PRODUCERS &
ENGINEERS WING

The Recording Academy's Producers & Engineers Wing

Recommendations For Surround Sound Production

Written by the
Producers & Engineers Wing Surround Sound Recommendations Committee

Chuck Ainlay
Joe Chiccarelli
Bob Clearmountain
Frank Filipetti
Leslie Ann Jones
Rory Kaplan
Jeff Levison
Bob Ludwig
George Massenburg
Howard Massey
Hank Neuberger
Phil Ramone
Elliot Scheiner
Eric Schilling
Al Schmitt
Jeff Skillen
Paul Stubblebine

Compiled and edited by Howard Massey

Graphics: Chuck Dahmer

Special thanks to:

Peter Cobbin
Lorr Kramer
Elliot Mazer
Steve Parr
Nick Peck
Ronald Prent
Jerry Stecking

© 2004 The National Academy of Recording Arts & Sciences, Inc.

Contents

Foreword

It has always been a dream of mine to create finer recordings. As a youngster I saw Cinerama and wished that we could someday bring its ideas to recordings. We tried Quad, and for some of us, discrete 6-track magnetic film, but it always seemed to elude the music world. Striving to have music sound like it does in our control rooms, we have achieved a lot in transposing the formats of the past. In front of us lies the best challenge in twenty years.

It's an honor to have the finest and most passionate engineers share their knowledge in this living document. This group offers standards with clarity and flexibility. Surround sound continues to grow from 5.1 upwards. The constant change deserves a platform and unity amongst us to drive forward. Rules are meant to be broken; however, we owe it to ourselves and the audio community to deliver music in all its spectacle and subtlety.

The group of contributors to this document share both practical and scientific data. We all wanted to translate studio and live events to a platform that still allows creative freedom. What follows is knowledge and experience, with never-ending striving for improvement.

My deepest thanks to the P&E Wing and the Recording Academy for making this paper possible.

Phil Ramone, Chairman, P&E Wing

Introduction

Perhaps the first rule of surround sound production is that there are no hard and fast rules. That may be the very reason why the medium is so attractive — and challenging — to audio professionals.

However, in order to achieve optimum results — that is, a surround sound mix that is aesthetically pleasing and translates well on the widest variety of playback systems — there are certain recommended practices that should be followed. That is precisely the purpose of this document. The Recording Academy's Producers & Engineers Wing consists of more than 5,000 members, including many of today's leading surround sound practitioners. It is our intention to provide a comprehensive set of guidelines and recommendations for the production of music and other types of audio (such as film and video postproduction, gaming, etc.) in surround sound.

With increasing numbers of producers, engineers and musicians making their initial foray into the world of surround sound, we also felt it was important to provide a primer of sorts — a summary of the current state of the art — as well as a description of current production practices, distribution formats, and encoding formats. Also included is a discussion of various relevant topics, including surround sound mastering considerations, sample rate conversion, downmixing, and upmixing. We intend for this to be a "living" document, with contents that will be periodically updated as new developments, technologies, and techniques are introduced. Comments and recommendations to the Surround Committee of the P&E Wing should be emailed directly to P&ESurroundSound@Grammy.com

It is our hope that this paper will not only serve as a foundation for the creation of more, and better, surround sound content but that it will also spark additional creative exploration in what is arguably the most exciting of all fields in audio today.

Summary Of Recommendations

- Discrete surround sound is the preferred method of delivering multichannel audio. *(Section 1.1, page 1-4)*
- The surround sound mixer should avoid mixing for one "sweet spot." *(Section 2, page 2-1)*
- Work should be evaluated on two or more distinctly different playback systems. *(Section 2, page 2-1)*
- The monitoring system in the professional mixing environment must deliver the full range of audible frequencies, and it must be positioned and calibrated correctly. *(Section 2.2, page 2-2)*
- Early reflections in the professional mixing environment should be suppressed. Appropriate amounts of low-frequency absorption should be deployed at least on the ceiling and on two of the four walls. As much diffusion as budget allows should be deployed. *(Section 3.1, page 3-2)*
- No two room dimensions should be equal. The ceiling height ideally should be greater than 11 feet. *(Section 3.1.1, page 3-2)*
- Background noise with all equipment powered on should not exceed 25dB SPL A-weighted. *(Section 3.1.2, page 3-2)*
- Appropriate amplification with sufficient headroom before clipping is required. *(Section 3.2, page 3-3)*
- Surround mixing should always be done on identical full range speakers of the same brand and model, plus a subwoofer. *(Section 3.2.1, page 3-3)*
- Surround mixes should be checked on a satellite speaker system, preferably one that emulates a typical consumer home theater environment. *(Section 3.2.1, page 3-3)*
- All speakers must be correctly calibrated so that they are not only equally matched in level, but so that their crossover frequencies are aligned to that of the subwoofer being used. *(Section 3.2.1, page 3-3)*
- All speakers must be wired in phase. *(Section 3.2.1, page 3-3)*
- Only direct radiator speakers should be used for music surround sound mixing. *(Section 3.2.2, page 3-4)*

- All five main speakers in a 5.1 configuration should be positioned along the circumference of an imaginary circle at whose center is the mixing position. *(Section 3.3.1, page 3-5)*

- It is absolutely critical that the signal coming from all five main speakers arrive at the mixing position (the sweet spot) at the same time. This is best accomplished by having all five speakers equidistant from the mixing position. If this cannot be achieved because of the physical layout of the room, disparity in arrival time can be corrected with the use of delay. *(Section 3.3.1, page 3-6)*

- The optimum speaker distance from the mix position is between 6.5 and 7.5 feet. *(Section 3.3.1, page 3-6)*

- The center speaker should be directly facing the center of the mix position. *(Section 3.3.1, page 3-6)*

- The front L and R speakers must be toed in, angled by approximately 30°. *(Section 3.3.1, page 3-6)*

- The front speakers should not be placed on the meter bridge of the mixing console and should all be at the same height — optimally at ear height of the mixing engineer, or approximately four feet off the ground. The rear speakers should be at the same height as the front speakers. *(Section 3.3.1, page 3-6)*

- The rear L and R speakers must be toed in, angled by 110° to 150°. Optimal rear speaker angling for most environments and genres of music is in the range of 135° - 150°. *(Section 3.3.1, page 3-6)*

- The subwoofer should be positioned in front of the mixing position, between the left and right speakers. *(Section 3.3.2, page 3-8)*

- Any low-pass filter internal to the subwoofer should be set no higher than 120Hz. *(Section 3.3.2, page 3-8)*

- Surround mixes should always be checked on a bass-managed satellite speaker system. *(Section 3.4, page 3-9)*

- Bass management during mixing is an option that may be employed at the discretion of the engineer. *(Section 3.4, page 3-9)*

- The bass management crossover frequency should be set to 80Hz. *(Section 3.4, page 3-9)*

- Reference listening level for surround sound production is in the range of 79 to 85dB C-weighted. *(Section 3.5, page 3-10)*

- An RTA is the preferred method of measurement when calibrating speakers in a surround sound system. If an SPL meter is instead used, set it to C-weighting on the slow scale. *(Section 3.6, page 3-10)*

- If compensatory delay is used, add .88 milliseconds of delay for each foot of distance disparity. The use of delay is not recommended unless absolutely necessary. *(Section 3.7, page 3-13)*

- Bus-to-channel allocation should be as follows: 1 = L, 2 = R, 3 = C, 4 = LFE, 5 = Ls, 6 = Rs. Tracks of a 5.1 master should be printed identically to the bus allocations. *(Section 4.2, page 4-2)*

- Whenever signal is placed into three, four, or five speakers, it should be decorrelated. *(Section 4.3, page 4-4)*

- The LFE channel should never be used to carry the bass content of the main speaker channels. *(Section 4.6, page 4-7)*

- The LFE channel should be low-pass filtered at 80 - 120Hz. In most cases, selecting a frequency between 80 and 100Hz will produce the best results. *(Section 4.6.1, page 4-8)*

- The surround mastering engineer must listen on a bass-managed system to check the results of combining low frequency signals. *(Section 5.3, page 5-9)*

- The center channel downmix coefficient should be set to -3dB in relationship to the left and right front channels and the rear channels must be at the same level as the front left and right channels. Surround mixes should always be checked in a typical downmix configuration. *(Section 5.5, page 5-11)*

- Producers and record labels should refrain from using upmixing tools except in situations where the original multitrack tapes are damaged, unavailable, or do not exist. *(Section 5.6, page 5-13)*

- A label should be placed prominently on all surround sound products to identify the sample rate and bit resolution of the original source material. *(Section 5.7, page 5-14)*

- A label should be placed prominently on all surround sound products to identify whether or not a true multichannel remix was done. Wherever a true multichannel mix has been done, the label should read: "The surround performance on this DVD / SA-CD is a full remix taken from the original multitrack masters." Wherever upmixing tools have instead been used, the label should read: "The surround performance on this DVD / SA-CD was electronically recreated from the original stereo source without the benefit of the multitrack masters." *(Section 5.7, page 5-14)*

1. Historical Perspective

The physiology of the human ear allows us to listen in surround; thus it is a more natural experience than mono or stereo can ever be. The history of surround sound production can be traced back to the Renaissance composers who wrote antiphonal church music designed to fill large cathedral spaces with sound, often to the extent of incorporating side-to-side or front-to-rear (choir / pipe organ) spatial effects. Centuries later, the Berlioz *Symphonie Fantastique* was scored for horns at the rear of the concert hall, and Wagner wrote mighty works for orchestras so large that musicians had to play under the stage, in the foyer, and scattered around in the audience!

Throughout the 1930s, scientists at Bell Laboratories experimented with various multichannel audio formats, including three-channel stereo (left, center, right). In 1938, Walt Disney conceived the idea of adding surround sound to his upcoming cinematic release, *Fantasia*. Accordingly, Disney engineers developed a technology called Fantasound, which stored three channels of audio and a control track on the film itself, with playback through five channels: three front speakers and two rear ones — a speaker configuration that, sans subwoofer, was remarkably prescient of the 5.1 arrangement in common use today. In the process of recording the film's soundtrack, those same engineers also — astonishingly — invented panning, multitrack recording, and overdubbing! Subsequent war efforts effectively curtailed further development of surround sound in entertainment, but the military did discover an important use for multichannel audio: dedicating independent speakers for different alarm sounds (i.e., bell, horn, klaxon) aboard warships in order to make it easier for sailors to instantly identify sonic cues in an emergency.

By the early 1950s, the dominance of the movie as the public's main source of entertainment was threatened by the growing popularity of television. In what proved ultimately to be a losing battle, the film studios fought back with an arsenal of technological advances, some of which were simply gimmicks (i.e., 3D glasses) and some of which eventually matured into meaningful improvements. In the latter category was wide-screen Cinemascope, accompanied by four-channel stereo (three front speakers and a fourth, switchable, "effects" speaker). This was soon followed by the introduction of 70mm 6-track film, which included five front channels (left, left/center, center, center/right, and right) and a dedicated effects channel situated in the auditorium. However, economic factors such as high production and theater conversion costs limited the acceptance of these technologies, especially since the rise of television was making significant inroads into the profit margins of the film industry.

Of course, ever since the development of the Edison recording system at the turn of the twentieth century, consumers had also been listening to music in their homes. In the 1920s, radio took over as the predominant means of delivering music, but the public never lost interest in their record players. By the 1960s, home "hi-fi" systems were all the rage, followed quickly by the acceptance of stereo sound (as opposed to monaural) as a standard. The early stereo recordings of a ping-pong ball bouncing between speakers served to whet the consumer's appetite for spatiality, effectively sowing the seeds for the rise of today's surround sound. By the 1970s, many audio professionals were beginning to experience frustration at the limitations of two speakers, leading to the first exploration into consumer surround sound, known as Quad.

Quad was an analog four-channel system beset by a multitude of problems, ranging from the technical (the vinyl medium was limited in its ability to carry four discrete signals without significant crosstalk and compromises in frequency response) to the economic (there were numerous competing, non-compatible formats) to the aesthetic (the psychoacoustics of four-channel sound were not well understood at the time, and so the few releases that saw the light of day were poorly executed). Though it ultimately failed, it was a pioneering idea that was simply ahead of its time.

But by the late 1970s, early digital audio devices were starting to make an appearance, moving slowly out of the research laboratories into the pro audio marketplace. The film industry was not standing still, either. Dolby Laboratories began their development of a matrixing system called "Dolby Stereo," whereby four channels of information (designed to be delivered to three front speakers and an array of rear speakers) were derived from two stereo channels printed optically onto film. Because the cinematic reproduction systems of the era were unable to generate loud low frequency signals without clipping, a separate "boom" subwoofer channel (called the "LFE channel," for Low Frequency Effects) was introduced. This had the added advantage of effectively increasing dynamic range in theatrical installations. By the mid 1980s, a standard of sorts for the audio portion of 70mm theatrical releases was in place: three full-range front channels (left/center/right), two full-range rear channels, and a subwoofer — a system which was named "5.1" (the LFE channel is the ".1," since it only carries low frequencies and therefore has approximately one-tenth the bandwidth of the other channels). This was seen by researchers as being the minimum number of speakers required to provide an immersive, enveloping experience for the listener, while still providing a sufficient degree of localization (the ability to perceive a sound as coming from a specific point in space). 5.1 still remains the basic configuration for surround sound today.

In the early 1990s, the European standards organization known as the ITU (International Telecommunications Union) began conducting research to determine optimum speaker placement in a 5.1 configuration. This culminated in a document published in 1994 entitled "Recommendation for Multichannel Stereophonic Sound System With and Without Accompanying Picture" (Rec. ITU-R BS.775-1), which was largely accepted as a de facto industry standard. However, it is worth noting that the ITU research was done almost entirely with classical music as source material, and that the rear speakers were characterized as "ambience" or “effects” speakers only. Because the ITU Recommendation was created well before the development of modern surround sound mixing methods which give equal importance to all five main speakers, it may not be ideal for many music applications. (See section 3.3.1 and section 4.)

Throughout the 1990s, the focus was on finding ways to incorporate surround sound into the second-generation digital audio systems that were proliferating. Digital audio workstations (DAWs) began to play a vital role in recording studios worldwide, digital audio for film was becoming commonplace, and plans were being drawn up for a digital television broadcast system that would evolve into High Definition Television (HDTV). In that same time period, no less than three major film digital surround sound encoding formats were introduced: Dolby® Digital™ (also known as AC-3), DTS™, and Sony SDDS™. In addition, the widespread introduction of laser discs (which had the capability of carrying encoded surround sound) served as a springboard to a nascent industry called home theater. Suddenly, people were beginning to add extra surround speakers and subwoofers to their home listening environments.

The introduction of the DVD in the mid-1990s — the fastest-growing consumer format ever released — finally provided a universally accepted medium capable of delivering surround sound to the mass public in the comfort of their living rooms. With the consumer delivery medium in place, professional audio manufacturers began offering a variety of surround sound production tools to complement the wide range of digital audio products that fired the rise of the home recording studio. Digital multitrack systems are now available for a fraction of the cost of their analog equivalents just a few years ago, and components like surround-ready mixing consoles, effects units and monitoring systems are finding their way into home and professional studios alike.

Today, surround sound has become a staple of home theater and is also starting to make its appearance in the automobile. Multichannel audio has found its way to digital television broadcast, into gaming consoles, and onto the Internet via streaming codecs such as Windows Media Audio 9 Professional. Perhaps most importantly, the advent of high definition formats such as DVD-Audio and SA-CD Multichannel allow the surround sound producer to distribute content without sonic compromise. It is fair to say that surround sound has finally come of age.

1.1 Discrete vs. Matrix

There are two basic ways of deriving surround sound: *discrete* and *matrixed*. In discrete surround sound, there are a number of individual channels, each containing unique audio information, and each devoted to a particular speaker (or group of speakers). For example, a discrete 5.1 system contains six channels: one for each of the five main speakers and a sixth ".1" channel dedicated to the subwoofer. The audio data streams on DVD-Video, DVD-Audio, and SA-CD (see section 5.1) are all discrete, as are the Dolby Digital (AC-3), DTS, MLP, and DST encoding formats (see section 5.2).

