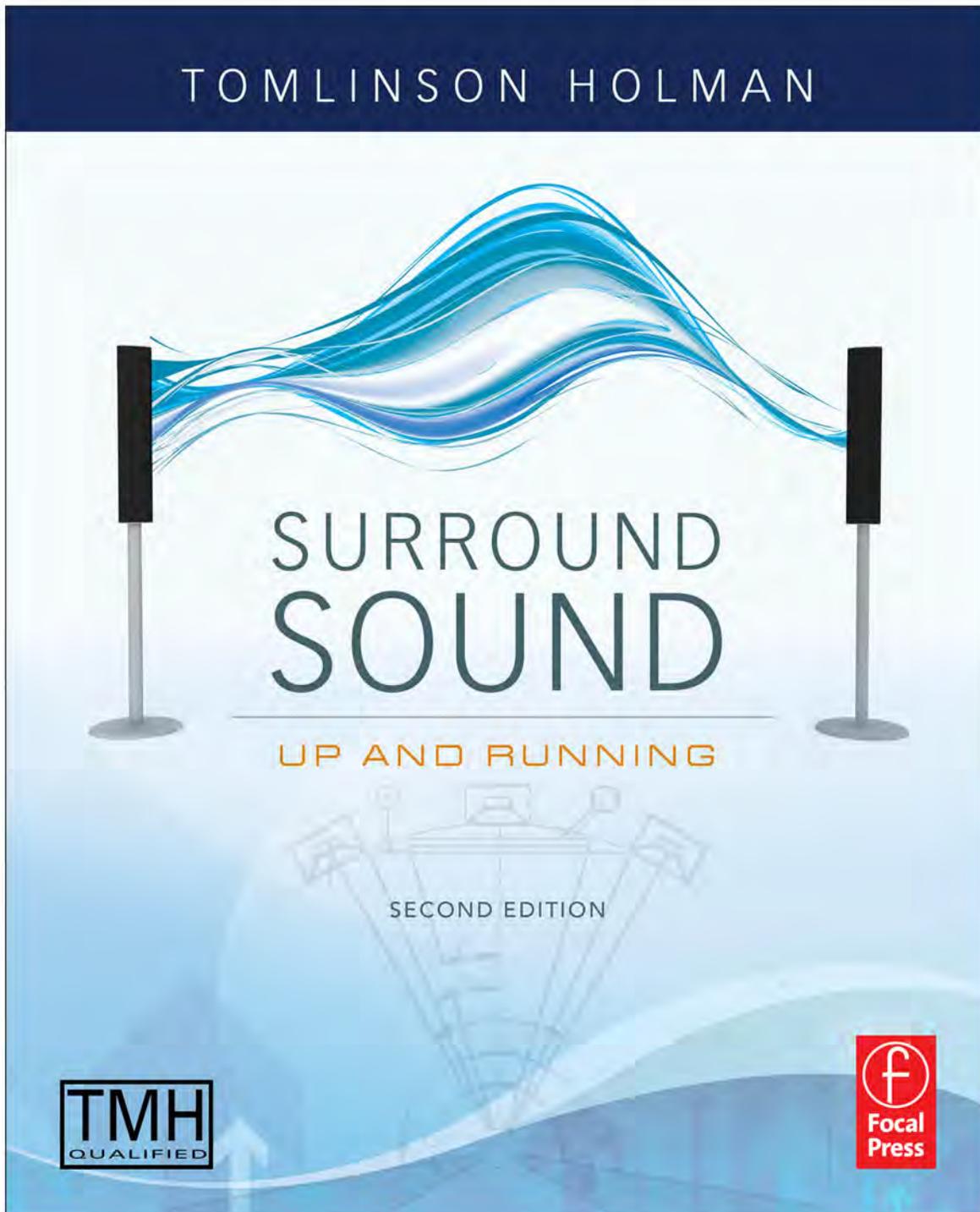




**Like what you see?  
Get the whole book at the  
Focal Press Bookstore**



**<http://focalbookstore.com/?isbn=9780240808291>**



# 1 Introduction

---

Recorded multichannel sound has a history that dates back to the 1930s, but in the intervening years up to the present it became prominent, died out, became prominent again, and died out again, in a cycle that's had at least four peaks and dips. In fact, I give a talk called "The History and Future of Surround Sound" with a subtitle "A Story of Death and Resurrection in Five Acts" with my tongue barely in my cheek. In modern times we see it firmly established in applications accompanying a picture for movies and television, but in other areas such as purely for music reproduction broad success has been more elusive. However using history as a guide to the future there may well be a broader place for surround music in coming years.

The purpose of this book is to inform recording engineers, producers, and others interested in the topic, about the specifics of multichannel audio. While many books exist about recording, post production, etc., few yet cover multichannel and the issues it raises in any detail. Good practice from conventional stereo carries over to a multichannel environment; other practices must be revised to account for the differences between stereo and multichannel. Thus, this book does not cover many issues that are to be found elsewhere and that have few differences from stereo practice. It does consider those topics that differ from stereo practice, such as how room acoustics have to be different for multichannel monitoring, for instance.

Five-point-one channel sound is the standard for multichannel sound in the mass market today. The 5 channels are left, center, right, left surround, and right surround. The 0.1 channel is called Low-Frequency Enhancement, a single low-frequency only channel with 10dB greater headroom than the five main channels. However, increasing pressure exists on the number of audio channels, since spatial "gamut"<sup>1</sup> (to borrow a term from our video colleagues) is easily audible to virtually

everyone. Seven-point-one channel sound is in fairly widespread use, and will be defined later. My colleagues and I have been at work on “next generation” 10.2-channel sound, for some years, and it also will be treated later.

## A Brief History

The use of spatial separation as a feature of composition probably started with the call-and-response form of antiphonal music in the medieval period, an outgrowth of Gregorian chant. Around 1550, Flemish composer Adrian Willaert working in Venice used a chorus in left and right parts for antiphonal singing, matching the two organ chambers oriented to either side of the altar at St. Mark’s Basilica. More complex spaced-antiphonal singing grew out of Willaert’s work by Giovanni Gabrieli beginning about 1585, when he became principal organist at the same church. He is credited with being the first to use precise directions for musicians and their placement in more than simple left-right orientation, and this was highly suited to the site, with its cross shape within a nearly square footprint (220 × 250’). In the developing polyphonic style, the melodic lines were kept more distinct for listeners by spatially separating them into groups of musicians. The architecture of other churches was affected by the aural use of space as were those in Venice, including especially Freiburg Cathedral in Germany, finished in 1513, which has four organ cabinets around the space. Medieval churches at Chartres, Freiburg, Metz, and Strasbourg had “swallow’s nest” organs, located high up in the nave (the tallest part of the church) “striving to lift music, embodied in the organ, high up into the light-filled interior, as a metaphor for *musica divina*”<sup>2</sup>

The Berlioz *Symphonie Fantastique* (1830) contains the instruction that an oboe is to be placed off stage to imply distance. The composer’s *Requiem* (1837) in the section “Tuba mirum” uses four small brass-wind orchestras called “Orchestra No. I to the North, Orchestra No. II to the East, Orchestra No. III to the West, and Orchestra No. IV to the South,” emphasizing space, beginning the problem of the assignment of directions to channels! The orchestration for each of the four orchestras is different.

