

MULTICANAL

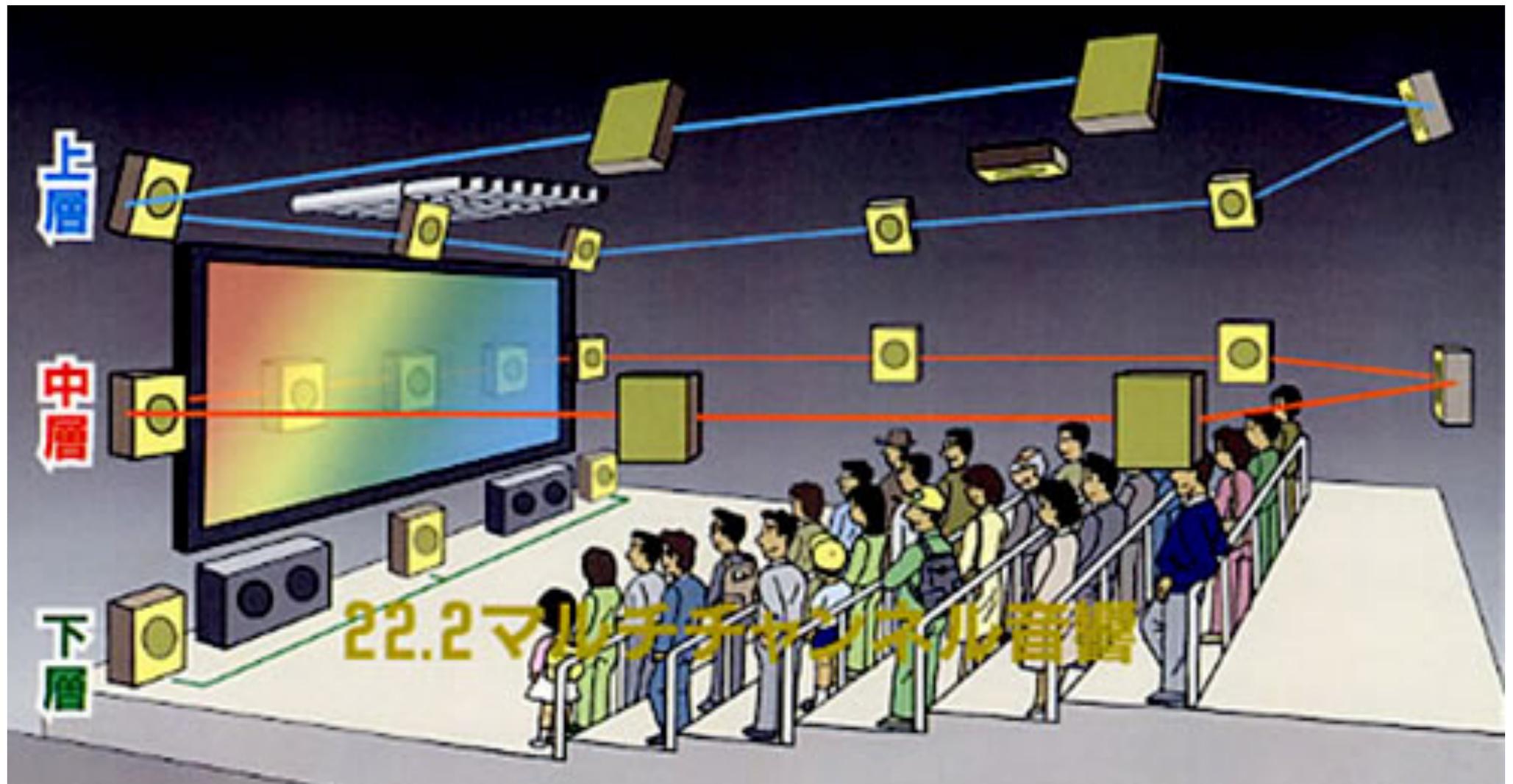
Down-mix du 22.2 vers le 5.1

Bernard Lagnel

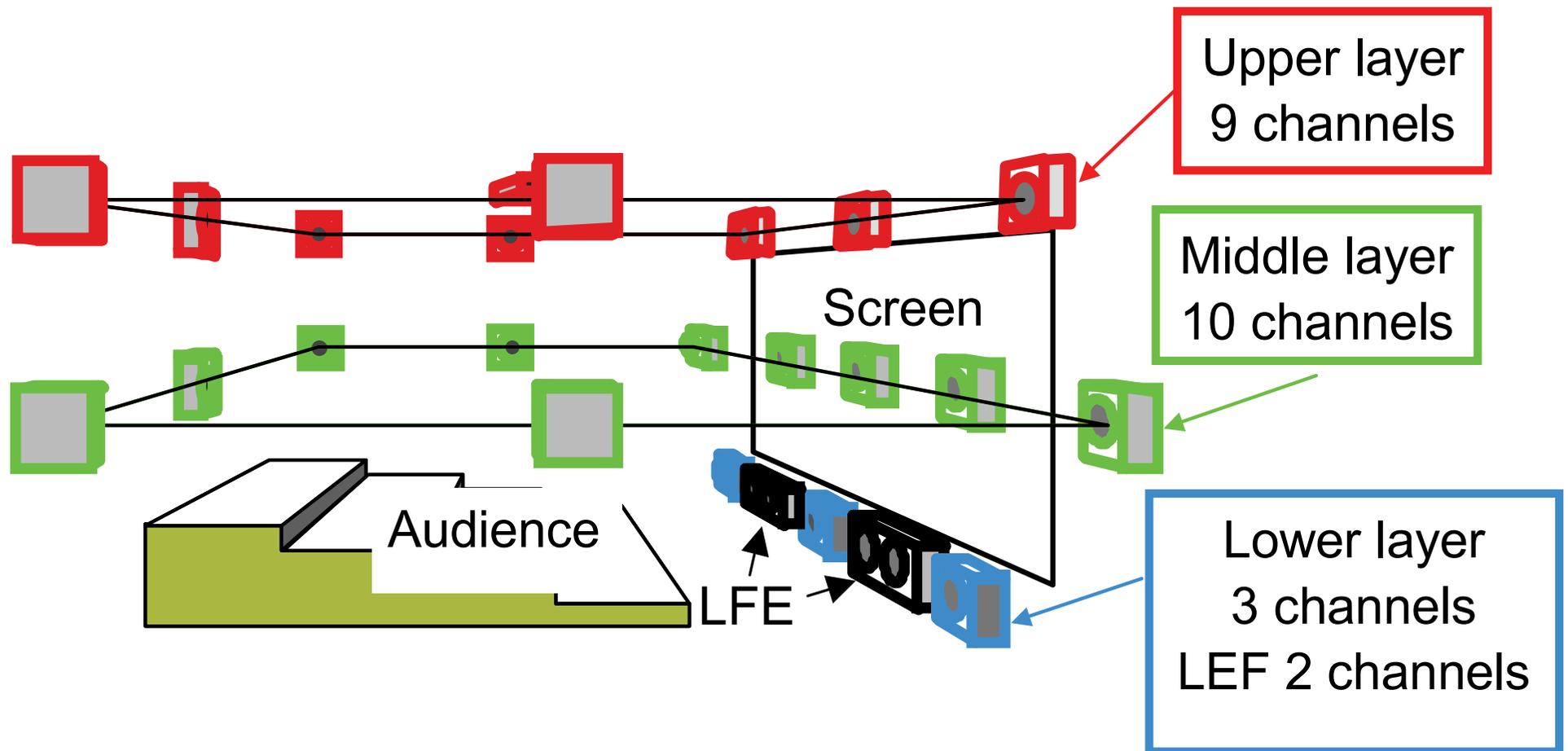
Octobre 2017

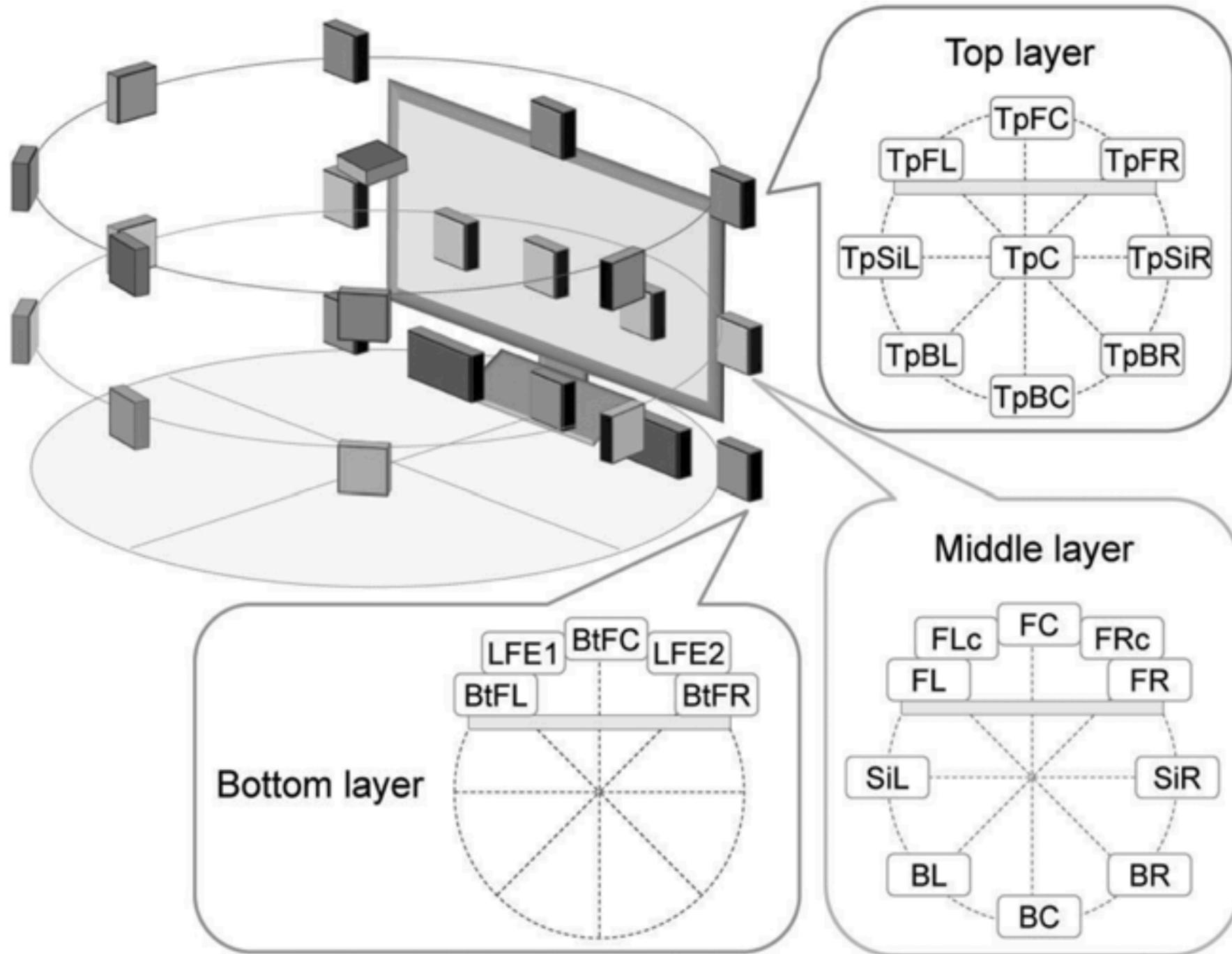
L'audio orienté objets pénètre le marché du broadcast...