In contrast, a matrixed system encodes extra channel audio information into a stereo signal and then recovers those channels through a mirror-image decoding process. Common matrix decoders found in home receivers include Dolby Surround Pro Logic® II (often simply called "Pro Logic"), SRS Labs Circle Surround, and DTS Neo:6. Although a stereo signal containing matrixed surround information can be played back through conventional two-channel equipment, if it is instead routed to a decoder, amplitude and phase differences between the left and right channels are used to recreate the original center and surround channel information. Matrixed surround is employed by analog television broadcasts, VHS tapes, laser discs and computer and video games.

There are two inherent limitations of a matrixed system, the first of which is limited separation (in other words, relatively high amounts of cross-talk) between channels, thus reducing the amount of localization possible. Secondly, there can be "steering" problems, where a signal shifts uncontrollably between speaker channels (or is routed to the wrong speaker altogether) because of complex phase correlations in the stereo signal (i.e., when there are similar signals in the left and right channels). Because it is free of these limitations, <u>discrete surround sound is the preferred method of delivering multichannel audio</u>.

1.2 Subwoofer vs. LFE

It is important to understand the distinction between the subwoofer and the LFE: the subwoofer is a physical speaker, while the LFE (the ".1" in "5.1") is a channel. One function of the subwoofer is to reproduce the information carried by the LFE channel. Another, optional, function is to also reproduce the low frequencies of the other channels. This latter application, called *bass management*, is described in detail in section 3.4.

Where a surround sound monitoring system consists of satellite speakers only (see section 3.2.1), a subwoofer is a necessary component, and bass management is required; without it, low frequencies will not be reproduced at all. In a professional surround sound mixing environment, which almost always uses full range speakers, a subwoofer is not necessary for sonic purposes, but is nonetheless required to reproduce the LFE channel. Section 4.6 describes the usage of the LFE channel in various applications.

1.3 Alternate Surround Sound Configurations

There are a number of alternate surround sound configurations that are variations on the standard 5.1 scheme.

These include the 6.1 format employed by Dolby-EX™ and DTS-ES™. (See sections 5.2.1 and 5.2.2) 6.1 surround adds a rear center speaker, removing the need to create a phantom rear image (thus making the precise positioning of the rear left and right speakers somewhat less critical), improving localization, and enabling enhanced panning options such as true flyovers from front to back, which can be accomplished simply by panning a signal from front center to rear center. Because that panning then occurs between two single point sources, comb filtering artifacts are also reduced. (See section 4.3)

One experimental variation on 6.1 is to use the additional channel as a height channel — actually mounting the speaker in the ceiling, facing downwards — instead of as a rear center. Some engineers even use the LFE channel of a 5.1 mix (which can carry full-range information if it is not Dolby Digital or DTS-encoded) for that purpose.

Another speaker arrangement in current use is the theatrical 7.1 format used by Sony SDDS™. This places five full-range speakers behind the screen — left, left center ("left extra"), center, right center ("right extra"), and right — along with two arrays of rear surround speakers, plus a subwoofer. The main advantage of this system is improved localization, particularly when applied to dialog. For example, two "talking heads" onscreen might be assigned to the left center and right center speakers, with the soundtrack panned hard left and right and sound effects routed to the center channel. Some consumer systems also use a 7.1 speaker arrangement, but with the two extra channels serving as sidefills.

There is also a proposal currently in development for a second-generation surround sound format consisting of ten main speakers and two subwoofers. This "10.2" system adds to the basic 5.1 configuration by providing stereo subwoofers, a rear center channel, left and right sidefills, and two "proscenium" speakers positioned on either side of the front center, but several feet higher, to provide height information.

Because all of these variations are relatively new and currently not well supported, this paper will focus on standard 5.1 techniques and practices.

2. Surround Sound Environments and Setup Diagrams

The art as well as the technology of surround sound is still young. For that reason, it is difficult to describe one "ideal" monitoring/evaluation space. Experience in multichannel mixing thus far, however, has underscored the importance of the following two criteria:

- <u>The surround sound mixer should avoid mixing for one "sweet spot"</u>
- <u>Work should be evaluated on two or more distinctly different playback systems</u>

Often, making just minimal compromises in the sonic presentation can yield a more effective presentation across different playback environments.

This section will provide generic descriptions of the five main surround sound environments:

- The professional mixing environment
- The consumer listening environment (home theater)
- Theatrical exhibition
- Automobile
- Club environments

Also included in this section are suggested setup diagrams.

2.1 Envelopment and Localization

Two key terms in defining surround sound environments are *envelopment* and *localization*. *Envelopment* refers to the perception of sound being all around the listener, with no definable point source. *Localization* refers to the ability to identify where a particular sound is coming from. The goal of a successful surround sound production is to establish an accurate balance between the two. One way to view the surround soundstage is to think of it as a blank canvas, on which can be painted both broad washes of sound (envelopment) and highly defined narrow brush strokes (localization) that add “dabs” of color.

2.2 Professional Mixing Environment

A professional mixing environment is, by definition, an acoustically tuned room that optimally provides flat frequency response across the full bandwidth. Of course, no room is perfect, and compromises can and often are made, but the goal is to provide the mixing engineer with an accurate picture of the sound that has been recorded and to allow a balanced blend between envelopment and localization. To that end, the monitoring system must deliver the full range of audible frequencies, and it must be positioned and calibrated correctly. (See sections 3.3 and 3.6)

All main speakers should be identical, of the same brand and model. Only full range direct radiator speakers should be used; satellite and dipole speakers have no place in the professional mixing environment (see section 3-2). Mid-field monitoring is usually preferred for surround mixing. (Unlike nearfield monitors, mid-field monitors are designed to be used free-standing and not placed on top of a console meter bridge.) In the interest of uniform frequency response, all main speakers should be placed on speaker stands; the front speakers should not be placed on top of the console meter bridge. The use of movable speaker stands can be helpful if the rear speakers are to be shifted or angled differently from project to project because of genre-specific considerations (see section 3.3.1).

At least one subwoofer must be used, ideally positioned along a boundary wall in front of the mix position (see section 3.3.2). Bass management in the professional mixing environment is optional and at the discretion of the engineer; however, depending upon the specific monitoring system being used and the room design, it may prove to be unsatisfactory and yield inaccurate results. We do, however, recommend that there be a separate bass-managed consumer system available (optimally installed in a room that emulates the home theater environment — see section 2.3) to check mixes on.

The illustration on the following page shows a typical professional 5.1 mixing environment.

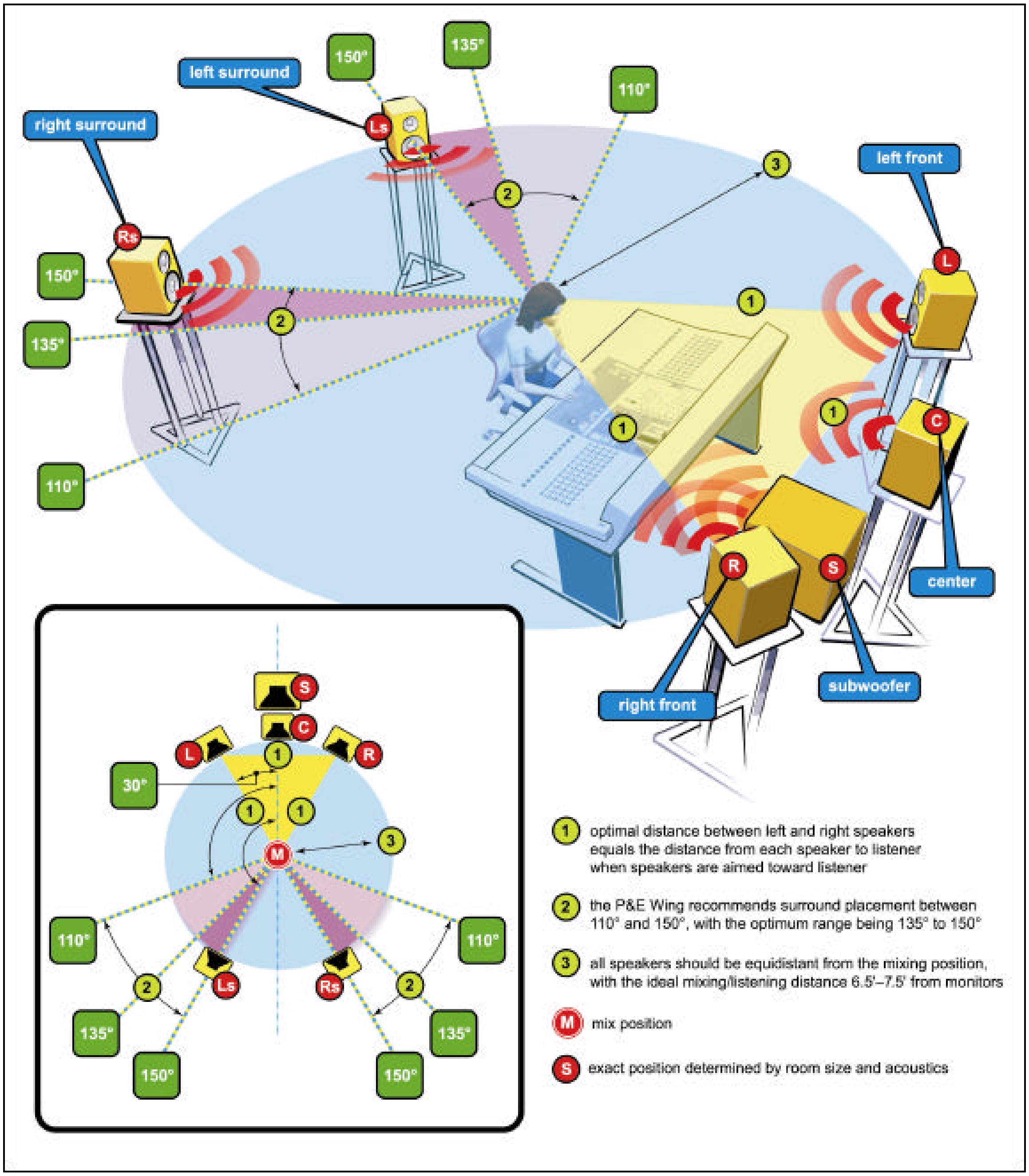


Figure 1 – Typical professional 5.1 mixing environment

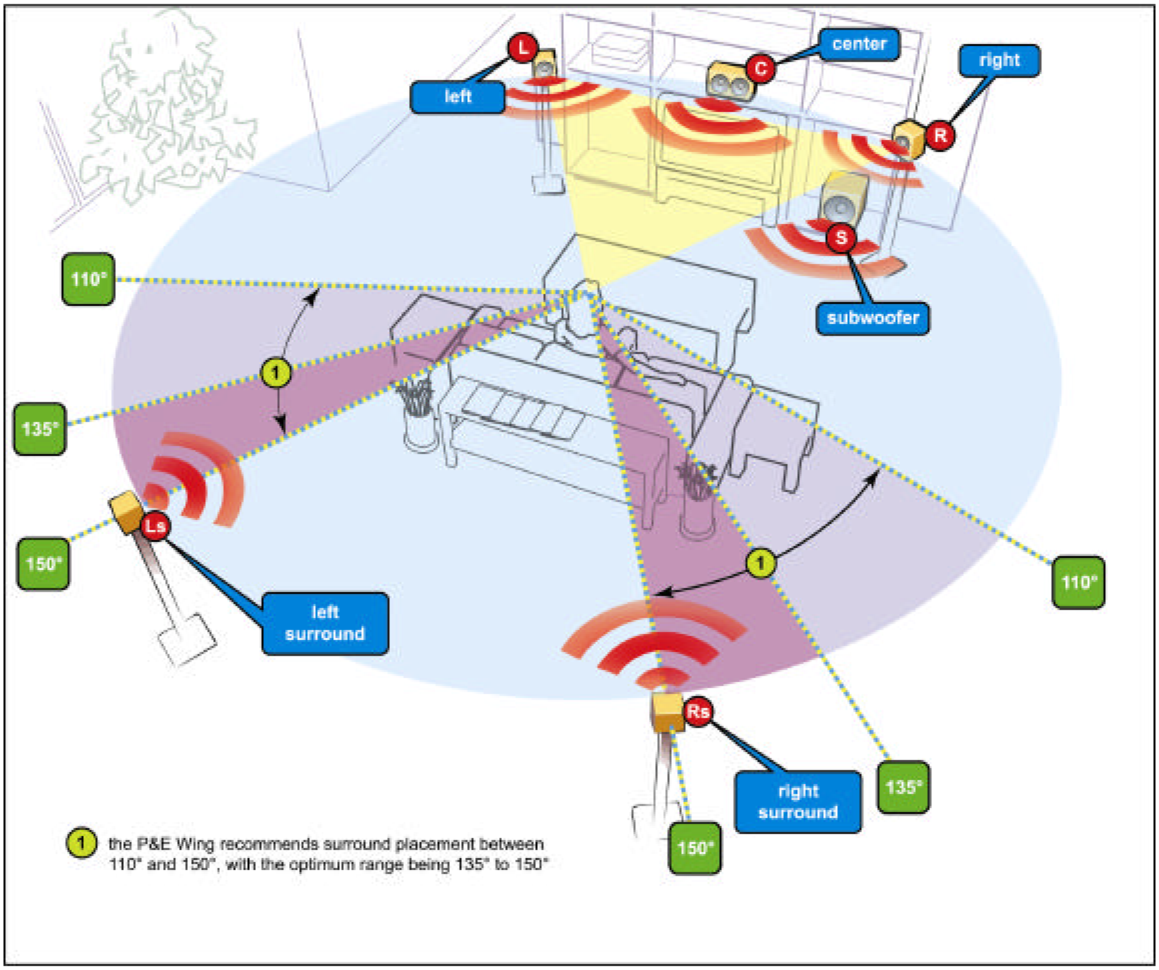
2.3 Consumer Listening Environment (Home Theater)

The typical home theater is installed in a rectangular room with no acoustic treatment. Despite the fact that the ceiling heights of average homes are more or less in the same range, room lengths and widths vary widely; thus the amount and nature of early reflections are unpredictable. The room will often be too absorptive due to the presence of carpeting and soft furniture; however, the furniture may not be soft enough to "deaden" the room reverberation at the most troublesome frequencies. In other circumstances, the room may be too reflective due to sheet rock walls and hardwood floors. Indeed, it is rare to find a home theater in a room that is acoustically well-tuned. In addition, in systems that allow the consumer to make interconnections with open wiring (as opposed to the use of jacks and plugs), there is tremendous potential for one or more of the speakers to be out of phase.

Home theaters are usually outfitted with satellite, as opposed to full range, speakers. Sometimes the rear speakers are dipoles, although in recent years the trend has been more towards direct radiator speakers. Many "home theater in a box" products offer extremely small satellites, with drivers as small as three inches — really more like computer speakers. The center speaker is usually horizontally oriented to accommodate a video monitor; the other satellites are usually vertically oriented. Sometimes the center and/or rear speakers are significantly smaller than the front left and right speakers, with differing frequency responses. Some home theater systems have two center speakers (one front and one rear), to permit playback of 6.1-encoded DVDs, and some do not even contain a single center speaker! The subwoofer can range in size from six to twelve inches; higher end home theater systems tend to provide larger subs. Bass management is almost always employed, often as a single filter after the summation of the main channels and LFE channel. The most typical crossover frequency used is 80Hz, though higher frequencies are sometimes utilized.

The positioning of the speakers in a home theater environment may be all over the map. Often, in a 5.1 configuration, the three front speakers are simply placed on top of the video monitor, in a straight line. Even if the left and right speakers are spread out on a bookshelf on either side of the video monitor, they are rarely angled in, and the center speaker will be placed either above or below the video monitor and not in line with the other front speakers. Rear speaker placement can be even more haphazard. Often, the rear speakers are positioned on the side of the listening area instead of behind it. If they are placed to the rear, they are often hidden behind furniture (thus producing a muffled sound) or mounted up on the wall behind the listening position, firing well above the listener's head. As a result, arrival times can vary so widely that there may be no "sweet spot" whatsoever where the listener can actually enjoy a correctly balanced blend of the speaker channels without phase smearing. Beyond all of that, the listeners' positions in a home theater change as they move around the room.