Off-stage brass usually played from the balcony is a feature of Gustav Mahler’s Second Symphony *Resurrection*, and the score calls for some of the instruments “in the distance”. So the idea of surround sound dates

---

<sup>1</sup>Meaning the range of available colors to be reproduced; here extended to mean the range of available space.

<sup>2</sup><http://www.koelner-dommusik.de/index.php?id=7&L=6>

back at least half a millennia, and it has been available to composers for a very long time.

The original idea for stereo reproduction from the U.S. perspective<sup>3</sup> was introduced in a demonstration with audio transmitted live by high-bandwidth telephone lines from the Academy of Music in Philadelphia to the then four-year old Constitution Hall in Washington, DC on April 27, 1933. For the demonstration, Bell Labs engineers described a 3-channel stereophonic system, including its psychoacoustics, and wavefront reconstruction that is at the heart of some of the “newer” developments in multichannel today. They concluded that while an infinite number of front loudspeaker channels was desirable, left, center, and right loudspeakers were a “practical” approach to representing an infinite number. There were no explicit surround loudspeakers, but there was the fact that reproduction was in a large space with its own reverberation, thus providing enveloping sound acoustically in the listening environment.

Leopold Stokowski was the music director of the Philadelphia Orchestra at the time, but it is interesting to note that he was not in Philadelphia conducting, but rather in Washington, operating 3-channel level and tone controls to his satisfaction. The loudspeakers were concealed from view behind an acoustically transparent but visually opaque scrim-cloth curtain. A singer walked around the stage in Philadelphia, and the stereo transmission matched the location. A trumpeter played in Philadelphia antiphonally with one on the opposite side of the stage in Washington. Finally the orchestra played. The curtain was raised, and to widespread amazement the audience in Washington found they were listening to loudspeakers!

Stokowski was what we would today call a “crossover” conductor, interested in bringing classical music to the masses. In 1938 he was having an affair with Greta Garbo, and to keep it reasonably private, wooed her at Chasen’s restaurant in Hollywood, a place where discretion ruled. Walt Disney, having heard “The Sorcerer’s Apprentice” in a Hollywood Bowl performance and fallen for it had already had the idea for *Fantasia*. According to legend, one night at Chasen’s Disney prevailed upon a waiter to deliver a note to Stokowski saying he’d like to meet. They turned out already to be fans of each other’s work, and Disney pitched Stokowski on his idea for a classical music animated film. Stokowski was so enthusiastic that he recorded the music for *Fantasia* without a fee, but he also told Walt he wanted to do it in stereo, reporting to him on the 1933 experiment. Walt Disney, sitting in his

---

<sup>3</sup>Alan Blumlein, working at EMI in England in the same period also developed many stereo techniques, including the “Blumlein” pair of crossed figure-8 microphones. However, his work was in two channel stereo.

living room in Toluca Lake some time later thought that during “The Flight of the Bumblebee,” not only should the bumblebee be localizable across the screen, but also around the auditorium. Thus surround sound was born. Interestingly “The Flight of the Bumblebee” does not appear in *Fantasia*. It was set aside for later inclusion in a work that Walt saw as constantly being revised and running for a long, long time. So even that which ends up on the proverbial cutting room floor can be influential, as it here invented an industry. The first system of eight different variations that engineers tried used three front channels located on the screen, and two surround channels located in the back corners of the theater—essentially the 5.1-channel system we have today. During this development, Disney engineers invented multitrack recording, pan potting, overdubbing, and when it went into theaters, surround arrays. Called Fantasound, it was the forerunner of all surround sound systems today.

While guitarist Les Paul is usually given the credit for inventing overdubbing in the 1950s, what he really invented was Sel-Sync recording, that is, playing back off the record head of some channels of a multitrack tape machine, used to cue (by way of headphones) musicians who were then recorded to other tracks; this kept the new recordings in sync with the existing ones. In fact, overdub recording by playing back an existing recording for cueing purposes over headphones was done for *Fantasia*, although it was before tape recording was available in the U.S. Disney used optical sound tracks, which had to be recorded, then developed, then printed and the print developed, and finally played back to musicians over headphones who recorded new solo tracks that would run in sync with the original orchestral recording. Then, after developing the solo tracks, and printing and developing the prints, both the orchestral and solo tracks were played simultaneously on synchronized optical dubbers (also called “dummies”) and remixed. By changing the relative proportions of orchestra and solo tracks in the mix the ability to vary the perspective of the recording from up-front solos to full orchestra was accomplished.

So multichannel recording has a long history of pushing the limits of the envelope and causing new inventions, which continues to the present day. By the way, note that multichannel is different from multitrack. Multichannel refers to media and sound systems that carry multiple loudspeaker channels beyond two. Multitrack is, of course, a term applied to tape machines that carry many separate tracks, which may, or may not, be organized in a multichannel layout.

*Fantasia* proved to be a one-shot trial for multichannel recording before World War II intruded. It was not financially successful in its first

run, and the war intervened too soon for it to get a toehold in the marketplace. The war resulted in new technology introductions that proved to be useful in the years following the war. High-quality permanent magnets developed for airplane crew headphones made better theater loudspeakers possible (loudspeakers of the 1930s had to have DC supplied to them to make a magnetic field). Magnetic recording on tape and through extension magnetic film was war booty, appropriated from the Germans. Post-war a pre-war invention was to come on the scene explosively, laying wreck to the studio system that was vertically integrated from script writing through theatrical exhibition and that had dominated entertainment in the 1930s—television. Between 1946 and 1954 the number of patrons coming to movie theaters annually fell by about one-half, largely due to the influence of television.

Looking for a post-war market for film-based gunner training simulators, Fred Waller produced a system shot on three parallel strips of film for widescreen presentation and having 7 channels of sound developed by Hazard Reeves called Cinerama.<sup>4</sup> Opening in 1952 with *This is Cinerama*, it proved a sensation in the few theaters that could be equipped to show it. The 7 channels were deployed as five across the screen, and two devoted to surrounds. Interestingly, during showing a playback operator could switch the surround arrays from the two available source tracks to left and right surrounds, or front sides and back surrounds, thus anticipating one future development by nearly 50 years.