22.2 NHK Japon



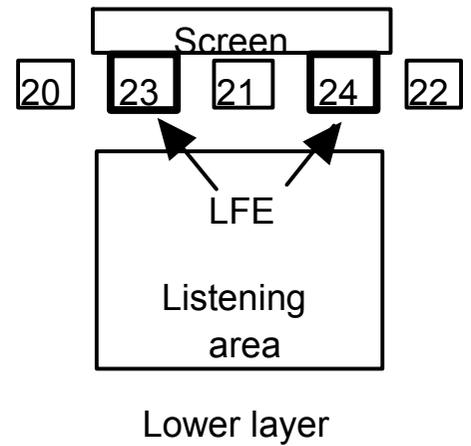
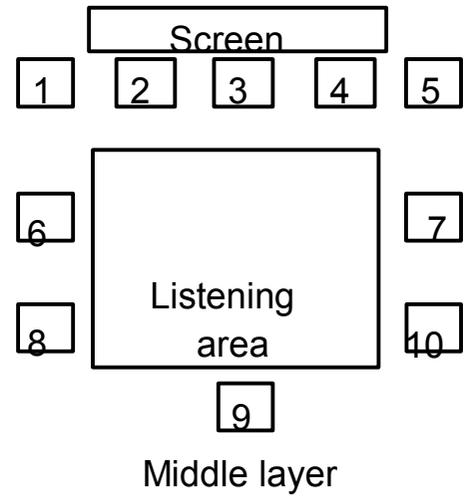
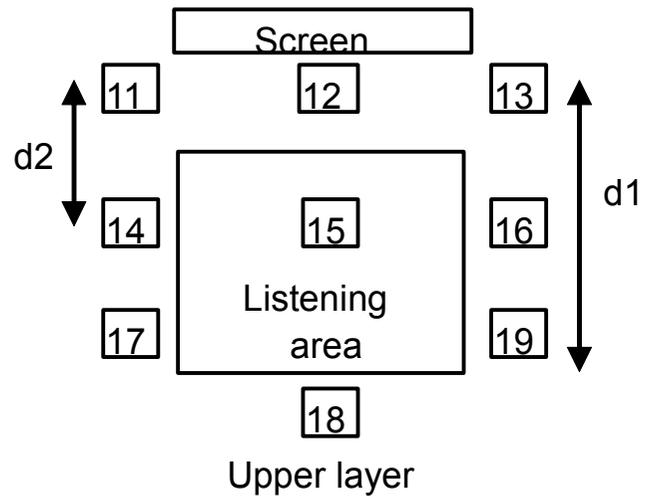
NHK Science & Technical Research Laboratories

