

Thomas Lund
TC Electronic A/S
DK-8240 Risskov, Denmark

**Presented at
the 109th Convention
2000 September 22-25
Los Angeles, California, USA**



AES

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd St., New York, New York 10165-2520, USA.

All rights reserved. Reproduction of this preprint, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

AN AUDIO ENGINEERING SOCIETY PREPRINT

Enhanced Localization in 5.1 Production

THOMAS LUND

TC Electronic A/S
Sindalsvej 34, DK-8240 Risskov, DENMARK
thomasl@tcelectronic.com

Power panning relies on the creation of phantom images. Outside a narrow listening sweet spot, there can be obtained only as many consistent directions as the number of speaker channels, i.e. positions not depending on phantom images.

Source localization in 5.1 format may be enhanced if panning is not based only on level but also involves room rendering utilizing multi-directional reflections. As an added benefit, the listening sweet spot can be widened thereby giving the whole 5.1 idea more advantages over traditional 2 channel reproduction.

The paper will discuss experiences from film and music production incorporating room rendering into the panning; especially how the image stabilization with integrated room rendering can be a tool to involve the listener in ways not possible before.

0. INTRODUCTION

In this paper we have conducted some preliminary studies of how phantom image localization as a function of interchannel level differences can be improved if perception based upon room rendering models and early reflection patterns are used to support the positional cues.

Enhanced localization can rely on direction and distance simulation possibilities not readily obtainable using discrete microphone techniques, and room color and geometry is subject to engineer manipulation leaving space for artistic freedom far beyond the constraints of natural environments.

Perceptual refinement of physical models can be turned into DSP code to provide the sound engineer more artistic freedom than power panning alone does. “In our experience, no semi-automatic physical modeling scheme, however elaborate, is likely to produce subjective results as good as those obtained by skilled people...”, [10].

Because human perception can be taken directly into account in DSP based localization and room models we typically have found the results to be more convincing than what would then appear to be the indirect approach: Multiple microphones in a real room.

Imaging and room properties may actually sound more natural than if a certain event is subjected to the compromises of distant miking, imperfect rooms, arrays of microphones and finally the end listeners' speaker configuration.

As part of the discussion about phantom imaging we will point out what seems to be different views between professionals from the music and film industry of how to incorporate (or not to incorporate) the center speaker channel.

Ways of how to work around those differences will be suggested.

1. MULTICHANNEL RECORDING AND MIXING

New mixing techniques and procedures for recording multichannel material are being developed these years. Several studies have pointed out the difficulties and limitations when using microphone arrays to capture well-localized sources, attaining precise perspective, avoiding excessive coloration and cross talk etc. Some evidence exist, though, that spaced microphone set-ups like the ORTF, KFM and others may achieve credible results in this respect [2], [3], [4], [5]. Still, no technique seems to transfer a room from microphones to 5.1 speakers in a truly precise and stimulating way.

Compared to artificial room modeling, we have found that microphone arrays simply cannot achieve as much as a dedicated, phycoacoustically based DSP model taking in several sources, e.g. spot microphones, and rendering a best-fit result for the 5 speakers.

The question arises: What kind of signal do we want to reproduce over our speakers? The whole principle of putting up a set of main microphones and making them produce stable phantom images etc. simply does not cut it for multichannel. We cannot ask the listener to sit in one precise spot to listen to a 5.1 mix with only little improvements when compared to stereo.

If we want to convince the consumers and ourselves that 5.1 is worthwhile; we should rather come up with something better. Aiming higher, we have to seek mixing, miking and processing techniques to unfold the true power of 5 discrete speakers. We also have to do it in a practical way: The recording and mixing process should not take longer - only the results should be better.

The key word is "precision" and that cannot be achieved without a high degree of control over source localization. The wider area we can cover with credible localization, the better. However, precision should not be understood as an attempt to completely reproduce the experience of being in a certain acoustical space [5], [10]. This is a loosing battle having only 5 main speakers. Credibility and predictability counts and the room geometries and source positions of the simulated room should be relatively, but not absolutely, accurate.

1.1 Center Channel

While the Center channel is the single most important one for film production, the music industry has reluctantly accepted 5-channel stereo. Many multichannel music mixes do not take advantage of the image stabilizing effect the center channel provide.

A typical phantom center image of a pop/rock music mix is lead vocal. Fear of exposing the singer may be the reason why she is not put in the Center, because most reverb and delay effects units until recently have had no center speaker output. In pop music tracking it is not practical to record even the lead vocal on 3 or 5 tracks because of the extensive editing being applied. Image shifts would surely result and many of the tools used for vocal manipulation would not work.