The biggest difficulty in creating content destined to be listened to in home theaters is their utter lack of consistency — there really is no "typical" consumer surround sound listening experience. Creating a surround sound mix that translates well in the broad range of home theater environments is indeed a lofty challenge!



Typical home 5.1 listening environment

2.4 Theatrical Exhibition

Discrete, as opposed to matrixed surround sound, has become commonplace in movie theaters, which are frequently equipped with 5.1, 6.1, and 7.1 speaker systems. One impetus in that direction is the recent trend towards live satellite broadcast of music concerts, beamed to theatrical exhibition spaces. Film is no longer the only medium enjoyed in cinemas.

Beyond the fact that cinemas are considerably larger than home theaters, theatrical exhibition spaces are characterized by the use of surround arrays. These arrays will consist of several speakers positioned on both to the sides and rear of the auditorium, typically evenly spaced in a region between 60° and 150°. Because hundreds of people may be seated in a movie theater (and since only a very few can occupy the soundstage mixing "sweet spot" at 2/3rds back and center), the goal here is envelopment rather than localization, and the use of many speakers aids greatly in distributing the sound evenly throughout the audience seating area.

The direct radiating speakers that comprise the surround arrays in cinemas are typically smaller than the front screen speakers. Surround arrays can have a reasonable frequency response due to the fact that there are usually many low frequency drivers present; frequencies below 125Hz will sum even though each individual speaker may have limited low frequency response. High frequencies radiating from surround arrays will be comb filtered and will therefore sound diffused and different than that coming from the front speakers.

The front speakers in a theatrical installation are located behind the screen and will generally exhibit some attenuation of the high frequencies and a resulting smear of the image in the upper octaves. These speakers are typically of a traditional two-way design, consisting of a 2-15" ported low frequency cabinet and a horn/compression driver, with the crossover most usually set to 500Hz. Theatrical subwoofers generally have a frequency response of approximately 35Hz - 120Hz; so-called "shaker" speakers are sometimes deployed to reproduce ultra low frequencies. Bass management is never used in standard theatrical sound systems.

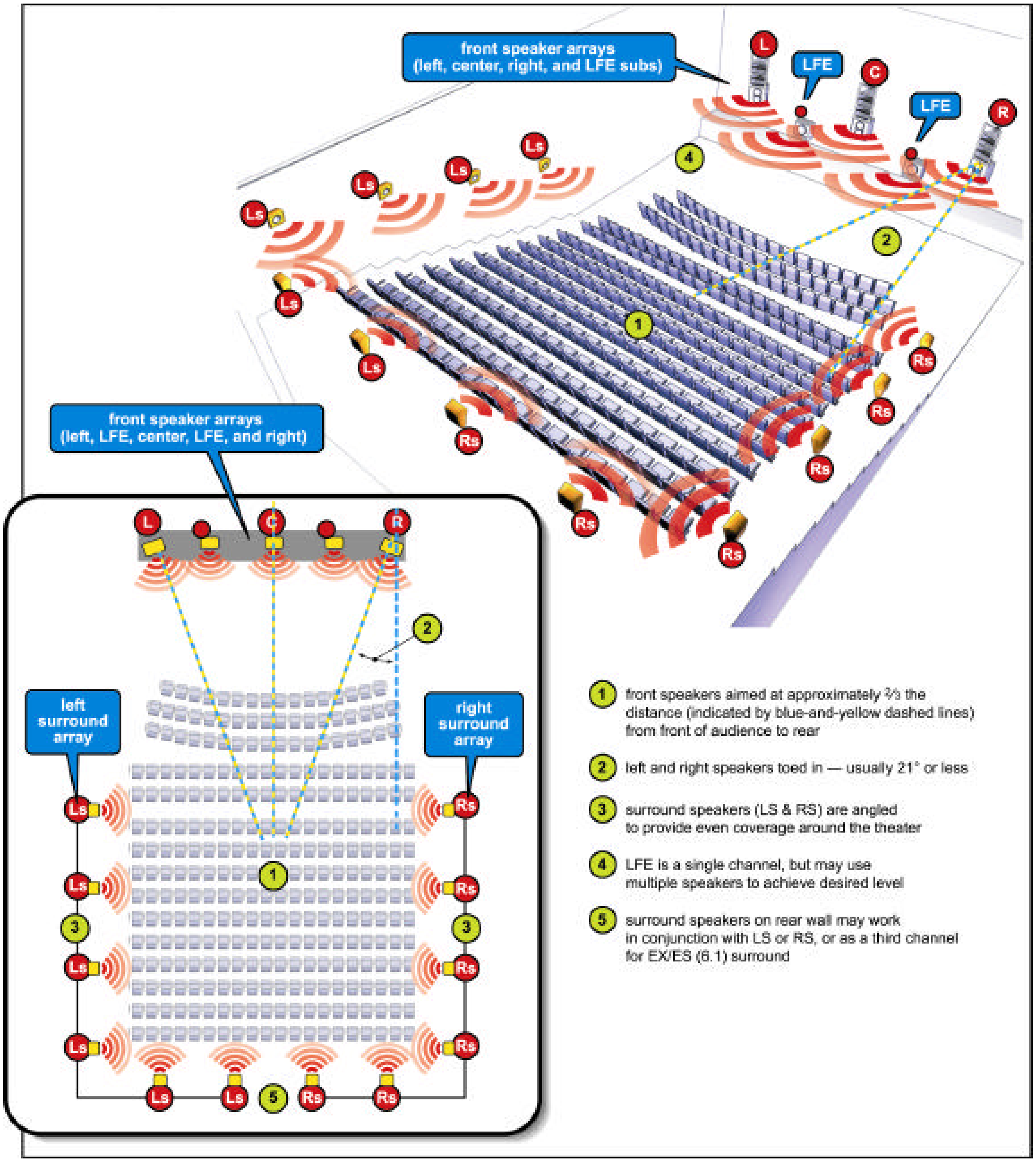
Virtually all cinema playback systems today are aligned using SMPTE 202M or ISO 2969. The two documents are virtually the same and both define the measuring method and the resulting frequency response known as the "X-Curve". This provides a uniform frequency response adjustment for all theaters throughout the world.

Cinema playback is also tuned to a specific level. All standard cinema systems are optimized for 85dB SPL (2/3rds back in the center) from each front channel and 82dB SPL for each of the rear channel arrays. All channels should have 20dB of headroom. The LFE channel is set at 10dB of in-band gain; that is, 10dB greater than the screen channels in each 1/3 octave frequency band. SPL ranges from 88dB to 92dB, depending upon the specific bandwidth of the LFE system in use.

Due to power issues, most front speakers will perform down to 45Hz, but the response below that point will fall off rapidly. Similarly, while the typical system will exhibit a frequency response out past 16kHz, the headroom will acoustically compress near 18dBfs. A slight amount of this compression will not necessarily sound bad unless abused. Mixers should view the frequency content in the highest couple of octaves with this in mind.

Delays in cinema are not coherent. The front speakers are typically mounted in a fairly flat screenwall, with the L and R speakers toed in at an angle of 21° or less. As a result, for most listeners, the center channel arrives approximately 1ms before the left and right channels. The surround channels have an additional delay of 10ms, plus the natural delay from the front channels to the listening position. For this reason, phantom image panning is not dependable from theater to theater, or from listening position to listening position. Signals shared between channels (sometimes called "shouldering") will combine only up to 600Hz or so. Mixes will translate best when decorrelated between channels (best described as taking a "multichannel mono" approach).

The illustration on the following page shows a typical theatrical surround sound environment.



Typical theatrical surround sound environment

2.5 Automotive Audio

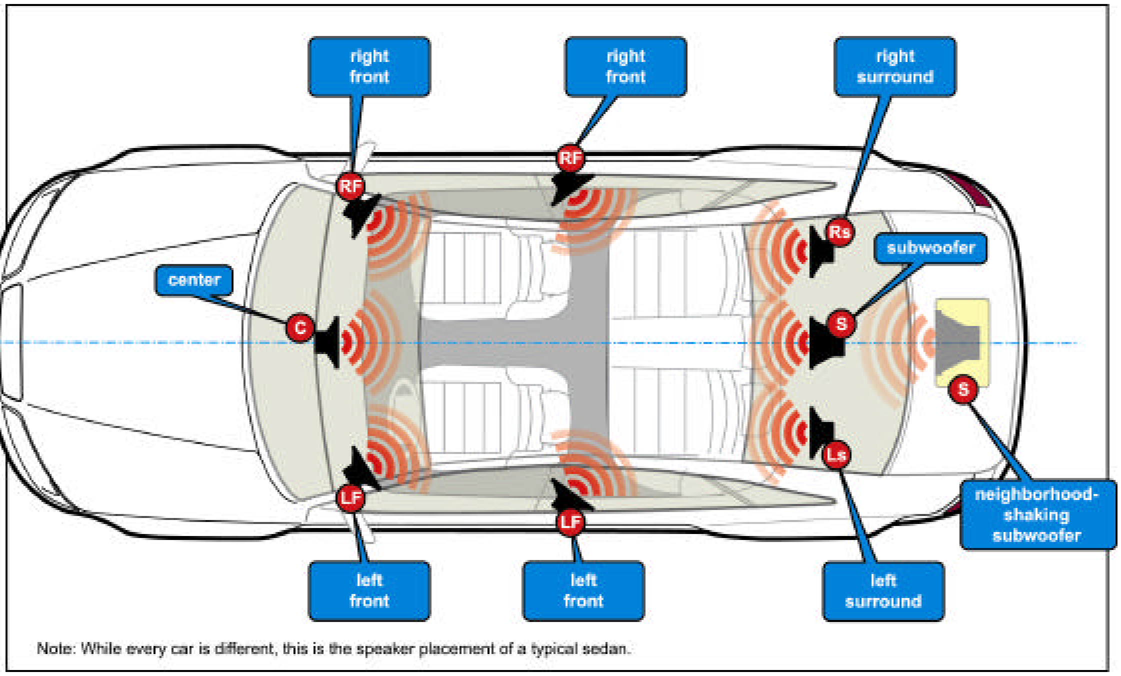
Like the theatrical exhibition space, the automotive environment provides a fixed, controllable set of elements. While the acoustic treatment in a car may be less than optimal (background noise especially can pose a problem), it will remain more or less constant, though, as in the cinema, it changes somewhat depending upon the number of audience members (passengers). More importantly, the speaker positions in a car are fixed. This ensures a consistent listening experience at both the driver's position (which serves as a highly focused "sweet spot" in most automotive audio designs) and in the various passenger positions.

The main criteria in the design of an automotive audio system is the limited real estate available. As a result, the speakers in automobiles tend to be quite small and are almost always satellites — there are few if any full range speakers which can fit! Typical main speaker driver sizes range from seven inches down to just an inch or two, with the center speaker sometimes smaller than the others.

Automotive audio components have to be squeezed in amongst the other instrumentation, and every manufacturer has their own idea about where speakers should be placed. The center speaker is usually installed in or near the center of the front dashboard, while the left and right speakers are most often placed in the front doors. In some instances, a split design is utilized, with the left and right tweeters placed on the sides of the dashboard and the woofers installed in the front doors. This kind of design tends to aid in dispersion and envelopment, as well as reducing the muffled sound that comes from listening to a speaker somewhere near calf-level. The rear speakers in an automotive surround sound system are either mounted on the rear deck or in the rear door panels.

Automobile subwoofers run the full gamut of sizes, from a modest six inches all the way up to a full eighteen inches (for the bone-rattling ride of your life). They are mounted either on the rear deck or in the trunk. Even the smallest automobile subwoofer is capable of producing low frequencies at significant amplitude, since they benefit from the extremely tight interior space, which facilitates the generation of standing waves free of ultra low frequency modes. Mode problems in small automotive spaces start much further up in the spectrum and can then be problematic.

Controls in an automotive surround sound system tend to be more rudimentary than in the home theater. Sometimes graphic equalizers are provided; more often, simple bass and treble controls are all that are available. Balance controls (between both left and right and front and back) are almost always present, and there are sometimes separate level controls for the center speaker and subwoofer. It is worth noting that most automotive surround sound systems deliver signal via the upmixing of stereo sources instead of providing a discrete multichannel player. See section 5.6 for more information.



Typical automotive surround sound environment

2.6 Club Environments

Clubs are becoming an increasingly visible area of multichannel reproduction. Collaboration with venues such as England's Ministry of Sound have resulted in learning specific ways of mixing music to cater for a dance market. Here it is important to mix in a non front/back-specific way: the listening audience is not necessarily stationary or facing the same direction and it can be important for rhythmic elements to be more homogenous through all the speakers than would be normally desired for home theatre or the automobile. Also, sound systems in clubs are normally full range — often even hyped in the low end — so LFE information can tend to make the mix sound abnormally boomy unless it is used with great discretion.

3. Surround Sound Monitoring In The Professional Mixing Environment

This section addresses the various issues concerning monitoring for surround sound mixing. To some degree, the recommendations provided here are application-specific; that is, the requirements for music mixing can vary greatly from that of film mixing. In fact, some recommendations are even genre-specific: for example, speaker positioning for pop music mixing, where the surround monitors are typically treated as "equal partners," often carrying key musical elements, differs significantly from that of classical music, where the rear speakers generally are used only for room ambience. Nonetheless, there are certain recommended practices that should be adhered to in most if not all circumstances.

Room design is discussed in section 3.1. Recommendations are made regarding dimensions, acoustic treatment, and minimum ambient noise requirements.

Of course, a key element in any professional mixing environment is an accurate monitoring system, one which is free of distortion and coloration. Definitions of popular speaker types used in surround mixing, with recommendations for their use, are provided in section 3.2.

Also, because the aim of a successful surround mix is to achieve a balance between envelopment and localization (see section 2.1), the placement of the monitors is an equally critical issue. Comprehensive recommendations for speaker positioning is provided in section 3.3.

It is important to remember that most consumers do not listen to surround sound on state-of-the-art monitoring. Surround mixes should therefore be referenced on systems that approximate the typical home theater experience and the effect of bass management (utilized by most consumer systems) should also be checked. Section 3.4 addresses bass management in detail.

Once positioned, the speakers in a surround mixing environment must be calibrated correctly. Recommended reference listening level is described in section 3.5 and the recommended procedure for speaker calibration is given in section 3.6.

Finally, the formulas for compensatory delay, when required, are given in section 3.7.

3.1 Room Design and Acoustic Treatment

From experience gained thus far, it seems that the average listener adapts over time to asymmetrical (both left-right as well as front-back) speaker positions. The goal of creating the best average presentation across the main speakers in a surround sound environment can best be attained by having more diffusion in the mixing environment rather than less. A neutral, as opposed to absorptive, monitoring room is optimum.

To as great a degree as possible, early reflections should be suppressed. Because the rooms that home theaters are typically installed in have no identifiable pattern of early reflections, significant early reflections in the professional mixing environment will be misleading. Appropriate amounts of low-frequency absorption should be deployed at least on the ceiling and on two of the four walls.

In addition, there should be as much diffusion as a budget will allow. From simply deploying everyday furniture and artifacts (shelves and bookcases, marble statuettes of one's significant other, etc.) to full-on quadratic residue diffusers, increasing diffusion helps flatten the spectral response of a room.

To summarize: the more uniform (diffuse) the ambience in the professional mixing environment, the more site-independent the resultant mixes will be.

3.1.1 Dimensions

No two room dimensions should be equal. Preferably, all three dimensions should vary by well-known ratios such that the low-frequency nodes generated in the three axes of a rectangular model are spread uniformly throughout the low three to four octaves. The ceiling height ideally should be greater than 11 feet.

3.1.2 Background Noise

Background noise with all equipment powered on should optimally not exceed 25dB SPL A-weighted.

3.2 Monitoring Recommendations

This section contains recommendations for speaker type. In all circumstances, appropriate amplification with sufficient headroom before clipping (at the recommended listening level; see section 3.5) is an absolute requirement. Self-powered mid-field monitors often provide the best solution for surround monitoring; in addition to the convenience factor, they can be easily moved and repositioned if necessary (see section 3.3). They also come with matching amplification and obviate the need for speaker cabling.