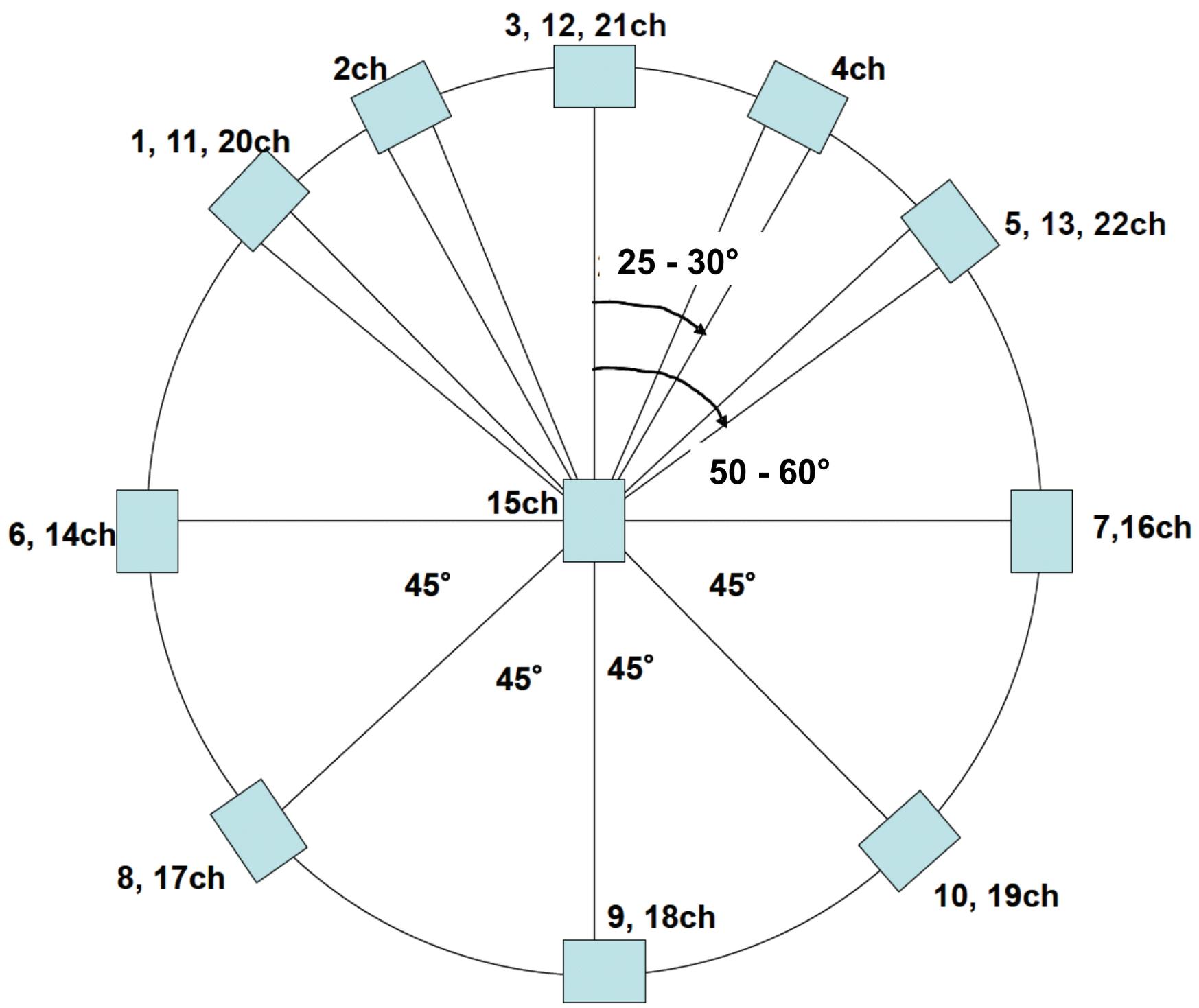


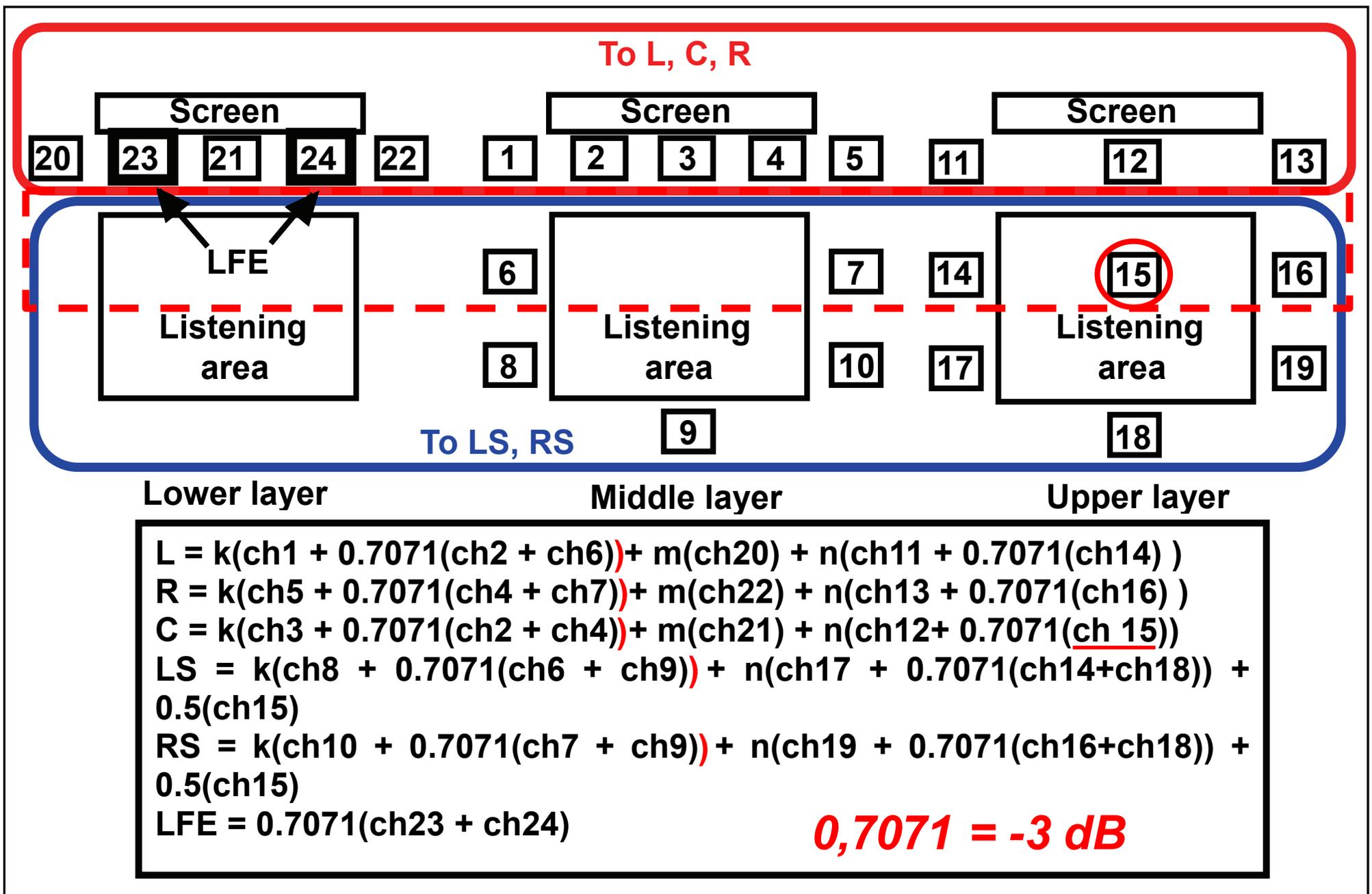


ITU 5.1

AES Pair No./Ch No.	Channel No.	Label	Name
1/1	1	FL	Front left
1/2	2	FR	Front right
2/1	3	FC	Front center
2/2	4	LFE1	LFE-1
3/1	5	BL	Back left
3/2	6	BR	Back right
4/1	7	FLc	Front left center
4/2	8	FRc	Front right center
5/1	9	BC	Back center
5/2	10	LFE2	LFE-2
6/1	11	SiL	Side left
6/2	12	SiR	Side right
7/1	13	TpFL	Top front left
7/2	14	TpFR	Top front right
8/1	15	TpFC	Top front center
8/2	16	TpC	Top center
9/1	17	TpBL	Top back left
9/2	18	TpBR	Top back right
10/1	19	TpSiL	Top side left
10/2	20	TpSiR	Top side right
11/1	21	TpBC	Top back center
11/2	22	BtFC	Bottom front center
12/1	23	BtFL	Bottom front left
12/2	24	BtFR	Bottom front right







Les coefficients (k, m, n) dans les équations de mélange ajustent le niveau sonore total et équilibrent les niveaux sonores dans les couches supérieure, intermédiaire et inférieure. Les valeurs k, m et n lors des évaluations subjectives ont été réglées à 1,0 pour maintenir le même volume sonore total dans les différents systèmes sonores multicanaux.

Downmixing Method for 22.2 Multichannel Sound Signal in 8K Super Hi-Vision Broadcasting

TAKEHIRO SUGIMOTO, AES Member, SATOSHI OODE, AND YASUSHIGE NAKAYAMA

Science & Technology Research Laboratories, NHK, 1-10-11 Kinuta, Setagaya-ku, Tokyo 1578510, Japan

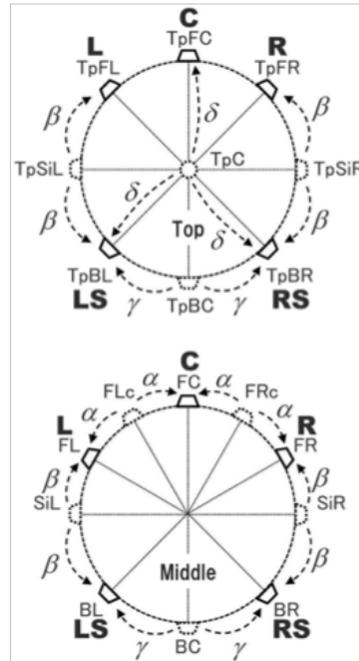


Fig. 2 Downmixing paths in the top and middle layers.

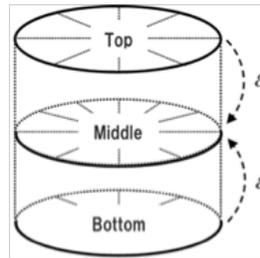


Fig. 3 Downmixing coefficient between the layers.

Fig. 2 shows the downmixing paths in the top and middle layers (first step). The channels in the bottom layer are directly converted into a 5.1 ch layout by converting BtFL to L, BtFC to C, and BtFR to R. The Greek letters in Fig. 2 correspond to the downmixing coefficients of the downmixing equations. Although the proposed paths cannot always render the channel position accurately, we endeavored to reduce the number of coefficients to as few as possible to simplify of the broadcasting system.

Fig. 3 shows the downmixing coefficient in the second step. The top and bottom layers are biased by ϵ when the middle layer is superimposed on them.

The downmixing equations from 22.2 ch to 5.1 ch are as follows:

$$L = FL + \alpha FLc + \beta SiL + \epsilon(TpFL + \beta TpSiL + BtFL), \quad (1)$$

$$R = FR + \alpha FRc + \beta SiR + \epsilon(TpFR + \beta TpSiR + BtFR), \quad (2)$$

$$C = FC + \alpha FLc + \alpha FRc + \epsilon(TpFC + \delta TpC + BtFC), \quad (3)$$

$$LFE = \zeta(LFE1 + LFE2), \quad (4)$$

$$LS = BL + \gamma BC + \beta SiL + \epsilon(\delta TpC + TpBL + \beta TpSiL + \gamma TpBC), \quad (5)$$

$$RS = +\gamma BC + \beta SiR + \epsilon(\delta TpC + TpBR + \beta TpSiR + \gamma TpBC). \quad (6)$$

The downmixing equations from 5.1 ch to 2 ch have been specified in Rec. ITU-R BS.775-3 [11] as

$$L_{2ch} = L + \frac{1}{\sqrt{2}}C + \frac{1}{\sqrt{2}}LS, \quad (7)$$

$$R_{2ch} = R + \frac{1}{\sqrt{2}}C + \frac{1}{\sqrt{2}}RS, \quad (8)$$

and they are widely adopted in current broadcasting systems.

Using Eqs. (1)–(8), a 22.2 ch signal is downmixed to a 2 ch signal in tandem as

$$L_{2ch} = FL + \frac{1}{\sqrt{2}}FC + \frac{1}{\sqrt{2}}BL + \left(1 + \frac{1}{\sqrt{2}}\right)\alpha FLc + \frac{1}{\sqrt{2}}\alpha FRc + \frac{1}{\sqrt{2}}\gamma BC + \left(1 + \frac{1}{\sqrt{2}}\right)\beta SiL + \epsilon \left\{ TpFL + \frac{1}{\sqrt{2}}TpFC + \sqrt{2}\delta TpC + \frac{1}{\sqrt{2}}TpBL + \left(1 + \frac{1}{\sqrt{2}}\right)\beta TpSiL + \frac{1}{\sqrt{2}}\gamma TpBC + \frac{1}{\sqrt{2}}BtFC + BtFL \right\}, \quad (9)$$

$$R_{2ch} = FR + \frac{1}{\sqrt{2}}FC + \frac{1}{\sqrt{2}}R + \left(1 + \frac{1}{\sqrt{2}}\right)\alpha FRc + \frac{1}{\sqrt{2}}\alpha FLc + \frac{1}{\sqrt{2}}\gamma BC + \left(1 + \frac{1}{\sqrt{2}}\right)\beta SiR + \epsilon \left\{ TpFR + \frac{1}{\sqrt{2}}TpFC + \sqrt{2}\delta TpC + \frac{1}{\sqrt{2}}TpBR + \left(1 + \frac{1}{\sqrt{2}}\right)\beta TpSiR + \frac{1}{\sqrt{2}}\gamma TpBC + \frac{1}{\sqrt{2}}BtFC + BtFR \right\}. \quad (10)$$

2.3 Initial Downmixing Coefficients

It is important in broadcasting that appropriate downmixing coefficients are selected by the sound engineer and transmitted in every program. However, there are several situations in which the downmixing coefficients are not provided. One such situation is a live broadcast without rehearsal. An unrehearsed program would not provide sufficient time to fix the downmixing coefficients. Another situation is when a linear pulse code modulation (PCM) multichannel stream is input to the AV receiver/amplifier. In this case, there is no common scheme for carrying the downmixing coefficients in the linear PCM multichannel stream. Therefore, the initial downmixing coefficients are required to prevent the case of no sound being reproduced due to the lack of downmixing coefficients.

2.3.1 Requirements for 5.1 ch and 2 ch Downmixing

We set the requirements for 5.1 ch and 2 ch downmixing to find compatible values for 5.1 ch and 2 ch, respectively. The requirements for 5.1 ch downmixing are as follows:

- The sound power at the listening point corresponding to each channel of 22.2 ch should be unchanged by downmixing in order to maintain the sound power balance of 22.2 ch in the 5.1 ch layout.
- The sound power of LFE should not exceed the sum of the sound powers of LFE1 and LFE2 even when an identical signal is assigned to both LFE1 and LFE2.

We set the requirements for 2 ch downmixing based on the philosophy of the downmixing method from 5.1 ch to 2 ch as follows:

- The sound power at the listening point corresponding to each frontal channel of 22.2 ch should be unchanged by downmixing in order to maintain the sound power balance among the 11 original frontal channels. This is required in the case that different sound objects, such as a voice and instrument, are assigned to the 11 individual frontal channels.
- The sound power at the listening point corresponding to each channel of 22.2 ch in the transverse plane should be less than or equal to that of each frontal channel after downmixing. This is because the sound relevant to the video, which is important compared with the other sounds, is often assigned to the frontal channels.
- The sound power at the listening point corresponding to each channel of 22.2 ch in the transverse plane should be greater than or equal to that of each rear channel after downmixing. This is because the loudness of sound from the transverse direction is liable to be perceived as greater than that of sound from behind [20, 21].

Table 1 shows the sound power ratio of each channel of 22.2 ch in downmixed 5.1 ch and 2 ch signals. Each ratio is

Table 1. Sound power ratio of each channel of 22.2 ch in downmixed 5.1 ch and 2 ch signals.

Channel label	Sound power ratio	
	5.1 ch	2 ch
FL	1	1
FR	1	1
FC	1	1
LFE1	ζ^2	0
BL	1	$\frac{1}{2}$
BR	1	$\frac{1}{2}$
FLc	$2\alpha^2$	$(2 + \sqrt{2})\alpha^2$
FRc	$2\alpha^2$	$(2 + \sqrt{2})\alpha^2$
BC	$2\gamma^2$	γ^2
LFE2	ζ^2	0
SiL	$2\beta^2$	$(\frac{3}{2} + \sqrt{2})\beta^2$
SiR	$2\beta^2$	$(\frac{3}{2} + \sqrt{2})\beta^2$
TpFL	ε^2	ε^2
TpFR	ε^2	ε^2
TpFC	ε^2	ε^2
TpC	$3\delta^2\varepsilon^2$	$4\delta^2\varepsilon^2$
TpBL	ε^2	$\frac{1}{2}\varepsilon^2$
TpBR	ε^2	$\frac{1}{2}\varepsilon^2$
TpSiL	$2\beta^2\varepsilon^2$	$(\frac{3}{2} + \sqrt{2})\beta^2\varepsilon^2$
TpSiR	$2\beta^2\varepsilon^2$	$(\frac{3}{2} + \sqrt{2})\beta^2\varepsilon^2$
TpBC	$2\gamma^2\varepsilon^2$	$\gamma^2\varepsilon^2$
BtFC	ε^2	ε^2
BtFL	ε^2	ε^2
BtFR	ε^2	ε^2

normalized by the sound powers of FL and FR, which are directly converted to 5.1 ch and 2 ch.

