TECHNICAL DOCUMENT

AES TECHNICAL COUNCIL
Document AESTD1001.1.01-10

Multichannel surround sound systems and operations

This document was written by a subcommittee (writing group) of the AES Technical Committee on Multichannel and Binaural Audio Technology. Contributions and comments were also made by members of the full committee and other international organizations. Writing Group: Francis Rumsey (Chair); David Griesinger; Tomlinson Holman; Mick Sawaguchi; Gerhard Steinke; Günther Theile; Toshio Wakatuki.

AUDIO ENGINEERING SOCIETY, INC.
INTERNATIONAL HEADQUARTERS
60 East 42nd Street, Room 2520 • New York, NY 10165-2520, USA
Tel: +1 212 661 8528 • Fax: +1 212 682 0477
E-mail: HQ@aes.org • Internet: http://www.aes.org

AUDIO ENGINEERING SOCIETY, INC.
INTERNATIONAL HEADQUARTERS
60 East 42nd Street, Room 2520, New York, NY 10165-2520, USA
Tel: +1 212 661 8528 ▪ Fax: +1 212 682 0477
E-mail: HQ@aes.org ▪ Internet: http://www.aes.org

Europe Services
B.P. 50, FR-94364 Bry Sur Marne Cedex, France
Tel: +33 1 4881 4632, Fax: +33 1 4706 0648
E-mail for membership and publication sales: euroservices@aes.org

United Kingdom
British Section, Audio Engineering Society Ltd.
P. O. Box 645, Slough, SL1 8BJ UK
Tel: +44 1628 663725, Fax: +44 1628 667002
E-mail: UK@aes.org

Japan
AES Japan Section
1-38-2 Yoyogi, Room 703, Shibuyaku-ku,
Tokyo 151-0053, Japan
Tel: +81 3 5358 7320, Fax: +81 3 5358 7328,
E-mail: japan@aes.org

AUDIO ENGINEERING SOCIETY, INC.
INTERNATIONAL HEADQUARTERS
60 East 42nd Street, Room 2520, New York, NY 10165-2520, USA
Tel: +1 212 661 8528 ▪ Fax: +1 212 682 0477
E-mail: HQ@aes.org ▪ Internet: http://www.aes.org

The Audio Engineering Society's Technical Council and its Technical Committees respond to the interests of the membership by providing technical information at an appropriate level via conferences, conventions, workshops, and publications. They work on developing tutorial information of practical use to the members and concentrate on tracking and reporting the very latest advances in technology and applications. This activity is under the direction of the AES Technical Council and its Committees.

The Technical Council and its first Technical Committees were founded by the Audio Engineering Society in 1979, and standing rules covering their activities were established in 1986, with the intention of defining and consolidating the technical leadership of the Society for the benefit of the membership. The Technical Council consists of the officers of the Technical Council, the chairs of the Technical Committees, the editor of the Journal, and as ex-officio members without vote, the other officers of the Society.

TECHNICAL COUNCIL
Wieslaw Woszczyk - Chair (WVW@AES.ORG)
Juergen Herre - Vice Chair (HRR@AES.ORG)
Robert Schulein - Vice Chair (RBS@AES.ORG)

TECHNICAL COMMITTEES
Acoustics & Sound Reinforcement
Archiving, Restoration and Digital Libraries
Audio Coding
Automotive Audio
High-Resolution Audio
Loudspeakers & Headphones
Microphones & Applications
Multichannel & Binaural Audio Technologies
Network Audio Systems
Optical Recording
Perception & Subjective Evaluation of Audio Signals
Signal Processing
Studio Practices & Production
Transmission & Broadcasting

AES TC TECHNICAL PUBLICATIONS
AES White Paper 1001: Networking Audio and Music Using Internet2
and Next-Generation Internet Capabilities

This and other Technical Council publications may be downloaded from its home page at:
WWW.AES.ORG/Technical

The Audio Engineering Society, now in its fifth decade, is the only professional society devoted exclusively to audio technology. Its membership of leading engineers, scientists, and other authorities has increased dramatically throughout the world, greatly boosting the Society's stature and that of its members in a truly symbiotic relationship.

Purpose: The Audio Engineering Society is organized for the purpose of uniting persons performing professional services in the audio engineering field and its allied arts; collecting, collating, and disseminating scientific knowledge in the field of audio engineering and its allied arts; advancing such science in both theoretical and practical applications; and preparing, publishing, and distributing literature and periodicals relative to the foregoing purposes and policies.

Membership: Individuals who are interested in audio engineering may become members of the AES and Sustaining Memberships are available to persons, corporations, or organizations who wish to support the Society. A subscription to the *Journal* is included with all memberships.

Publications and Events: The AES publishes a *Journal*, Convention Papers, Conference Proceedings, Standards, Information Documents, and Technical Documents, and organizes two International Conventions per year and numerous International Conferences on late-breaking audio topics.

For membership applications, dues information, publications sales, and details on conventions and conferences, contact any AES office or go to the AES website at www.aes.org

Multichannel surround sound systems and operations

Multichannel surround sound systems and operations

1 BACKGROUND

This document, issued by the AES Technical Committee on Multichannel and Binaural Audio Technology, is intended to report developing practices concerning the configuration and use of multichannel surround sound systems based on the 3/2- or 5.1-channel model described in ITU-R BS.775-1 [1]. It is intended to acquaint studios with these developments, but it should also be useful, to a large extent, for consumer equipment. It is not an AES standard or information document and has not been subject to AES due process to determine a consensus on its content. However, it aims to summarize some of the most important features of existing international standards relating to multichannel sound recording and reproduction, as well as to report good practice based on contributions from expert members of the AES and other international groups.

While there will always be debate over what constitutes good practice in recording and reproduction, the Technical Council believes that one of its important functions is to convey the current views of its members and to educate the audio industry at large. Consequently this document represents the committee's best attempt at describing the current state of the art, and may be open to revision, resulting in further versions as new knowledge becomes available. In some cases, where international standards are clear about the way in which systems should be set up or used, and where little disagreement exists, this information has been related directly. (The document is intentionally biased toward an acceptance of existing AES, ITU, EBU, and SMPTE standards where they exist.) Where there is more uncertainty, or where standards lag evolving industry practice, the differing approaches have been described so that the reader is aware of alternative points of view.

The committee welcomes additional input, corrections, and proposed content for this document. Correspondence details are provided at the end of the document.

2 INTRODUCTION TO 3/2- OR 5.1-CHANNEL STEREO

Although multichannel stereophony is not limited to a specific number of channels in principle, international agreement was reached some years ago on a configuration that represented a compromise between the need for optimum spatial enhancement of reproduction and the need for an approach that was practicable and compatible with conventional two-channel reproduction. The solution has become known colloquially as "5.1-channel" reproduction owing to its use of five full-bandwidth channels plus an optional, limited-bandwidth, low-frequency-extension (LFE) channel (the "0.1" channel). It has its origins in configurations designed for film sound reproduction where a center "dialogue" channel is considered of prime importance. In order to maintain compatibility between the reproduction of film sound in the cinema or home and other types of surround sound program material, the same configuration was adopted for all applications.

The standard configuration is also referred to as "3/2 stereo," in recognition of the position of this configuration in a hierarchy of multichannel stereo systems ranging from mono to many channels. In this hierarchy, a distinction is made between the number of front channels and the number of rear/side "effects," "surround," "room impression" or "ambience" channels. (The designation 3/2 therefore refers to the use of three front channels—left, center, right—and two rear/side channels—left surround and right surround. This is described further in the following, noting that the 3/2 format can also be extended to accommodate an LFE channel.)