3.2.1 Full Range vs. Satellite

For the purposes of this document, "full range" speakers are defined as those which are capable of reproducing frequencies of at least 18kHz or higher at the high end, and 40Hz or lower at the low end. Surround mixing should always be done on identical full range speakers of the same brand and model, plus a subwoofer.

"Satellite" speakers are typically found in consumer home theater systems. They are much smaller in size than full range speakers, and have a very limited low end response, relying instead on a subwoofer to deliver bass frequencies. Surround mixes should be checked on a satellite speaker system, preferably one that emulates a typical consumer home theater environment. (See section 2.3)

All speakers must be correctly calibrated so that they are not only equally matched in level, but so that their crossover frequencies are aligned to that of the subwoofer being used in order to ensure a flat frequency response. This is important whether bass management is being used or not. Of course, as with stereo mixing, it is critical that all speakers be wired in phase. See section 3.6 for detailed instructions on how to correctly calibrate all surround monitors, including the subwoofer.

3.2.2 Direct Radiator vs. Dipole Speakers

Most loudspeakers designed for professional applications are of the direct radiator variety, where the speaker cone fires in one direction only. This type of design provides the smoothest frequency response as well as the best localization qualities — it is easy to perceive exactly where the sound is coming from.

Despite the fact that most theatrical installations use direct radiator speakers (albeit in arrays — see section 2.4), some consumer home theater systems employ dipole speakers in an effort to emulate the wash of sound that occurs in a large room such as a cinema. Dipole speakers radiate signal in opposite directions at the same time, at equal energy but with opposite polarity. Thus, they can provide better envelopment than direct radiator speakers, but with very poor localization. Also, due to their construction, dipole speakers have limited bass response (unless they are quite large) and so generally require the addition of a subwoofer, making them fall into the category of "satellite" speakers. (See section 3.2.1)

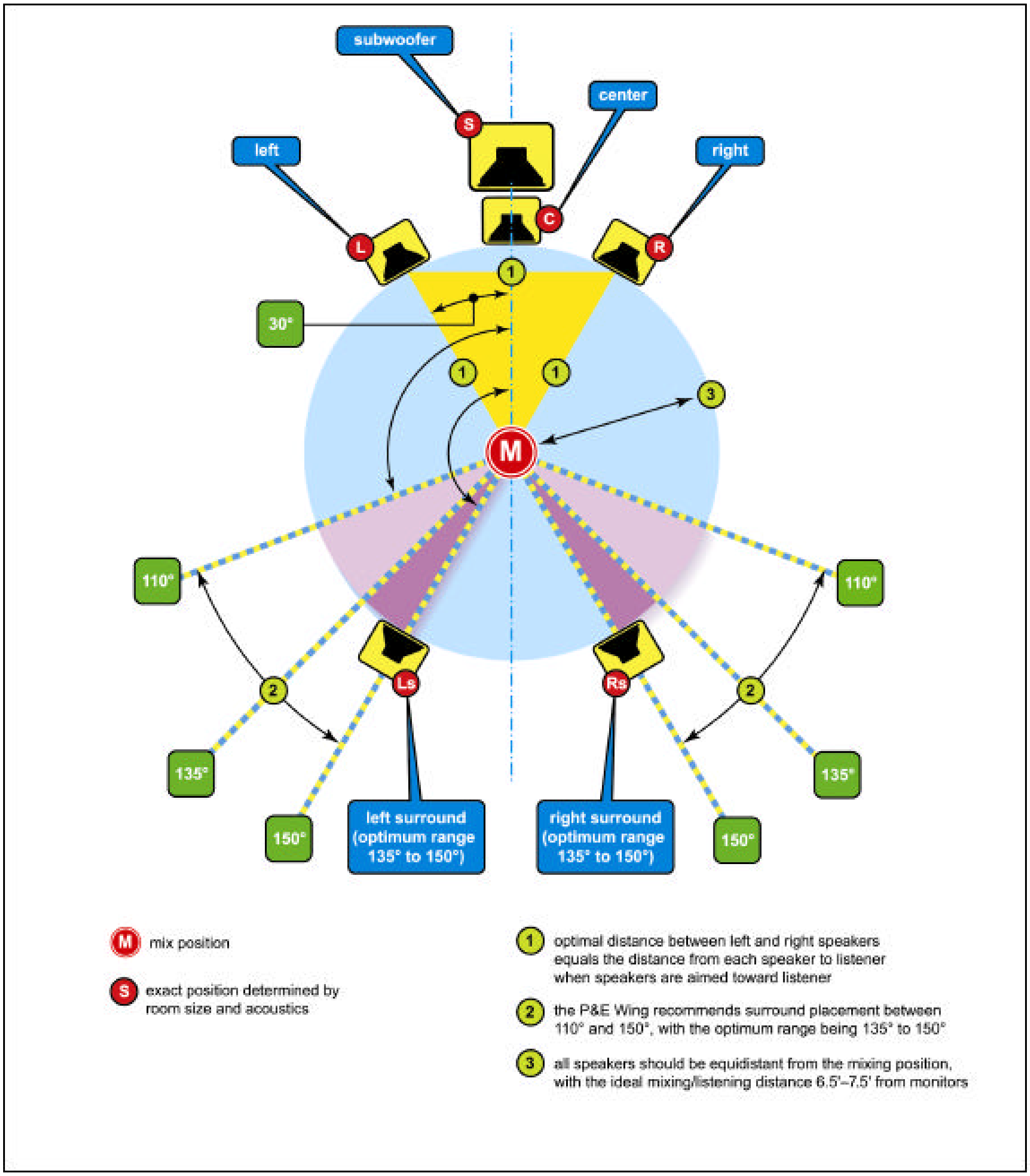
<u>Only direct radiator speakers should be used for music surround sound mixing</u>. Because of their dependence upon reflective surfaces and the phase anomalies inherent in their design, dipole speakers will produce inaccurate results when placed in an acoustically tuned professional mixing environment. However, it may be helpful to check a surround mix on a consumer system equipped with dipole speakers, provided the system is physically located in a room that simulates a typical home theater listening environment. (See section 2.3)

3.3 Speaker Placement Recommendations

The goal of monitoring is, of course, to provide the engineer with as much accuracy as is possible. Much of this is down to the quality of the speakers and amplification being used, but in the world of surround sound, speaker positioning is an equally critical element. Moving speakers by just a few inches in distance or rotating them by just a few degrees in direction can cause a significant change in imaging, in frequency response, in envelopment / localization characteristics, or in some combination of all three.

That said, there are often physical conditions (such as room size or dimension) beyond the control of the engineer that dictate speaker placement in the professional mixing environment, and there is certainly no way to control how a consumer listening to surround sound situates *their* speakers. For these reasons, there is no intrinsically "correct" way to position speakers for surround sound production. Instead, it is our intention to provide a range of acceptable options, with a discussion of the applications best suited by each approach.

3.3.1 Main Speaker Placement



Recommended 5.1 surround sound speaker placement

The illustration above indicates the range of recommendations for speaker placement. As shown, <u>all five main speakers in a 5.1 configuration should be positioned along the circumference of an imaginary circle at whose center is the mixing position</u> (sometimes called the "sweet spot").
In order to avoid phase cancellation and comb filtering problems, <u>it is absolutely</u>

critical that the signal coming from all five main speakers arrive at the mixing position at the same time. This is best accomplished by having all five speakers equidistant from the mixing position. If this cannot be achieved because of the physical layout of the room, disparity in arrival time can be corrected with the use of delay. (See section 3.7.)

The optimum distance from the mix position is between 6.5 and 7.5 feet, depending upon the particular monitors being used and the size of the room.

In a 5.1 configuration, the "front wall" consists of the left, center, and right speakers (L, C, R). The center speaker should be directly facing the center of the mix position, with the left and right speakers toed inwards, with their axes oriented towards the mixing position. The L and R speakers should be angled by approximately 30°. This configuration is not only suitable for surround mixing but also offers compatibility with stereo monitoring and works well for many home theater systems, where the L and R speakers are not far from the video screen. Some surround mixing engineers prefer to work with a steeper angle (up to 45°) between the L and R speakers, while shallower angles (such as the 21° typically used in cinemas) can be useful for checking mixes destined for theatrical exhibition spaces. (See section 2.4.) If for some reason the L and R speakers cannot be toed in along a semi-circular arc and thus create a "flat" front wall, delay can be added to the center speaker in order to compensate for its earlier arrival time. (See section 3.7)

All three speakers in the front wall should be at the same height — optimally at ear height of the mixing engineer, or approximately four feet off the ground. The three front speakers should not be placed on the console meter bridge. If a video monitor is present, it should be raised or lowered (or positioned off to one side) as necessary, rather than moving the center speaker. The "rear wall" of surround speakers (Ls, Rs) should be at the same height as the front wall. Some surround mixing engineers prefer to raise the rear wall speakers slightly higher than the front wall, angled downward and directed to the mix position.

As shown in the speaker placement illustration on the previous page, the Ls and Rs speakers should also be toed inwards, with their axes oriented towards the mixing position. The acceptable range of angles for the rear speakers is 110° to 150° relative to the center of the mix position. This range of position options reflects the need to facilitate various mixing practices and thus the respective use of the surround speakers. Recent analysis of current practices shows a consistent relationship between the content in the surround channels and the accompanying surround speaker placement. This relationship can be best described as the *mix perspective*. A mix perspective where the listener perceives himself or herself as being “in the audience” often places only room ambience or effects information in the surround speakers. In this case, the wash of sound created when the surround speakers are placed more to the side, at angles of approximately 110°, may be preferable. Conversely, many popular music mixes utilize an “in the band” perspective in which the listener appears to be part of the ensemble and featured performers are placed in the rear surround channels as well as the front channels. These kinds of mixes benefit from the improved rear phantom imaging achieved by placing the surround speakers further back behind the listener, angled in at 135° - 150°. In most cases, this will yield the most satisfying music listening experience while still providing home theater aficionados with an excellent cinematic experience.

Another commonly used technique is to slightly offset the convergence angles of the front left and right speakers relative to the rear left and right speakers so as to create a wider "sweet spot." This typically involves aiming the front left and right speakers so they converge centrally one foot *behind* the mix position, and aiming the rear speakers so they converge centrally one foot in *front of* the mix position. A laser pointer, or the more low-tech approach of using two identical length pieces of string, can be useful in creating such a setup.

It is worth noting that the ITU (International Telecommunications Union) has recommended that the rear speakers be positioned more to the side of the listener, angled between 100° and 120°. (See section 1.) This recommendation, which was originally developed for television broadcast and not for music, represents a compromise between envelopment, which is strongest at 90°, and rear phantom imaging, which improves as the rear speakers are moved further back and are angled in more steeply. Although this practice has been adopted by many surround professionals in the film and broadcast industries, it may not be suitable for music applications where the rear speakers are used to carry vital musical information and not just ambience or effects. Speaker placement, therefore, need not be restricted to the ITU recommendation alone: at the engineer’s discretion, more extreme rear speaker anglings (at up to 150°) are acceptable, depending upon the program content being mixed.

3.3.2 Subwoofer Placement

The optimum location for the subwoofer in a surround sound system depends almost entirely upon the room dimensions and design, as well as the physical layout of other equipment in the room. The key to its successful placement is that its location not be apparent. If the listener is able to perceive the direction that the low frequencies are coming from, that is a sure sign that the subwoofer is not positioned correctly.

That said, <u>it is our recommendation that the subwoofer be positioned in front of</u>

the mixing position, between the left and right speakers. The reason for this is that bass-heavy elements such as bass guitar and kick drum are most often placed in one or more of the front wall speakers. Positioning the subwoofer off to the side or behind the mix position therefore compromises imaging and can contribute to phase smearing.

There are two basic techniques for identifying the optimum position for a subwoofer. One is to listen at the mixing position while an assistant physically moves the subwoofer around the room while signal with significant low-frequency content is being played. (If program material with sufficient low-frequency content is not available, an 80 or 100Hz sine wave, or band-limited pink noise, low pass filtered at 80 or 100Hz, can be substituted.) The second technique involves temporarily placing the subwoofer itself at the mixing position while the *listener* moves around the room — much easier on the back! Then simply re-position the subwoofer to the spot in the room (optimally along a boundary wall in front of the mix position, between the left and right speakers) that yields the strongest and smoothest bass response. Measurements taken by a trained acoustician can also aid in identifying the best position for the subwoofer in a particular room.

When positioning a subwoofer, beware of standing waves! This is most likely to occur if the subwoofer is placed in a symmetrical location, such as directly under the center speaker. Shifting it slightly off to the side can often reduce or remove standing waves and yield better sonic results.

Phase correlation with the main speakers is also an important consideration. If the subwoofer itself provides a phase switch, experiment with different settings. Alternatively, try rotating the subwoofer in 90° increments until the smoothest bass response is heard. (See section 3.6.)

In some rooms, particularly larger ones, more even bass dispersion and improved imaging can be achieved by adding a second subwoofer. Be sure not to position the two subwoofers so that they are firing directly across from one another, or phase anomalies will result.

Remember that by its very design, the subwoofer is band-limited and carries only low frequencies. As its crossover frequency rises, the ability to localize its position increases. It is therefore recommended that any low-pass filter internal to the subwoofer should be set no higher than 120Hz. This is distinct from the bass management crossover point, which is typically set at 80Hz (see Section 3.4).

3.4 Bass Management

The term *bass management* refers to the redirection of low frequencies from the main channels to the subwoofer, so that it reproduces all the low frequencies in a surround mix, including the dedicated ".1" LFE channel (see section 1.3). Because most consumer home theater systems use satellite speakers instead of full range speakers, their amplifiers almost inevitably employ crossover networks for this purpose. Therefore, surround mixes should always be checked on a bass-managed satellite speaker system.

Creating a surround sound mix without bass management is sometimes referred to as mixing "direct to sub." Bass management is not required for full range speakers and may in fact significantly change the overall sound when switched in, so it is an option that may be employed during surround mixing at the discretion of the engineer. However, it is worth noting that many studio monitors, including some termed "full range," are not flat to the bottom octaves and will not reproduce sounds heard through bass-managed home systems. Bass management therefore allows the detection of unwanted low frequency information such as rumble which may otherwise be inaudible in the studio — but which the consumer may well hear in his or her home theater environment! It also allows the surround mixer to check the interaction between the LFE channel and redirected low frequencies from the main channels. Potential phase cancellation or phase smearing problems in the consumer system can thus be avoided.

Accurate calibration of both the main speakers and subwoofer is an absolute requirement for proper bass management. (See section 3.6) This will ensure a seamless transition between the high frequencies being radiated by the main speakers and the low frequencies being output by the sub. We recommend setting a bass management crossover frequency of 80Hz. This is not only the frequency most commonly used by consumer home theater systems, but its use also significantly reduces localization of the subwoofer and therefore adds flexibility in terms of its placement (see section 3.3.2).

It is worth noting that some consumer home theater systems offer a variety of bass management options. Some allow only the low frequency information from selected main channels (instead of all of them) to be redirected to the subwoofer. Other bass management options allow for playback through systems that contain no subwoofer at all, redirecting the LFE content to the front left and right speakers instead. It is worth checking the effect of each of these processes on a mix in progress in order to ensure a satisfactory result even when these kinds of options are engaged.

3.5 Reference Listening Level

The recommended reference listening level for surround sound production is in the range of 79 to 85dB C-weighted. However, it is important to check mixes at varying levels, from very soft (as low as 40dB) to quite loud (not to exceed 92dB, however, and only for short periods of time).

3.6 Speaker Calibration Procedure

Along with physical placement, correct speaker calibration is the single most important factor in ensuring an accurate monitoring environment. The tools required for calibrating a surround sound speaker system are much the same as those required for a stereo system: a source for pink noise and/or tones, and a Real Time Analyzer (RTA) or Sound Pressure Level (SPL) meter, with a reasonable quality omnidirectional condenser microphone. RTA metering is the preferred method for calibrating a surround sound speaker system because an SPL meter will yield a less accurate result since only the peak of one band is measured. If an SPL meter is used, set it to C-weighting on the slow scale.
It is also helpful to have a number of well-recorded commercial surround sound releases on hand to play back in order to do final “tweaking” by ear.