This limitation is about to disappear because of the dedicated 5.1 equipment now being offered to the users. Tracking can continue with mono microphones, but a suitable 3 or 5 channel representation can quickly be realized using integrated multichannel mixing or effects systems, [9], [10].

Another differing view between the Music and Film industry came to our attention while doing 5.1 room simulation and effects developments for both camps. In music production the engineer has been so used to phantom images that the same signal produced through a single center speaker sounds “wrong”.

Because of the interaural delay and head shadowing, a phantom image does not sound like a real single point source. This is the old truth behind the well-known music expression “pan before you equalize”. To make the 5.1 transition with its center speaker more eatable for the music industry, we are giving users the option of “contaminating” the center channel signal as if it was a good old phantom.

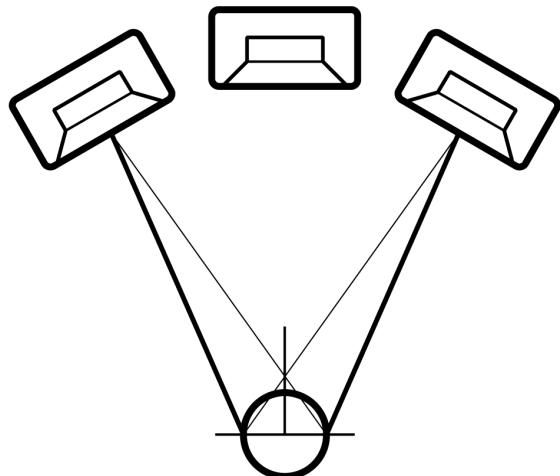


Figure 1

Figure 1: Center image being produced as Phantom using Left and Right speakers. Signals to the opposite ear shown as thin lines.

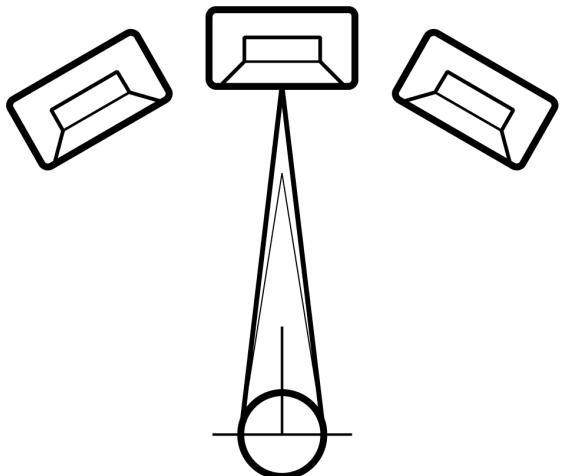


Figure 2

Figure 2: Center image being produced as real single source. The signal may or may not be “contaminated” with simulated cross-talk signal shown as thin lines.

Some music engineers tell us that “now the Center does not jump out on you” and they can still hear the stabilization benefits, while of course film engineers find the principle awful because presence is lost.

2. MULTICHANNEL FORMATS

Many qualified studies have pointed to the ITU-R BS 775-1 Recommendation, [1], as being the best arrangement if five speakers are employed and the orientation of the listener is known.

The set-up, being the best compromise between 360 degree imaging and stimulation of the listener from lateral angles while maintaining the long used equilateral triangle for two channel stereo between left and right, is also worth aiming at for standardized interchanging and meaningful evaluation of multi-channel material.