In 1953 20th Century Fox, noting both the success of Cinerama but also its uneconomic nature, and finding the market for conventional movies shrinking, combined the 1920s anamorphic photography invention by French Professor Henri Chrétien, with four-track magnetic recordings striped along the edges of the 35 mm release print, to produce a wide screen system that could have broader release than Cinerama films. They called it Cinemascope. The 4-channel system called for three screen channels, and one channel directed to loudspeakers in the auditorium. Excess noise in the magnetic reproduction of the time was partially overcome by the use of high-frequency standardized playback rolloff called generically the Academy curve. When this proved to be not enough filtering so that hiss was still heard from the “effects” loudspeakers during quiet passages, a high-frequency 12 kHz sine-wave switching tone was recorded on the track in addition to the audio to trigger the effects channel on and off. Played with a notch filter for the audio so that the tone would not be audible, simultaneously a controlling parallel side chain was peaked up at the tone frequency and

---

<sup>4</sup>A good documentary about the subject is *Cinerama Adventure*, <http://www.cineramaadventure.com/>

used to decide when to switch these speakers on and off. Today these speakers form what we call the surround loudspeakers.

In 1955, a six-track format on 70 mm film was developed by the father and son team Michael Todd and Michael Todd, Jr. for road show presentations of movies. Widescreen photography was accomplished for Todd AO<sup>5</sup> without the squeezing of the image present in Cinemascope through the simple expedient of making the film twice as wide. The six tracks were deployed as five screen channels as in Cinerama, and one surround channel. Interestingly, both the Todds had worked at Cinerama producing the most famous sequence in *This is Cinerama*, the roller coaster ride. They realized that the three projectors and one sound dubber and all the operators needed by Cinerama could be replaced by one wide format film with six magnetic recordings on four stripes, with two tracks outside the perforations and one inside, on each side of the film, thus lowering costs of exhibition greatly while maintaining very high performance. The idea was described as "Cinerama from one hole." In fact the picture quality of the 1962 release of *Lawrence of Arabia* shot in 65 mm negative for 70 mm release has probably never been bested.

While occasional event road shows like *2001* (1968) continued to be shot in 65 mm, the high cost of striping release prints put a strong damper on multichannel recording for film, and home developments dominated for years from the mid-1950s through the mid-1970s. Stereo came to the home as a simplification of theater practice. While three front channels were used by film-based stereophonic sound, a phonograph record only has two groove walls that can carry, without complication, two signals. Thus, two channel stereo reproduction in the home came with the introduction of the stereo LP in 1958, and other formats followed because there were only two loudspeakers at home. FM radio, various tape formats, and the CD followed the lead established by the LP.

Two-channel stereo relied upon the center of the front sound stage to be reproduced as what is called a phantom image. These sounds made a properly set-up stereo seem almost magical, as a sound image floated between the loudspeakers. Unfortunately, this phantom effect is highly sweet-spot dependent, as the listener must sit along the centerline of the speakers, and the speakers and room acoustics for the two sides must be well matched. It was good for a single listener experience, and drove a hobbyist market for a long time, but was not very suitable for multiple listeners. The reasons for this are described in Chapter 6 on Psychoacoustics.

---

<sup>5</sup>The Todds partnered with the American Optical company for Todd AO.

The Quad era of the late 1960s through 1970s attempted to deliver four signals through the medium of two tracks on the LP using either matrices based on the amplitude and phase relations between the channels, or on an ultrasonic carrier requiring bandwidth off the disc of up to 50kHz. The best known reasons for the failure of Quad include the fact that there were three competing formats on LP, and one tape format, and the resistance on the part of the public to put more loudspeakers into their listening rooms. Less well known is the fact that Quad record producers had very different outlooks as to how the medium should be used, from coming closer to the sound of a concert hall than stereo by putting the orchestra up front and the hall sound around the listener, to placing the listener “inside the band,” a new perspective for many people that saw both tasteful, and some not so tasteful, presentations. Also, the psychoacoustics of a square array of loudspeakers were not well understood at the time, leading to the absurdity that some people still think: that Quad is merely stereo reproduced in four quadrants, so that if you face each way of a compass in turn you have a “stereo” sound field. The problem with this though is that we can’t face the four quadrants simultaneously. Side imaging works much differently from front and back imaging because our ears are at the two sides of our heads. The square array would work well if we were equipped by nature with four ears at 90° intervals about our heads! When it came to be studied by P.A. Ratliff at the BBC,<sup>6</sup> it is fair to say that the quadraphonic square array got its comeuppance.

The misconception that Quad works as four stereo systems ranks right up there with the inanity that since we have two ears, we must need only two loudspeakers to reproduce sound from every direction—after all, we hear all around don’t we? This idea neglects the fact that each loudspeaker channel does not have input to just one ear, but two, through “crosstalk” components consisting of the left loudspeaker sound field reaching the right ear, and vice versa. Even with sophisticated crosstalk cancellation, which is difficult to make anything other than supremely sweet-spot sensitive, it is still hard to place sounds everywhere around a listener. This is because human perception of sound uses a variety of cues to assign localization to a source, including interaural level differences, that is, differences in level between the two ears; interaural time differences; the frequency response for each angle of arrival associated with a sound field interacting with the head, outer ears called the pinnae, shoulder bounce, etc. and collectively called the head-related transfer function (HRTF); and dynamic variations on these static cues since our heads are rarely clamped down

---

<sup>6</sup><http://www.bbc.co.uk/rd/pubs/reports/1974-38.pdf>

to a fixed location. The complexity of these multiple cue mechanisms confounds simple-minded conclusions about how to do all-round imaging from two loudspeakers.

While home sound developed along a two channel path, cinema sound enjoyed a revival starting in the middle 1970s with several simultaneous developments. The first of these was the enabling technology to record 4 channels worth of information on an optical sound track that had space for only two tracks and maintain reasonable quality. Avoiding the expensive and time-consuming magnetic striping step in the manufacture of prints, and having the ability to be printed at high speed along with the picture, stereo optical sound on film made other improvements possible. For this development, a version of an amplitude-phase matrix derived from one of the quadraphonic systems but updated to film use was used. Peter Scheiber's amplitude-phase matrix method was the one selected. Dubbed Dolby Stereo in this incarnation, it was a fundamental improvement to optical film sound, combining wider frequency and dynamic ranges through the use of companding noise reduction with the ability to deliver two tracks containing 4 channels worth of content by way of the matrix. In today's world of some 30 years later, the resulting dynamic range and difficulties with recording for the matrix make this format seem dated, but at the time, coming out of the mono optical sound on film era, it was a great step forward.