3 HIERARCHY OF COMPATIBLE MULTICHANNEL SOUND SYSTEMS FOR BROADCASTING AND RECORDING

The 3/2 system is embedded within a hierarchy of multichannel sound formats. For such a hierarchy, down compatible as far as the monophonic format, simple matrixing conditions are given in [1] for the addition of partial signals at the transmission and storage or reproduction stages of a signal chain, facilitating basic intercompatibility between channel formats. (It is noted that the compatibility matrixing approaches recommended in [1] are relatively crude, involving the simple folding down of the rear channels and the center channel into the front channels with a given level of attenuation. Alternative approaches to the downmixing of multichannel stereo to two-channel stereo may be more subjectively satisfactory.)

3/1-matrix formats (three frontal signals, one surround signal) are integrated in the hierarchy and may be reproduced with the 3/2 configuration, in which case the monophonic surround signal feeds the two surround loudspeakers and the gain of the surround channels is reduced by 3 dB. Japan accepted the 3/1 format within the ITU standard as an exception because it is used in Japan with the MUSE transmission system.

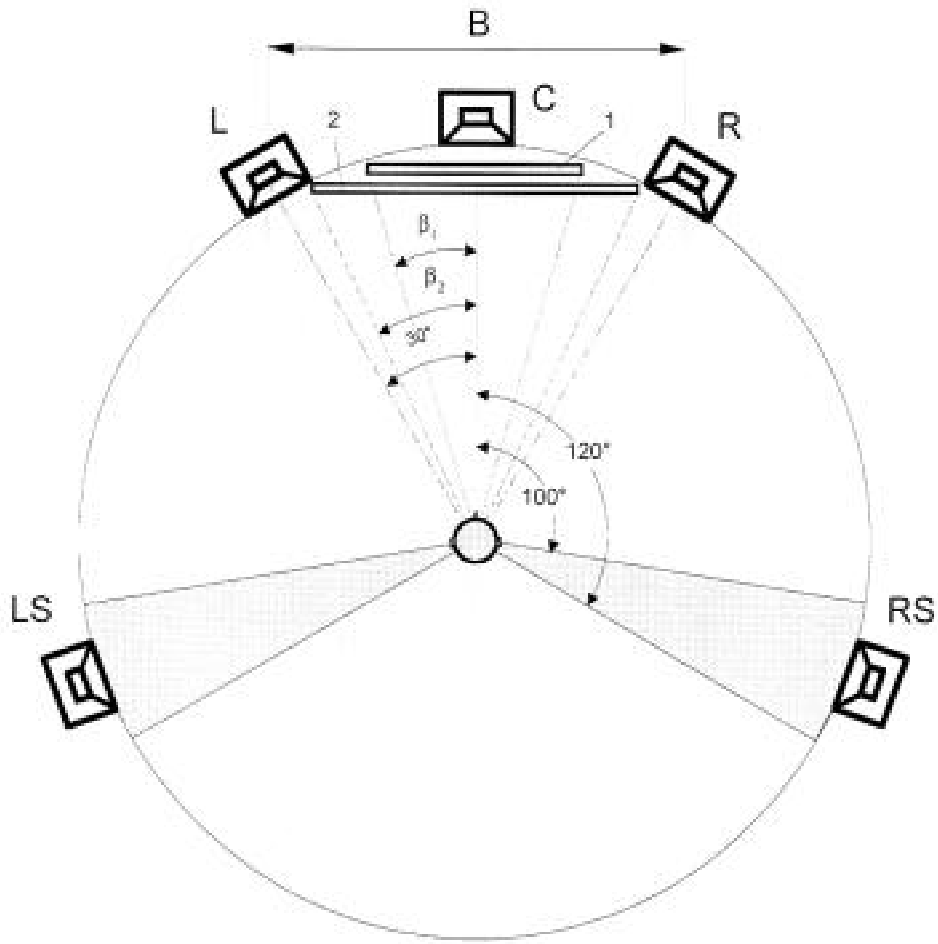
Systems with more channels than the 3/2 format are possible and can be matched—such as, 5/2, 5/4, and so on. These formats are not included within the ITU standard and are not recommended for material intended to be reproduced under home conditions. The format with five frontal loudspeakers is used in the film domain under certain circumstances (and is an option for the DVD), but it should be produced in such a way that it is also down compatible with the 3/2 and 2/0 formats.

For all the other possible format combinations the reference configuration has to be the basis (see later). Further loudspeakers can be attached to the reference configuration, but with the same number of source signals, to increase the enveloping effect and/or to render discrete sound sources in more positions. This should be handled in such a manner that up and down compatibility can be ensured.

4 REFERENCE CONFIGURATION

The reference arrangement (basic reproduction configuration) has the 3/2 format with three front signals or channels (L = left, C = center, R = right) plus two so-called surround channels—room and ambience channels (LS and RS = left and right surround). This principal reproduction standard is totally independent of the applied transmission system and recording processes, and should not be confused with different coding formats (e.g., ISO/MPEG or Dolby Digital).

For setting up the five loudspeakers, Figure 1 shows an arrangement based on the recommendations in ITU-R BS.775-1 [1] and SMPTE [2] If a loudspeaker setup on the circumference of a circle is not possible, these recommendations imply that the loudspeakers inside this circumference should be delayed accordingly.



Screen 1: Listening distance = 3*H* (2ß$_1$ = 33°)
Screen 2: Listening distance = 2*H* (2ß$_2$ = 48°)
H: Screen height
B: Loudspeaker basis width

Figure 1. Reference loudspeaker setup with loudspeakers L/C/R and LS/RS, in combination with picture reproduction installation (in accordance with ITU-R BS. 775-1)

Acoustical Center	***Angle***	***Height***	***Tilt***
C	0°	1.2m*	0°*
L, R	±30°	1.2m	0°
LS, RS	±100–120°	≥1.2m	≤ 15°

* Depending on shape, type, and size of screen.

To create a larger listening zone and/or improved envelopment by means of the room ambience information reproduced with the 3/2 format, one may add more surround loudspeakers to the two standard channels LS and RS. For larger reproduction rooms (such as cinemas) this is necessary and usually done anyway. In this case a sufficient decorrelation of the added loudspeaker channels is desirable, for example, by appropriate delay, and connected via suitable signal distributors (matrixes) or processors.

5 LOW-FREQUENCY EXTENSION

In order to avoid confusion a clear distinction is made here between a low-frequency-extension (LFE) signal, that can be carried over a separate LFE channel in a transmission or recording system, and the separate radiation of low-frequency program content through so-called subwoofers. Although these may seem to be one and the same, they need not be. Indeed, it is this confusion about low-frequency management that causes a large number of problems in practical situations.

5.1 LFE signal and channel

In the film domain the use of a special channel was introduced in the bass range from 20 to about 80–120 Hz as a low-frequency extension, which—according to ITU-R BS.775-1—can be used optionally as a supplement to the formats in the studio or home. The designation is abbreviated as "0.1" or "../1" because of the small frequency range used. Therefore the designations 3/2/1 or 5.1 and 5/2/1 or 7.1 are common.

According to [1], optionally, one additional channel for the enhanced bass region is permitted, with a frequency range of 20–80 Hz (up to 120 Hz maximum), which is the norm in cinemas with motion pictures. In consumer audio systems, the LFE channel is also considered optional in reproduction. Media should be prepared that conform to this recommendation so that they sound satisfactory even if the LFE channel is not reproduced.