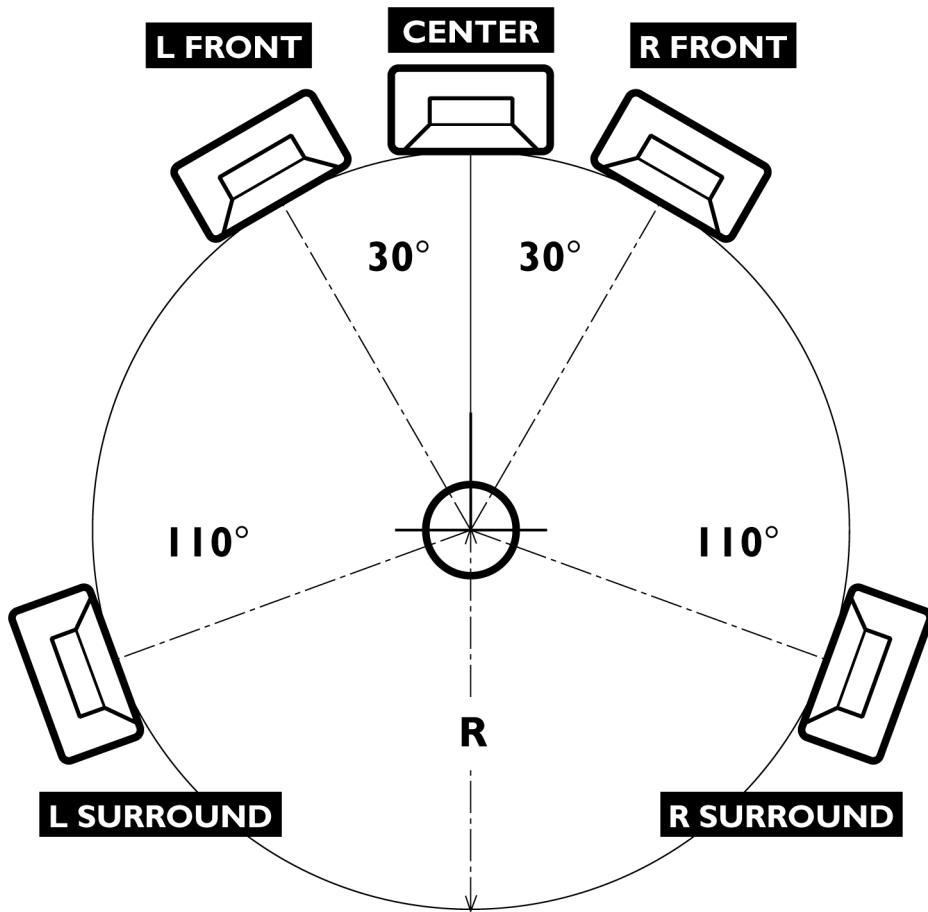


Figure 3

ITU-R BS 775 Recommendation. Placement of 5 identical main channel speakers [1].

In this study we have therefore used the ITU speaker set-up exclusively. We believe only this scheme should be focused on when describing 5.1 reproduction to the public, and not the Quad arrangement with its ± 45 and ± 135 degree angles.

However, most 5.1 mixes for film are performed while monitoring through a 3rd type of speaker arrangements; a Cinema style set-up with diffused speaker arrays for the surrounds. For large audiences, the ITU speaker arrangement of course is not suited.

Under Cinema conditions, the listening angles are less predictable: Much of the audience will not even agree if a signal appearing only in one of the surrounds is from the front, the side or behind you.