Referring to Table 1, the requirements for 5.1 ch can be expressed by setting the sound power ratio to 1 except for LFE as follows:

$$2\alpha^2 = 1, \quad (11)$$

$$2\beta^2 = 1, \quad (12)$$

$$2\gamma^2 = 1, \quad (13)$$

$$3\delta^2\varepsilon^2 = 1, \quad (14)$$

$$\varepsilon^2 = 1, \quad (15)$$

$$2\zeta^2 \leq 1, \quad (16)$$

where the calculated values are shown in Table 2.

Table 2. Calculated downmixing coefficients based on the requirements for 5.1 ch and 2 ch downmixing, and proposed initial downmixing coefficients compatible with both 5.1 ch and 2 ch downmixing.

Downmixing coefficient	5.1 ch		2 ch		Proposed values	
	Antilog	Level [dB]	Antilog	Level [dB]	Antilog	Level [dB]
α	$\frac{1}{\sqrt{2}}$	-3.0	$\sqrt{1 - \frac{1}{\sqrt{2}}}$	-5.3	$\frac{1}{2^{\frac{3}{4}}}$	-4.5
β	$\frac{1}{\sqrt{2}}$	-3.0	$\sqrt{2} - 1$ to $2 - \sqrt{2}$	-7.7 to -4.6	$\frac{1}{2^{\frac{3}{4}}}$	-4.5
γ	$\frac{1}{\sqrt{2}}$	-3.0	$\frac{1}{\sqrt{2}}$	-3.0	$\frac{1}{\sqrt{2}}$	-3.0
δ	$\frac{1}{\sqrt{3}}$	-4.8	$\frac{1}{2\sqrt{2}}$ to $\frac{1}{2}$	-9.0 to -6.0	$\frac{1}{2}$	-6.0
ε	1	0.0	1	0.0	1	0.0
ζ	$\leq \frac{1}{\sqrt{2}}$	≤ -3.0	N.A.	N.A.	$\frac{1}{\sqrt{2}}$	-3.0

The requirements for 2 ch can be expressed by setting the sound power ratio to 1 for the frontal channels, to 1/2 to 1 for the channels in the transverse plane, and to 1/2 for the rear channels as follows:

$$(2 + \sqrt{2})\alpha^2 = 1, \quad (17)$$

$$\frac{1}{2} \leq \left(\frac{3}{2} + \sqrt{2}\right)\beta^2 \leq 1, \quad (18)$$

$$\gamma^2 = \frac{1}{2}, \quad (19)$$

$$\frac{1}{2} \leq 4\delta^2\varepsilon^2 \leq 1, \quad (20)$$

$$\varepsilon^2 = 1, \quad (21)$$

where the calculated values are listed in Table 2.

2.3.2 Proposed Initial Downmixing Coefficients

Because different values are obtained for α , β , and δ for 5.1 ch and 2 ch, the next step is to find compatible values for 5.1 ch and 2 ch. An additional restriction is imposed that the downmixing coefficients must be selected at intervals of 1.5 dB starting from 0 dB in MPEG-4 AAC [18]. The interval between the downmixing coefficients of MPEG-2 AAC [22], which is the current audio coding of digital broadcasting in Japan, is 3.0 dB. Thus, an interval of 1.5 dB increases the flexibility of downmixing because it extends the range of choices of the downmixing coefficients. Under these conditions, we proposed compatible values, which are given in the last two columns of Table 2, as a result of the following consideration.

We selected -4.5 dB for α because it is between -3.0 dB for 5.1 ch and -5.3 dB for 2 ch. For β , the value for 2 ch is about -6.0 dB; thus, a compatible value between -3.0 dB for 5.1 ch and -6.0 dB for 2 ch is considered to be -4.5 dB. γ is -3.0 dB because both 5.1 ch and 2 ch require the same value. For δ , the value for 2 ch is about

Table 3. Downmixing coefficients from 22.2 ch to 5.1 ch for program production.

Downmixing coefficient	Antilog	Level [dB]
α	$\frac{1}{\sqrt{2}}$	-3.0
β	$\frac{1}{\sqrt{2}}$	-3.0
γ	$\frac{1}{\sqrt{2}}$	-3.0
δ	$\frac{1}{\sqrt{2}}$ [LS, RS]	-3.0
	$\frac{1}{2}$ [C]	-6.0
ε	1 or $\frac{1}{\sqrt{2}}$	0.0 or -3.0
ζ	$\frac{1}{\sqrt{2}}$	-3.0

-7.5 dB; thus, a compatible value between -4.8 dB for 5.1 ch and -7.5 dB for 2 ch is considered to be -6.0 dB. ε is 0.0 dB because both 5.1 ch and 2 ch require the same value. We selected -3.0 dB for ζ to prevent the sound power of LFE from exceeding the sum of the sound powers of LFE1 and LFE2 even when an identical signal is assigned to both LFE1 and LFE2.



Fig. 6 Acoustic evaluation room at NHK. This room strictly adheres to Rec. ITU-R BS.1116-1.

Table 4. Geometric parameters of loudspeaker arrangement.

	Top	Middle	Bottom
Height	2.9 m	1.4 m	0.3 m
Radius	2.5 m	2.5 m	2.5 m
Elevation angle	31°	0°	-23°

We carried out the subjective evaluation in an acoustic evaluation room at NHK (Fig. 6). The design of the room strictly adheres to Rec. ITU-R BS.1116-1 [24]. The reverberation time is 0.38 s at 500 Hz, the room dimensions are 6.4 m (W) × 8.0 m (D) × 4.5 m (H), and the room complies with NR-10.

230 m³

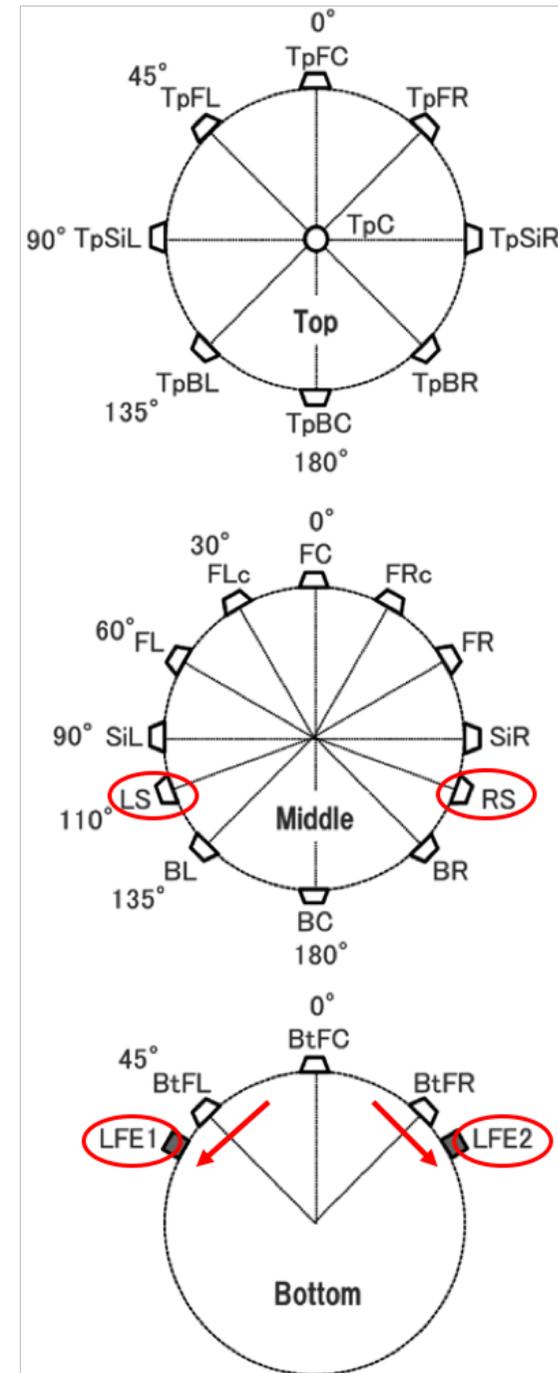


Fig. 7 Azimuthal angle of the loudspeaker arrangement in each layer.

Down-mix 22.2 vers 5.1 et stéréo 2.0 :

$$L = FL + \frac{1}{2^{3/4}} FLc + \frac{1}{2^{3/4}} SiL + TpFL + \frac{1}{2^{3/4}} TpSiL + BtFL$$

$$R = FR + \frac{1}{2^{3/4}} FRc + \frac{1}{2^{3/4}} SiR + TpFR + \frac{1}{2^{3/4}} TpSiR + BtFR$$

$$C = FC + \frac{1}{2^{3/4}} FLc + \frac{1}{2^{3/4}} FRc + TpFC + \frac{1}{2} TpC + BtFC$$

$$LFE = \frac{1}{\sqrt{2}} (LFE1 + LFE2)$$

$$LS = BL + \frac{1}{2^{3/4}} SiL + \frac{1}{\sqrt{2}} BC + TpBL + \frac{1}{\sqrt{2}} TpBC + \frac{1}{2^{3/4}} TpSiL + \frac{1}{2} TpC$$

$$RS = BR + \frac{1}{2^{3/4}} SiR + \frac{1}{\sqrt{2}} BC + TpBR + \frac{1}{\sqrt{2}} TpBC + \frac{1}{2^{3/4}} TpSiR + \frac{1}{2} TpC$$

$$\frac{1}{\sqrt{2}} = 0,7071 = -3 \text{ dB}$$

$$\frac{1}{2^{3/4}} = 0,5946 = -4,5 \text{ dB}$$

$$\frac{1}{2} = 0,5 = -6 \text{ dB}$$

Downmix 7.1.4 vers 5.1

$$L = L + (-3 \text{ dB}) \text{ Side L} + \text{Tp Front L}$$

$$R = R + (-3 \text{ dB}) \text{ Side R} + \text{Tp Front R}$$

$$C = C \text{ et } \text{Sub} = \text{Sub}$$

$$Ls = Ls + (-3 \text{ dB}) \text{ Side L} + \text{Tp Back L}$$

$$Rs = Rs + (-3 \text{ dB}) \text{ Side R} + \text{Tp Back R}$$

5.1.4 vers 5.1

$$L = L + \text{Tp Front L}$$

$$R = R + \text{Tp Front R}$$

$$C = C \text{ et } \text{Sub} = \text{Sub}$$

$$Ls = Ls + \text{Tp Back L}$$

$$Rs = Rs + \text{Tp Back R}$$

5.1 vers Stéréo

$$L = L + Ls + (-3 \text{ dB}) C + (-4,5 \text{ dB}) \text{Sub}$$

$$R = R + Rs + (-3 \text{ dB}) C + (-4,5 \text{ dB}) \text{Sub}$$



Car régie 22.2 NHK Japon

Object-based audio

The next big turn

Matthieu Parmentier



Matthieu Parmentier has recently been made cochair of the AES Technical Committee on Transmission and Broadcasting. He is France Television's research and development coordinator, as well as being chairman of the EBU strategic project "Future of Audio Formats and Renderers."