The technical developments would probably not have been sustained—after all, it cost more—if not for particularly good new uses that were found immediately for the medium. *Star Wars*, with its revolutionary sound track, followed in six months by *Close Encounters of the Third Kind*, cemented the technical developments in place and started a progression of steady improvements. At first Dolby A noise reduction was used principally to gain bandwidth. Why a noise reduction system may be used to improve bandwidth may not be immediately obvious. Monaural sound tracks employed a great deal of high-frequency rolloff in the form of the Academy filter to overcome the hiss associated with film grain running past the optical reader head. Simply extending the bandwidth of such tracks would reveal a great deal of hiss. By using Dolby A noise reduction, particularly the 15dB it achieved at high frequencies, bandwidth could be extended from effectively about 4 to 12kHz while keeping the noise floor reasonable and bringing film sound into the modern era. Second generation companding signal processing, Dolby SR, was subsequently used on analog sound tracks starting in 1986, permitting an extension to 16kHz, greater headroom versus frequency, lower distortion, and noise performance such that the sound track originated hiss remained well below the noise floor of good theaters.

With the technical improvements to widespread 35mm release prints came the revitalization of the older specialty format, 70mm. Premium prints playing in special theaters offered a better experience than the day-to-day 35mm ones. Using six-track striped magnetic release prints, running at 22.5ips, with wide and thick tracks, the medium was unsurpassed for years in delivery of wide frequency and dynamic ranges, and multichannel, to the public at large. In the 1970s Dolby A noise reduction allowed a bandwidth extension in playback of magnetic tracks, just as it had with optical ones, and some use was made of Dolby SR on 70mm when it became available.

For 70mm prints of *Star Wars*, the producer Gary Kurtz realized along with Dolby personnel including Steve Katz, that the low-frequency headroom of many existing theater sound systems was inadequate for what was supposed to be a war (in space!). Reconfiguring the tracks from the original six-track Todd AO 70mm format, engineers set a new standard: three screen channels, one surround channel, and a "Baby Boom" channel, containing low-frequency only content, with greater headroom accomplished by using separate channels and level adjustments. An extra low-frequency only channel with greater headroom proved to be a better match to human perception. This is because hearing requires more energy at low frequencies to sound equally as loud as the mid-range. *Close Encounters of the Third Kind* then used the first dedicated subwoofers in theaters (*Star Wars* used the left-center and right-center loudspeakers from the Todd AO days that were left over in larger theaters), and the system became standard.

Two years after these introductions, *Superman* was the first film since Cinerama to split the surround array in theaters into two, left and right. The left/right high-frequency sound for the surrounds was recorded on the left-center and right-center tracks, while the boom track continued to be recorded on these tracks at low frequencies. Filters were used in cinema processors to split bass off from treble to produce a boom channel and stereo surrounds. For theaters equipped only for mono surround, the wide range surround track on the print sufficed. So one 70mm stereo-surround print was compatible for playback in both mono- and stereo-surround equipped theaters. In the same year, *Apocalypse Now* made especially good artistic use of stereo surround, and most of the relatively few subsequent 70mm releases used the format of three screen channels, two surround channels, and a boom channel.

In 1987, when a subcommittee of the Society of Motion Picture and Television Engineers (SMPTE) looked at putting digital sound on film, meetings were held about the requirements for the system. In a series of meetings and documents, the 5.1-channel system emerged as being the minimum number of channels that would create the sensations

desired from a new system, and the name 5.1 took hold from that time. In fact, this can be seen as a codification of existing 70mm practice that already had five main channels and one low-frequency only, higher headroom channel.

With greater recognition among the public of the quality of film sound that was started by *Star Wars*, home theater began, rather inauspiciously at first, with the coming of two channel stereo tracks to VHS tape and Laser Disc. Although clearly limited in dynamic range and with other problems, early stereo media were quickly exploited as carriers for two-channel encoded Dolby Stereo sound tracks originally made for 35mm optical film masters, called Lt/Rt (left total, right total, used to distinguish encoded tracks from conventional stereo left/right tracks), in part because the Lt/Rt was the only existing two channel version of a movie, and copying it was the simplest thing to do to make the transfer from film to video.

Both VHS and Laser Disc sound tracks were enhanced when parallel, and better quality, two channel stereo recording methods were added to the media. VHS got its "Hi-Fi" tracks, recorded in the same area as the video and by separate heads on the video scanning drum as opposed to the initial longitudinal method, which suffered greatly from the low tape speed, narrow tracks, and thin oxide coating needed for video. The higher tape-to-head speed, FM recording, and companding noise reduction of the Hi-Fi tracks contributed to the more than 80dB signal-to-noise ratio, a big improvement on the "linear" tracks, although head and track mismatching problems from recorder to player could cause video to "leak" into the audio and create an annoying variable buzz, modulated by the audio. Also, unused space was found in the frequency spectrum of the signals on the Laser Disc to put in a pair of 44.1kHz, 16-bit linear PCM tracks, offering the first medium to deliver digital sound in the home accompanying a picture.

The two higher-quality channels carried the Dolby Stereo encoded sound tracks of more than 10,000 films within a few years, and the number of home decoders, called Dolby Pro Logic, is in the many tens of millions. This success changed the face of consumer electronics in favor of the multichannel approach. Center loudspeakers, many of dubious quality at first but also offered at high quality, became commonplace, as did a pair of surround loudspeakers added to the left-right stereo that many people already had.

Thus, today there is already a playback "platform" in the home for multichannel formats. The matrix, having served long and well and still growing, nevertheless is probably nearing the end of its technological lifetime.

Problems in mixing for the matrix include:

- Since the decoder relies on amplitude and phase differences between the 2 channels, interchannel amplitude or phase difference arising from errors in the signal path leads to changes in spatial reproduction: a level imbalance between Lt and Rt will “tilt” the sound towards the higher channel, while phase errors will usually result in more content coming from the surrounds than intended. This problem can be found at any stage of production by monitoring through a decoder.
- The matrix is very good at decoding when there is one principal direction to decode at one time, but less good as things get more complex. One worst case is separate talkers originating in each of the channels, which cause funny steering artifacts to happen (parts of the speeches will come from the wrong place).
- Centered speech can seem to modulate the width of a music cue behind the speech. An early scene in *Young Sherlock Holmes* is one in which young Holmes and Watson cross a courtyard accompanied by music. On a poor decoder the width of the music seems to vary due to the speech.
- The matrix permits only 4 channels. With three used on the screen, that leaves only a monaural channel to produce surround sound, a contradiction. Some home decoders use means to decorrelate the mono-surround channel into two to overcome the tendency of mono surround to localize to the closer loudspeaker, or if seated perfectly centered between matched loudspeakers, in the center of your head.