In addition to console oscillators and onboard pink noise generators, there are a number of surround sound calibration tapes and CDs available which can be very helpful in calibrating speakers. These range from professional releases (such as those available from MLSSA, Tomlinson Holman, and Dolby) to numerous consumer DVDs and CDs. The key is to have access to both full bandwidth and band limited pink noise (low-pass filtered at 80 - 120Hz) and a sine wave at the subwoofer crossover frequency (80Hz in most instances).

The need for precise speaker calibration should be obvious: Only if the system plays back accurately can it be used to produce surround sound mixes which will translate well in other listening environments. The converse is also true: If the speaker system used for mixing is not correctly calibrated, there is a greater likelihood that the resulting surround mixes will *only* sound good in the studio in which they were mixed, and nowhere else.

After making sure that the signal path from the console is in proper electrical phase to all speakers (a near-certainty if using powered monitors and correct balanced wiring), the first, and perhaps most important calibration is the acoustic phase alignment of the subwoofer(s) at its crossover point (again, 80Hz in most instances); incorrect alignment will cause a drop in the frequency response of the entire system at the crossover point. Some subwoofers have built-in phase matching controls which allow adjustment in precise 90° steps.
Following is the recommended procedure for subwoofer phase alignment: (Note

that this procedure assumes that all main speakers in the system are full range and not satellite.)

1. Route a sine wave at the crossover frequency (generally 80Hz) at a moderate listening level to the left front and right front speakers and to the subwoofer.

2. Using an RTA or SPL meter, note the signal level at the mix position.

3. If the subwoofer provides phase controls, toggle the switches, noting the signal level at the mix position each time. Leave the switch at the position which yields the maximum signal level.

4. If the subwoofer does not provide phase controls, simply rotate it by hand at 90° increments until the signal level at the mix position is highest.

The next step is to set the reference level for the main speakers. We recommend a nominal reference level of 79 - 85dB SPL (see section 3.5). The important thing here is not so much the actual SPL chosen, but that all five main speakers are set to that same chosen level. <u>Following is the recommended procedure for reference level adjustment for the main speakers</u>: (for purposes of illustration, a reference level of 85dB is assumed)

1. Turn off all speakers and subs except the front left speaker.

2. Place the calibration microphone at the center of the mix circle, at ear height, facing directly towards the center speaker. If an RTA is being used for measurement, set it to read 85dB. If an SPL meter is being used for measurement, set it to C-weighting on the slow scale, and then set it to read 85dB.

3. Route pink noise at 0 vu to the front left speaker and raise the speaker's amplifier level until all bands of the RTA (or the SPL meter) read 85dB.

4. Continue by routing pink noise to each remaining speaker in turn (front right, then center, then rear left, then rear right), adjusting their amplifier levels so that the RTA or SPL meter reads 85dB for each. In each instance, leave the calibration microphone at the same fixed position as in step #2 above.

Some professionals instead suggest pointing the calibration microphone towards the phantom center when measuring the rear speakers. Others recommend pointing it at 90° left and 270° right when calibrating the rear speakers. Though these techniques will yield slightly different results, the fundamentals remain the same.

The final step is to set the level of each subwoofer relative to that of the main speakers. Common practice is to calibrate the subwoofer approximately 4dB above the reference level of the main speakers. This procedure differs somewhat depending upon whether the subwoofer is receiving the LFE channel only or whether bass management is being utilized to route signal to it from some or all of the main channels.

<u>The recommended subwoofer calibration procedure when no bass management is being used is as follows</u>:

1. Turn off all five main speakers.

2. Route band-limited pink noise (low-pass filtered at 80 - 120Hz) at 0 vu via the LFE channel bus to the subwoofer and raise its amplifier level until the RTA or SPL meter reads +4 dB over the selected reference level (i.e. 89dB if the selected reference level for the main speakers is 85dB).

3. Turn on the front left and right speakers.

4. Route full frequency pink noise at 0 vu to the front left and right speakers as well as to the sub. Adjust the subwoofer amplifier so that the gain boost when adding the subs to the mix does not exceed 4 - 6db, as measured by the RTA or SPL meter.

<u>The recommended subwoofer calibration procedure when bass management *is* being used is as follows</u>:

1. Route band-limited pink noise (low-pass filtered at 80 – 120Hz) at -10 vu via the LFE channel bus to the subwoofer. (The 10dB of attenuation compensates for approximately 10dB of "in-band gain" in the LFE channel as compared with a main channel.)

Note: "In-band gain" refers to the fact that the level in each 1/3-octave band within the frequency range of the subwoofer is 10dB above the level in each 1/3 octave band in each of the main channels, averaged across the main frequency range. This does not mean that the LFE channel is 10dB higher in SPL than the main channels, however, due to the broader bandwidth (and correspondingly greater energy) in the main channels.

2. At the same time, route band-limited pink noise (low-pass filtered at

80 - 120Hz) at 0 vu via a single bus that is sending signal to the subwoofer via the bass management circuitry.

3. Raise the amplifier level of the subwoofer until all bands of the RTA (or the SPL meter) read at the selected reference level.

Note that, following this calibration procedure, the subwoofer level may need to be adjusted up or down by a dB or two to compensate for heavy bass trapping (or lack thereof) in the room. The best way to do this is to listen critically to some commercially recorded surround sound music while seated at the mix position. Some subwoofers have DIP switches that allow their level to be adjusted in fine increments of plus or minus 1 or 2dB.

Also note that, following speaker calibration, the overall gain of the system is mathematically up to 12db louder, excluding the sub (i.e., a reference level of 85db for each individual speaker actually sums to 97db at the mix position with all five main speakers driven).

3.7 Use of Delay

As noted in section 3.3.1, delay may be used to compensate for non-coincident (differing) arrival times if it is not possible to place all five main speakers equidistant from the mix position, or if it is not possible to angle them correctly (for instance, if there is a "flat" front wall, with the L, C, and R speakers in a straight line).

The following formula should be applied: <u>For each foot of distance disparity, add .88 milliseconds of delay</u>. (For each meter of distance disparity, add 2.94 milliseconds of delay). For example, if the rear speakers are two feet further away from the mix position than the front speakers, the signal going to the front speakers should be delayed by 1.76 milliseconds. In the case of a "flat" front wall, delay should be added to the center speaker.

It is important to note that <u>the use of delay is not recommended unless absolutely necessary</u>. It provides a far less satisfactory solution than actually positioning and angling speakers correctly!

4. Surround Sound Mixing Techniques

This section describes various production practices utilized by many surround sound professionals. As noted in the introduction to this paper, there is no one "correct" way to mix surround sound. Instead, we offer a variety of acceptable techniques that have proven to yield consistently good results.

4.1 Music vs. Sound For Picture

Human beings tend to rely more on visual than aural cues; we focus more on what we are seeing than what we are hearing, which is why we often close our eyes when listening critically to music. Multichannel audio can be used to accompany picture (i.e., film scoring and sound design or soundtrack accompaniment to concert video) or it can be used to present music on its own (i.e., DVD-Audio or SA-CD Multichannel release). The production approach to each can be quite different.

When mixing multichannel sound for picture, the audio follows the action on screen, and not vice-versa. The goal is to match the sound to what you are seeing; thus, in a concert video, if a musician is featured prominently onscreen, the tendency is to raise the level of that instrument. Conversely, if an instrumentalist is at the back of the stage or largely offscreen, the tendency will be to tuck that sound in. In feature film production, wherever there is dialog (which most often will be anchored in the center speaker), care must be taken to reduce the overall level of accompanying music and effects so that the spoken word can be clearly heard. In some cases, equalization has to be applied to midrange content of those tracks in order to carve out frequency notches which might be fighting the dialog. Separate mixes may be undertaken for cinematic exhibition and home theater release. In the case of the former, the music mix is normally delivered as stems so that they can be rebalanced by the film mixer against the FX and dialogue stems; in the case of the latter, provision may be made for the fact that the center and/or rear speakers may not be full range.

Perhaps most importantly, the audio content in sound for picture projects is determined ultimately by the director or producer of the film and not by an audio specialist. This can have both positive and negative implications. On the one hand, there is little fear of using radical panning or dynamic fly-overs from speaker to speaker; on the other, there may be less attention paid to musical content than there is to dialog and effects.

Because there are no visual cues to follow, mixing surround sound for music only can be a much more freeing experience. However, initial music-only multichannel mixes were relatively conservative compared to today's mixes. The rear speakers were often used only to carry room ambience or effects returns, and the center channel often carried only one or two select instruments, with other instrumentation placed statically in the front left-right speakers as if it were a standard stereo production. Recent years have seen a trend towards a more inventive approach, with more use of the rear speakers to carry significant musical content, greater flexibility in instrument placement and panning, an increased awareness of the importance of creating ambient spaces through decorrelated effects, and more willingness to pump up the LFE channel where musically appropriate.

4.2 Bus- and Track-To-Channel Allocations

<u>We recommend the usage of the following 5.1 bus-to-channel allocation, which conforms to the recommendation made by SMPTE and ITU:</u>

Bus	**1**	**2**	**3**	**4**	**5**	**6**
Channel	**L**	**R**	**C**	**LFE**	**Ls**	**Rs**

Other allocations are permissible. However, <u>it is vital that the final surround mix be labeled clearly and correctly</u> so that the mastering engineer knows exactly which track is carrying which channel. One alternate allocation, used for DTS encoding, is as follows:

Bus	**1**	**2**	**3**	**4**	**5**	**6**
Channel	**L**	**R**	**Ls**	**Rs**	**C**	**LFE**

Both of these allocations offer logical, phase-coherent pairs (i.e., L/R, C/LFE, Ls/Rs), an important factor when utilizing stereo converters. However, another commonly used allocation serves as the film standard in Europe:

Bus	**1**	**2**	**3**	**4**	**5**	**6**
Channel	**L**	**C**	**R**	**Ls**	**Rs**	**LFE**

It is worth noting that one popular digital audio workstation (DAW) records 5.1 data as a single, grouped track, displaying the channels in this format ("L, C, R, Ls, Rs, LFE"). Although this display cannot be changed, the input/output routing should be altered to default to the SMPTE/ITU allocation described above.

<u>We recommend that the tracks of a 5.1 master be printed identically to the bus allocations</u>, so that bus 1 (carrying the L channel information) is printed to track 1, bus 2 (carrying the R channel information) is printed to track 2, etc. Tracks 7 and 8 are often used to carry a time-aligned stereo mix of the same program material.

4.3 Imaging and Panning

Audio engineers mixing in stereo have long relied on the ability to create a "phantom" center by placing equal amounts of signal in the left and right channels. This creates the illusion of the sound appearing mid-way between the two speakers, but it is a fragile image that shifts closer to one speaker or the other as the listener moves out of the sweet spot. Even directly in the sweet spot, a phantom center will exhibit a change in amplitude and frequency response due to a phenomenon called *comb filtering*. Because the signal is coming from two speakers, each at a slightly different distance from each ear, the resulting offsets will cause certain sonic components to cancel out one another. The addition of a true center channel in surround sound eliminates these problems by providing an "anchor"; signal routed to that channel always appears to come from the center — and with a consistent frequency response — regardless of how the listener moves around in the room. Many surround mixing engineers opt to use both the true center and a phantom center. An important control called *divergence* allows the precise determination of the relative amounts of center-panned signal routed to the center channel versus that routed equally to the left and right channels.

The presence of a dedicated center channel also allows the surround mixing engineer to create phantom images between the left and center and right and center speakers, thus further enhancing localization. Because the distance between those speakers is much shorter (exactly half if the speakers are positioned correctly), these kind of "in-between" phantom images are not as strong; however, they can nonetheless be helpful in spot placement of sounds across the front wall. In applying this technique, however, the mixer must take into consideration the fact that the end user monitoring system may contain a smaller and/or poorly positioned center speaker — or perhaps no center speaker at all!

Similarly, a phantom rear center can be created by routing equal amounts of signal to the left rear and right rear speakers. If the rear speakers are correctly positioned and angled, this image can be nearly as stable as a front phantom center. The presence of a rear center speaker in 6.1 systems allows additional anchoring of a rear center signal, and also allows for the creation of "halfway" phantom images (i.e., between rear left and rear center, or rear center and rear right).

However, side phantom images, between the front left and rear left speakers or front right and rear right speakers, are very weak and unstable — they seem to jump around with even the slightest head movement. This is due to the simple fact that our heads get in the way! Mathematical formulas called Head Related Transfer Functions (HRTFs) factor in the size, shape and density of the human head as well as the position of the ears. They are used by acousticians to predict the ability to perceive various sounds originating from different points around us.

These formulas show that it is easy to discern slight panning movements from left to right, and even easier to detect panning from left to center, and then center to right (the recommended way of doing left-right pans in a surround sound system). However, they also demonstrate that smooth panning from front to rear is not possible, even with the presence of a dedicated rear center speaker added by 6.1 systems. This is because the human ear's frequency response to sounds coming from the rear is radically different from that of sounds originating from directly in front, so there's little cohesion; in essence, a front-rear "flyover" is really a fade, not a pan. However, as noted in section 1.3, front-rear flyovers in 6.1 systems do not suffer as much from comb filtering problems since the panning is being done between two single point sources.

Also, because the size of the head remains constant even though different frequency components generate different wavelengths, localization is very much frequency-dependent; higher frequencies localize much better than lower frequencies. In fact, it is quite difficult to localize very low-frequency sounds, which is why a single subwoofer can be used to handle all the bottom end in a surround system.

One potential problem that can arise from routing a signal into two or more speakers is the danger of increased, and increasingly complex, comb filtering. This problem multiplies as more speakers are engaged and can become critical if downmixing is ever employed by the playback system. Therefore, many experienced surround mixers selectively turn off channels when bringing a sound "inside" the surround bubble or when dynamically panning a sound from one area in the surround space to another. <u>It is recommended that whenever signal is placed into three, four, or five speakers, it be decorrelated</u> (see section 4.7).

It is worth noting that, while most "surround-ready" mixing consoles provide four- or five-channel panpots (often in the form of joysticks), it is also possible — and sometimes preferable — to accomplish pinpoint positioning with the judicious use of delay lines instead. For example, by routing a signal to all five channels and then slightly delaying the rear channels only, the sound can be "spotted" front of center, for all listening positions.

4.4 Use of the Center Channel

In the arena of surround sound mixing, there is probably no area that prompts more debate than the use (or abuse) of the center channel. As noted in section 4.3, its primary function is to provide hard center anchoring for key components (such as dialog in film postproduction, or lead vocals or solo instruments in music applications) with greater stability than phantom centering, and without any of the comb filtering problems that occur with phantom centers.

However, too much reliance on the center channel alone can be problematic due to the fact that the center speaker in many home theater systems is smaller than the main left and right speakers. As a result, signals routed to the center channel alone can be severely compromised in terms of their frequency spectrum during playback. (Some consumer surround sound systems don't provide a center speaker at all; however, most consumer receivers provide an option to route center channel information at equal level to the left and right speakers if no center speaker is connected.)

Another problem stems from the fact that most playback systems — even the most rudimentary consumer systems — allow each channel to be heard in isolation. Placing a lead vocal "naked" in the center channel, without other instrumentation to help mask poorly intonated notes, "auto-tuning" glitches, or bad drop-ins, can therefore potentially expose weaknesses in a performance and consequently incur the wrath of the recording artist and record label.

For these reasons, most surround sound music mixers treat the center channel with caution, rarely if ever using it to carry any mix components exclusively. Instead, those instruments routed to the center channel (most often lead vocal, bass, snare drum, kick drum and/or instrument solos) are also generally routed to other speakers as well. Placing selected instruments in the center channel and one or both front speakers helps emphasize their sound within the front wall and also aids in localization if the listener moves around the room. Conversely, creating a virtual triangle by placing selected instruments in the center channel and one or both rear speakers can yield an interesting psychoacoustic effect where the sound appears to come out into the room, closer to the listener. However, care must be taken to decorrelate such signals in each speaker (most often, by slightly altering equalization, delay times or pitch — see section 4.7); otherwise, masking and/or phase cancellation problems can occur.