As a young broadcast sound engineer in the 1990s, I was told to think d.d.digital, I am now invited into the era of audio objects.

In the 1990s I was dreaming of 5.1 production on various TV shows, but the digital mixers I was in charge of setting up—while our good-old-analog consoles were sold to vintage studios—were all chosen for their advanced snapshots library. Then we were able to chain different programs with a simple snapshot recall during a commercial break, without spending hours dealing with a patchbay full of spaghetti. Immersive and natural audio was absolutely not a priority.

The ITU 5.1 multichannel format was a great opportunity to raise the quality of DVD soundtracks and develop home cinema sales. But what about TV 5.1 productions, beyond the broadcasting of movies? The 1998 FIFA World Cup was entirely mixed in surround sound, inspiring producers to rethink the sound of sport events, already looking for better immersion and emotion. Within a decade followed various

cultural programs, dramas, and premium documentaries.

Even if all the content that reasonably deserves immersive sound can today benefit from tools, methods, and experienced engineers to create it, this is not matched by the landscape around the immersed listener. Home cinemas are rarely correctly set up, even less connected to regular broadcast signals. The audience escapes on mobile networks, listens to over-the-top content that is easy to find and share, but highly compressed in terms of either dynamics and bitrates. The audience wears sharp earbuds or even expensive boom-bass headphones. All these gaps make multichannel sound broadcasting miss the target, despite the efforts of engineers. When will it be possible to achieve innovations that really educate ears and raise the sound quality of the listening experience?

HOW OBJECTS MATTER FOR THE FUTURE OF AUDIO

As an older broadcast sound engineer, living in the era of objects, I had to redefine

my priorities. While movie theaters may be considered as more or less "controlled environments," in terms of acoustics and quality of the reproduction chain, TV and radio consumers cope every day with a large variety of uncontrolled listening situations. Object-based audio systems brought new tools to cinema mixers for an even greater control of the reproduction, more and more adapted to each theater's constraints. For us broadcasters, object-based audio production opens minds, and naturally lets us dream of the same ability to adapt our content to the listening conditions. But when we consider the acoustical and technical uniqueness of each of our reproduction chains, the way to achieve perfect control of them is far far longer, and it is more realistic to think of personalizing the listening experience.

Thanks to audio objects it is first possible to enlarge accessibility services—providing real dialog enhancement, easier and cost-effective language substitution, simple add-on description commentaries, and even cleverer adaptation to environmental noise. Because

object-based broadcasting needs a renderer embedded in the receiver, this is also the first time that a broadcaster can control the receiver. By developing its own web players and other native apps for smartphones and tablets, any broadcaster can shape its user interface, offering services never seen before, and enter the world of big data, where it is even possible to personalize the design and functionalities of a player for each subscribed user. This is where broadcasters go with objects, adjusting the mix with respect to the original dynamic and the desired output level, matching the needs of hard-of-hearing listeners, describing video for blind people, raising the dialog level when the intelligibility declines. In a way we all experience these sorts of disabilities when attempting to listen on the move, or trying to learn a new language with foreign dramas.

INTRODUCING OBJECT-BASED BROADCASTING

To make this future real, broadcasters face a new bunch of challenges, involving new miking techniques, production workflows, delivery formats, and audio renderers. Each introduction of a new technology leads to a “chicken and egg” situation. Fifteen years ago, it was not difficult to convince a few sound engineers to start playing with more channels, and feel closer to any Hollywood mixer. But it took years to build the whole broadcast chain and allow 5.1 reception of live feeds, mostly never connected to any home cinema reserved for DVD/Blu-ray sound.

Far from multi-channel sound, a bit of DVD “midnight mode” and other Dynamic Range Controls, until the great success of loudness regulation, have shown the real appetite of consumers for what has been finally elected as “better audio.” To move to the next generation of audio systems, technically and commercially speaking, today the first challenge is probably the development of end-user receivers, offering the cleverer features described above. Before asking sound producers to mix and deliver programs differently, we need a little audience able to experience them. Indeed, this is possible thanks to nonlinear distribution of content, which first concerns premium programs such as dramas, documentaries, and cultural productions. With a noticeable benefit for end-users, the motivation of sound engineers usually overtakes their mistrust, and this is a perfect experimental field for web players, offering statistics and user feedback.

NEW TOOLS TO CREATE

The choice of the number and parameters of audio objects belong to the producer and the broadcast chain characteristics. With over a hundred objects



22.2-channel immersive audio studio at France TV, in which sound objects can be panned in three dimensions. Photo © France TV

making up a movie mix, object-based broadcasting certainly requires a pre-rendering of most of them, leading to a more reasonable number of objects such as dialog(s), music, ambiance, and commentary(ies). Hence, broadcast sound engineers need rendering units to control these objects, and then appreciate the result by simulating an end-user interface. Today various rendering engines exist, but they don't share any input standard, nor guidelines for basic process. These things are exactly the goals of the International Telecommunications Union standardization works concerning immersive audio. The recent draft ITU-R BS.2076 has been defining a standard for exchanging object-based audio sessions. It opens a large door for the spread of advanced audio content, builds new bridges between the cinema and broadcast fields, and gives a strong support for the definition of a baseline renderer in the broadcast context.

This is the very exciting period we enter now, when our tools and methods are redefined, when it is time to rethink every basic part of our job. In the Internet age, one of the key words is “disruption,” meaning that sometimes new technologies do not improve an industry, but simply create a shortcut and kill former economic models. With objects, the sound engineer of the near future is invited to directly serve a demanding audience. In this case of disruption, object-based broadcasting keeps a strategic seat for the sound engineer, that deserves our keen interest.

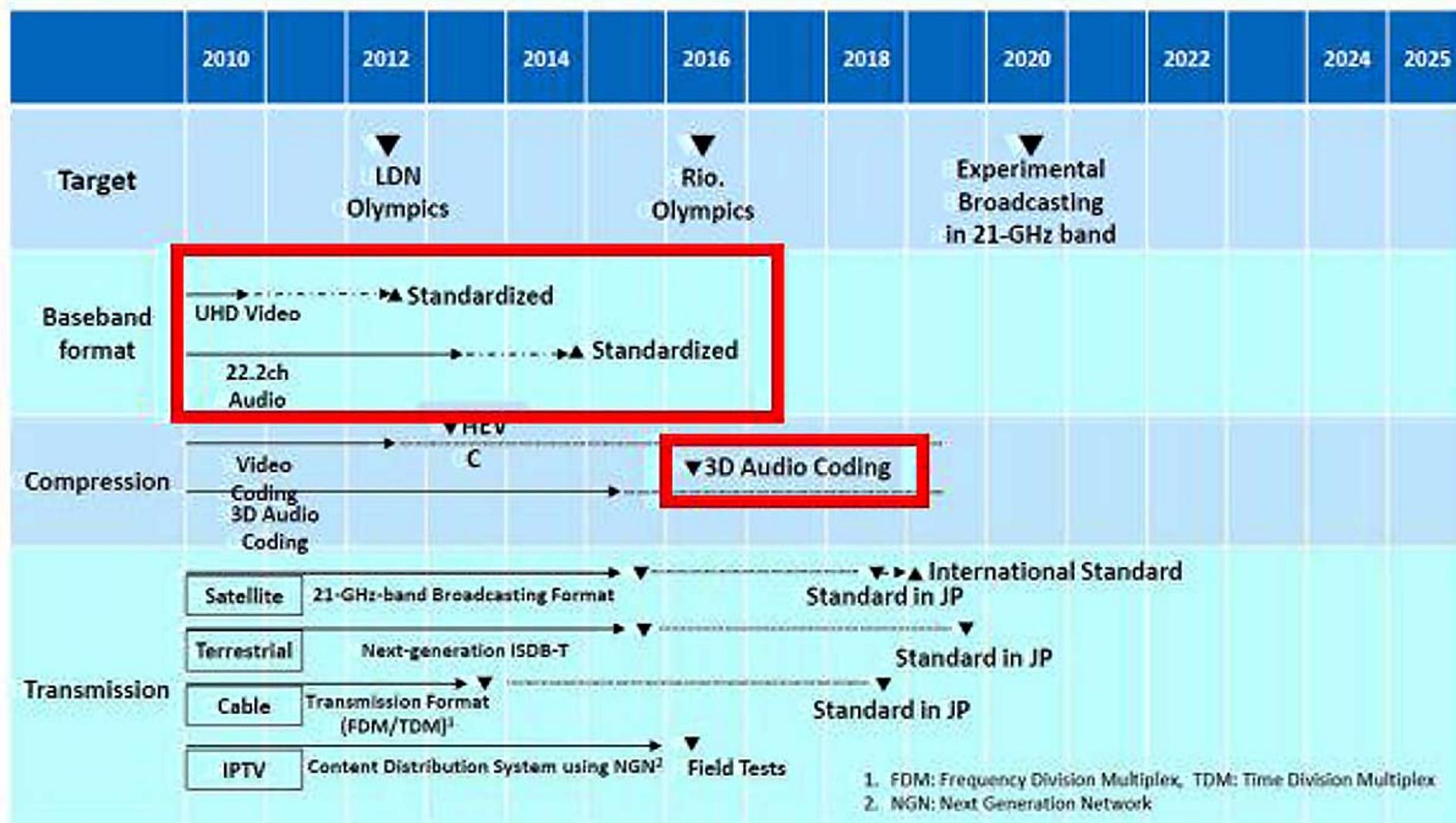


Studio interface for panning objects in 3D using IRCAM Spat. Photo © IRCAM



Cabine 22.2 France Télévision

Roadmap for Standardization



- [1] T. Hatada et al., "Psychophysical analysis of the sensation of reality induced by a visual wide-field display", SMPTE Journal, Vol. 89, pp. 560-569 (1980)
- [2] F. Okano, M. Kanazawa, K. Mitani, K. Hamasaki, M. Sugawara, M. Seino, A. Mochimaru, and K. Doi, "Ultrahigh-definition Television System with 4000 Scanning Lines", presented at NAB 2004, Las Vegas, USA (2004)
- [3] K. Hamasaki, K. Hiyama, T. Nishiguchi, and K. Ono., "Advanced multichannel audio systems with superior impression of presence and reality", AES 116th Convention, Berlin, Germany, Convention paper 6053 (2004)
- [4] K. Hamasaki, K. Hiyama, and R. Okumura, "The 22.2 multichannel sound system and its application", AES 118th Convention, Barcelona, Spain, Convention paper 6406 (2005 May 28-31)
- [5] C.E. Osgood, G.J. Suci, and P.H. Tannenbaum, "The Measurement of Meaning", University of Illinois Press, Urbana (1957)
- [6] K. Hamasaki, T. Nishiguchi, K. Hiyama, and R. Okumura, "Effectiveness of height information for reproducing presence and reality in multichannel audio system", presented at AES 120th Convention, Paris, France, Convention paper 6679, 2006 May 20-23.
- [7] K. Hamasaki, T. Nishiguchi, K. Hiyama, and R. Okumura, "Influence of picture on impression of three-dimensional multichannel sound", presented at AES 28th International Conference, Piteå, Sweden, 2006 June 30-July 2.