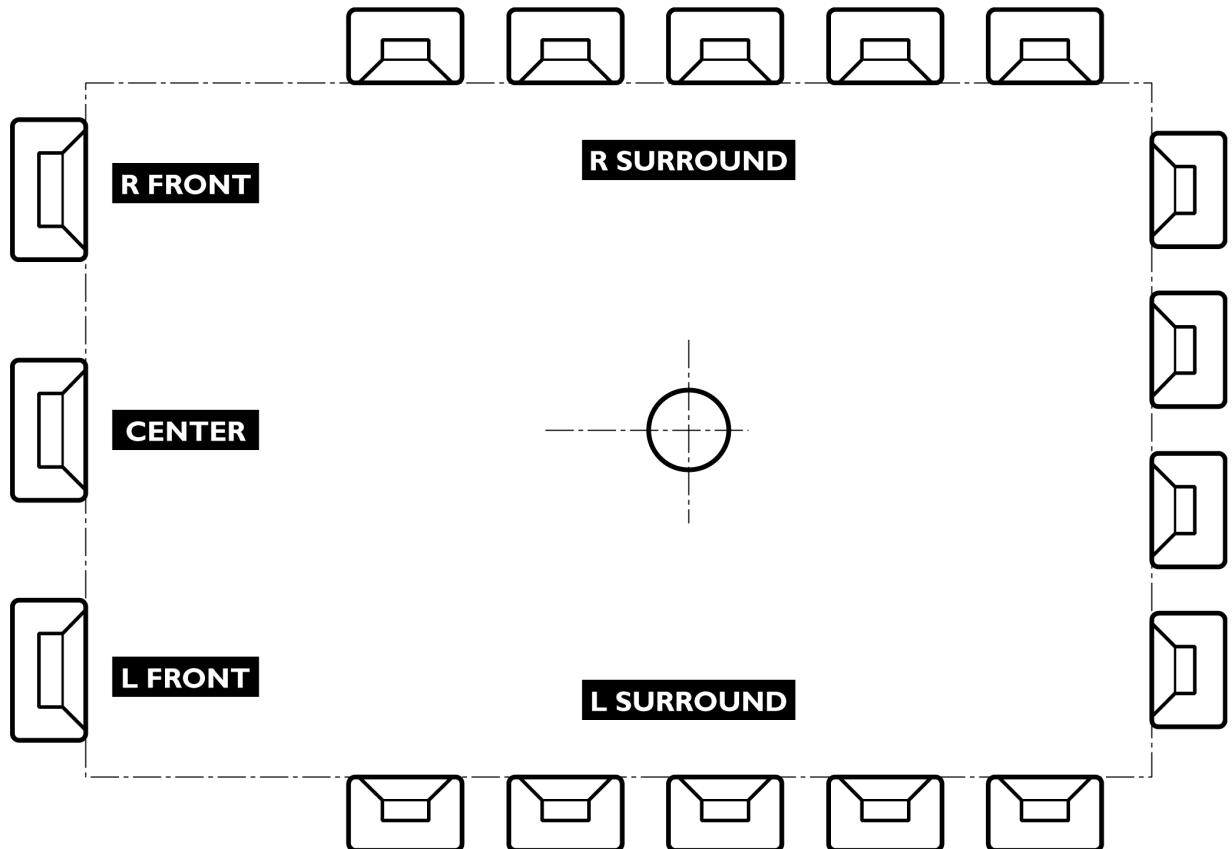


Figure 4

Cinema style speaker arrangement with surround arrays.

In Dolby EX rear surround speakers can be fed with signal decoded from Left Surround and Right Surround.

The reason for the Dolby EX treatment seems well indicated, but it would have made more sense to wrap a 6th main channel into the digital compression scheme instead of ending up with analog decoding, inferior channel separation, custom mix monitoring systems etc.

Future reproduction techniques may involve an even larger number of speakers and dedicated signal processing per speaker. Currently this is not used outside the sound laboratory and not aimed towards the end user under normal domestic restrictions. On the contrary, such technology will initially find its way into large entertainment systems, ambitious movie theatres etc.

For several years to come it seems we will either use 5.1, or have to make do with stereo.

If we settle for recommending the practical 5.1 standard, the two most dominant problems to solve are

1. Consistent Localization having only 5 speakers.
2. Size of the useable Sweet Spot.

2.1 Localization

Control of perceived direction and distance are crucial factors to enable localization, which is one of the important means to directly engage the listener.

Over the years, many studies have described the difficulties with using phantom sources to render robust directional information, [6], [7], [8].

While 2-channel stereo can produce a fairly consistent imaging over a ± 30 degree angle with the listener precisely centered; lateral direction is not even predictable when listeners are sitting in the sweet spot. Deviation between listeners is very pronounced regarding side imaging.

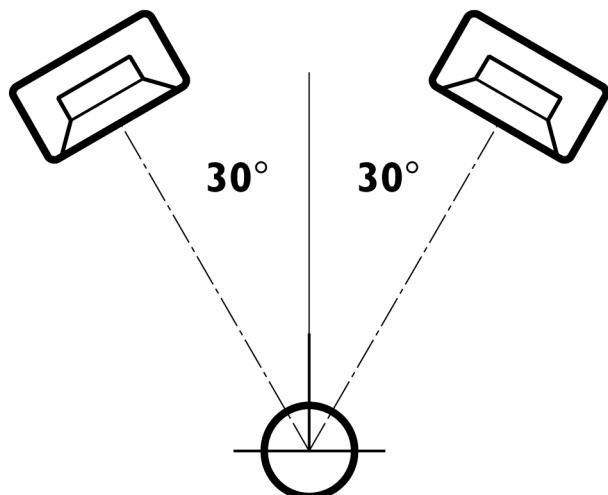


Figure 5

Figure 5: Direction is quite well defined using power panning between the two speakers if the listener is precisely located in the sweet spot.

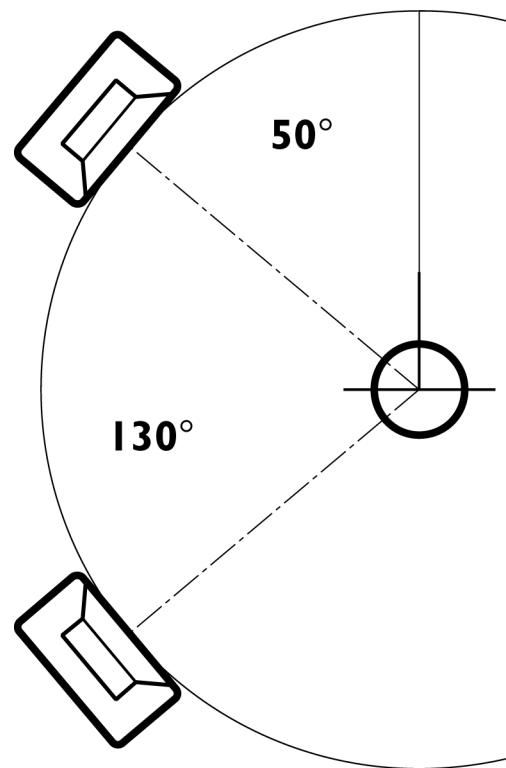


Figure 6

Figure 6: There is considerable uncertainty of direction using power panning between the two speakers. Lateral imaging based upon power panning works poorly using directional speakers pointing at the listener located in the sweet spot, [7].