Some surround mixers prefer to leave the center channel dry (free of reverberation), while others opt to add a small amount of decorrelated reverb in order to prevent the signal from feeling too disembodied. If a decision is made to route reverb to the center channel, early reflections and/or reverbs with short delays (i.e., "room" presets) are generally a better choice than long reverb tails.

4.5 Use of the Rear Channels

As described in section 4.1, modern surround sound music production techniques make much greater utilization of the rear channels than ever before. Whereas the Ls and Rs channels were once used almost exclusively for room ambience and effects returns, engineers today often place significant musical content in the rear speakers, frequently adopting an "in the band" mix perspective (see section 3.3.1). Some engineers characterize opening up the soundfield this way as making things "larger than life."

The rear speakers are also often used to carry transient events, such as percussive accents or sound effects. This introduces the element of surprise into the surround sound listening experience and temporarily draws the listener's attention away from the overall wash of sound, perhaps even making them turn their heads (sometimes called the "exit sign effect").

Another important usage of the rear speakers, as noted in sections 4.3 and 4.4, is to pull a center channel signal out "into" the surround space. With careful balancing and by selectively routing it to center speaker and both the Ls and Rs speakers, the sound can appear to float directly in front of the listener's face. By routing it to the center speaker and either the Ls or Rs speaker (not both), again with careful balancing, a somewhat fragile yet interesting phantom image can be created where the sound appears to be floating just behind the listener's left or right shoulder. The addition of slight delays or phase offsets can help enhance such imaging.

Where signals have been recorded in multichannel format with the use of microphone arrays, the rear speakers serve the important function of carrying spatial positioning information. Even when the signal source is mono or stereo, ambient room mics can be routed to the rear speakers, thus enhancing the apparent size of the image. Similarly, effects returns such as reverb that are routed to the rear speakers (decorrelated from any that are placed in the front speakers; see section 4.7) can aid in making a sound appear bigger.

Multichannel bus compression and equalization tools have appeared in recent years which can help the front and rear channels be better integrated into a coherent soundfield. Their use by the mixing and/or mastering engineer is optional and subjective, depending upon the program material. See section 5.3.

4.6 Use of the LFE Channel

As noted in section 1, the LFE ("Low Frequency Effects") channel was originally introduced by the film industry because early theatrical speaker systems were unable to generate loud low frequency signals without clipping. Sometimes referred to as the "boom" channel, it is used in film applications to add dramatic effect, almost exclusively carrying the rumble of volcanic eruptions, spaceships thundering into view, and bombs and planets exploding.

In terms of multichannel music production, however, there is some debate as to whether the LFE channel is necessary at all. One can argue that the home theater experience is heightened by having the walls shake whenever rocket launchers are fired, but is there really that much value in having the listener feel every bass drum hit in such a similarly dramatic fashion?

There is no clear-cut answer, but the mere presence of the LFE channel almost dictates that it be used. The danger lies in overuse, because too much reliance on the LFE channel to carry bass information can result in the loss of low end altogether on incorrectly configured or poorly designed home theater systems. In addition, the LFE channel is discarded by most matrixed encoding systems (such as Dolby Pro Logic; see section 1.1) and downmixing algorithms (see section 5.5), including those used for HDTV broadcast. Therefore, <u>the LFE channel should never be used to carry the bass content of the main speaker channels</u> — that is the job of bass management. Because bass management is employed by almost every consumer home theater system, placing too much information in the LFE channel will effectively result in double management — total bass overload and probable distortion.

Instead, the LFE channel is best approached with caution. Only modest amounts of signal from specific instruments with significant low frequency content — kick drum, tympani, bass guitar, acoustic bass, low organ or piano notes — should be routed to the LFE, and in all instances those instruments should also be printed full range to the desired main channels as well. An experienced mastering engineer can help in correctly assessing the relative level of the LFE channel as compared with the main channels.

4.6.1 Filtering the LFE Channel

The subwoofer is often the most inaccurately configured component in the typical home theater system. Bass management schemes in consumer receivers vary widely, and most employ filtering after the summation of main channel and LFE information. A mixer who is not monitoring through a bass managed system can inadvertently create a mix that plays back with phase cancellation problems or even the entire loss of some low frequency information. Even in cinemas with fixed installations done by professionals, the sub can be poorly set up. As a result, it is our recommendation that <u>the LFE channel should be low-pass filtered at 80 - 120Hz</u>. This should be considered even if the project is destined for release on SA-CD or DVD-Audio, despite the fact that the spec does not require such filtering. Because the optimum frequency to be used is program-dependent, such filtering should be done by ear; simply experiment with different frequencies while listening carefully. <u>In most cases, selecting a frequency between 80 and 100Hz will produce the best results</u>. Use the steepest filter available (24dB/octave or higher), and choose one that maintains the most accurate phase correlation. While monitoring during mixing through a bass managed system is imperative, if low pass filtering of the LFE channel is not possible or desired, it can be printed full range and the job left to the mastering engineer.

4.7 Creating Realistic Ambience

Unless one is working with material that was recorded with multiple ambient microphones (or a multichannel microphone array), the surround mixer often has to deal with single channel sources that need to be clearly positioned within an open, defined space. There is often also a need to move the sound around freely in that space — and at the same time have the acoustics work correctly. For such precision panning to work realistically, the accompanying ambience needs to change as well.

The key to accomplishing this is the use of *decorrelated* effects, where the output of each channel is similar, but not exactly the same (if they were exactly the same, they would be *correlated*). Decorrelation is achieved when one or more basic parameters — typically filter settings, delay times, and/or pitch — are very slightly offset in each channel. In any two-channel pair, the sound of a left-right decorrelated pair is distributed evenly across the field; in contrast, an exactly correlated pair sounds monophonic. The sum of any two correlated signals is 6dB higher in their center, while the sum of any two perfectly decorrelated signals is only 3dB higher in the center. More importantly, decorrelated signals yield a sense of realistic ambience, because real spaces are decorrelated. In the case of a decorrelated surround sound reverb, instead of yielding a sense of a different reverb in the back, the listener gets the sense of a real space.

While dedicated multichannel reverberation algorithms and hardware devices provide an easy means for decorrelation, it can also be accomplished with stereo, or even mono reverbs. In the case of stereo reverbs, which often are programmed with identical parameters in the left and right channels, care must be taken to alter one or more basic parameters in one side only. Some engineers prefer to use multiple mono reverbs instead, routing each to a different speaker and creating decorrelation by calling up the same preset in each, then slightly altering selected settings.

As with most mixing techniques, the best way to approach the creation of realistic multichannel ambience is to listen carefully and experiment. Take the time to listen to a sound in a live acoustic space, then go back into the studio and try to duplicate what you are hearing.

5. Related Issues

This section discusses surround sound production issues not covered previously.

Section 5.1 describes the various surround sound distribution formats currently in use.

Section 5.2 describes the various surround sound encoding formats currently in use.

Section 5.3 describes the special demands of surround sound mastering and identifies several potential problem areas.

Section 5.4 presents a discussion about sample rate conversion and resampling issues.

Section 5.5. discusses the impact of automatic downmixing on a surround mix.

Section 5.6 discusses the important issue of automatic upmixing (sometimes called "faux 5.1"), with specific recommendations as to when it is appropriate, and when it is not.

Section 5.7 contains recommendations for labeling surround sound product.

5.1 Distribution Formats

There are currently five means of distributing discrete surround sound commercially:

- DVD-Video
- DVD-Audio
- SA-CD
- Encoded CD
- Windows Media Audio 9 Professional (WMA9 Pro)

This section describes the main features of each delivery method.

5.1.1 DVD-Video

The introduction of the DVD ("Digital Versatile Disc") in 1996 was a vital first step in the acceptance of surround sound by the mass public. Although exactly the same diameter and nearly the same thickness as a CD, the DVD contains smaller laser “pits” so that more data can be stored. Unlike CDs, DVDs can utilize an optional second layer of data, and are available in both single-sided and double-sided versions. This makes for much higher data capacity — even a single-layer, single-sided DVD holds 4.7 gigabytes of data (more than seven times the capacity of a CD), and a dual-layer, double-sided DVD can store 17 gigabytes of data (equivalent to almost 32 CDs).

There are several different forms of DVD, some of which are recordable, and two of which — DVD-Video and DVD-Audio (see section 5.1.3) — are designed as read-only entertainment media. DVD-Video, which was the first version to come to market, has already enjoyed the most successful launch of any new consumer electronics product in history, far surpassing equivalent early sales of audio cassette players and VCRs. As its name implies, it reserves most of its storage area for digital video data, with only a relatively small portion set aside for digital audio. However, the key point is that, with the use of Dolby Digital (mandatory) and DTS (optional) encoding (see section 5.2.1 and 5.2.2), DVD-Video allows for the storage of not just two, but multiple channels of audio data.

As many as eight audio data streams can be present on a DVD-Video; the data in stream 1 is typically used by DVD players as the default setting for that disc. These streams can include linear PCM (up to eight channels of 48kHz / 16-bit data; up to four channels of 48kHz / 24-bit data; or up to two channels of 96kHz / 24-bit data), or up to six channels of Dolby Digital- or DTS-encoded data (see sections 5.2.1 and 5.2.2). In addition, up to eight channels of MPEG-2 encoded audio data may be present, although this is rarely used in Region One (North America) releases. Maximum bit rate for audio data streams is 6.144Mbps. The available number of streams, maximum tracks, bit rate, and coding options are all variables which are dependent upon picture quality.

5.1.2 DVD-Audio

The DVD-Audio specification was only finalized in the year 2000; as a result, it is still a relatively new delivery media for surround sound. Unlike DVD-Video, the vast majority of its storage space is earmarked to carry audio data; in fact, only still pictures and limited motion MPEG-2 encoded video can be added.

Maximum bit rate for a DVD-Audio disk is 9.6Mbps, and a variety of sample rates (up to 192kHz) and word lengths (up to 24 bits) are supported. What's more, sample rate and word length can vary from track to track. The selected sample rate and word length determines the maximum number of channels and playback time. Audio streams can included unencoded linear PCM data or data encoded with Dolby Digital, DTS, MPEG-2, or MLP (see section 5.2.3), or any combination thereof. Up to eight channels of linear PCM data can be played back simultaneously (a discrete 5.1 mix plus a stereo mix) at a 96kHz sampling rate. With the use of MLP encoding, more than an hour of eight-channel data at 96kHz / 24-bits can be stored.

One DVD-Audio feature that surround sound producers should be aware of is its built-in downmixing (multichannel to stereo) capability, called SMART (an acronym for System-Managed Audio Resource Technique). This allows the producer to specify mixdown gain coefficients on a track-by-track basis. (See section 5.5.) When the data is MLP-encoded, dynamic mixdown capability is added so that levels can be changed over time; however, this is rarely implemented.

DVD-Audio disks can only be played on legacy DVD-Video players if a Dolby Digital- or DTS-encoded audio stream is present. The only way to enjoy the full range audio it contains is to play it either on a dedicated DVD-Audio player (which must contain MLP decoding capabilities) or on a "universal" player capable of playing both DVD-Video and DVD-Audio disks. Universal players usually have the capability of playing back CDs and multichannel SA-CDs as well. (See section 5.1.3)

5.1.3 SA-CD

SA-CD is an acronym for "Super Audio CD." Introduced by Philips and Sony in 1999, SA-CDs are the same physical size as both CDs and DVDs, but are capable of storing data in an entirely different format known as DSD ("Direct Stream Digital"), as opposed to PCM. DSD is a single-bit delta-sigma modulation coding method which uses an extremely high sample rate of 2.8224MHz, resulting in an extended high frequency response.

The SA-CD specification is provided by a document called the “Scarlet Book,” which describes three disk format options: single-layer DSD, dual-layer DSD, or dual-layer hybrid. The latter includes a standard CD “Red Book” layer (see section 5.1.4) that can be played on any existing CD player, in addition to a high-density layer that has the capacity to deliver eight channels of DSD. It is worth noting that multichannel SA-CDs (marketed as "SA-CD Multichannel") can only be played in dedicated SA-CD players with multichannel capability, or in universal players.

Other Scarlet Book specifications include: *Super Bit Mapping Direct*, a proprietary DSD-to-PCM conversion method that enables improved audio when the disks are played on an ordinary CD player; Direct Stream Transfer (DST), a lossless encoding process that increases data capacity (see section 5.2.4); and a digital watermark to protect against piracy. DST is required for multichannel SA-CDs and is optional for stereo SA-CDs. There is an overriding logical limit of 256 minutes per area per disc. For a single layer (or hybrid) disc, this results in approximately 110 minutes of stereo audio uncompressed, or 70 minutes of 8-channel audio compressed (six channels of surround sound plus stereo). However, SA-CD Multichannel discs are not required to necessarily carry all six channels, so this time may vary. The maximum playback times for a dual layer disc may be considerably longer.

It is worth reiterating that producing surround sound content for SA-CD Multichannel release requires that the final mix be recorded in (or converted to) DSD format. At the time of this writing, there are only a limited number of DSD editing and processing tools available on the market.

5.1.4 Encoded CD

Standard compact discs (CDs) utilize a fixed bit rate of 1.411Mbps (as defined in the "Red Book" specification). While this is sufficient to carry two channels of 16-bit PCM audio at 44.1kHz, it is insufficient to carry unencoded surround sound audio. However, the application of DTS encoding (see section 5.2.2) allows the delivery of 5.1 or 6.1 channels of 24-bit 44.1kHz audio at this same bit rate. These "encoded CDs" are playable on standard CD players; however, unless the player has a built-in DTS decoder (or unless it provides an S/PDIF digital output for connection to an external DTS decoder), only high-level noise will be heard. Up to 74 minutes of 5.1 audio can be stored on a DTS-encoded CD — exactly the same amount of stereo signal that can be stored on an unencoded audio CD.

5.1.5 Windows Media Audio 9 Professional

Windows Media 9 (WM9) describes a family of codecs developed by Microsoft for the delivery of various kinds of media formats over the Internet, as well as via high-definition television and satellite broadcast. One of these is Windows Media Audio 9 Professional (WMA Pro), a scalable lossy codec which allows the streaming of up to 8 channels of 96kHz / 24-bit audio data. WMA Pro uses variable bit rates ranging from 128 to 768kbps and is thus capable of delivering 5.1 surround sound at the same bit rate MP3 uses for stereo playback. The use of higher bit rates results in improved quality, and there are multiple encoding modes, including a 2-pass variable bit rate lossless mode.

Another codec within this family is Windows Media Audio 9 Lossless. Developed for archiving and mastering applications, this provides mathematically lossless encoding. Compression ratio is approximately 2:1 for 44.1kHz / 16-bit source material. This increases (up to 4:1) with source material at a higher sample rate, bit resolution and/or channel count. A free tool is available that allows WMA9-encoded files to be converted back to standard WAV format.

Automatic downmixing is provided when a user attempts to play back an encoded multichannel 96kHz / 24-bit file but does not have a system or sound card that supports multichannel or high resolution audio. Downmix coefficients are provided for center, rear, and LFE channel levels. These can be entered either prior to, or after the encoding process. Default coefficients are -3dB for the center and rear channels, and -12dB for the LFE channel. To prevent clipping, the resulting stereo volume is normalized to the sum of all channels.

The peak and average values of the audio signal are calculated during encoding, and placed in the header of the file. During playback, users can specify three dynamic range settings using a "Quiet Mode" feature. This is useful for maintaining voice intelligibility in movie content that has a wide dynamic range. In addition, a provided editor allows the user to specify different peak and average values than those that were calculated during encoding.

During playback, the audio is automatically resampled to match the capabilities of the audio card. In addition, the bit depth is re-quantized to 16 bits, if required by the sound card. If a user has a two-channel sound card, but also has a Dolby Pro Logic-style decoder downstream, matrix encoding (Surround Sound mode) can be enabled to derive four-channel LCRS output from the stereo output. This configuration is popular with many game systems.

Another important component of Windows Media Audio is Digital Rights Management (DRM). DRM is a data encryption technology that limits access to only those people who have acquired a proper license to play the content. Coupled with WMA9 Pro, it enables the secure delivery and distribution of surround sound content on the Internet. DRM can also be used to facilitate the distribution of multichannel tracks, mixes and masters at various stages of production.

