spatial presentation upon listener placement. When the listener is not located within the stereo "sweet spot" stereo, as well as mono, content suffers in the areas of dialog intelligibility, image localization and listener envelopment. Efforts to quantify the errors have historically taken the form of HRTF descriptors (1) as follows:

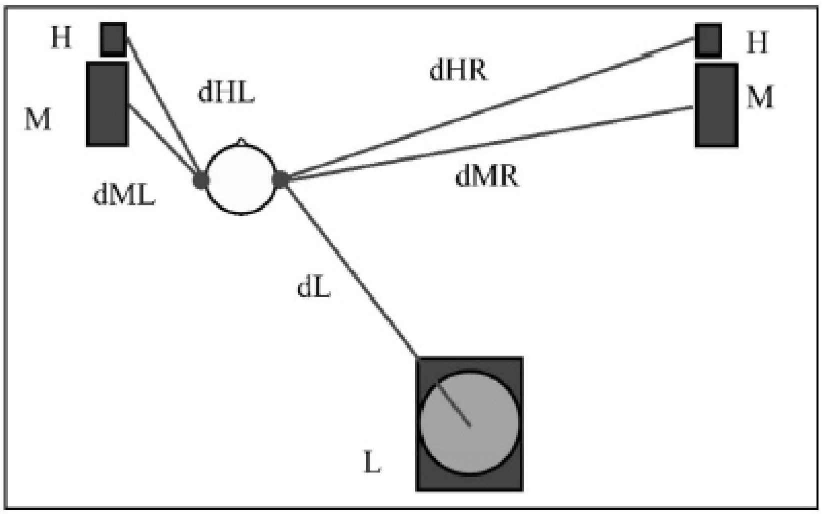
***Interaural Arrival Time Difference (IATD):*** The temporal difference in arrival time between signals at the ears. Dependent only on the position of the sound event, which dictates the path differences around the head. Most effective up to 630 ms of delay with no effect when greater than 1000 ms. Only impacts frequencies below 1.6 kHz.

***Interaural Phase Difference (IPD):*** The temporal difference in phase between signals at the ears. It is dependent on the relationship between the ITD and the frequency of the signal. Most effective for frequencies between 20-800 Hz for up to 630 ms of phase difference. Minimized effect for frequencies 800-1600 Hz with phase differences up to the half period (T/2). Only impacts frequencies below 1.6 kHz.

***Interaural Envelope Time Difference (IETD):*** The temporal difference in signal envelopes at the ears. This applies to modulated signals and depends only on the envelope of the signal and not any fine temporal differences. Similar effect as IATD for frequencies above 1.6 kHz: maximum at 630 ms and no effect over 1000 ms of delay. Unpredictable for frequencies content under 1.6 kHz.

***Interaural Time Difference (ITD):*** A generic term used to describe any of the above time differences. Typically refers to the one that dominates the frequencies under discussion. Below 1.6 kHz are dominated by the IATD (if impulsive) while periodic sounds are dominated by the shorter IATD or IPD cue. IETD dominates frequencies above 1.6 kHz.

It was theorized that because of the spatial and spectral changes in the automotive audio system, designers should make some basic adjustments. For instance, knowing that ILDs are the primary cue at high frequencies, the gain should be adjusted between the two tweeters. Low frequencies are time sensitive, so the door speakers and subwoofers should be adjusted with time delays. However, it is near impossible to adjust for those frequency-selective cues from the pinna and head.




The descriptors were then used to nominate closed, stable errors with solutions based on time and level correction. Conclusion were drawn that distortion of the stereophonic image was a function of inequities in energy and arrival time at the listener's head. No mention of the effects of the proposed virtual source alignment on other (non driver) passengers were made.

Other research suggests the treatment of different parts of the image differently. This research recommends time alignment of the difference (L-R) portion of the image with the driver's seating position, restoring the sense of "spaciousness". Still, there is no consideration of the other passengers in the auto.

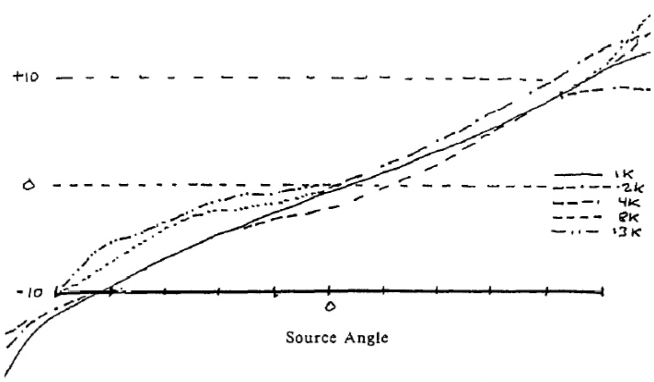
Finally, it was suggested that through proper content management and presentation, an automotive "spatial environment" could be achieved. It was further suggested that a satisfactory trade off could be made between preserving image depth at the expense of ideal lateral fine structure.