Without improvements in this area we believe the incitement to put up with more speakers and new reproduction equipment may not exist with the average consumer. In home theatres the 5.1 set-up significantly has to add to the illusion presented by the picture to be justified. For music reproduction indisputable benefits over 2 channel stereo are needed, but the ability to present a consistent direction even for the comfy chair in the sweet spot is actually worse in 5.1 than in stereo if mixing only takes advantage of power panning.

With the amount of signal processing power and memory bandwidth now being available, it seems an obvious choice to integrate panning with reflection patterns and reverberation to gain more and integrated control over Direction and Distance. Using different numbers of reflections, directions, diffusion and equalization, qualified studies have been taken on several independent places in the audio community, e.g. [9].

2.2 Sweet Spot

The sweet spot - or in some cases the sweet area - is the area in which listeners get a good and consistent impression of the directions and distances to individual sources of a mix. The sweet spot is also the reference point of the loudspeaker arrangement.

In a two channel stereo set-up, credible reproduction is only obtained within a relatively small listening sweet spot.

As we have seen, phantom imaging relying on power panning, even under ITU 775 reproduction conditions and even within a narrow sweet spot, produces very variable and uneven imaging results.

In a 5.1 speaker reproduction system based on power panning, satisfactory defined directions only exist at the 5 main speaker positions. It would appear important to offer the sound engineer possibilities of widening the sweet spot in a way not obtainable using normal microphone technique or panning schemes.

3. INTEGRATED ROOM SIMULATION AND PANNING

It is our assumption that we in general decode information from our sensory system in an adaptive way. As long as the information is not a complete contradiction, we assemble our sensory puzzle in a best-fit manner. This thesis is the foundation for building extensive, source dependent early reflection generators to support localization even outside a listening sweet spot.

From a perceptual point of view, the way power panning works is not very nature-like and probably therefore not very suitable to extract valid auditory information from. In nature it is highly unlikely that well correlated sounds occur from several sources simultaneously.

To provide the listener with more (maybe) useful auditory information when reproduced through a 5.1 speaker system, we used an integrated positioning and room-rendering model for these experiments.

The model chosen was explained in a previous paper, [10].

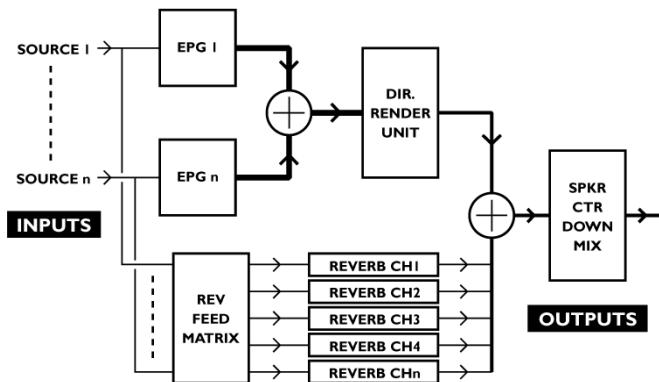


Figure 7

Block diagram of the real-time DSP structure used for the experiments.

It provided early reflection patterns with several levels of diffusion initially rendered as 18 different directions and internally converted in the Direction Rendering Unit to the 5 main channels.