EBU and SMPTE documents on multichannel sound [3], [4] contain some remarks on the use of the LFE channel. This is from the SMPTE document [3]:

> When an audio programme originally produced as a feature film for theatrical release is transferred to consumer media, the LFE channel is often derived from the dedicated theatrical subwoofer channel. In the cinema, the dedicated subwoofer channel is always reproduced, and thus film mixes may use the subwoofer channel to convey important low frequency programme content. When transferring programmes originally produced for the cinema over television media [e.g. DVD], it may be necessary to re-mix some of the content of the subwoofer channel into the main full bandwidth channels. It is important that any low frequency audio which is very significant to the integrity of the programme content is not placed into the LFE channel. The LFE channel should be reserved for extreme low frequency, and for very high level <120 Hz programme content which, if not reproduced, will not compromise the artistic integrity of the programme.

With cinema reproduction the in-band gain of this channel is usually 10 dB higher than that of the other individual channels. According to SMPTE [3] this will be compensated by a level increase of the reproduction channel, not by an increased recording level. This has to be observed in the studio domain and also with home reproduction, for reasons of compatibility. (It does not mean that the broad-band or weighted sound pressure level of the LFE loudspeaker should measure 10 dB higher than that of any of the other channels when aligned using broad-band pink noise—in fact it will be considerably less than this as its bandwidth is narrower.)

5.2 Separate low-frequency loudspeakers (subwoofers) within the standard configuration

It may be useful, in addition to the main loudspeakers (L/C/R/LS/RS), to use separate bass radiators (subwoofers) for the extension of the lower frequency range, so that the lower limit frequency of the five main loudspeakers can be raised to about 80 Hz and their volumes consequently reduced.

In this case it is possible to use several subwoofers for specific individual channels (for example, frontal and/or surround channels), or one single subwoofer to supplement the low-frequency range of all of the five main loudspeakers. All the loudspeakers are connected via crossover circuits. (Limit frequencies of 80–160 Hz are common in the consumer industry; more efficient is in the vicinity of 80 Hz.)

The configuration has to be regarded as a 3/2 format, but it is possible that separate bass equipment can be configured such that both 5.1-channel motion pictures and 3/2-format material without a separate LFE channel can be handled, according to Figure 2.

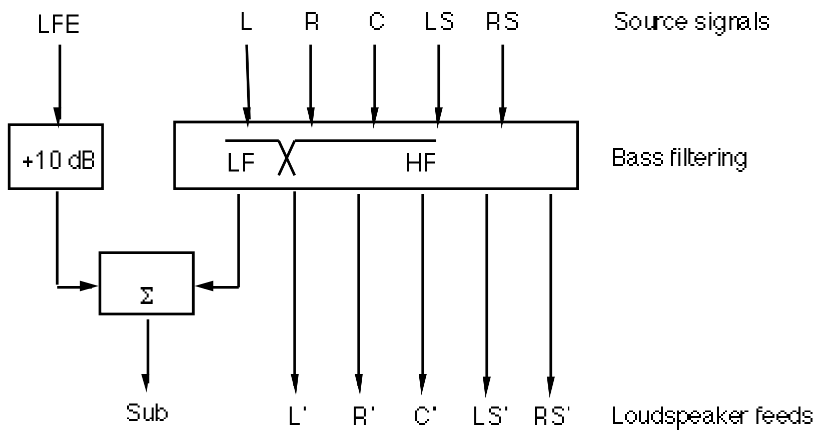


Figure 2. Derivation of combined subwoofer and LFE signals.

The resulting quality, including the operational sound level response, is also dependent on the position of the loudspeakers in relation to the listening position as well as on the nonlinear distortions of the subwoofers, by which localization errors can occur (see more details in [4]).

There appears to be little agreement about the optimum location for a single subwoofer in a listening room, although measurements have been published suggesting that a corner location for a single subwoofer provides the most extended, smoothest low-frequency response [5]. In choosing the optimum locations for subwoofers it is noted that loudspeakers placed in corners tend to give rise to a noticeable bass boost, and couple well to most room modes (because they have antinodes in the corners). Some subwoofers are designed specifically for placement in particular locations whereas others need to be moved around until the most subjectively satisfactory result is obtained. Some artificial equalization may be used to obtain a reasonably flat overall frequency response at the listening position. Phase shifts or time-delay controls are sometimes provided to enable some correction of the time relationship of the subwoofer to the other loudspeakers, but this will necessarily be a compromise with a single unit. A subwoofer phase shift is sometimes used to optimize the sum of the subwoofer and the main loudspeakers in the crossover region for a flat response.

Although substantial measured differences have been found between subwoofer positions, in terms of frequency response, it may be difficult to detect the differences subjectively when listening to a range of multichannel program material with subwoofers in different positions [6]. Positioning such loudspeakers in front of the wall at which the frontal loudspeakers are installed has been found to be useful. In comparison to the use of a single subwoofer in different positions with stereo subwoofers placed under the main two-channel loudspeakers, it was found that the detectability of a difference varied with the program material, location, and crossover frequency. It was most noticeable once the crossover frequency rose much above 120 Hz [7]. Informal tests in reference listening rooms have shown that the position of separate subwoofers can often be detected. It may therefore be suggested that separate subwoofers should be located very near the corresponding frontal or surround loudspeakers. The reasons for the separate detectability of a subwoofer location can be various. Some have shown that port noise, distortion, and information above 120 Hz radiating from the subwoofer position can make it localizable, whereas otherwise it would not be. A centrally located subwoofer is likely to suffer from being at the null of lateral standing-wave modes. An offset might therefore be considered acoustically desirable.

There is some evidence to suggest that multiple low-frequency drivers generating decorrelated signals from the original recording create a more natural spatial reproduction than monaural low-frequency reproduction from a single driver [8]. According to this proposal, if monaural low-frequency content is reproduced it is better done through two units placed to the sides of the listener, driven 90° out of phase, to excite the asymmetrical lateral modes more successfully and improve low-frequency spaciousness.

5.3 Considerations regarding the channel allocation of low-frequency program content

The "0.1 channel" sometimes creates confusion for users of the standards described in the preceding when mixing sound that is not related to cinema applications. In such cases the assumption that it is necessary to generate a separate LFE signal in order to "conform to the standard" may be a distraction. It should be stressed that the generation of a separate LFE signal is entirely optional, and that in many music applications its use may even work against the requirement to generate an optimum degree of low-frequency envelopment.

Recent research suggests that optimum envelopment at low frequencies is achieved through the use of adequately decorrelated loudspeaker signals. Such decorrelation could be generated artificially in the consumer replay chain, as part of a consumer system's low-frequency management, but this removes control from the recording engineer. If the recording engineer's intentions in this respect are to stand a chance of being conveyed to the listener, it follows that low-frequency content intended to create stereophonic envelopment should not be allocated to a monophonic LFE channel but should be retained within the full bandwidth channels. As the standards note, any low-frequency content that is crucial to the success of the mix should be routed to the main channels rather than the LFE. The LFE signal is only really suitable for optional "effects," and it should not matter whether or not the consumer is able to replay this channel.

Some multichannel audio encoders sample the LFE channel at a low sampling rate such as 240 Hz, thereby low-pass filtering any content routed through such a channel to an upper limit of <120 Hz. This emphasizes the importance of checking any mixes to be encoded using such systems by monitoring via any encode–decode chain that is envisaged.

6 MONITORING ENVIRONMENTS

6.1 Listening conditions—general notes

Reference listening rooms are designed for the critical comparison of program material, where the facilitation of interchangeable judgments between sites is a primary aim. It is recognized that in many practical sound-monitoring environments it will be difficult to approach these ideal conditions, especially where there are large items of equipment in the room. Nonetheless the information is provided as a guideline for good practice. The approaches found in the literature and given here relate primarily to small- and medium-sized rooms. The conditions and criteria for large film sound mixing rooms may differ considerably from these approaches in some respects.