International Telecommunication Union

ITU-R
Radiocommunication Sector of ITU

Recommendation ITU-R BS.2051-1
(06/2017)

**Advanced sound system for
programme production**

BS Series
Broadcasting service (sound)



Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from <http://www.itu.int/ITU-R/go/patents/en> where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Recommendations

(Also available online at <http://www.itu.int/publ/R-REC/en>)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.

Electronic Publication
Geneva, 2017

© ITU 2017

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without written permission of ITU.

RECOMMENDATION ITU-R BS.2051-1

Advanced sound system for programme production

(Question ITU-R 135/6)

(2014-2017)

Scope

This Recommendation specifies an advanced sound system with and without accompanying picture, beyond systems specified in Recommendation ITU-R BS.775. An advanced sound system uses audio data in combination with an appropriate set of metadata to specify a sound scene to be delivered/broadcasted. The specifications include requirements for signalling the properties of advanced sound content and loudspeaker layout to be used in content production for advanced sound systems. The advanced sound system can apply to the sound component of television and expanded large screen digital imagery (LSDI) programmes, as well as for sound only programmes.

Keywords

Advanced sound system, channel-based sound system, object-based sound system, scene-based sound system, multichannel audio, loudspeaker layouts, immersive audio

The ITU Radiocommunication Assembly,

considering

- a)* that Recommendation ITU-R BS.775 – Multichannel stereophonic sound system with and without accompanying picture, specifies a multichannel stereophonic sound system with three front channels and two rear/side channels together with an optional low frequency effect (LFE) channel, as the highest level in a hierarchy of multichannel sound systems that range from 1/0 (monophonic) up to 3/2;
- b)* that Recommendation ITU-R BT.1769 – Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange, specifies an expanded hierarchy of LSDI image formats for production and international programme exchange;
- c)* that Recommendation ITU-R BT.709 – Parameter values for the HDTV standards for production and international programme exchange, specifies HDTV system image parameters;
- d)* that Recommendation ITU-R BT.2020 – Parameter values for ultra-high definition television systems for production and international programme exchange, specifies UHD TV image system parameters;
- e)* that the wide field of view image presentation of expanded-LSDI, HDTV, and ultra-high definition television (UHD TV) programmes benefit from spatially enhanced sound beyond the 5.1 channel sound;
- f)* that Recommendation ITU-R BS.1909 – Performance requirements for an advanced multichannel stereophonic sound system for use with or without accompanying picture, specifies the requirements for an advanced sound system with or without accompanying picture;
- g)* that Report ITU-R BS.2159 – Multichannel sound technology in home and broadcasting applications, includes the results of subjective evaluation experiments on loudspeaker layout to meet the requirements described in Recommendation ITU-R BS.1909,

recommends

- 1 that static or dynamic metadata/descriptors meeting the requirements described in Annex 1 should be used to signal the properties of all audio signals used in an advanced sound system in order to fully represent the desired audio content;
- 2 that the loudspeaker positions and configurations described in Annex 1 should be considered for production of advanced sound programmes;
- 3 that the appropriate number of audio elements¹ and loudspeaker layout configurations in programme production should be a choice by agreement between the producer and the recipient in the programme exchange;
- 4 that consumer interactions with the received audio should be facilitated in the production and broadcast sound system and that its characteristics should be a choice by agreement between the producer and the recipient in the programme exchange,

further recommends

- 1 that further work should be conducted to provide information about the features of any advanced sound system according to this Recommendation to fulfil the quality requirements of Recommendation ITU-R BS.1909;
- 2 that only new sound systems to be used in broadcasting content production should be added in Annex 1. New sound systems should be a clear expansion from the other sound systems already specified in Annex 1 and their common parts should be as compatible as possible. Such sound systems should be specified by the loudspeaker layout (positions and their ranges), the channel labels and their ordering.

Annex 1 **(normative)**

Advanced sound system for programme production

1 Introduction

The sound system specified in this Recommendation is defined as an advanced sound system which allows the metadata associated with each audio stream to be static or dynamic for the duration of a programme. This, for example, allows a programme to be represented by elements made by a combination of object signals and channel signals.

Channel signals are defined as audio signals that are mixed to a predefined number of channels and accompanied by metadata being static for the duration of a programme. Each of these channels is associated with a specific loudspeaker. The channel is reproduced by routing the channel to the associated loudspeaker, if present, or is routed to one or more available loudspeakers (e.g. via a

¹ An audio element is considered to be a signal with metadata which is either static for the duration of a programme or is dynamic. This enables the delivery of channel-based, object-based and scene-based content.

channel downmix) so as to best represent playback on the intended loudspeaker. The production workflows, broadcasting networks and reproduction systems are defined by a set of loudspeaker positions. Examples are systems according to Recommendation ITU-R BS.775.

Object signals are audio signals that, when accompanied by dynamic metadata for the duration of a programme, represent audio elements allowing a renderer to play back the audio objects in a way most appropriate to the playback system and listening environment. An object-based approach also may allow users to interact with the audio content.

Object-based elements and channel-based elements can be associated with each other or exist independently. To allow any combinations of channel or object-based elements, all signals should be accompanied by necessary metadata/descriptors, including time-independent (static) and/or time-dependent (dynamic) spatial position of the desired auditory event. These signals can be reproduced via a configuration of loudspeakers using a variety of rendering and/or mapping techniques.

Advanced sound programmes consist of the audio signals and the accompanying metadata.

Section 2 specifies requirements for metadata for sound content of the advanced sound system.

Section 3 describes loudspeaker layouts for the systems in production environments. Because a rendering or mapping process is needed for the reproduction of audio signals, the numbers and positions of loudspeakers are required to be well defined. This information enables rendering of the audio signals according to a predefined loudspeaker configuration in a reproduction scenario.

2 Requirements for metadata/descriptors for sound content of advanced sound system content

All audio files and streams used in an advanced sound system will require suitable metadata to accompany them. Unlike simple fixed channel-based systems where channel ordering is often enough to define the channels, the advanced system will need complete descriptions for all the audio elements used to ensure they are handled correctly. Therefore, a metadata model that has been standardized by a standards organization is required to provide consistent definitions for the audio. This model should have the following requirements:

- Contain all information required to reproduce/render a programme in all reproduction scenarios given by Recommendation ITU-R BS.1909 based on a single representation.
- Be able to describe the format of any channel-, object- and scene-based audio element.
- Be flexible enough to describe any combination of elements.
- The metadata items should be fully described such that they can be used by any renderer.
- Be specified in an open XML schema to allow the metadata to be represented in XML (as its primary method, it could of course be translated to other formats such as JSON).
- Be able to be added to an existing audio file format.
- To allow commonly used definitions (in particular existing channel-based configurations) to be open and freely accessible from a reference set of definitions.

3 Loudspeaker configurations for advanced sound system

Channel-based signals (including those in an advanced sound system) require a loudspeaker setup where the number and positions of the loudspeakers are well defined. Object-based signals can be reproduced via loudspeakers configured for channel-based signals or additional loudspeakers for other advanced rendering systems.

To ensure the loudspeaker configuration for the advanced multichannel sound system has a consistent definition, a set of parameters have been defined that specify each loudspeaker label, its position, and associated loudspeaker configurations as described in Table 1:

- SP Label: denoted by the initial of the layer name and the three-digit azimuth angle. ‘+/-SC’ indicates a loudspeaker pair at the left and right edge of the screen (see Attachment 2 to Annex 1). The centre of the screen should be at an azimuth of 0 degrees.
- Azimuth: the azimuth angle expressed in degrees, positive values rotate to the left when facing the front.
- Elevation: the elevation angle expressed in degrees, positive values go up from the horizontal plane.

TABLE 1

**List of possible loudspeaker positions for advanced sound system, identification of the loudspeaker layouts
in form of “Upper + Middle + Bottom loudspeakers”**

SP Label	Azimuth	Elevation	A	B	C	D	E	F	G	H	I	J
			0+2+0	0+5+0	2+5+0	4+5+0	4+5+1	3+7+0	4+9+0	9+10+3	0+7+0	4+7+0
M+000	0	0		X	X	X	X	X	X	X	X	X
M+022	+22.5	0										
M-022	-22.5	0										
M+SC	Left edge of display	0							X			
M-SC	Right edge of display	0							X			
M+030	+30	0	X	X	X	X	X	X	X	X	X	X
M-030	-30	0	X	X	X	X	X	X	X	X	X	X
M+045	+45	0										
M-045	-45	0										
M+060	+60	0								X		
M-060	-60	0								X		
M+090	+90	0						X	X	X	X	X
M-090	-90	0						X	X	X	X	X
M+110	+110	0		X	X	X	X					
M-110	-110	0		X	X	X	X					
M+135	+135	0						X	X	X	X	X
M-135	-135	0						X	X	X	X	X
M+180	+180	0								X		
U+000	0	+30								X		
U+022	+22.5	+30										
U-022	-22.5	+30										
U+030	+30	+30			X	X	X					
U-030	-30	+30			X	X	X					

TABLE 1 (*end*)

SP Label	Azimuth	Elevation	A	B	C	D	E	F	G	H	I	J
			0+2+0	0+5+0	2+5+0	4+5+0	4+5+1	3+7+0	4+9+0	9+10+3	0+7+0	4+7+0
B-135	-135	-30										
B+180	+180	-30										
LFE1	+45	-30		X	X	X	X	X	X	X	X	X
LFE2	-45	-30						X		X		

NOTE – The loudspeakers are assumed to be on a sphere. When this is not the case they should preferably be time aligned (at the central listening position) with an accuracy of 100 μ s.