Due to these problems, professionals in the film industry thought that a discrete multichannel system was desirable compared to the matrix system.

The 1987 SMPTE subcommittee direction towards a 5.1-channel discrete digital audio system led some years later to the introduction of multiple digital formats for sound both on and off release prints. Three systems remain after some initial shakeout in the industry: Dolby Digital, Digital Theater Systems (DTS), and Sony Dynamic Digital Sound (SDDS). Dolby Digital provided 5.1 channels in its original form, while SDDS has the capacity for 7.1 channels (adding two intermediate screen channels, left center, and right center). While DTS units could be delivered initially with up to 8 channels as an option, most of the installed base is 5.1. The relatively long gestation period for these systems was caused by, among other things, a fact of life: there was not enough space on either the release prints or on double-system CD-ROM followers to use conventional linear PCM coding. Each of the

three systems uses one method or another of low-bit-rate coding, that is, of reducing the number of bits that must be stored compared to linear PCM.

The 5.1 channels of 48kHz sampled data with 18-bit linear PCM coding (the recommendation of the SMPTE to accommodate the dynamic range necessary in theaters, determined from good theater background floor measurements to the maximum undistorted level desired, 105dB SPL per channel) requires:

$$\begin{array}{l} 5.005 \text{ channels} \times 48\text{k samples/s} \\ \text{for 1 channel} \times 18 \text{ bits/sample} \end{array} = 4,324,320 \text{ bits/s.}$$

(The number 5.005 is correct; 5.1 was a number chosen to represent the requirement more simply. Actually, 0.005 of a channel represents a low-frequency only channel with a sample rate of 1/200 of the principal sample rate.)

In comparison, the compact disc has an audio payload data rate of 1,411,200 bits/s. Of course, error coding, channel coding, and other overhead must be added to the audio data rate to determine the entire rate needed, but the overhead rate is probably a similar fraction for various media.

Contrast the 4.3 million bits per second needed to the data rate that can be achieved due to the space on the film. In Dolby Digital, a data block of 78 bits by 78 bits is recorded between each perforation along one side of the film. There are 4 perforations per frame and 24 frames per second yielding 96 perforations per second. Multiplying  $78 \times 78 \times 96$  gives a data capacity off film of 584,064 bits/s, only about 1/7 that needed just for the audio, not to mention error correcting overhead, synchronization bits, etc. While other means could be used to increase the data rate, such as shrinking the bit size, using other parts of the film, or use of a double system with the sound samples on another medium, Dolby Labs engineers chose to work with bits of this size and position for practical reasons. The actual payload data rate they used on film is 320,000 bits/s, 1/13.5 of the representation in linear PCM. Bit-rate reduction systems thus became required, in this case because the space on the film was so limited. Other media had similar requirements, and so do broadcast channels and point-to-point communications such as the Internet.

While there are many methods of bit-rate reduction, as a general matter, those having a smaller amount of reduction use waveform-based methods, while those using larger amounts of reduction employ psychoacoustics, called perceptual coding. Perceptual coding makes use of

the fact that louder sounds cover up softer ones, especially those close by in frequency, called frequency masking. Loud sounds not only affect softer sounds presented simultaneously, but also mask those sounds that precede or follow them through temporal masking. By dividing the audio into frequency bands, then processing each band separately, just coding each of the bands with the number of bits necessary to account for masking in frequency and time domains, the bit rate is reduced. At the other end of the chain, a symmetrical digital process reconstructs the audio in a manner that may be indistinguishable from the original, even though the bit rate has been reduced by a factor of more than 10.

The transparency of all lower than PCM bit-rate systems is subject to potential criticism, since they are by design losing data.<sup>7</sup> The only unassailable methods that reveal differences between each of them and an original recording, and among them, are complex listening tests based on knowledge of how the coders work. Experts must select program material to exercise the mechanisms likely to reveal sonic differences, among other things.

There have been a number of such tests, and the findings of them include:

- The small number of multichannel recordings available gave a small universe from which to find programs to exercise the potential problems of the coders—thus custom test recordings are necessary.
- Some of the tradeoffs involved in choosing the best coder for a given application include audio quality, complexity and resultant cost, and time delay through the coding process.
- Some of the coders on some of the items of program material chosen to be particularly difficult to code are audibly transparent.
- None of the coders tested is completely transparent all of the time, although the percentage of time in a given channel carrying program material that will show differences from the original is unknown. For instance, one of the most sensitive pieces of program material is a simple pitch pipe, because its relatively simple spectrum shows up quantizing noise arising from a lack of bits available to assign to frequencies “in between” the harmonics, but how much time in a channel is devoted to such a simple signal?
- In one of the most comprehensive tests, film program material from *Indiana Jones and the Last Crusade*, added after selection

---

<sup>7</sup>Of course no data is deliberately lost, but rather lossy coders rely on certain redundancies in the signal and/or human hearing perceptual mechanisms, to reduce the bit rate. Such coders are called “lossy” because they cannot reconstruct the original data.

of the other particularly difficult program material, showed differences from the original. This was important because it was not selected to be particularly difficult, and yet was not the least sensitive among the items selected. Thus experts involved in selection should use not only specially selected material, but also “average” material for the channel.

Bit-rate reduction coders enabled the development of digital sound for motion pictures, using both sound-on-film and sound-on-follower systems. During the period of development of these systems, another medium was to emerge that could make use of the same coding methods. High-definition television made the conceptual change from analog to digital when one system proponent startled the competition with an all-digital system. Although at the time this looked nearly impossible, a working system was running within short order, and the other proponents abandoned analog methods within a short period of time. With an all-digital system, at first two channels of audio were contemplated. When it was determined that multichannel coded audio could operate at a lower bit rate than two channels of conventional digital audio, the advantages of multichannel audio outweighed concerns for audio quality if listening tests proved adequate transparency for the proposed coders. Several rounds of listening tests resulted in first selection, and then extensive testing, of Dolby AC-3 as the coding method. With the selection of AC-3 for what came to be known as Digital Television, Dolby Labs had a head start in packaged media, first Laser Disc, then DVD-Video. It was later dubbed simply Dolby Digital in the marketplace, rather than relying on the development number of the process, AC-3. DTS also provided alternate coding methods on Laser Disc, DVD-Video, and DTS CD for music.