The overall listening conditions and the achievable quality of the sound field associated with them are determined by:

- The geometric and acoustical properties of the listening room
- The properties and arrangement of the loudspeakers in the listening room
- The listening position or the listening zone.

These suggested listening conditions for a high-quality listening room are intended to allow neutral and critical monitoring of the sound signal such that the characteristics and deficiencies can be clearly recognized, and the listening events can be unimpaired. Furthermore, the reproduction of a high-quality sound signal can give a technically and aesthetically satisfactory impression. The approaches described in Section 6.2 are only minimum suggestions to ensure that a high quality of program exchange can be achieved, and are based on international standards for reference listening conditions. They are not yet adequate to describe optimal arrangements or to guarantee an adequate conformity between different listening rooms.

6.2 Parameters and values for reference listening conditions

6.2.1 Suggestions for reference listening room (Table 1)

Table 1. Suggestions for reference listening room.

Parameter	*Units/Conditions*	*Value*
Room size (floor surface area) Mono/2-channel stereo Multichannel	S [m^2]	 >30 >40
Room proportions	l = length w = width h = height	$1.1\, w/h \le l/h \le 4.5\, w/h - 4$, with $l/h < 3$ and $w/h < 3$ (Ratios within ±5% of integer values are considered unsatisfactory.)
Base width 2-channel stereo Multichannel	B [m]	 2.0–4.0 2.0–4.0
Basis angle 2-channel stereo Multichannel	[deg] referred to L/R	 60 60
Listening distance 2-channel stereo Multichannel	D [m]	 2 m–1.7 B
Listening zone (radius) 2-channel stereo Multichannel	R [m]	 0.8 0.8
Loudspeaker height (from acoustic center) 2-channel stereo Multichannel (all)	h [m]	 ≈1.2 ≈1.2
Distance to surrounding reflecting surfaces 2-channel stereo Multichannel	d [m]	 ≥1 ≥1

The literature suggests that a volume of 300 m^3 should not be exceeded for studio listening rooms. In order to obtain a suitable distribution of room modes, dimensions according to the values in column 3 are suggested. The room shape suggested is largely symmetrical around the listening direction, and with regard to the distribution of the absorption material, especially around the loudspeakers, doors, windows, and technical equipment, so that any acoustical discontinuities can be avoided. Also the surface of any mixing desk can cause disturbing reflections.

6.2.2 Suggestions for the reference sound field at listening position (Table 2)

Table 2. Suggestions for reference sound field at reference listening position.

Parameter	*Units/Conditions*	*Value*
Direct sound Amplitude/frequency response	 Free-field propagation measurements	 For tolerance borders see Table 3 (reference monitor)
Reflected sound Early reflections	 0–15 ms (in region 1–8 kHz)	 < –10 dB relative to direct sound
Temporary diffusion of reverberant sound field	Avoidance of significant anomalies in sound field	No flutter echoes, no sound coloration, etc.
Reverberation time	T_m [s] = nominal value in region of 200 Hz to 4 kHz V = listening room volume V_0 = reference room volume (100 m^3)	$\approx 0.25\ (V/V_0)^{1/3}$ (Reverberation time decay and tolerance borders are shown in Figure 3.)
Stationary sound field Operational sound level curve	 50 Hz–2 kHz 2 kHz–16 kHz	 ±3 dB ±3 dB from –3 to –6 dB (in accordance with tolerance field, see Figure 4)
Background noise		Ideally <NR10; never >NR15
Reference listening level (relative to defined measurement signal)	Input signal: pink noise, –18 dBFS (rms)	78 dBA (rms slow) (per channel)*

* This is an international standard recommendation for listening test program comparison. NHK proposed 78 ± 2 to 85 ± 2 dBC per channel, depending on room size and application. Higher monitor levels are also common in film mixing environments, where –18 dBFS is normally aligned for 85 dBC. This is covered in greater detail in Section 7.4.

Direct sound is sound without influence from the listening room in the form of reflections and reverberation. The quality is determined by the characteristics of the appropriate loudspeakers (see Table 3).

Reflected sound (reverberation field) is split into *early reflections,* within the first 15 ms, in the region of 1–8 kHz, and *diffuse sound* of the reverberant sound field (linear decay).

A tolerance field for the *reverberation time* is shown in Figure 3. The measurements are made with the loudspeakers used and with one-third-octave-band filtering. T_m is the arithmetic average of the measured reverberation time T in the one-third-octave bands from 200 Hz to 4 kHz. The literature suggests between 0.2 and 0.4 s, depending on the room size (see Table 2), in order to allow a "natural" spatial perception. Large film sound mixing rooms sometimes exceed these limits.

The frequency response for the reverberation time is suggested to be steady and continuous. Sudden or strong breaks influence the operational sound level curve. This condition can be achieved if such deviations in adjoining one-third-octave bands in the region of 200 Hz to 8 kHz do not exceed ±0.05 s; and below 200 Hz if 25% of the longest reverberation time is not exceeded.

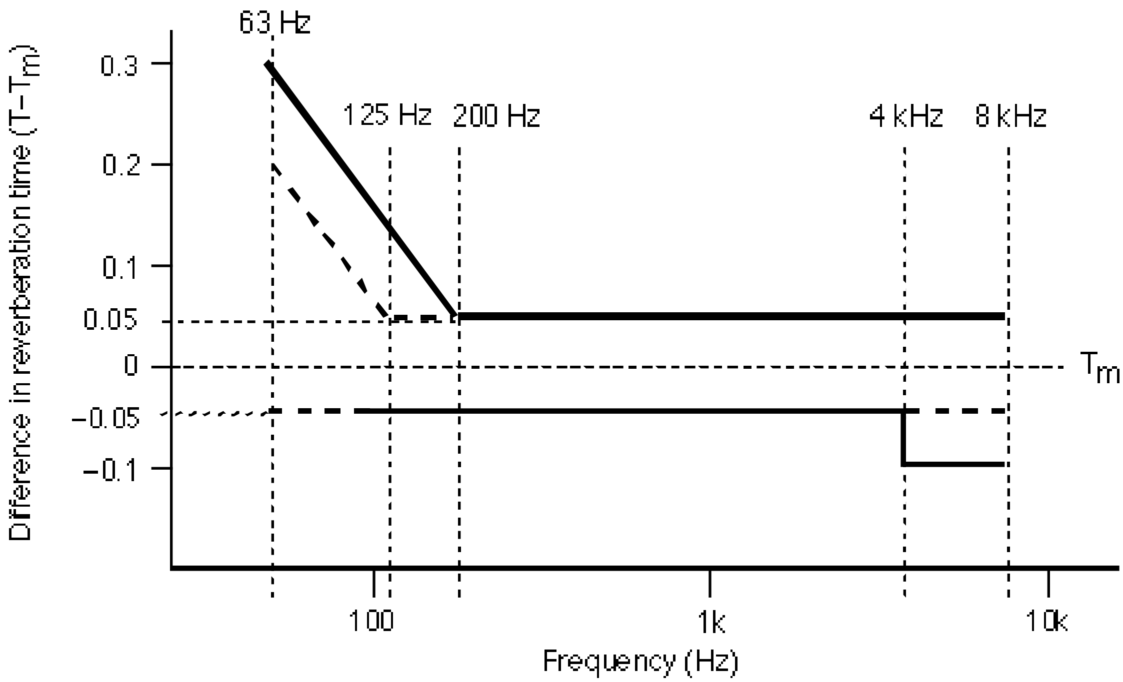


Figure 3. Tolerance mask for reverberation time, relative to arithmetic average value T_m. (Based on international recommendations, but extended to lower frequencies, with smaller tolerances in the range of 63–125/200 Hz.)