The loudspeaker layouts shown in Table 1 are illustrated in Table 2.

TABLE 2

Loudspeaker layouts for advanced sound system

<p>Lower layer Includes LFE loudspeakers</p>	
<p>Middle layer Same level as the listener</p>	
<p>Upper layer Includes overhead loudspeaker (T+000)</p>	

3.1 Loudspeaker positions in production environments

As various conditions may need to be considered for the practical design of loudspeaker set-ups in production studios such as room size, accompanying picture or other constraints, loudspeaker positions may have some azimuth angle and elevation angle variations. To ensure proper adaptation and rendering of channel-based elements, the loudspeaker positions used in the particular production environment should be stored as part of the metadata according to § 2. When the content is transferred to a different reproduction system/location, it should be ensured that the programme is adapted if necessary so that all quality requirements of the advanced sound system are met. The quality requirements are specified in Recommendation ITU-R BS.1909.

However, loudspeakers should be placed within the sectors defined by the azimuth and elevation range as given in Tables 3 to 12 to lessen the sound quality variations by loudspeaker positions differences. A pair of loudspeakers with a nominal azimuth whose magnitude is greater than 45 degrees should have positions that are perfectly symmetrical or symmetrical within 10 degrees for azimuth and elevation. Other pairs of loudspeakers should be perfectly symmetrical. Regardless of any asymmetry, actual loudspeaker positions should still be within the range specified. The notations “a .. b” in the Tables should be taken to mean the smaller of the two sectors that might be obtained by rotating clockwise or anti-clockwise between angle “a” and angle “b”.

TABLE 3

Loudspeaker configuration for Sound system A (0+2+0)

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30	0
M-030	R	Right	-30	0

TABLE 4

**Loudspeaker configuration for Sound system B (0+5+0)
(from Recommendation ITU-R BS.775)**

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30	0
M-030	R	Right	-30	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	–	–
M+110	Ls	Left surround	+100 .. +120	0 .. +15
M-110	Rs	Right surround	-100 .. -120	0 .. +15

TABLE 5

Loudspeaker configuration for Sound system C (2+5+0)

SP Label	Channel		Loudspeaker location, Polar	
			Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30	0
M-030	R	Right	-30	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	–	–
M+110	Ls	Left surround	+100 .. +120	0 .. +15
M-110	Rs	Right surround	-100 .. -120	0 .. +15
U+030	Ltf	Left top front	+30 .. +45	+30 .. +55
U-030	Rtf	Right top front	-30 .. -45	+30 .. +55

TABLE 6

Loudspeaker configuration for Sound system D (4+5+0)

SP Label	Channel		Loudspeaker location, Polar	
			Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30	0
M-030	R	Right	-30	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	–	–
M+110	Ls	Left surround	+100 .. +120	0
M-110	Rs	Right surround	-100 .. -120	0
U+030	Ltf	Left top front	+30 .. +45	+30 .. +55
U-030	Rtf	Right top front	-30 .. -45	+30 .. +55
U+110	Ltr	Left top rear	+100 .. +135	+30 .. +55
U-110	Rtr	Right top rear	-100 .. -135	+30 .. +55

TABLE 7
Loudspeaker configuration for Sound system E (4+5+1)

SP Label	Channel		Loudspeaker location, Polar	
			Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30	0
M-030	R	Right	-30	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	–	–
M+110	Ls	Left surround	+100 .. +120	0
M-110	Rs	Right surround	-100 .. -120	0
U+030	Ltf	Left top front	+30 .. +45	+30 .. +55
U-030	Rtf	Right top front	-30 .. -45	+30 .. +55
U+110	Ltr	Left top rear	+100 .. +135	+30 .. +55
U-110	Rtr	Right top rear	-100 .. -135	+30 .. +55
B+000	Cbf	Centre bottom front	0	-15 .. -30

TABLE 8
Loudspeaker configuration for Sound system F (3+7+0)

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+000	C	Centre	0	0
M+030	L	Left	+30	0
M-030	R	Right	-30	0
U+045	LH	Left height	+30 .. +45	+30 .. +45
U-045	RH	Right height	-30 .. -45	+30 .. +45
M+090	LS	Left side	+60 .. +150	0
M-090	RS	Right side	-60 .. -150	0
M+135	LB	Left back	+60 .. +150	0
M-135	RB	Right back	-60 .. -150	0
UH+180	CH	Centre height	180	+45 .. +90
LFE1	LFE1	Left low frequency effects	+30 .. +90	-15 .. -30
LFE2	LFE2	Right low frequency effects	-30 .. -90	-15 .. -30

TABLE 9

Loudspeaker configuration for Sound system G (4+9+0)

SP Label	Channel		Loudspeaker location, Polar	
			Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30 .. +45	0
M-030	R	Right	-30 .. -45	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	–	–
M+090	Lss	Left side surround	+85 .. +110	0
M-090	Rss	Right side surround	-85 .. -110	0
M+135	Lrs	Left rear surround	+120 .. +150	0
M-135	Rrs	Right rear surround	-120 .. -150	0
U+045	Ltf	Left top front	+30 .. +45	+30 .. +55
U-045	Rtf	Right top front	-30 .. -45	+30 .. +55
U+135	Ltb	Left top back	+100 .. +150	+30 .. +55
U-135	Rtb	Right top back	-100 .. -150	+30 .. +55
M+SC	Lsc	Left screen	Left screen edge	0
M-SC	Rsc	Right screen	Right screen edge	0

The angle α between two surround loudspeakers on the same side (i.e. left or right) should be in the range $30^\circ \leq \alpha \leq 60^\circ$, for the middle layer.

TABLE 10

Loudspeaker configuration for Sound system H (9+10+3)

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+060	FL	Front left	+45 .. +60	0 .. +5
M-060	FR	Front right	-45 .. -60	0 .. +5
M+000	FC	Front centre	0	0 .. +5
LFE1	LFE1	Low frequency effects-1	+30 .. +90	-15 .. -30
M+135	BL	Back left	+110 .. +135	0 .. +15
M-135	BR	Back right	-110 .. -135	0 .. +15
M+030	FLc	Front left centre	+22.5 .. +30	0 .. +5
M-030	FRc	Front right centre	-22.5 .. -30	0 .. +5
M+180	BC	Back centre	+180	0 .. +15
LFE2	LFE2	Low frequency effects-2	-30 .. -90	-15 .. -30
M+090	SiL	Side left	+90	0 .. +15
M-090	SiR	Side right	-90	0 .. +15
U+045	TpFL	Top front left	+45 .. +60	+30 .. +45
U-045	TpFR	Top front right	-45 .. -60	+30 .. +45
U+000	TpFC	Top front centre	0	+30 .. +45
T+000	TpC	Top centre	-	+90
U+135	TpBL	Top back left	+110 .. +135	+30 .. +45
U-135	TpBR	Top back right	-110 .. -135	+30 .. +45
U+090	TpSiL	Top side left	+90	+30 .. +45
U-090	TpSiR	Top side right	-90	+30 .. +45
U+180	TpBC	Top back centre	+180	+30 .. +45
B+000	BtFC	Bottom front centre	0	-15 .. -30
B+045	BtFL	Bottom front left	+45 .. +60	-15 .. -30
B-045	BtFR	Bottom front right	-45 .. -60	-15 .. -30

TABLE 11

Loudspeaker configuration for Sound system I (0+7+0)

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30 .. +45	0
M-030	R	Right	-30 .. -45	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	–	–
M+090	Lss	Left side surround	+85 .. +110	0
M-090	Rss	Right side surround	-85 .. -110	0
M+135	Lrs	Left rear surround	+120 .. +150	0
M-135	Rrs	Right rear surround	-120 .. -150	0

The angle α between two surround loudspeakers on the same side (i.e. left or right) should be in the range $30^\circ \leq \alpha \leq 60^\circ$, for the middle layer.

TABLE 12

Loudspeaker configuration for Sound system J (4+7+0)

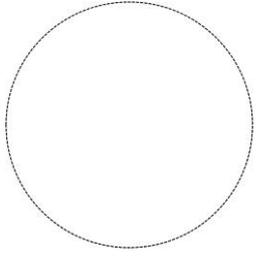
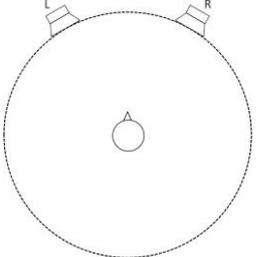
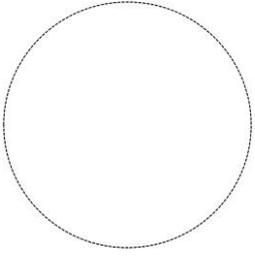
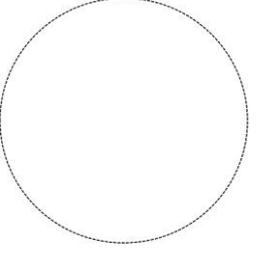
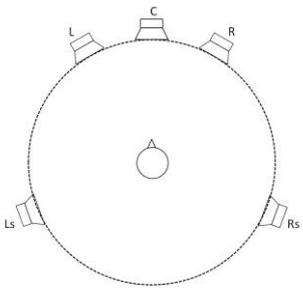
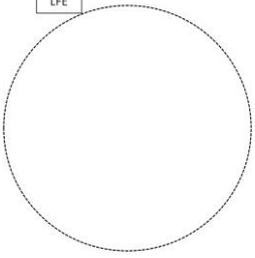
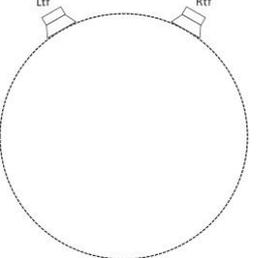
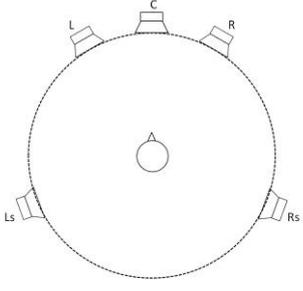
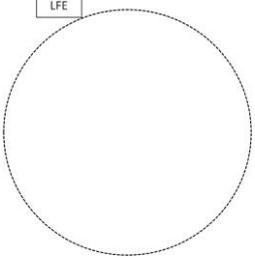
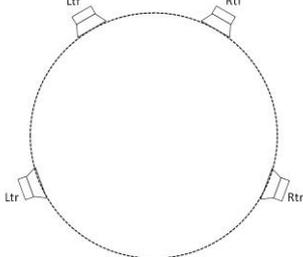
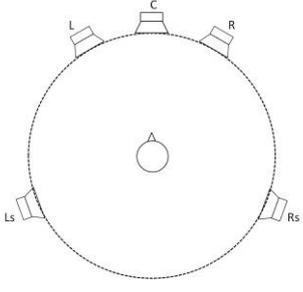
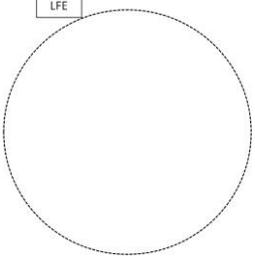
SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30 .. +45	0
M-030	R	Right	-30 .. -45	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	–	–
M+090	Lss	Left side surround	+85 .. +110	0
M-090	Rss	Right side surround	-85 .. -110	0
M+135	Lrs	Left rear surround	+120 .. +150	0
M-135	Rrs	Right rear surround	-120 .. -150	0
U+045	Ltf	Left top front	+30 .. +45	+30 .. +55
U-045	Rtf	Right top front	-30 .. -45	+30 .. +55
U+135	Ltb	Left top back	+100 .. +150	+30 .. +55
U-135	Rtb	Right top back	-100 .. -150	+30 .. +55