With the large scale success of DVD, introduced in 1997, and with music CDs flagging in sales due to, among other things, lack of technical innovation, two audio-mostly packaged media were introduced: DVD-A in 2001 and Super Audio Compact Disc (SACD) in 2003. Employing different coding methods, insofar as surround sound operates the two are pretty similar, delivering up to six wide range channels, usually employed as carriers for either 2.0- or 5.1-channel audio. Unfortunately, the combined sales of the two formats only added up to about the sales of LP records in 2005, some 1/700th of the number of CDs. There are probably several reasons for this: one is the shift among early adopters away from packaged media to downloadable media to computers and portable music devices. While this shift does not preclude surround sound distribution to be rendered down to two channel headphones as needed, nonetheless the simplicity of two channel infrastructure has won out, for now. Perhaps this parallels

the development of surround for movies, where two-channel infrastructure won for years until a discrete 5.1-channel system became possible to deliver with little additional cost.

Over the period from 2000 to 2006 U.S. sales show a decline in CD unit sales from 942.5 million to 614.9 million. The corresponding dollar value is from \$13,214.5 million to \$9,162.9 million, not counting the effect of inflation. Some of this loss is replaced by digital downloads of singles, totaling 586 million units in 2006, and albums, totaling 27.6 million units. The total dollar volume though of downloads of \$878.0 million, added to the physical media sales and mobile and subscription sales, still only produces a total industry volume at retail of \$11,510.2 million, an industry shrinkage of 45.7% in inflation adjusted dollars over the period from 1999 to 2006. Accurate figures on piracy due to at first download sites such as Napster and later peer-to-peer networks are difficult to come by, but needless to say contribute to the decline of sales. One survey by market research firm NPD for the RIAA showed 1.3 billion illegal downloads by college students in 2006, and the RIAA has filed numerous lawsuits against illegal downloaders. The survey also claims that college students represent 26% of the total of P2P downloads, so the total may be on the order of 5.2 billion illegal downloads. College students self-reported to NPD that two-thirds of the music they acquired was obtained illegally.

To put the 2006 size of the music business in perspective, the U.S. theatrical box office for movies was \$9,490 million in that year, and DVD sales were \$15,650 million. Rental video added another \$7,950 million, bringing the total annual spent on movies and recorded television programs<sup>8</sup> in the USA to \$33,090 million. So adding the various retail values together the motion picture business, much of which is based on 5.1-channel surround sound, is 2.87 times as large as the music industry in the USA in dollar volume.<sup>9</sup> Illegal downloads of movies also affect the bottom line, but less so than that of the music industry simply because the size of the files for movies is much larger and requires longer download times. The Motion Picture Association of America (MPAA) estimates that U.S. losses in 2005 were \$1.3 billion, divided between Internet downloads and DVD copies.

A second reason for the likely temporary failure of surround music—mostly media is that the producers ignored a lesson from consumer

---

<sup>8</sup>Also counts DVD Music Video discs with sales of about \$20 million.

<sup>9</sup>The source for these statistics is the 2006 Year-End Shipment Statistics of the Recording Industry Association of America, the Motion Picture Association of America, and an article on January 5, 2007 by Jennifer Netherby on [www.videobusiness.com](http://www.videobusiness.com).

research. The Consumer Electronics Association conducts telephone polling frequently. A Gallup poll for CEA found that about 2/3 of listeners prefer a perspective from the best seat in the house, compared to about 1/3 that want to be “in the band.” This is a crucial difference: while musicians, producers, and engineers want to experience new things, and there are very real reasons to want to do so that are hopefully explained later in this book, nonetheless much of the audience is not quite ready for what we want to do experimentally, and must be brought along by stages. What music needs is the effect that the opening of *Star Wars* had on audiences in 1977—many remember the surround effect of the space ship arriving coming in over your head, with the picture and sound working in concert, as their first introduction to surround sound.

Further consumer research in the form of focus groups listening to surround music and responding was conducted by the Consumer Electronics Association. With a person-on-the-street audience, surround music was found to range from satisfying through annoying, especially when discrete instruments were placed in the surround channels. Harkening back to the introduction of *Fantasia* when some people left the theater to complain to management that there was a chorus singing from behind them, insecurity on the part of portions of the audience drive them to dislike the experience.

A third reason, and probably the one most cited, is that two competing formats, like the Betamax versus VHS war, could not be sustained in the marketplace.

Meanwhile the Audio Engineering Society Technical Committee on Multichannel and Binaural Audio Technology wrote an information guideline called *Multichannel Surround Sound Systems and Operations*, and published it in 2001. It is available as a free download at <http://www.aes.org/technical/documents/AESTD1001.pdf>. The document was written by a subcommittee of the AES committee.<sup>10</sup> Contributions and comments were also made by members of the full committee and other international organizations.

HD-DVD and Blu-ray high-definition video discs are now on the market. These second generation packaged media optical discs offer higher picture quality, so enhancements to sound capability was also on the agenda of the developers. Chapter 5 covers these media.

With 5.1 channels of discrete digital low-bit-rate coded audio standard for film, and occasional use made of 7.1 discrete channels including

---

<sup>10</sup>The Writing Group was: Francis Rumsey (Chair), David Griesinger, Tomlinson Holman, Mick Sawaguchi, Gerhard Steinke, Günther Theile, and Toshio Wakatuki.

left-center and right-center screen channels, almost inexorably there became a drive towards more channels, due to the sensation that only multidirectional sound provides. In 1999, for the release of the next installment in the Star Wars series, a system called Dolby Surround EX was introduced. Applying a specialized new version of the Dolby Surround matrix to the two surround channels resulted in the separation of the surround system into left, back, and right surround channels from the two that had come before. Later, since a 4:2:4 matrix was used but 1 channel was not employed in Dolby Surround EX, one film, *We Were Soldiers*, was made using 4 channels of surrounds: left, right, back, and overhead. DTS also developed a competitive system with 6.1 discrete channels, called DTS ES.

With 5.1-channel sound firmly established technically by the early 1990s and with its promulgation into various media likely to require some years, the author and his colleagues began to study what 5.1-channel audio did well, and in what areas its capabilities could be improved. The result is a system called 10.2. All sound systems have frequency range and dynamic range capabilities. Multichannel systems offer an extension of these capabilities to the spatial domain. Spatial capabilities are in essence of two kinds: imaging and envelopment. These two properties lie along a dimension from pinpoint imaging on the one hand, to completely diffuse and thus enveloping on the other. For the largest seating area of a sound system, imaging is improved as the number of channels goes up. In fact, if you want everyone in an auditorium to perceive sound from a particular position the best way to do that is to put a loudspeaker with appropriate coverage there. This however is probably inconvenient for all sound sources, so the use of phantom images, lying between loudspeakers, becomes necessary. Phantom images are more fragile, and vary more as one moves around a listening space, but they are nonetheless the basis for calling stereophonic sound "solid," which is the very meaning of stereo.