The *stationary sound field* is represented by the operational sound level curve shown in Figure 4. This constitutes an important criterion for the interaction between room and loudspeakers and for the quality of the listening conditions achieved. It is measured as the frequency response of the sound pressure level at the reference listening position. Measurement signals are band filtered pink noise. The tolerances are checked separately for each loudspeaker. Consistency with the operational sound level curve is particularly important for the front loudspeakers. (See also further advice in [4].) The region from 20 to 30 Hz is yet to be completed.

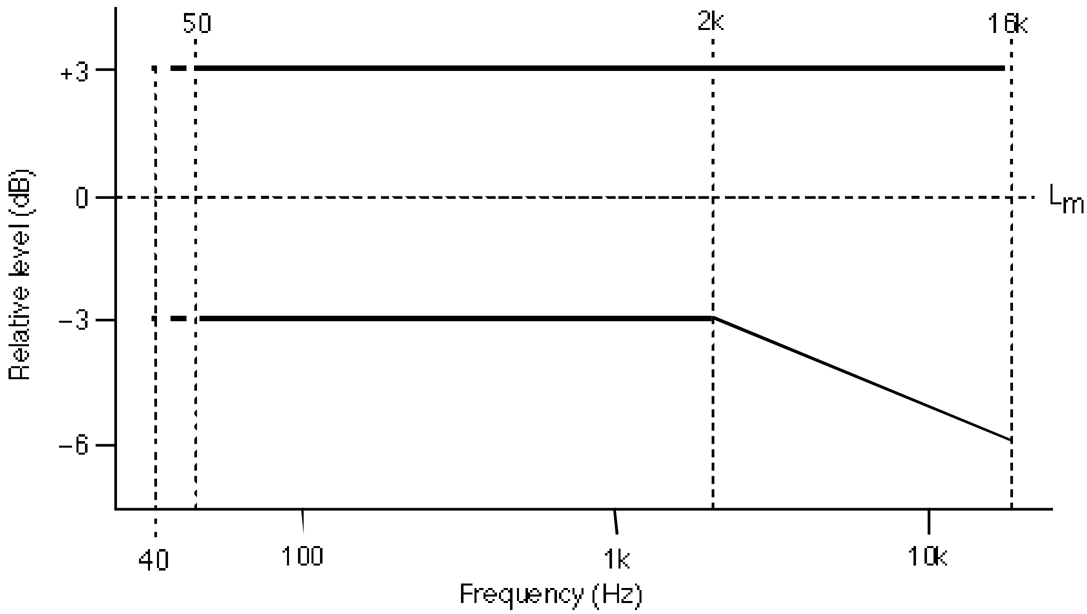


Figure 4. Tolerance limits of operational room response curve, relative level. (Based on international recommendations but extended to lower frequencies.)

6.2.3 Background noise

The continuous noise level (from air conditioning or other external or internal sound sources) is given in form of the one-third-octave band sound pressure level $L_{pFeq,\ T\text{-}30s}$ (rms, slow), in accordance with ISO noise rating curves [9] for the one-third-octave band averaged frequencies from 50 Hz to 10 kHz by a table or a curve. The information for single values is not sufficient. The literature prefers that the NR 10 curve not be exceeded and proscribes exceeding the NR 15 curve.

Note: In many international documents the NR curves are given as octave-band averages. $NR10_{Oct}$ then gives a value at 1 kHz of 10 dB. The corresponding values for one-third-octave measurements are on average 5 dB lower.

6.2.4 Suggestions for reference monitor loudspeakers

The specifications in Table 3 include the objective minimum conditions for a reference monitor loudspeaker. It must be mentioned that there are loudspeakers which comply with these recommendations that are not necessarily suitable as reference loudspeakers for all program genres. To be able to fully perform these critical functions, the conclusive selection and decision is formed on the strength of investigative subjective tests and the resulting criteria and attributes.

For measurement conditions, known guidelines (that is, [4]) are used to relate the measurement distances to the dimensions of the loudspeaker casing (usually distances are >2m). According to IEC 60268-5, the result has to be referred to the nominal distance of 1 m. For electrical measurements it specifies the guaranteed value to be within ±0.2 dB; for acoustic measurements the measurement error margin is to be less than ±1 dB in the total frequency range.

The *amplitude/frequency response* is measured under free-field conditions with pink noise for the one-third-octave band averaged frequencies in the range of 31.5 Hz to 16 kHz at 0°, ±10° and ±30°. The permitted tolerances and differences are given in the table. The correct characteristics are preferred to be symmetrical around the reference axis.

Table 3. Suggestions for reference monitor loudspeakers and advice for home loudspeakers.

Parameters	*Units/Conditions*	*Value*
Amplitude/frequency response	40 Hz–16 kHz	
	0°	Tolerance 4 dB
	±10°	Deviation to 0°, 3 dB
	Horizontal ±30°	Deviation to 0°, 4 dB
Difference between front loudspeakers	In the range >250 Hz to 2 kHz	0.5 dB
Directivity index	250 Hz–16 kHz	8 dB ±2 dB
Nonlinear distortion attenuation (SPL = 96 dB)	<100 Hz >100 Hz	–30 dB (=3%) –40 dB (=1%)
Transient fidelity Decay time t_s, for reduction to a level of $1/e$, i.e., 0.37 of output level	t_s[s]	<5/f [Hz] (preferably 2.5/f)
Time delay Difference between stereo loudspeakers	∂t	≤10 μs
System dynamic range Maximum operating level (measurement acc. to IEC 60268, § 17.2, referred to 1 m distance)	$L_{\text{eff max}}$	>112 dB (at IEC 60268 program simulation noise or special condition)
Noise level	L_{noise}	≤10 dBA

The *directivity index* can also be derived from the one-third-octave band measurements. It can either be calculated from the directional characteristics or derived from the difference between the free field measurements and the diffuse field measurements. According to ITU-R BS.1116-1 [10] a directivity index of >6 dB with a steady slow increase toward higher frequencies is desirable. So-called omnidirectional radiators are regarded as unsuitable for the front loudspeakers. Among other things, a diffuse radiation might be desirable for the surround loudspeakers, this being dependent on the program material, but for the time being current standards recommend only uniform (equal) loudspeakers for all five channels for compatibility reasons.

6.3 Alternative conditions for multichannel mixing rooms

While the recommendations described are based on international standards for reference listening rooms, the practical requirements of small- and medium-sized multichannel mixing rooms have also been suggested by the Japanese HDTV multichannel sound forum. These are given in Table 4 for information. It will be noted that they are broadly similar to the previous criteria in the majority of respects, but they provide slightly greater operational flexibility.