"Codec pre-conditioning" was cited as an integral part of maintaining a predetermined set of spatial aspect "targets" necessary for efficient operation of lossy data compression prevalent in the digital broadcast industry. An additional benefit of maintaining the spatial aspect targets of the content was higher performance of so-called matrix decoders in the form of reduced dialog leakage, better separation and reduced detector core drift at low levels.

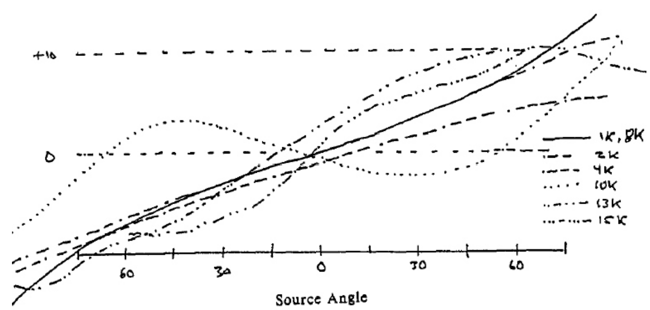
**Realities**

This is not to say that there is no skill or science in the capture and presentation of a spatial event. There has been much research in this area as there are many implemented techniques of capturing the stereo image with various degrees of success.

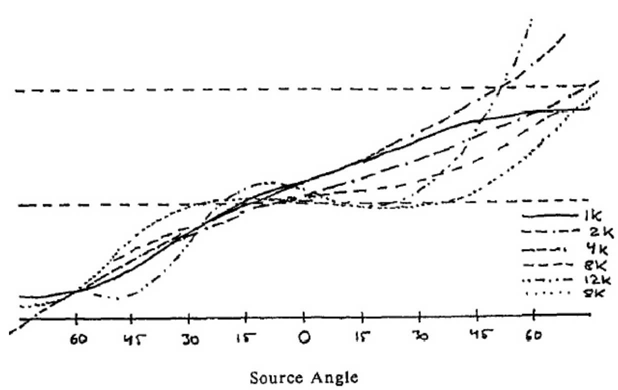
Figs. 2 to 4 illustrate the image variances as a function of spectra of typical stereo microphone arrangements. None are bad. All are different. You can imagine how variable the rest of content production can be.




L/R ratio for 90 degree coincident Cardioid pair - 20mm diameter




L/R ratio for 90 degree coincident Cardioid pair - 17mm diameter




L/R ratio for 90 degree coincident Cardioid pair - Soundfield

After the content is captured, there is the matter of mass distribution mediums, that is, getting the content from it’s pristine origin to the consumer.

Fig. 5  A short list of possible sources:

1) Uncompressed mono content (cassette and CD based “talking books”, analog AM etc.)

2) Compressed mono content (“mp3" based talking books, digital AM, satellite talk and news)

3) Uncompressed stereo content (CD music, analog FM, cassette music, etc.)

4) Compressed stereo content (digital AM and FM music, satellite music, “mp3" music, etc.)

5) Uncompressed 5.1 content (MLP, DTS and other lossless 5.1 formats)

6) Compressed 5.1 content (AC3, PAC, and other “lossy” 5.1 formats)

Fig. 6 Various means (lossy and lossless) of conveying the content:

Analog tape

Vinyl (what’s that??)

PCM

Amplitude Modulation

Frequency Modulation

MPEG-1 Layers 2&3

AAC

AACplus

HDC

ADPCM

Lossless Packing

CELP

MELP

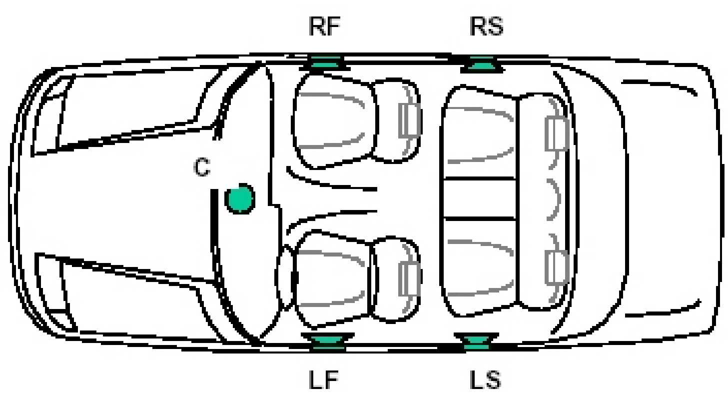
After overcoming the variabilities and shortcomings of lossy data compression, various physical medium and the interferometrics of terrestrial or space broadcast over the vast expanse of all known broadcast wavelengths...what’s left of the content is presented to the automotive engineer.