Digital Cinema is now a reality, with increasing number of screens so equipped in the U.S. Audio for it consists of up to 16 linear PCM coded audio tracks, with a sample rate of 48kHz and a 24-bit word length. Provision is made in the standards to increase the sample rate to 96kHz, should that ever become required.

Other systems entered the experimental arena too. NHK fielded an experimental system with 22.2 channels, while Europe's Fraunhofer Institute introduced its Iosono wavefield synthesis system harkening back to Bell Labs infinite line of microphones feeding an infinite line of loudspeakers, with 200 loudspeakers representing infinity in one cinema installation.

All in all, upwards pressure on the number of audio channels is expected to continue into the future. This is because the other two competitors<sup>11</sup> for bits, sample rate representing frequency range, and word length also known as resolution and bit depth representing dynamic range, both are what engineers call saturating functions. With a saturating function, when one has enough there is simply no point in adding to the feature. For instance, if 24-bit audio were actually practical to be realized, setting the noise floor at the threshold of hearing would produce peaks of 140 dB SPL, and the entire hearing audience would revolt. I tried playing a movie with relatively benign maximum levels, *Sea Biscuit*, to a full audience at its original Hollywood dubbing stage level in the midwest, and within 10 minutes more than five people had left the full theater to complain to the management about how loud it was, and this was at a level that could not exceed 105 dB SPL/channel, and almost certainly never reached that level. So there can be said to be no point in increasing the bit depth to 24 from, say 20, as the audience would leave if the capability were to be used.

The number of loudspeaker channels on the other hand is not a saturating function, but rather what engineers call an asymptotic function: the quality rises as the number of channels increases, but with declining value as the number of channels goes up. Many people have experienced this, including the whole marketplace. The conversion from mono to stereo happened because all of the audience could hear the benefit; today stereo to 5.1 is thoroughly established as an improvement. Coming at the question from another point of view, English mathematician Michael Gerzon has estimated that it would take 1 million channels to transmit one space at each point within it into another space, while James Johnston then at AT&T Labs estimated that 10,000 channels would do to get the complex sound field reproduced thoroughly in the vicinity of one listener's head. With actual loudspeakers and interstitial phantom images, Jens Blauert in his seminal book *Spatial Hearing* has said that about 30 channels would do placed on the surface of 1/2 of a sphere in the plane above and around a listener for listening in such a half-space.

The sequence 1, 2, 5, 10 is familiar to anyone who has ever spent time with an oscilloscope: the voltage sensitivity and sweep time/div knobs are stepped in these quantized units. They are approximately equal steps along a logarithmic scale. G.T. Fechner showed in 1860 that just noticeable differences in perception were smaller at lower levels and

---

<sup>11</sup>This is not strictly true because various low-bit-rate coding schemes compared to linear PCM ameliorate the need for bits enough that they make practical putting sound on film optically for instance; but in the long run, it is really bandwidth and dynamic range that are competitors for space on the medium.

larger at bigger ones. He concluded that perception was logarithmic in nature for perceptual processes. Applying Fechner's law today, one would have to say that the sequence of channels 1, 2, 5 10, 20 ... could well represent the history of the number of channels when viewed in the distant future.

A summary of milestones in the history of multichannel surround sound is as follows:

- Invention in 1938 with one subsequent movie release *Fantasia* (3 channels on optical film were steered in reproduction to front and surround loudspeakers).
- Cinerama 7 channel, Cinemascope 4 channel, and Todd AO 6 channel introduced in short order in the 1950s, but declining as theater admissions slipped through the 1960s.
- Introduction of amplitude-phase matrix technology in the late 1960s, with many subsequent improvements.
- Revitalization in 1975–1977 by 4-channel matrixed optical sound on film.
- Introduction of stereo surround for 70 mm prints in 1979.
- Introduction of stereo media for video, capable of carrying Lt/Rt matrixed audio, and continuing improvements to these media, from the early 1980s.
- Specification for digital sound on film codified Five point one channel sound in 1987.
- Standardization on 5.1 channels for Digital Television in the early 1990s.
- Introduction of three competing digital sound for film systems in the early 1990s, see Fig. 1-2 for their representation on film.
- Introduction of 5.1-channel digital audio for packaged media, in the late 1990s.
- Introduction of matrixed 3-channel surround separating side from back surround using two of the discrete channels of the 5.1-channel system with an amplitude-phase matrix in 1999.
- Introduction of 10.2-channel sound in 1999.
- Consumer camcorder recording 5.1-channel audio from built-in microphone, Christmas 2005.
- Blu-ray and HD-DVD formats including multichannel linear PCM among other standards, 2006.
- Digital Cinema, 2006.

Fig. 1-1, a timeline, gives additional information.

In the history of the development of multichannel, these milestones help us predict the future; there is continuing pressure to add channels because more channels are easily perceived by listeners. Anyone