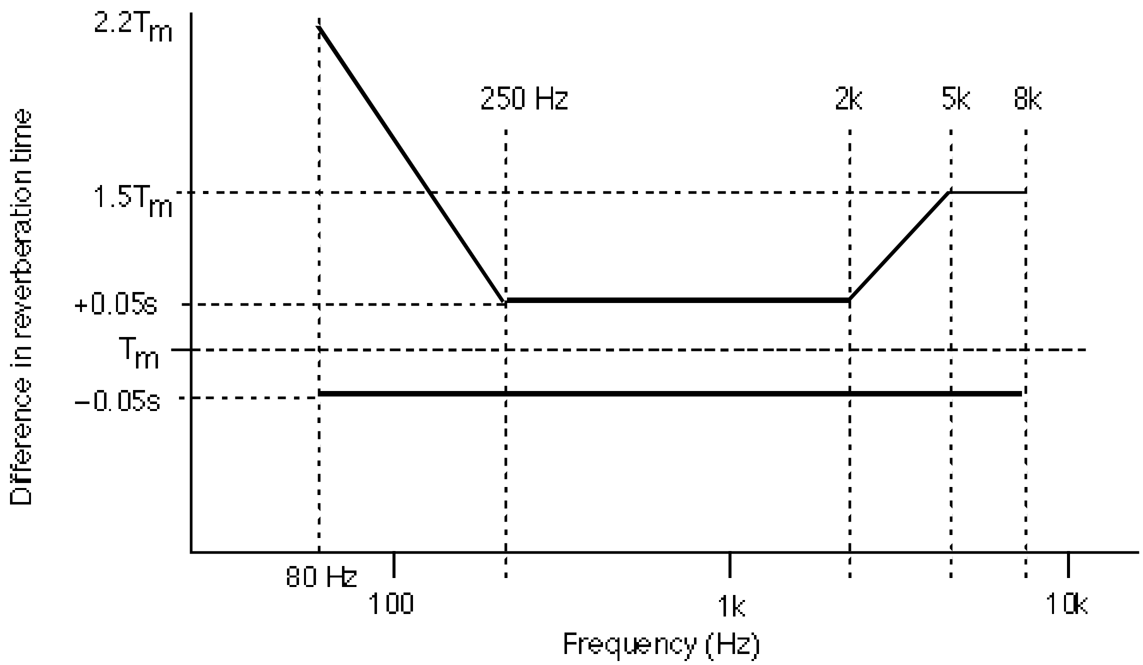


Figure 5. Multichannel mixing room RT characteristics according to Japanese HDTV forum.

Table 4. Multichannel mixing room specifications according to Japanese HDTV surround forum documents.

Parameters		***Design Guideline***	
		Small Rooms	***Medium Rooms***
Display		CRT, 36 inches	Acoustically transparent (perforated) screen, 140 inches
Room	Floor area [m^2]	50 ± 20	100 ± 30
	Room volume [m^3]	≥80	≥200
	Room shape	Nonrectangular (avoid parallel surfaces)	
	Dimensional ratios	Avoid ratios with simple integral numbers; h:w:l = 1:1.59 ± 0.7:2.52 ± 0.28... are desirable	
	Room height [m]	3.0–4.0	4.0–6.0
Interior finish		Uniform absorbent/diffusively reflective treatment to avoid strong reflections from specific directions	
Acoustical properties	Reverberation time [s]	0.2 ± 0.05 (at 500 Hz)	0.3 ± 0.1 (at 500 Hz)
	Mean absorption coefficient	0.4–0.6 (at 500 Hz)	
	Reverberation characteristics	See Figure 5	
	Static transfer frequency response	±3 dB (one-octave band) between 125 Hz and 4 kHz; up to 2 bands may be within ±4 dB	
	Early reflections	Any reflections within 15 ms after direct sound should be 10 dB lower relative to direct sound	
	Interaural cross correlation	Not specified (under consideration)	
	Distribution of SPL	Uniform SPL within listening area, including mixing point	
Noise	Air-conditioning noise	Noise criterion curve NC15 (NR15 would be desirable)	
	Equipment or background noise	Noise criterion curve of NC20 (NR20 would be desirable) (fan noise of video projector, etc., should be reduced)	

Loudspeaker arrangement			
L/R	Setting	Flush mounting is desirable to avoid reflections from rear walls, etc; eliminate these reflections when free standing	
	Axis direction (reference point)	Mixing position or 0 to 1 m in rear	
	Distance (L–R) [m]	3.0–6.0	5.0–8.0
	Height [m][a]	1.2–2.0[b]	Center of screen[b]
	Distance to reference point	All distances from L/C/R/SL/SR loudspeakers to reference point would be desired to be equal	
	Subtended angle against room centerline [deg]	30	
C	Setting	Flush mounting is desirable to avoid reflections from rear walls, etc., eliminate these reflections when free standing	
	Axis direction (reference point)	Mixing position or 0 to 1 m in rear	
	Height [m]	Same height as L/R is desirable[c]	Center of screen[c]
	Distance to reference point	All distances from L/C/R/SL/SR loudspeakers to the reference point would be desired to be equal	
Parameters		***Design Guideline***	
		Small Rooms	***Medium Rooms***
SL/SR	Number	≥2	≥4
	Setting	Flush mounting is desirable but being attached to wall is acceptable because of room shape	
	Axis direction (reference point)	Mixing position or 0 to 1 m in rear	
	Height [m]	Same or higher than L/R is desirable L/R[d] (0.9–1.4)[d]	
	Distance to reference point	All distances from L/C/R/SL/SR loudspeakers to reference point would be desired to be equal	
	Subtended angle against room centerline [deg]	120 ± 10	>110 (symmetrically dispersed at regular intervals)
Monitoring level		85 ±2 dB (C weighted)/ch (pink noise) at –18 dBFS for large loudspeaker	
		80 ±2 dB (C weighted)/ch (pink noise) at –18 dBFS for medium loudspeaker	
		78 ± 2dB (C weighted)/ch (pink noise) at –18 dBFS for small loudspeaker	

Monitor loudspeaker			
Maximum sound pressure level[e]	L/C/R	≥117 dB	≥120 dB
	2	≥114 dB	≥117 dB
	4	≥111 dB	≥114 dB
	8	≥108 dB	≥111 dB
Amplitude versus frequency response	L/C/R	See Figure 6	
Effective frequency range[f]	L/C/R	40 Hz to 20 kHz	
	SL/SR	Same as L/C/R; at least 80 Hz to 20 kHz	
Nonlinear distortion[g]	L/C/R	<3% for 40 Hz to 250 Hz; <1% for 250 Hz to 16 kHz	
	SL/SR	Same as L/C/R; at least <3% for 80 Hz to 250 Hz; <1% for 250 Hz to 16 kHz	
Transient fidelity	L/C/R/SL/SR	Decay time to level of 1/*e* (approximately 0.37) from original level should be less than 5/*f* (where *f* is frequency)	
Phase, Group delay	L/C/R/SL/SR	Either of them would desirably be indicated	
Directivity index[h, i]	L/C/R/SL/SR	6–12 dB (ITU-R BS.1116-1)	
Impedance	L/C/R/SL/SR	>3.2 Ω	
Deviation of frequency response	L/C/R/SL/SR	<1.5 dB for 100 Hz to 10 kHz; peak/dip narrower than one-third octave shall be neglected	
Efficiency[j]	L/C/R/SL/SR	Should be indicated	

Notes:

a Loudspeakers height: height of acoustical center of loudspeaker from floor level at mixing position.

b More than 1.2 m is recommended. But height may be 1.7 m to avoid meter bridge of high console shadowing direct sound, and that may be 1.9 m when loudspeakers are set above window.

c When C loudspeaker is set below the CRT, its height may be lower than L/R loudspeakers.

d Same as L/C/R is desirable, but it could be 2.2–2.7 m because of doors on side or rear walls.

e Maximum sound pressure level = rated output sound pressure level + maximum input level.

f Effective frequency range: frequency range –10 dB.

g Absolute sound level is measured at 1 m from loudspeaker.

h Directivity index of front loudspeakers depends on program or software.

i Difference of overall impressions caused by directivity index of rear loudspeakers is rather small.

j Efficiency is indicated by rated output sound pressure level at 1 m,1 W.