At this time, Windows Media Audio 9 Professional is the only readily available delivery format suitable for real-time streaming of discrete surround sound over the Internet.

5.2 Encoding Formats

This section describes the four most commonly used formats for discrete encoding of surround sound:

- Dolby Digital (AC-3) and Dolby-EX
- DTS and DTS-ES
- MLP
- DST (Direct Stream Transfer)

Note that the first three of these operate on PCM (Pulse Code Modulation) audio only, where the data is stored as 16-, 20-, or 24-bit words. Their purpose is to reduce the size of the data so that multichannel information can be stored on CD, DVD-Video, or DVD-Audio — all of which use the PCM format exclusively.

SA-CD stores data in an entirely different format called DSD (an acronym for "Direct Stream Digital"). DSD data is single bit only (though at an extremely high sample rate), and neither Dolby nor DTS nor MLP encoding can be applied; instead, on multichannel SA-CDs (see section 5.1.3), a proprietary encoding scheme called DST (Direct Stream Transfer) is used.

Both Dolby Digital and DTS (and their 6.1 variations: Dolby Digital-EX and DTS-ES) are lossy encoding schemes where some audio data is discarded, while MLP and DST are lossless "packing" algorithms that provide bit for bit accuracy. Every lossy encoding scheme has its advantages and disadvantages, resulting in greater or lesser audibility of artifacts, depending upon a number of variables. Higher bit rates always result in increased audio fidelity and so the highest bit rate possible should be used when applying lossy encoding to multichannel audio. It is strongly advised that engineers personally evaluate the coding method and bit rate to determine what is most suitable for the material.

5.2.1 Dolby® Digital™ (AC-3) and Dolby-EX™

Dolby Digital (sometimes referred to by its technical name, AC-3) is a 16- or 20-bit perceptual (lossy) encoding scheme for both stereo and 5.1 channel audio. Widely used in film applications and on DVD-Video (as well as for HDTV broadcast, and in games written for the Microsoft Xbox), Dolby Digital supports sampling rates of 32, 44.1, and 48kHz and provides variable bit rates from as low as 32kbps (for a single mono channel) to as high as 640kbps, resulting in data compression ratios of up to 12:1. Although the maximum bit rate allowed for DVD-Video players is 448kbps, the most commonly used bit rate is 384kbps (for a data compression ratio of 10:1). Dolby Digital is considered a "mandatory" requirement for DVD-Video, which means that it must be used as the primary encoding scheme if one is included. It may also be present as an "authorized optional" encoded audio stream on DVD-Audio.

Dolby Digital also provides meta-flags for downmixing coefficients. These are automatically applied by the decoder whenever fewer than 5.1 channels are connected to the receiver. See section 5.5.

Dolby Digital Surround EX™ (usually simply called "Dolby-EX") is a 6.1 format that adds a rear center speaker channel. This extra channel is not discrete but is matrix-encoded onto the left and right rear channels of a standard 5.1 mix; hence, no information is lost when the audio is played back in 5.1. A special receiver with either Dolby-EX or THX Surround EX decoding is required to derive the extra channel for playback in 6.1 format.

5.2.2 DTS and DTS-ES

DTS Digital Surround (usually referred to simply as "DTS") is a scalable lossy encoding scheme for multichannel audio (up to eight channels are supported) that operates on up to 24-bit words at sampling rates of up to 96kHz. When employed on DVD-Video (see section 5.1.1), two bit rates are offered: 1.509Mbs and a half-rate of 754kpbs; the latter yields reduced audio quality but increased storage capacity. When employed on encoded CD (see section 5.1.4), bit rate is 1.235Mbps, for a compression ratio of approximately 3:1 (for 16-bit data) or 4:1 (for 24-bit data). DTS is considered an "authorized optional" encoding scheme for both DVD-Video and DVD-Audio. It contains no downmixing controls.

DTS-ES always operates at a 48kHz sampling rate. It is a 6.1 format that adds a rear center channel in either a discrete or matrixed fashion. If the source material was created using matrix encoding for rear center playback (for example, in film soundtracks), a flag is included indicating this. Decoders with the ES option will detect this flag and engage a matrix decoder such as DTS Neo:6. In discrete encoding, the rear center channel is added to the left and right surrounds but reduced in level by 3db. The front, LFE and modified rear channels are then encoded creating a “core.” The rear center channel is then encoded individually and attached to the end of the core. To ensure backward compatibility, DTS-ES adds two flags to the meta-data of the 5.1 core signal: one flag that identifies the presence of a discrete rear center channel and another flag that indicates it may be decoded with a matrix decoder. Devices that employ a DTS-ES decoder chip recognize the presence of this additional channel and route it to the physical rear center speaker, simultaneously reducing it in level by 3dB and subtracting it from the left and right rear speakers. This is a linear process which requires no logic or steering. DTS-ES matrix decoders ignore the extra encoded channel but engage the matrix decoder for the left and right rear channels. Legacy decoders operating in 5.1 only ignore DTS-ES flags and extra channel data.

5.2.3 MLP

Because the DVD-Audio standard specifies that lossy encoding not be mandated, there was a need to implement a lossless encoding scheme in order to accommodate extended playback times of high-resolution (i.e., 96kHz / 24-bit) multichannel audio. After an extended evaluation period, MLP (an acronym for "Meridian Lossless Packing") was selected for that purpose. It is a lossless coding system for PCM data that is bit for bit accurate. Able to support sample rates of up to 192kHz and word lengths of up to 24 bits, MLP typically provides a 2:1 compression ratio. Maximum bit rate is 9.6Mbps.

It is not necessary to use MLP on programs with lower sampling rates and word lengths, although a producer may elect to do so. However, MLP is required for a single layer DVD-Audio disc to deliver a CD's worth (74 minutes) of 192kHz / 24-bit stereo or 5.1 channels of 96kHz / 24-bit data. All DVD-Audio players and universal players with DVD-Audio capability must provide MLP decoding capability. The MLP technology is licensed by Dolby Laboratories.

In addition to extending playing time, MLP also adds dynamic mixdown capability to DVD-Audio, so that the levels of individual channels can be changed over time.

5.2.4 DST (Direct Stream Transfer)

Because the single-bit DSD data format employed by SA-CD operates at such a high sampling rate (2.8Mbps per channel), approximately four times more raw audio data is generated than for equivalent CD signals. Data reduction is therefore necessary to store all eight (two stereo and six multichannel) high-density channels on a single layer. Although the sigma-delta modulated DSD signal inherently has a very noisy structure, lossless coding can be successfully applied in order to reduce the required capacity for storage or transmission. Direct Stream Transfer (DST) is the lossless coding method developed by Philips for SA-CD. Coding parameters such as prediction coefficients are optimized once per frame, resulting in a variable bit rate; buffer control is employed to convert this to the constant bit rate equivalent to linear disk playback speed.

Through the use of advanced data framing, adaptive prediction and entropy encoding methodology, DST typically achieves a 50% reduction in bit rate, with no loss in data integrity. SA-CD players are required to provide DST decoding capability; however, as noted in section 5.1.3, there are currently only a limited number of DSD editing and processing tools (including DST encoders) available on the market.

5.3 Surround Sound Mastering

Surround sound mastering is a highly specialized skill which differs significantly from stereo mastering. Although the purpose of mastering in surround is the same as that in stereo — to maximize the musical content inherent in the mix — it requires a somewhat different approach, as well as specialized equipment such as multichannel dynamics processors with flexible channel linking. Mastering engineers are also sometimes called upon to add reverb or other effects to a mix. In those cases, it is important to have access to equipment that can deliver both correlated and decorrelated multichannel processing (see section 4.7). In addition, the monitoring setup in a surround mastering facility is considerably more demanding, since the mastering engineer has to listen even more closely for phase anomalies.

Another important difference between surround sound mastering and stereo mastering — especially those stereo releases targeted for radio play — is that it is not necessary to highly compress the mixes in order to have them stand out. In the world of surround sound, compression should only be used to make the mix more exciting, not to raise the overall level. Because there are six or more speakers doing the work instead of just two, maximum dynamic range can and should be preserved. Similarly, mastering engineers also report that, in general, they need to apply less equalization to surround material than stereo material.

Even on projects destined for release on SA-CD or DVD-Audio, the surround mastering engineer should filter the LFE channel, despite the fact that the specification does not require such filtering. Limiting the frequency range of the LFE channel to only the bottom few octaves allows for more consistent playback in different environments. (See section 4.6.1) However, this does raise the spectre of potential latency issues, since the main channels remain unfiltered. If the LFE channel is delayed even slightly, it can completely cancel redirected bass content from the main channels when monitored on a bass-managed system. Therefore, <u>the surround mastering engineer must listen on a bass-managed system</u> to check the results of combining the low frequency signals. In some cases, it may be necessary to time-shift the main channels in order to ensure that all channels remain in absolute perfect phase after the filtering process.

Time code issues can also present problems when mastering surround sound that has to be locked to picture (such as music concerts presented on DVD-Video). (See section 5.1.1) For those projects, the mastering facility may have to serve almost as a postproduction house, with good quality video monitors, a source of extremely stable, jitter-free blackburst sync, and the ability to reconform.

Lastly, the job of encoding multichannel audio in Dolby Digital, DTS, MLP, and/or DST formats is often done by the surround sound mastering engineer, requiring additional critical listening and the presence of both encoding and decoding equipment (for playback monitoring purposes).

5.4 Sample Rate Conversion and Resampling

Music can be distributed in many different formats. The well-known Red Book CD — the traditional "CD" — is a single-layer optical disk with room for approximately 700 - 800 mbytes of data; that data is invariably stored as 16-bit words at 44.1 kilosamples per second and streamed at a constant rate of 1.411 Mpbs (Megabits per second).

As described in section 5.1, distribution formats have expanded markedly in the last few years. Some of them are described in the table below:

Distribution format	**Encoding format**	**Sample rate / Word length / Bit rate**
CD (Traditional "Red Book") Single stream	2-channel discrete PCM (unencoded)	44.1kHz 16-bit Constant-rate, 1.411Mbps stream
DVD-V Multiple streams:	2-channel discrete PCM (unencoded)	48 or 96 kHz 16, 20, 24-bit Constant-rate, 6.144Mbps max stream
	Dolby Digital (AC-3)	32, 44.1, 48kHz 16, 20-bit Constant-rate, 448kbps max stream
	DTS	48, or 96kHz 16, 20, 24-bit Constant-rate, 1.5Mbps max stream
DVD-A Multiple streams:	2-channel through 6- (5.1) channel discrete PCM (unencoded) or MLP	44.1kHz – 192kHz 16, 20, 24-bit, Variable-rate (MLP) or constant-rate (PCM), 9.6Mbps max stream
	Dolby Digital (AC-3) (contained in video area only)	32, 44.1, 48kHz 16, 20-bit Constant-rate, 448kbps max stream
	DTS (contained in video area only)	48 or 96kHz 16, 20, 24-bit Constant-rate, 1.5Mbps max stream
SA-CD Single or multiple streams:	2-channel and 6-channel, single-bit, high-rate encoding.	Release format (64FS = 2.8224Mbps bit rate DSD format)
Windows Media Audio 9, MP3 (MPEG-2 Level 3), AAC, RealAudio	Various bit-rate reduction methods	From very low to reasonably high rates

The mixing engineer may assume three things. First, the highest practical rate should be chosen for original recording and mastering. Second, *all* known resampling (sample-rate conversion) techniques are to some degree audible, although recent work has shown that with great care and provided with considerable DSP resources, one may resample, both up and down, very effectively. And third, an *integer ratio conversion* (i.e., 96kHz to 48kHz, or

88.2kHz to 44.1kHz) is more likely to have fewer resampling artifacts.
It must always be understood that when *upsampling* audio (converting to a higher sample rate, i.e., from 48kHz to 96kHz, or 44.1kHz to 96kHz), one will never restore the upper octave filtered out at the original A/D conversion (i.e., when one converts at 44.1kHz, there will be little signal above 20kHz *from then on*). However, there are possible benefits to be gained from upsampling when playback occurs. In any event, sample rate is not the sole, nor even the most significant factor in perceived audio quality; the design and quality of the A/D and D/A converters makes the most important sonic difference. A state-of-the-art audiophile 44.1kHz / 16-bit converter can easily sound better than a 192kHz / 24-bit converter built using an inferior chipset.

PCM audio may be converted to DSD (SA-CD) format and vice-versa. It is, however, not clear at this time exactly what PCM rate might be equivalent in quality to DSD (SA-CD) recording and reproduction.

5.5 Downmixing

Downmixing (sometimes called "fold-down") refers to the process of automatically reducing a multichannel mix to two channels — i.e., 5.1 playback to stereo. A downmix is almost always inferior to an original stereo mix; however, because almost all consumer receivers include the feature, it is a necessary part of surround sound production. Both Dolby Digital and MLP encoding, as well as the WMA9 delivery format, provide for metadata settings known as *downmix coefficients* to describe how the 5.1 mix will be converted to stereo. Note that if *no* downmixing coefficients are specified, MLP will create a downmix consisting of only the L and R channels! With DTS encoding and the SA-CD format, no downmixing is possible; indeed, SA-CDs always include a dedicated stereo mix.

When encoding for Dolby Digital, MLP, or WMA9, <u>the Center channel downmix coefficient should be set to -3dB in relationship to the Left and Right front channels</u>. However, the rear channel coefficients are largely program-dependent. Many of today's surround productions are mixed with an "in the band" perspective. (See section 3.3.1) In those instances, <u>the rear channels must be at the same level as the front left and right channels</u>. However, the rear channels *may* be lowered -3dB if they contain only ambience or audience material. The stereo sum of the 5.1 mix will usually need to be lowered in level in order to avoid creating signals so loud that they create digital "overs" (distortion). While the Dolby Digital encoder/decoder chain can handle the occasional "over," the MLP encoder has no built in compressor or limiter, so a single "over" during the course of the entire musical program will cause the critical MLP verification step to fail.

Typical settings for downmix coefficients are: Left Front = -6dB, Right Front = -6dB, Center = -9dB, LFE usually OFF or to taste (never more than -9dB in this scenario), Left Surround = -6dB, Right Surround = -6dB. Surround mixes should always be checked in this typical downmix configuration. The mix engineer should indicate to the mastering engineer whether these coefficients are suitable or whether any changes are necessary.

If surround sound material is Dolby Digital- or MLP-encoded, there is no way to completely avoid the possibility that a consumer might listen to a downmix instead of a true stereo mix. However, there are three ways to reduce the chances of that occurring. One is to ensure that a separate dedicated stereo mix is included on the disk; in the case of legacy material being repurposed, this should preferably be the original stereo mix. The second is to specify to the mastering engineer that the disk be authored with the stereo mix as the default audio stream. This will, of course, require that the consumer proactively select "5.1" if they want to hear the surround sound mix. However, if this is not done — if the disk is instead authored with the 5.1 mix as the default — then consumers who have only two speakers attached to their receiver will hear the downmix and not the stereo mix, even if one exists.

Finally, in the case of DVD-Audio, when authoring in "pgc block" mode (where there is a dedicated stereo mix present but the 5.1 audio stream is the default), the MLP encoding should have the downmix function turned OFF. While there are some legacy DVD-Audio players that do not recognize this flag, its use is nonetheless recommended since all new and future machines are designed to conform to it.

5.6 Upmixing ("Faux 5.1")

The process of building a multichannel mix is considerably more complex than stereo mixing, requiring a great deal of expertise and decision-making. For example, ambient spaces need to be created within the surround soundfield — something that is normally done with the application of precisely customized reverbs and delays. In addition, a multichannel mix has to be carefully crafted so that it maintains its integrity even when heard outside the "sweet spot" (i.e., when the consumer goes off to the kitchen to make a cup of coffee). And when a new multichannel mix is being made of a recording that was previously released in stereo (a process known as *repurposing*), it must match the aesthetic of the original. These are things that can only be achieved with the human ear and mind.