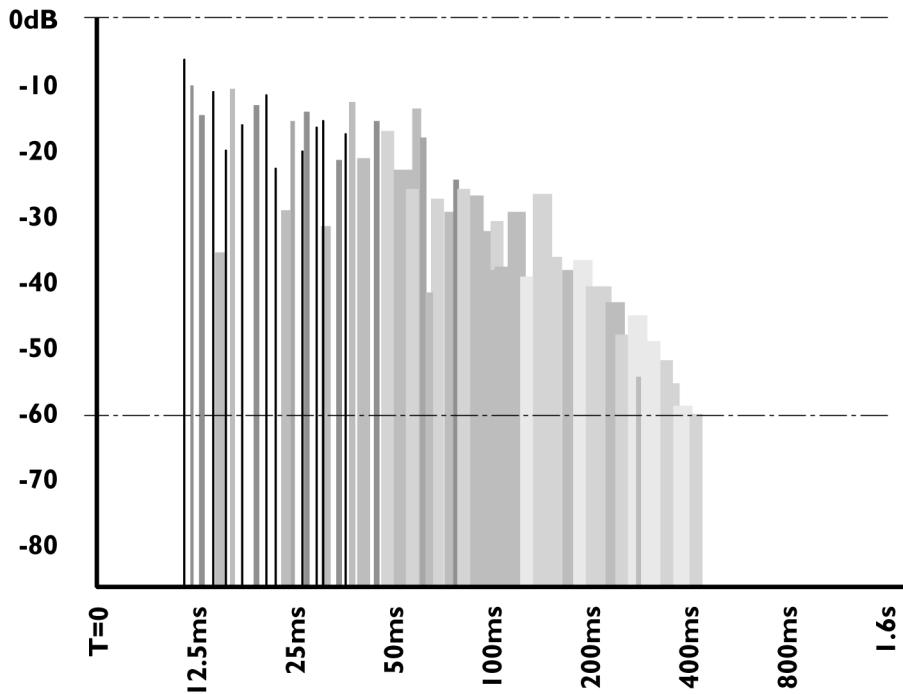


Figure 8

Example of an early reflection pattern from one source to one of the main channels. Shades of gray are used to visualize different ratios of diffusion. Reflections are created using 24-bit precision RAM storage.

The late part of the reverb was produced using 5 uncorrelated diffuse field generators and adjustable delays, level control and filtering from each source.

4. RESULTS

In these preliminary tests, 3 experienced and 2 non-experienced multichannel listeners were subjected to a variety of signals including speech, sound effects, solo instruments and fully orchestrated music.

Image consistency experiments were conducted using sources at 0, 7, 15, 30, 45, 90, 115 and 180 degree angles. The source angles 45 and 90 degrees were chosen to be able to compare lateral localization with the well known results from power panning directional experiments.

The 5 output channels from the real-time digital processor, capable of doing both power panning with different focus settings and early reflection based positioning, were fed to the speaker systems.

Power panning was done using standard cosine constant power amplitude control with no attempt of crosstalk cancellation etc.

The radius of the ITU-R BS 775 speaker set-up used (see Figure 3) was 2.5 meters. The environment was non-anechoic, resembling a real-world production or listening situation.

Two different ITU 775 speaker set-ups were tried, one with DynAudio BM15A, the other with Genelec 1031 Speakers. Results were identical using the two different types of speaker systems.

An arbitrary Consistency scale was used to determine image precision and robustness. Values were appointed according to these criteria:

Consistency Score	Certainty of angle	Robustness	Diffusion
5	No doubt	Very	None
4	High	High	Slight
3	Good	Some	Some
2	Some	Low	Mostly
1	Poor	Poor	Very

Generally, Consistency scores of 3 and above should be regarded as useable in a production situation.

4.1 Direction

In the first test, Consistency was measured with the test subjects sitting in the sweet spot.

Because of the low number of subjects, scores should be taken as a guide only, but as indicated there was little deviation in the results between the testers.

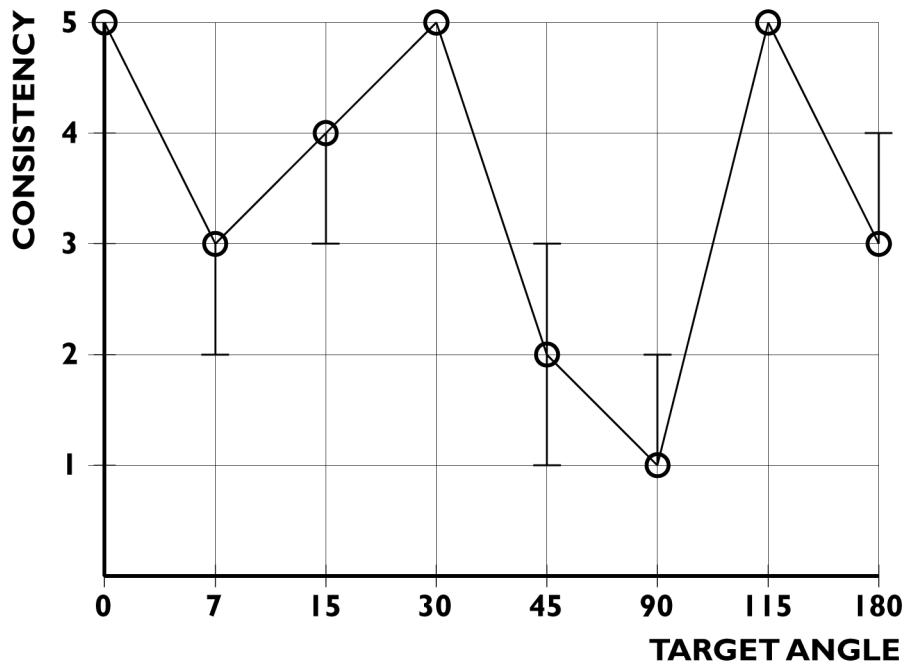


Figure 9

Image consistency in the sweet spot using power panning to describe direction.