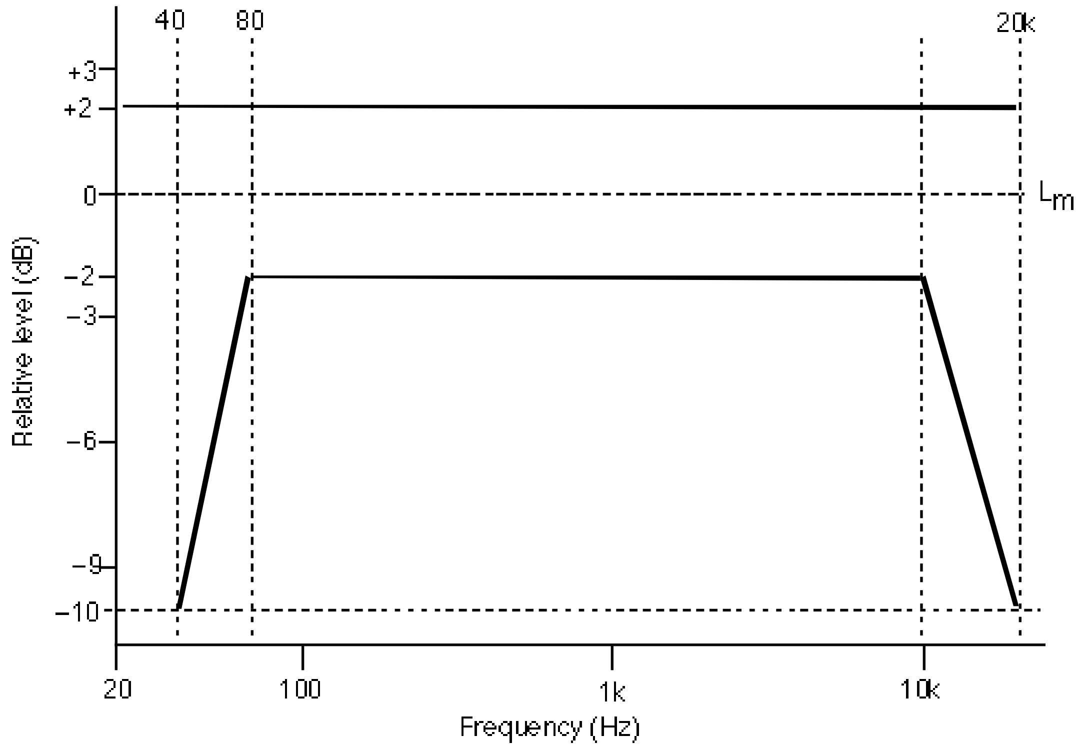


Figure 6. Amplitude/frequency characteristics of loudspeakers, measured in an anechoic chamber, according to Japanese HDTV forum.

7 PROGRAM INTERCHANGE

This section offers guidelines based on international recommendations currently in force, designed to enable the interchange of multichannel program material between sites. This is to be read in conjunction with Section 7.4 on monitor level alignment and Section 6 on studio acoustics. Although it is impossible to ensure that a multichannel program sounds the same in every environment, some attention to these issues will bring about a degree of acoustic compatibility. It is acknowledged that local operational practices may differ in certain respects and that alternatives have been proposed and used. For information, some of the issues are presented here.

In individual cases the designated use of tracks 4, 7, and 8 is stated on the recording medium.

In many film studios a different order of track allocation or listening buttons is being used as the normal practice, namely, L—C—R—LS—RS. However, new mixing desks follow the international standard as recommended in Table 5. After ongoing debate over the practices of the international organizations of the ITU, this recommendation was formed and is binding for sound broadcasting and television, but not for film studios.

7.1 Track allocation in an eight channel recording format (Table 5)

Table 5. Track allocation in eight-channel recording format [3], [11], [12]

Track[a]	Signal	Comments	Color[b]
1	L Left		Yellow
2	R Right		Red
3	C Center		Orange
4	LFE Low-frequency extension	Additional subbass and effects signal for subwoofer, optional[c]	Grey
5	LS Left surround	–3 dB in case of mono surround	Blue
6	RS Right surround	–3 dB in case of mono surround	Green
7	In program exchange free use[d]	Preferably left signal of 2/0 stereo mix	Violet
8	In program exchange free use[d]	Preferably right signal of 2/0 stereo mix	Brown

a The term "track" is used to denote either tracks on magnetic tape, or virtual tracks on storage media where no real tracks exist.

b This color coding is at present only a proposal of the German surround sound forum at present, and not internationally standardized.

c Preferably, used in film sound, but is optional for home reproduction. If no LFE signal is being used, track 4 can be used freely, e.g., for commentary. In some regions a mono surround signal MS = LS + RS is applied, where the levels of LS and RS are decreased by 3 dB before summing.

d Tracks 7 and 8 can be used alternatively, for example, for commentary, for additional surround- signals, or for half–left/half–right front signals (e.g., for special film format), or rather for the matrix format sum signals Lt/Rt.

7.2 Recording Levels

Practice regarding alignment levels and maximum recording levels varies. In broadcasting and some studio recording operations, where program interchange compatibility is of primary importance, it is normal to work to international standard guidelines that define an alignment level, $L_{\mathbf{AS}}$, and a permitted maximum signal level $L_{\mathbf{PMS}}$. ITU and EBU recommendations, among others, specify a digital alignment signal level of –18 dBFS, whereas SMPTE recommendations specify –20 dBFS (1-kHz tone, rms measurement). Both are likely to be encountered in operational practice, and it is therefore important to indicate clearly which alignment level is adopted, in order to avoid subsequent confusion.

The $L_{\mathbf{PMS}}$ is normally 9 dB below the digital clipping level, and is intended to be related to the measurement of program signals on quasi-peak meters that have an integration time of 10 ms, thereby ensuring that short transients are not clipped. True peak-reading meters will exceed this indication on some program material, whereas VU meters will typically underread this indication as they have a long integration time. In mastering and some film sound operations it is common to use the entire recording level range up to 0 dBFS. In such circumstances it is important to use true peak-reading meters in order to avoid clipping on digital media.

7.2.1 Recording levels in film sound

In film sound environments it is the norm to increase the recording level of the surround channels by 3 dB compared with that of the front channels. This is in order to compensate for the –3-dB alignment of each surround channel's sound pressure level with respect to the front, which takes place in dubbing stages and movie theaters. It is important to be aware of this discrepancy between practices, as it is the norm in music mixing and broadcasting to align all channels for equal level, both on recording media and for acoustical monitoring. Transfers from film masters to consumer or broadcast media may require a 3-dB alteration in the gain of the surround channels.

7.3 Alignment signals

It has been ideal practice to record, preceding the program section of the recording, a level checking section, with two reference signals for each track used. These signals are usually recorded at the alignment level L_{AS}:

- A 1-kHz sine tone to check the alignment signal level
- Random noise, noncorrelated, to check the sound pressure levels.

Recording of the noise signal is superfluous if measurements and test recordings such as those described here are used as a standard method throughout the world. At present, because of varying international standards, measurement signals and therefore the resulting sound pressure levels are not handled uniformly.

7.4 Reproduction system alignment

Reproduction level alignment is the subject of considerable debate in the industry, particularly concerning the nature of the test signal, the type of measurement, and its weighting. This is a subject requiring further debate and experimentation before a consensus can be reached. Nonetheless, for information purposes it is possible to outline a number of practices in common use today, together with a discussion of some of the concerns expressed, and to present Table 6 comparing some of the alternatives.

It is noted that the subjective alignment of monitor balance is sometimes practical, balancing in a pairwise fashion for a central image of a test signal between loudspeakers when facing the center of the pair. This becomes less practical in multichannel configurations because of the wide angle subtended by some pairs of loudspeakers at the listening position, or when the loudspeakers are not arranged in a symmetrical fashion. It is also sometimes necessary to adjust the monitor gain for a certain absolute loudness level, in which case methods such as those given in the following will be helpful.

