

Downmixing, Blasphemy or Necessity?

Robert Reams, Neural Audio

Why?

Why downmix? After all, there are other alternatives:

- a) Deploy a totally digital world system of unlimited bandwidth, where the cost of content transport is free.
- b) Teach consumers to enjoy waiting on long downloads.
- c) Make them buy gigabit storage media.
- d) Remove every mono, stereo and matrix device from their home, car, game and person.
- e) Throw away their existing content collection.
- f) Have them purchase all new equipment that is compatible with all known new formats.

An alternative is to downmix.

So, the sane engineer may ask, “what are the pros and cons?”

A Short History of Downmixing

Spatial compression is both old and new. The earliest versions of downmixers being close cousins to the “quad” era utilizing all-pass filters, summers and differencers. With a heavy dose of managed expectations and training, skilled operators were able to crank out pleasing downmixes that would adequately “upmix” through the complementary consumer matrix decoder and remain esthetically pleasing when auditioned through stereo. These downmixes didn’t just “happen”, they were the work of much planning, experience, trial and error. All matrix mixes were planned compromises designed create an entertaining environment, not to be confused with conveying the artist’s unbridled creativity. Downmixers for matrix playback were not easy to use and there were artists known to be critical of the “sound”. As such, the adoption of downmixing by the music industry was historically resisted for a variety of subjective, however valid, reasons.

Dullness

Dullness or “dulling down” is a subjective complaint that seems to be attributed to the use of the allpass networks used in the matrix encoding process. Anecdotally, individual allpass components are “inaudible” to expert listeners during the encoding of stationary (tone like) content. The same expert listeners find that allpass networks “color” non-stationary (transient) content. When used in networks, allpass components may not interact well with each other, producing subtle but audible modifications of the content timbre as a function of image “position”. Some work better

than others.

Downmixing Dangers

The majority of perceptual challenges “caused by downmixing” aren’t really caused by the allpass filter networks in the downmixer at all. The real challenge with downmixing is 5.1 content that isn’t mixed with stereo or mono compatibility in mind.

Non Co-Incident Stereo Pairs

Content element pairs that are similar yet non-coherent can cause problems in downmixing. Elements that sound natural with separate image origins can sound quite unnatural downmixed to a common image sector. A good example of this phenomenon is a voice element localized in the front stereo pair with an “echo” of the voice element localized in the rear stereo pair. Upon downmixing, both the original element and the delayed element co-exist in a common image sector. If the delay is short enough this will result in the element sounding as if it has passed through a comb-filter when rendered in stereo...far from the intent of the artist.

Image Concatenation

Without adaptive normalization (more on this later), weaknesses within “watermark” based intensity/coherence mapping leave the encoding vulnerable to the pre-existence of spatial relationships between content element pairs prior to downmix. Watermarking depends upon statistical weighting of intensity and coherence as a function of spectra within the 2-D (stereo) fine structure pair. Concatenation of the watermark with spatially related element pairs may result in inconsistent recovery of spatial location upon rendering.

Side information based upmixing may not fare much better. Although the spatial relationship between a pair of content elements may be appropriately mapped, linear summation in the downmix stage could cause combing, partial or full cancellation.

Error Prevention

All laterally related input sets (L/C/R and Ls/Rs) must be spatially normalized via “image packing”. What this does is that it transforms *excess phase* that may exist between coherent elements into intensity and coherence descriptors. These offsets are generally accidental or well intentioned, caused by certain types of non-coincident “microphone arrays” or delays placed artistically without testing for stereo/mono compatibility. This stage of image packing transforms lateral inter channel time differences to perceptually equivilant intensity ratio’s and coherence.

Following the normalization of *excess phase*, *simple phase* must then be normalized. This prevents inadvertent rendering of image “depth” that could be misconstrued as “crosstalk” within the rendering (upmix) process. The combination of both processes effectively reduce or remove time and phase differences between similar mix elements reducing the azimuth descriptor of the element pair to a simple energy ratio (with modified coherence). This allows for more controlled placement of “stereo” mix elements regardless of origin or pedigree (and artifact free

downmixing).

Depth related input sets (L/Ls and R/Rs) are treated similarly and then spatially de-correlated prior to downmixing operations. This part of the process removes problematic “front-to-back” time offsets that may exist between coherent elements and then de-correlates them through fractional Hilbert transformation. This reduces the depth descriptor of an element to simple inter-channel coherence.

The correlation and fractional Hilbert transformation lattice preserves the original timbre, balance and intended dimension of the original 5.1 (4.1?) mix in a presentation suitable for 2.0 stereo playback.

The reduction of image complexity to two-dimensional (2-D) intensity and coherence descriptors sets the stage for a more accurate translation of the image when rendered back to multichannel, not to mention making great sounding stereo.

Cost of image data

There is difficulty meeting even entertainment grade quality at real world data rates. When constrained to a bit rate of 64kBs it seems that dedicating 32-48kBs to the fine structures is not a good idea. Watermark based downmix encoding may be more technically challenging than a side data approach, but it allows for 20% to 30% more data for the fine structures. There is no need to “burn side data” even though there is no benefit (broadcasting “original stereo content”).

The 5225 downmix does not use “side information”, that is, data multiplexed into the bitstream to aid or enable spatial reconstruction of discrete 5.1 content. Instead, the 5225 embeds the spatial reconstruction descriptors directly into the waveform of the audio. This allows the mixer to treat the downmixed audio as (time domain) “audio”. The mixer can cut and paste the downmixed 5.1 just like any other stereo content without the fear of “ticks”, “pops” and artifacts caused by editing bitstream based files.

Benefits

Additional benefits of the 2-D downmixing are:

- 1) production of content that is 100% compatible with lossy compression
- 2) generation of 5.1 content that may be edited, mixed with and archived in 2.0 stereo
- 3) easy and intuitive to use
- 4) allows use with favorite “old school” equalization and dynamics processors
- 5) natural sounding downmixes, no matter what

Summary

2-D downmixing ushers in a new era of content creation possibilities. It is a methodology that allows the creation of versatile mix/edit tools necessary for the creation of 5.1 downmixes either from scratch or from existing 5.1 original source content. The 5225 facilities would also allow the user to easily downmix 5.1 content and test for upmix compatibility. The 2-D downmixing

enables the user to produce content quickly and confidently without fear of stereo or mono compatibility.

This is a powerful mix/edit tool that solves future downmix needs and provides tools that allow creation or editing of content that originates as 5.1 or 2.0, creatively blend discrete and upmixed elements and store the content in 2.0 or 5.1 formats. The 2-D downmixed content is renderable to 5.1 to provide perceptual confidence monitoring that insures that the art will be heard just the way it was designed.