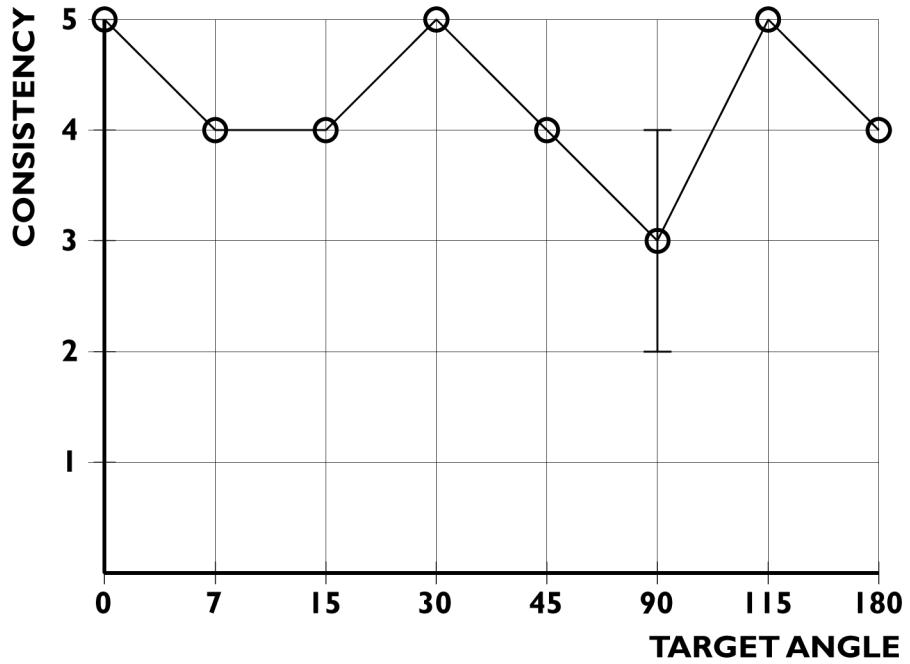


Figure 10

Image consistency in the sweet spot using supportive room simulation. Scores shown for pan environment “Jazz Club”. Similar consistency scores were achieved using pan environments “THX Cinema CC” and “Concert Hall CG”.

4.2 Sweet spot

In the second test, Consistency was measured with the test subjects sitting at different listening locations.

Listening areas where most of the directions scored 4 or 5 are shown in dark gray.

Listening areas where most of the directions scored 3 are shown in light gray.

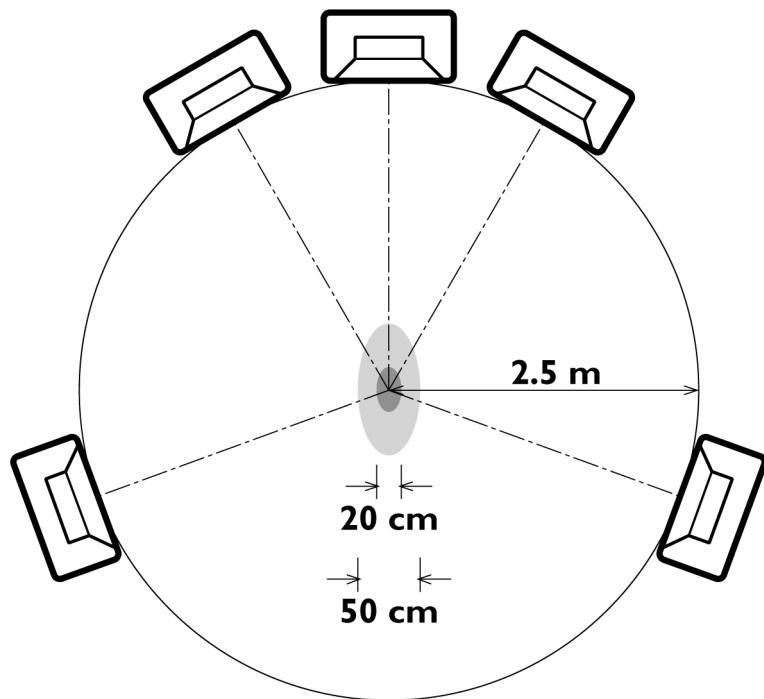


Figure 11

*Size of Sweet Spot: Image consistency as a function of listening position.
Consistency scores based on power panning.*