In recent years, a number of audio signal processors and computer programs have been developed that allow for automatic *upmixing* (sometimes called "unwrapping"); that is, they artificially convert stereo mixes to surround sound mixes. These tools provide some degree of control and can be of immense benefit to the engineer when repurposing music that was not originally recorded on multitrack, or when the multitrack tapes are damaged or unavailable. However, no upmixing tool can substitute for a true surround sound remix done from the original multitracks, for the following reasons:

- A true surround mix allows for more differentiated content. The end result is an expansion of the soundstage — there is far more access to instruments, with increased discrete placement of individual signals. In some cases, the engineer may discover musical parts on the multitrack that were not even used in the original stereo mix, and can thus create a "value-added" remix.

- Because individual instruments can be accessed in a true surround mix, entire new perspectives can be created: a lead vocal can be dry and upfront one moment, then wet and behind the listener the next, or a guitar solo can swirl above the listener's head. The end result can be an entirely new level of excitement — one that no computer program or signal processor can create.

- Through the use of discrete reverbs and delays, a skilled engineer can create a much wider sweet spot, so the listener doesn't have to be anchored in one area to hear the whole mix.

- Perhaps most importantly, with a full, original surround sound remix, there's a clear musical goal, as determined by the artist and record producer. No tool, however sophisticated, can possibly match the creativity of the human mind.

In short, a true surround sound mix created by a skilled engineer provides a better product than any computer program or signal processor can. For that reason, <u>we strongly recommend that producers and record labels refrain from using upmixing tools provided by signal processors or computer programs except in situations where the original multitrack tapes are damaged, unavailable, or do not exist</u>.

5.7 Labeling

As noted in section 5.4, sample rate is not the sole, nor even the most significant factor in perceived audio quality (the design and quality of the A/D and D/A converters makes the most important sonic difference). It is nonetheless important for consumers to be able to differentiate between high resolution (sample rates of 88.2kHz, 96kHz or higher, and 24-bit word lengths) and "CD-quality" (44.1 or 48kHz / 16-bit) surround sound productions.

Therefore, <u>we recommend the placement of a prominent label on all surround sound products to identify the sample rate and bit resolution of the original source material</u>, in the following format:

Audio Track	**Sample Rate**	**Bit Resolution**	**Compatibility**
(DVD-V PCM 2.0 Stereo, Dolby Digital 5.1 Surround, Dolby Digital-EX 6.1 Surround, DTS 5.1 Surround, DTS-ES 6.1 Surround, DVD-A PCM 5.1 Surround, DVD-A MLP 5.1 Surround, DVD-A PCM 2.0 Stereo, DVD-A MLP 2.0 Stereo SA-CD Multichannel, SA-CD Stereo)	(44.1kHz, 48kHz, 88.2kHz, 96kHz, 176.4kHz, 192kHz, DSD)	(16 bits, 20 bits, 24 bits, DSD)	(DVD-Audio, DVD-Video, SA-CD)

In addition, <u>we recommend the placement of a prominent label on all surround sound products to identify whether or not a true multichannel remix was done</u>.

Wherever a true multichannel mix has been done, we recommend the following terminology:

> "The surround performance on this DVD / SA-CD is a full remix taken from the original multitrack masters."

Wherever upmixing tools have instead been used, we recommend the following terminology:

> "The surround performance on this DVD / SA-CD was electronically recreated from the original stereo source without the benefit of the multitrack masters."

Glossary Of Terms:

5.1 - Describes the standard speaker setup used for surround sound playback: Five main (full-range) speakers (left, right, center, rear left, rear right), plus a subwoofer (the ".1"), which reproduces the LFE channel as well as (in most consumer systems) the low frequency information from the five main channels. (See "Bass management," "LFE" and "Subwoofer")

6.1 - An alternative variation on 5.1 which adds a center rear channel. Both Dolby-EX™ and DTS-ES are 6.1 formats.

7.1 - An alternative variation on 5.1 which adds two additional speakers. In some consumer systems, the extra speakers act as sidefills. Sony SDDS™ is a cinematic 7.1 configuration which places the extra speakers behind the movie screen.

AC-3 - MPEG-2 Level 3, Dolby Laboratories' third-generation audio encoding algorithm. (See "Dolby Digital")

Bass management - Describes a routing scheme (employed in most consumer 5.1 systems) whereby the low frequency information from all five main channels is routed to the subwoofer, along with the LFE channel information. (See "LFE" and "Subwoofer")

Bit - An acronym for Binary Digit — literally a one or a zero. Computers store all data in the form of bits; digital audio systems are simply highly specialized computers that store audio data in the form of bits. (See "Bit rate" and "Bit resolution")

Bit rate - A measurement of how many bits of data are streamed per second. Depending upon the distribution format, bit rates may be fixed or variable. (See "Mbps")

Bit resolution - (Sometimes called "word length") Describes how many bits are used to describe each sample of audio data in a digital audio system. The greater the number of bits used, the greater the dynamic range (the difference between the softest and loudest sections); the greater the dynamic range, the more realistic the sound. CDs are 16-bit systems while DVD-Audio disks are capable of carrying 24-bit signals (both in PCM format); SA-CDs carry only 1-bit signals (in DSD format), but at very high sample rates. Most professional digital recording systems today employ 24-bit PCM technology. (See "Bit," "DSD," "DVD-Audio," "Dynamic range," "PCM," and "SA-CD")

CD-DA - An acronym for Compact Disc Digital Audio, the audio-only version of CD (as opposed to CD-ROM and other formats for data storage on CD).

CD-quality - A term used to describe 16-bit digital audio recorded at a 44.1kHz sampling rate. Warning: Due to differences in circuitry, software design, and the skill of the recording engineer, not all "CD-quality" audio is the same! (See "Bit resolution," "kHz," and "Sampling rate")

Codec - An algorithm that describes both en<u>co</u>ding and <u>dec</u>oding. (See "WM9")

dB - Short for deciBel. A unit of measurement that, in audio, describes sound pressure level or amplifier gain. The decibel scale is logarithmic, meaning that an increase of 10dB represents a tenfold increase in power: it takes ten times as much energy to produce an SPL of 80dB as one of 70dB, even though 80dB is subjectively only twice as loud as 70dB. Doubling the power raises the SPL by about 3dB; cutting the power in half reduces the volume about 3dB.

Dipole - A loudspeaker design whereby signal is radiated in two directions simultaneously — forwards and backwards — with equal energy but with opposite polarity. Dipole speakers provide better envelopment than direct radiator speakers, but with very poor localization. (See "Direct radiator")

Direct radiator - The most commonly used loudspeaker design, where the speaker cone fires in one direction only. This type of design provides the smoothest frequency response as well as the best localization qualities. (See "Dipole")

Decoding - The process of restoring an encoded signal. (See "Encoding")

Delta - Refers to change in a signal. DSD is a one-bit delta encoding system, where each bit indicates either a positive change in waveform amplitude (a value of one) or a negative change in waveform amplitude (a value of zero).

Dolby® Digital™ - A lossy encoding scheme developed by Dolby Laboratories which greatly reduces the size of digital audio data by discarding information that, in theory, cannot be perceived. Technically called "AC-3." DVD-Video disks released in North America must carry a Dolby® Digital™ encoded surround sound track. (See "DVD-Video," "Encoding," and "Lossy encoding")

Dolby-EX™ - See "6.1"

Downmixing - (sometimes referred to as "Fold-down") The process of reducing a multichannel mix to stereo or mono. Downmixing may be automatically done by some consumer playback equipment, making it critical that its integrity first be checked by human ears! (See "Upmixing")

DSD - An acronym for Direct Stream Digital. A digital recording process designed by Sony and Philips that uses only a single bit, combined with an extremely high sample rate (2.8MHz, or 2.8 *million* samples per second), to provide very high quality audio. (See "Bit resolution," "SA-CD," and "Sample rate")

DTS - An acronym for Digital Theater Systems. Commonly refers to the company's lossy encoding scheme, which greatly reduces the size of digital audio data by discarding information that, in theory, cannot be perceived. (See "DVD-Video," "Encoding," and "Lossy encoding")

DTS-ES - See "6.1"

DVD-Audio - A specialized type of DVD where most of the available storage space is dedicated to digital audio (as opposed to digital video) data, allowing for very high fidelity sound. By using a lossless encoding scheme called MLP, a DVD-Audio disk can accommodate up to eight channels of audio (typically, a stereo mix plus a 5.1 mix), at sample rates of up to 192kHz, and with a bit resolution of up to 24 bits. (See "Bit Resolution," "Lossless encoding," "MLP," and "Sample rate")

DVD-Video - The most common type of DVD. (There are actually five types of DVD, the other four being DVD-Audio, DVD-ROM, DVD-RAM, and DVD-WO [Write Once].) DVD-Video uses most of the storage area for digital video data, with only a small portion set aside for digital audio. DVD-Video disks released in North America must carry at least one PCM and one encoded track (usually in Dolby Digital format). Optionally, they may carry additional audio tracks (such as DTS-encoded surround sound) if space permits. (See "Dolby Digital," "DTS," "DVD-Audio," "Encoding," and "PCM")

Dynamic range - The difference between the softest perceived sound and the loudest perceived sound. The dynamic range of human hearing is approximately 120dB. DSD coding is said to deliver a full 120dB of dynamic range. In PCM systems, the dynamic range is a function of bit resolution. 16-bit systems deliver approximately 96dB of dynamic range; 20-bit systems deliver approximately 120 dB of dynamic range; and 24-bit systems deliver approximately 144dB of dynamic range. (See "Bit resolution," "dB," "DSD," and "PCM")

Encoding - The process of altering a signal, usually in order to conserve space for storage purposes. An encoded signal must be decoded in the same way by the playback system in order to be restored. (See "DTS," "Decoding," "Dolby Digital," "Lossless Encoding," "Lossy Encoding," and "MLP")

Envelopment - A term used to describe the degree to which a signal is perceived as being all around the listener, with indeterminate localization. (See "Localization")

Faux 5.1 - A tongue-in-cheek term (as in "four, five, one") for 5.1 mixes automatically created by signal processors or computer algorithms.

Fold-down - See "Downmixing."

Hz - Short for Hertz, the measurement of a sound's wavecycle. One Hertz equals one wavecycle per second. The range of human hearing is approximately 20Hz to 20,000Hz (20kHz).

kbps - Short for kilobits (thousands of bits) per second. A unit of measurement that defines the bit rate. (See "Bit rate" and "Mbps")

kHz - Short for KiloHertz, or 1,000 Hertz. (See "Hz")

ITU - An acronym for the International Telecommunications Union, an international standards organization based in Europe.

LFE - An acronym for "Low Frequency Effects" — the ".1" channel in 5.1 (because it carries only approximately one-tenth of the frequency range of the main channels). In postproduction, this channel is used to provide rumble from explosions and the like; in music, it is sometimes used to carry very low frequency information from kick drums and/or bass, or sometimes not at all.

Localization - A term used to describe the degree to which the location of a signal can be accurately perceived. (See "Envelopment")

Lossless encoding - A method of reducing the size of data by "packing" it in a more efficient manner without actually discarding any information. MLP, used in DVD-audio, is a lossless encoding scheme. (See "DVD-Audio," "Encoding," "Lossy encoding," and "MLP")

Lossy encoding - A method of reducing the size of data by discarding unnecessary information. Sometimes called "Perceptual encoding" because, in theory, the data that is discarded would not be perceived anyway. Dolby Digital and DTS are forms of lossy encoding used in digital audio. (See "Dolby Digital," "DTS," "DVD-Video," "Encoding," and "Lossy encoding")

Mbps - Short for Megabits (millions of bits) per second. A unit of measurement that defines the bit rate. (See "Bit rate" and "kbps")

Midfield monitors – Full-range studio speakers that are designed to be used free-standing. (See “Nearfield monitors”)

MLP - An acronym for Meridian Lossless Packing. An optional (but often-used) lossless encoding scheme for DVD-Audio that greatly reduces the size of digital audio data by "packing" the data in a more efficient format without actually discarding any of it. (See "DVD-Audio," "Encoding," and "Lossless encoding")

Nearfield monitors – Studio speakers that are designed to be placed on top of a mixing console meter bridge in order to aid their low-frequency response. (See “Midfield monitors”)

PCM - An acronym for Pulse Code Modulation. A common way of storing unencoded digital audio data (used in audio CD and both DVD-Video and DVD-Audio), PCM describes each instant of audio signal at a specified sample rate, and with a specified bit resolution. The information on an audio CD is stored in PCM format, as 16-bit words with a sample rate of 44.1kHz. The audio information on a DVD is also stored in PCM format, and may be as 16-, 20-, or 24-bit words, with a sample rate of up to 192kHz. (See "Bit resolution," "CD-quality," "Encoding," "kHz," and "Sample rate")

Perceptual encoding - See "Lossy encoding"

Red Book - The official audio CD (CD-DA) specification, which includes definitions of sample rate (44.1kHz), bit length (16 bits) and bit rate (1.411Mbps). (See "Bit length," "Bit rate," "CD-DA," and "Sample rate")

Repurposing - A term used to describe the reformatting of existing recorded music; i.e., repurposing an existing catalog of mono and stereo recordings to 5.1.

SA-CD - An acronym for Super Audio CD. Developed by Sony and Philips, the SA-CD is the same physical size as a CD or DVD, although it is capable of storing data in DSD, as opposed to PCM, format. The SA-CD specification (called the “Scarlet Book”) provides three disk format options: single-layer DSD, dual-layer DSD, or dual-layer hybrid. The latter includes a standard CD “Red Book” layer that can be played on any existing CD player, in addition to a high-density layer that has the capacity to deliver eight channels of DSD. To access DSD audio (which can be stereo or multichannel), SA-CDs must be played back in specialized or universal players. (See "CD-quality," "DSD," "PCM," "Red Book," “Scarlet Book” and "Universal players.")

Sample rate - Describes the number of "snapshots" (samples) taken by a digital recorder in a given time period. The higher the sample rate, the greater the frequency range of the resulting audio (and the better the sound). CDs use a rate of 44,100 samples per second (44.1kHz), while the sample rate most commonly used by professional studio-quality recorders is 96kHz. Other commonly used sample rates are 22.05kHz (used in Internet streaming and multimedia presentations), 48kHz, and 88.2kHz. The highest sample rate available today, though rarely used, is 192kHz.

Satellite speakers - A term used by some consumer electronics systems to describe the small speakers in a 5.1 system. Satellite speakers generally have extremely limited low frequency response, and so a bass management scheme employing the subwoofer is usually required in order to reproduce a full range listening experience. (See "5.1," "Bass management," and "Subwoofer")

Scarlet Book – The official SA-CD specification. (See “Red Book” and “SA-CD”)

Sony SDDS - A cinematic 7.1 surround sound configuration which places five full-range speakers behind the screen — left, left center (“left extra”), center, right center (“right extra”), and right — along with two rear surround speakers and a subwoofer. (See "7.1")

SMPTE - An acronym for the Society of Motion Picture and Television Engineers, an international membership and standards organization.

Subwoofer - A special kind of speaker designed to carry low frequency information only (generally signals up to approximately 120Hz). Because the human ear cannot easily determine the directionality of low frequency signals, a single subwoofer is generally sufficient to carry low frequency information from all main channels, and its physical location within a room is not especially critical. (See "Bass management" and "Hz")

Surround sound - A catch-all term used to describe any multichannel audio system that requires more than two speakers for playback. The most common surround sound speaker configuration is 5.1. (See "5.1")

Universal players - Devices that can play CDs as well as SA-CDs and both kinds of DVDs (DVD-Video and DVD-Audio). (See "DVD-Audio," "DVD-Video," and "SA-CD")

Unwrapping - A commonly used term to describe the use of a signal processor or computer program to automatically upmix a stereo signal to 5.1. (See "Upmixing")

Upmixing - Increasing the number of channels in a mix, i.e., from mono to stereo, or from stereo to 5.1. (See "Downmixing")

WMA9 Pro - Short for "Windows Media Audio 9 Professional," a lossy high-resolution stereo / multichannel codec developed by Microsoft for Internet audio streaming. The WMA family also includes a stereo-only codec (WMA9) and WMA9 Lossless, an archiving/mastering codec. (see "Codec," “Lossless Encoding” and "Lossy Encoding")

Word length - See "Bit resolution"

X-curve - A 3dB/octave slope starting at 2.5kHz designed to compensate for room acoustics and screen absorption so as to provide a uniform frequency response adjustment for all theaters throughout the world.