7.4.1 Reference listening level $L_{LISTref}$

The reference listening level L_{LISTref} allows the specified listening level, or volume, to be set correctly during the reproduction of program material under specified reproduction conditions, as well as during the reproduction of the same program material under different conditions. The measurement takes place for each individual reproduction channel, from the reference listening position. Each channel is to be played through one fader and one monitor loudspeaker at a time. The measuring signal used to set each channel is pink noise, band-pass filtered.

The ideal bandwidth of the noise signal is the subject of some debate. While it is generally agreed that some low frequency roll-off is desirable in order to avoid that the measurement is dominated by the effect of room modes, there is no agreement on the precise frequency of this roll-off. Furthermore, while some proponents have also recommended band-limiting at high frequency (e.g., 2 or 4 kHz), others have proposed no high-frequency limit (noise extending to 20 kHz). Variations in measurement weighting also exist, with both A weighting (ITU, EBU) and C weighting (SMPTE, Japanese HDTV Forum) commonly being recommended.

Deleted page September 2001.

In ITU and EBU standards the level of the each reproduction channel (excluding the LFE channel) is set so that the sound level (rms slow) at the reference listening position is

$L_{\mathbf{LISTref}} = 85 - 10 \log n$ [dBA]

Where n is the number of reproduction channels in the relevant configuration. So if one channel has a reference listening level $L_{\mathbf{LISTref}}$ = 78 dBA, then the five combined channels of the 3/2 multichannel stereo configuration have a resulting reference listening level $L_{\mathbf{LISTref}}$ = 85 dBA. Level differences between any two channels not exceeding 1 dB and, where it is possible to measure more precisely than this, less than 0.5 dB are suggested.

It sometimes occurs that the listening level needs to be adapted individually to suit the content of a program. This value can be given with reference to the reference listening level, and so stated on the recording medium. For example, the replay of a special program over a 3/2 stereo configuration with a level of –10 dB, compared to the reference listening level, means that the total level from all five replay channels measured at the reference listening position, using a noncoherent pink noise signal, will be 75 dBA.

Some standard recommendations for level alignment recommend the use of broad-band pink noise, or pink noise band limited from 200 Hz to 20 kHz. This has been criticized by some for involving too much low-frequency content, and thereby making the measurement highly dependent on room mode response, as well as being very direction-dependent at high frequency. Such measurements, though, are normally made with A-weighting filters, which reduce the extreme low- and high-frequency components considerably.

It is common in the film sound domain and certain other operations, including the Japanese recommendations for multichannel mixing rooms for HDTV, to use C-weighting instead of A-weighting for monitor level alignment (C-weighting is a somewhat “flatter” curve than A-weighting which approximates the equal-loudness contours at higher levels). An alternative “film-style” recommendation uses pink noise, band-limited between 500 Hz and 2 kHz, at the SMPTE standard alignment level of –20 dBFS. This signal is aligned for a sound presssure level of 83 dBC (slow) at the monitoring position, when setting the level of each channel individually. (Note: A –18 dBFS test signal would then read 85 dBC.) In movie theaters and on film dubbing stages it is common practice to align the surround channels with a –3-dB offset in gain with respect to the front channels. The recording levels of stereo surround channels are correspondingly increased, as noted before. The Japanese HDTV mixing room recommendation appears to use broad-band pink noise with C-weighted measurement, giving different sound pressure level recommendations depending on the size of the loudspeaker in use. These methods of alignment are unlikely to result in the same monitoring level at the listening position as the first method given, but in each case (with the exception of film theaters) all channels are aligned individually, using a noise signal, for an equal-weighted sound pressure level at the listening position.

Recent research attempted to find correlations between the subjective alignment of channel loudness and a variety of objective measurements, using a wide range of different test signals [13]. There is some evidence that the low-frequency content of test signals is ignored when subjectively aligning channel gain, and that constant specific loudness noise signals may be preferred over other noise signals with regard to subjective/objective correlation. Further work is required to determine what high-frequency roll-off (if any) is ideal for noise signals to be used in system alignment.

8 ON THE DISCRIMINATION OF REPRODUCTION FORMATS AND CODING FORMATS

Current practice of multichannel sound engineering makes use of a variety of distribution and presentation formats. These practices are mostly not standardized but introduced as proprietary formats. Therefore the incorrect impression is given that these formats are system solutions of the recording and reproduction formats, presented earlier in this document (which was written on the basis of international standards according to ITU-R BS.775-1 [1]).

For short connections, e.g., between recording rooms and control rooms, direct lines are normally available, without any coding in between. Therefore the format to be reproduced can be defined clearly. On the other hand, for the delivery and recording of multichannel sound signals for consumer applications, digital bit-rate reduction coding or analog matrixing methods are often necessary, particularly with the constraints of limited delivery and data capacities. Therefore some distinction is required:

- *Multichannel reproduction formats:* Represented by the reference loudspeaker layout 3/2 or 3/2/1 (and including 2/2, 3/1, 5/2, etc.), as described in Sections 3 and 4;
- *Coding and delivery formats:* For recording, delivery or transmission, and connection of multichannel signals with different media.

In the latter case the number of delivery channels always needs to be considered separately from the format. For example, the code "4-2-4" denotes a matrixing format, with which the four signals (L, C, R, S) are delivered or recorded on two channels, and reproduced later in the 3/1 format.

9 REFERENCES

[1] ITU-R BS.775-1, "Multichannel stereophonic sound system with and without accompanying picture," Rec., International Telecommunications Union, Geneva, Switzerland (1992ñ1994).

[2] SMPTE RP-173, "Loudspeaker placements for audio monitoring in high definition electronic production," Rec., SMPTE N 15.04/152-300B, Society of Motion Picture and Television Engineers (1991).

[3] SMPTE "Channel assignments and levels on multichannel audio media," Proposed Standard for Television, ITU Information doc. ITU-R 10C/11 and 10-11R/2 (1998 Mar. 16).

[4] EBU R22, "Listening conditions for the assessment of sound programme material," Rec., European Broadcasting Union (2000). For details see EBU Tech3276 with suppl. 1.

[5] T. Nousaine, "Multiple subwoofers for home theater," presented at the 103rd Convention of the Audo Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 45, p. 1015 (1997 Nov.), preprint 4558.

[6] N. Zacharov, S. Bech, and D. Meares, "The use of subwoofers in the context of surround sound program reproduction," *J. Audio Eng. Soc. (Abstracts),* vol. 46, pp. 276–287 (1998 Apr.).

[7] C. Kügler and G. Theile, "Loudspeaker reproduction: study on the subwoofer concept," presented at the 92nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts*), vol. 40, p. 437 (1992 May), preprint 3335.

[8] D. Griesinger, "Spatial impression and envelopment in small rooms," presented at the 103rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 45, pp. 1013–1014 (1997 Nov.), preprint 4638.

[9] ISO1996, "One-third octave band background noise level limits noise rating curves (NR)," Rec., International Standards Organization, Geneva, Switzerland (1972).

[10] ITU-R BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," Rec., International Telecommunications Union, Geneva, Switzerland (1997).

[11] EBU R 91, "Track allocations and recording levels for the exchange of multichannel recording," Rec., European Broadcasting Union (1998); EBU R 96 "Formats for production and delivery of multichannel programme," Rec., ibid. (2000).

[12] ITU-R BS.1384, "Parameters for international exchange of multi-channel sound recording," International Telecommunications Union, Geneva, Switzerland (1998).
[13] N. Zacharov and S. Bech, "Multichannel level alignment, Part iv: the correlation between physical measures and subjective level calibration," presented at the 109th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts),* vol. 48, p. 1110 (2000 Nov.), preprint 5241.