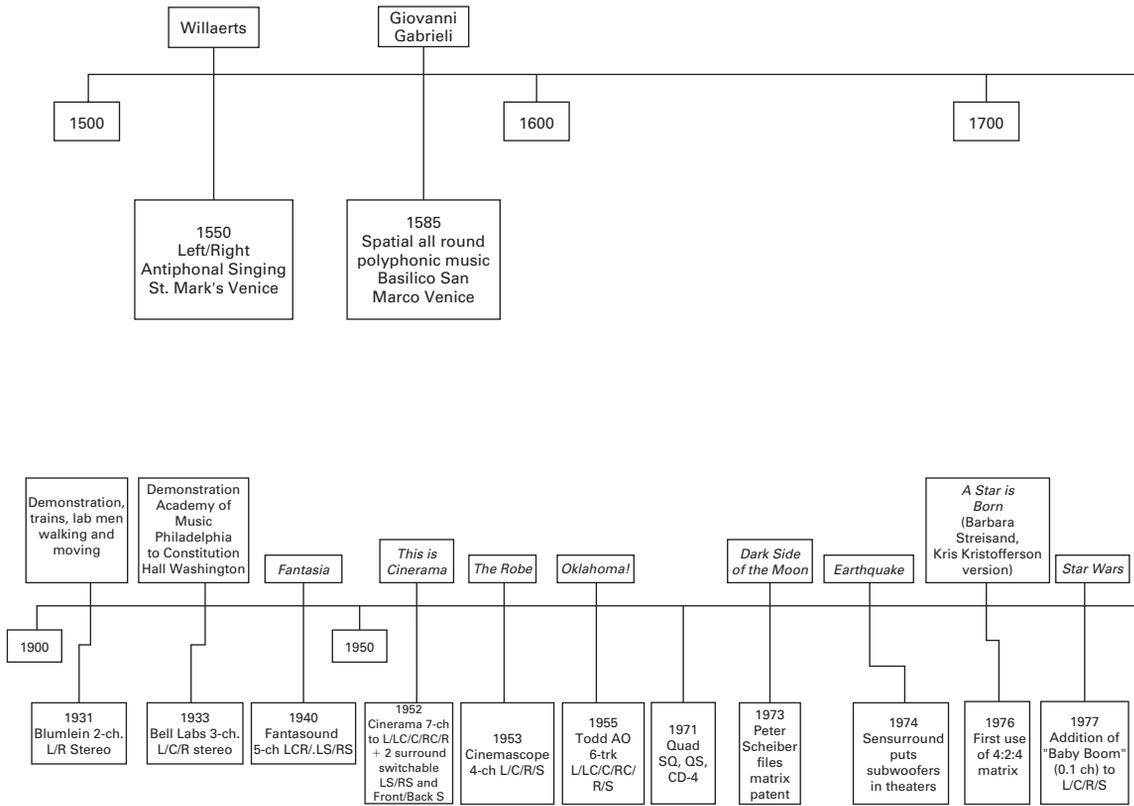
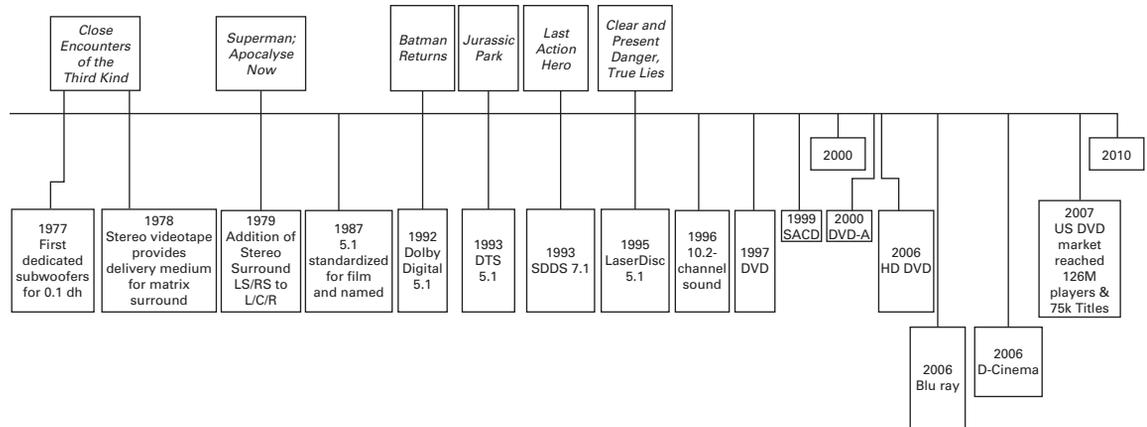
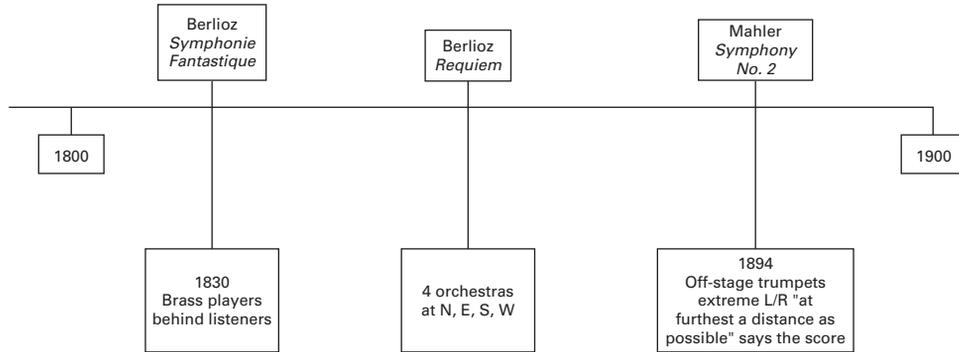
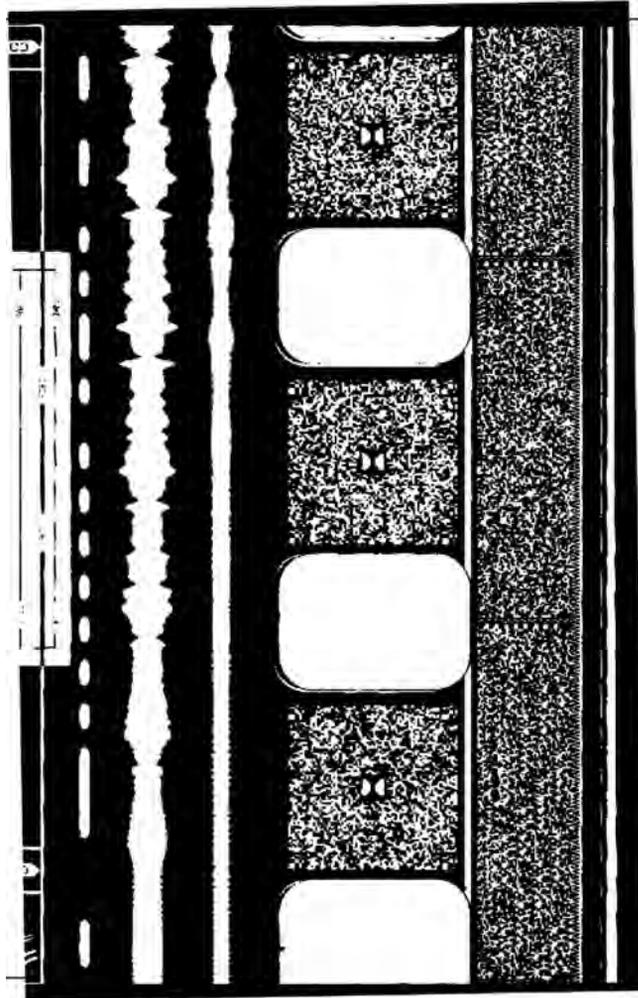


Fig. 1-1 Timeline of surround developments.





**Fig. 1-2** The analog sound track edge of a 35mm motion picture release print. From left to right is: (1) the edge of the picture area, that occupies approximately four perforations in height; (2) the DTS format time code track; (3) the analog Lt/Rt sound track; (4) the Dolby Digital sound on film track between the perforations; and (5) one-half of the Sony SDDS digital sound on film sound track; the corresponding other half is recorded outside the perforations on the opposite edge of the film.

with normal hearing can hear the difference between mono and stereo, so too can virtually all listeners hear the difference between 2- and 5.1-channel stereo. This process may be seen as one that does have a logical end, when listeners can no longer perceive the difference, but the bounds on that question remain open. Chapter 6 examines the psychoacoustics of multichannel sound, and shows that the pressure upwards in the number of channels will continue into the foreseeable future.