Automotive engineers have long deplored the gap between what is considered “ideal” and what is considered “practical” for stereo imaging.  The luxuries of an excellent (abiet variable) stereo image are elusive, to say the least, even in the most controlled environments. Other than obviously limited acoustic treatment options, physical symmetry, as other authors have pointed out, is practically non-existent within the automotive environment.  The resultant intensity and temporal offsets naturally warp the lateral image fine structure.  Less obvious, however, is that because of this asymmetry there is no mechanism for reproducing the depth structure of the stereo image. It's not just that the image depth fine structure is warped; it's completely non-existent.

While the nature of this mechanism is hotly discussed, the need is sorely noticed as demonstrated by the plethora of "blind" matrix decoders, artificial reverberation and content-less* 5.1 channel systems.  This is not to say that these technologies are without merit, rather, none seem offer a complete solution to the automotive environment designer as they are "fixed" and require the automotive environment to conform to the specification of the audio system, not the other way around.

There are many possible (and required) permutations of the automotive "spatial environment" as well as all of the possible sources (hard, soft and broadcast) of content sources.  Choices for loudspeaker number and  placement are seldom ideal and always budget and aesthetics driven.

Fig. 5 proposed loudspeaker element placements within the automotive spatial environments




1) L front door, R front door , S rear deck (LFE to S rear deck)

2) L front door, R front door , LS  rear deck, RS  rear deck (LFE to LS & RS rear deck)

3) L front door, R front door , LS  rear door, RS  rear door (LFE to L & R rear deck)

4) L front door, R front door , LS  rear door, RS  rear door (LFE to all)

5) L front door, C dash,  R front door , LS  rear door, RS  rear door (LFE to L & R rear deck)

6) L front door, C dash,  R front door , LS  rear deck, RS  rear deck (LFE to L & R rear deck)

7) L front door, C dash,  R front door , LS  rear deck, RS  rear deck (LFE to sub)

8) L front door, R front door , S rear deck (LFE to sub)

9) L front door, R front door , LS  rear deck, RS  rear deck (LFE to sub)

10) L front door, R front door , LS  rear door, RS  rear door (LFE to sub)

11) L front door, C dash,  R front door , LS  rear door, RS  rear door (LFE to sub)

12) L front door, C dash,  R front door , LS  rear deck, RS  rear deck (LFE to sub)

13) L front door, C dash,  R front door , S  rear deck (LFE to sub)

14) L front door, C dash,  R front door , S  rear deck (LFE to S rear deck)

15) etc., etc....

**Solutions**

This seems to be a  complicated scenario at best, but it is solvable.

First, all content must be "downmixed" to a common spatial denominator so that it may be effectively distributed to all practical (automotive) spatial environments through all practical hard, soft and broadcast medium.  This is possible with modern signal processing and watermarking technologies.

Second, there must be a reasonable limit to unnatural or hostile artifacts as produced within the spatial environment.  This means that spectral extension and image fine structure may be "artfully" manipulated to minimize the audibility of said artifacts.  This also means that there will be conditions where *only a very limited* spatial environment is possible because of highly aggressive (40-50:1) data compression or analog broadcast multipath conditions.

Lastly, there needs to be a dedicated "engine" to resolve spatial redistribution of the common spatial denominator within the various automotive environments.  This "engine" should be easy for the designers to use and portable to the various DSP environments.  As such, the ideal spatial environment would be flexible and allow for limited "customization" to achieve as high of performance as possible within the limitations of driver number and placement.

This is not to say that this solution has no application with the consumer "home theater" environment.  Many clever designers may see this as a closed solution for simplifying existing content compatibility issues for consumer electronics including satellite, cable and terrestrial broadcast.

**Benefits**

There are a number of advantages to management of content within the constraints of a common spatial denominator.  The first, of course, is the ability to broadcast the content at significantly lower data rates.  This could translate into more content channels, more robust transmission, higher data capacity or in many cases much faster downloads.   Secondly, these factors mean better service for the consumer regardless of the content origin or consumption environment.

Thirdly, higher revenues as well as additional mechanisms for delivery of data for the content provider. The value of compatibility with the existing two channel storage and broadcast medium cannot be emphasized enough.

**Summary**

In the past there has been a great deal of variability in the image structure of content, distribution medium and playback resources within the automotive environment. Methods for managing spatial content and storage/transmission medium consistency have been vastly improved over the years. It is proposed that a spatial environment development system and engine be defined to provide the final link between real world content and the real world automotive consumer.

**P.S.**

Backward compatibility *is* a driver for adoption. The success of a format isn't always driven completely by the technical "straw man" of sonic superiority under tightly controlled conditions. Factors as economy, stability, convenience, speed to market and compatibility are always considered by the mature systems architect.