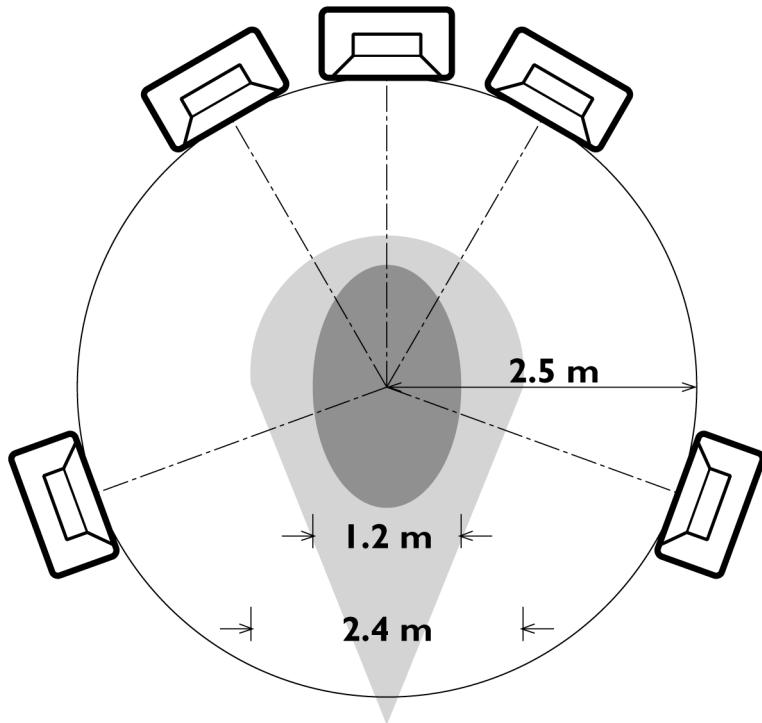


Figure 12

Size of Sweet Spot: Image consistency as a function of listening position.

Consistency scores based on using supportive room simulation. Scores shown for pan environment “Jazz Club”. Similar consistency scores were achieved using pan environments “THX Cinema CC” and “Concert Hall CG”.

5. CONCLUSION

A considerable widening of the listening sweet spot, and generally better localization results in a 5.1 channel reproduction systems, seem to be within reach utilizing some of the principles described in this paper.

To support direction and distance illusions of sources in a mix, early reflection patterns adjusted to suit individual sources, or groups of sources, appear to be an improvement in two channel but especially in 5 channel reproduction systems.

More precision and consistency with how we hear a sound source in a room translated to a 5.1 speaker set-up in a best-fit manner may actually be achieved employing new or adjusted mixing techniques.

Clearly, these preliminary studies indicate that more experiments should be conducted to clarify precisely what can be achieved with regards to localization and widening of the sweet when the compromises of a 5 main channel speaker system are taken into account.

This test gives no evidence against our thesis: Our ears will adapt to most listening conditions without contradictions and extract meaningful information whenever we give them a chance to do so. That is probably the reason why power panning does not work, but a more information-intensive model like the one described does.

7. REFERENCES

- [1] ITU-R BS. 775: Multichannel Stereophonic Sound System With and Without Accompanying Picture (Geneva, 1994).
- [2] Akira Fukada, Kiyoshi Tsujimoto & Shoji Akita: “Microphone Techniques for Ambient Sound on a Music Recording”. AES preprint 4540, 1997.
- [3] D.G. Kirby, N.A.F. Cutmore & J.A. Fletcher: “Programme origination of 5-channel Surround Sound”. AES preprint 4430, 1997.
- [4] Russell Mason & Francis Rumsey, “An Investigation of Microphone Techniques for Ambient Sound in Surround Sound Systems”. AES preprint 4912, 1999.
- [5] Günther Theile: “Multichannel Natural Recording based on Psychoacoustic Principles”. AES preprint 5156, 2000.
- [6] Geoff Martin, Wieslaw Woszczyk, Jason Corey & René Quesnel: “Controlling Phantom Image Focus in a Multichannel Reproduction System”. AES preprint 4996, 1999.
- [7] Günther Theile & Georg Plenge: “Localization of Lateral Phantom Sources”. Journal of the Audio Engineering Society, Vol. 25, No. 4, 1977.
- [8] Geoff Martin, Wieslaw Woszczyk, Jason Corey & René Quesnel: “Sound Source Localization in a Five-Channel Surround Sound Reproduction System.” AES Preprint no. 4994, 1999.
- [9] Ulrich Horback: “New Techniques for the Production of Multichannel Sound”. AES preprint 4624, 1997.
- [10] Knud Bank Christensen & Thomas Lund: “Room Simulation for Multichannel Film and Music”. AES Preprint no. 4993, 1999.