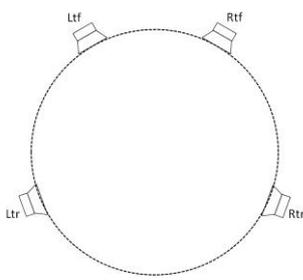
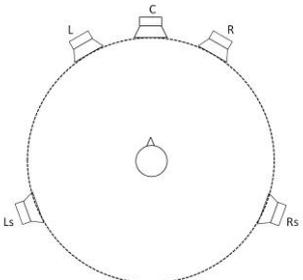
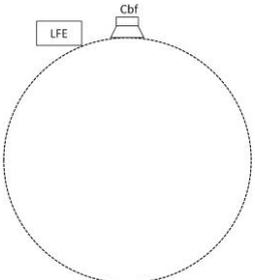
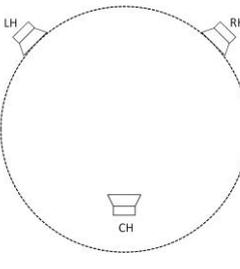
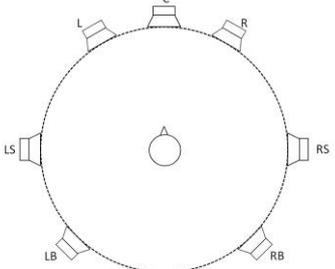
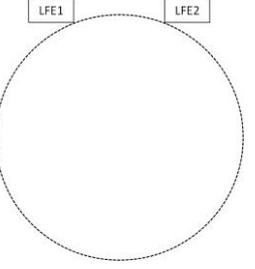
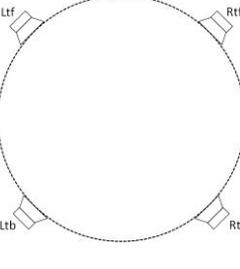
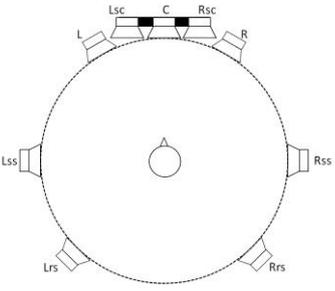
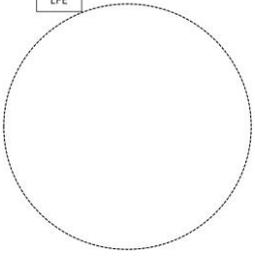
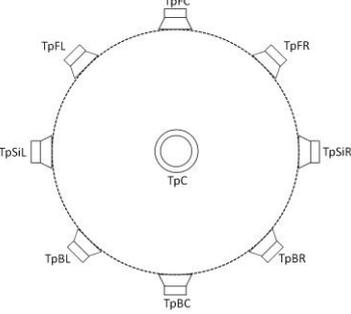
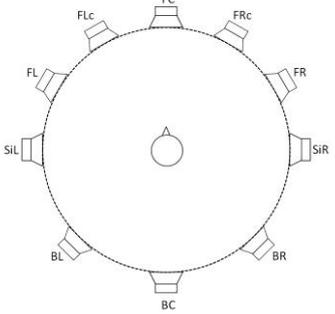
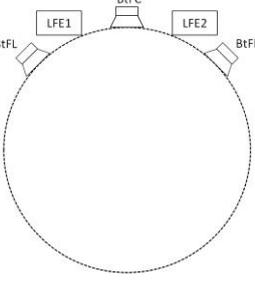
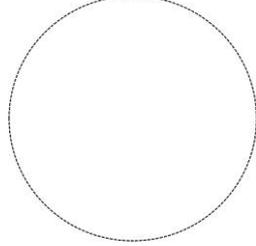
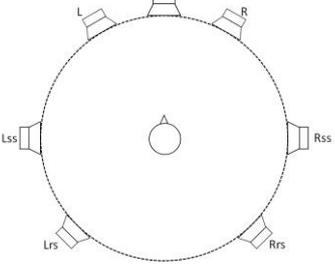
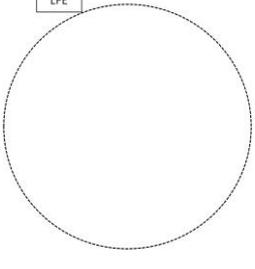
The angle α between two surround loudspeakers on the same side (i.e. left or right) should be in the range $30^\circ \leq \alpha \leq 60^\circ$, for the middle layer.

The loudspeaker layouts shown in Tables 3 to 12 are illustrated in Table 13.

TABLE 13

Loudspeaker layouts for advanced sound system

Sound system	Upper layer	Middle layer	Bottom layer
A (0+2+0)			
Upper layer 0/0/0 (Note)			
Middle layer 2/0/0			
Bottom layer 0/0/0			
B (0+5+0)			
Upper layer 0/0/0			
Middle layer 3/0/2			
Bottom layer 0/0/0.1			
C (2+5+0)			
Upper layer 2/0/0			
Middle layer 3/0/2			
Bottom layer 0/0/0.1			
D (4+5+0)			
Upper layer 2/0/2			
Middle layer 3/0/2			
Bottom layer 0/0/0.1			

Sound system	Upper layer	Middle layer	Bottom layer
E (4+5+1)			
Upper layer 0/0/0			
Middle layer 3/0/2			
Bottom layer 0/0/0.1			
F (3+7+0)			
Upper layer 2/0/1			
Middle layer 3/2/2			
Bottom layer 0/0/0.2			
G (4+9+0)			
Upper layer 2/0/2			
Middle layer 5/2/2			
Bottom layer 0.0.0.1			
H (9+10+3)			
Upper layer 3/3/3			
Middle layer 5/2/3			
Bottom layer 3/0/0.2			
I (0+7+0)			
Upper layer 0/0/0			
Middle layer 3/2/2			
Bottom layer 0.0.0.1			

Sound system	Upper layer	Middle layer	Bottom layer
J (4+7+0)			
Upper layer 2/0/2			
Middle layer 3/2/2			
Bottom layer 0.0.0.1			

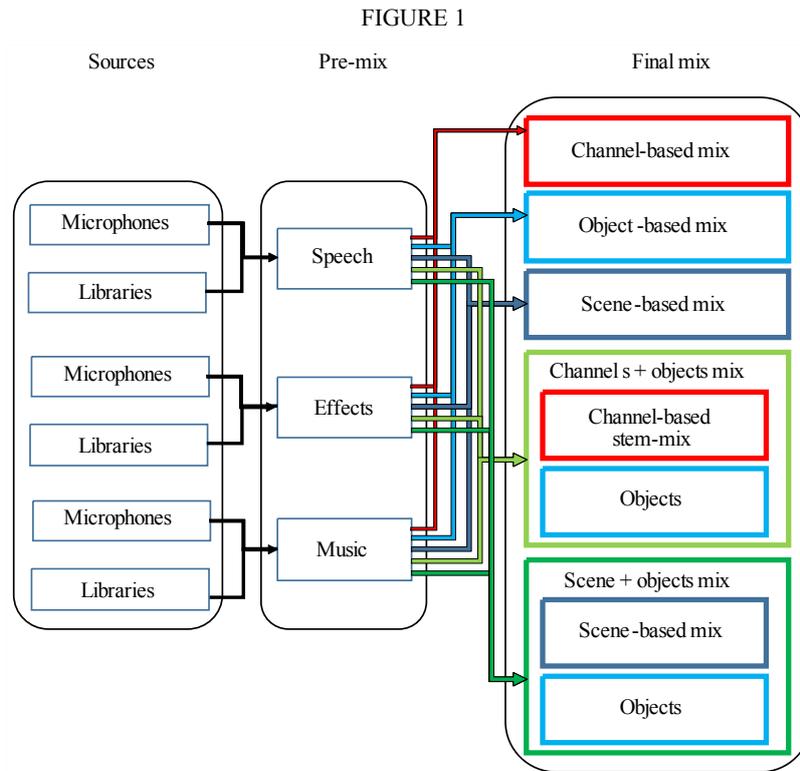
NOTE – X/Y/Z.LFE identifies the number of front/side/rear (plus LFE) loudspeakers in each layer.

Attachment 1 to Annex 1

Sound programme production and background of advanced sound system

In the process of mixing, a number of elements that depend on the nature of the production and signal acquisition process, typically determined by the audio engineer, are reduced to a representation of the creator intent, keeping only a reduced number of elements separate. The difference in production between a channel-based, object-based, scene-based and approaches which combine them is the method of mixing for the pre and final mixes. The channel-based approach means mixing all the elements into a predefined set of channels, whereas the hybrid channel-based + object-based approach allows the objects to be either mixed into channels or kept as separate objects. Similarly, in the hybrid scene-based + object-based approach the elements are either stored in a scene-based format (e.g. HOA) or kept as separate objects. In a pure object-based approach, all elements to recreate a certain experience are kept separate.

Existing production, live mixing and the post-production workflow of channel-based systems are the same as the hybrid model of channel-based stem-mix and objects. The delivery format of the final mix is different between the channel-based model and the hybrid model which delivers channel-based stem-mix signals and object signals.



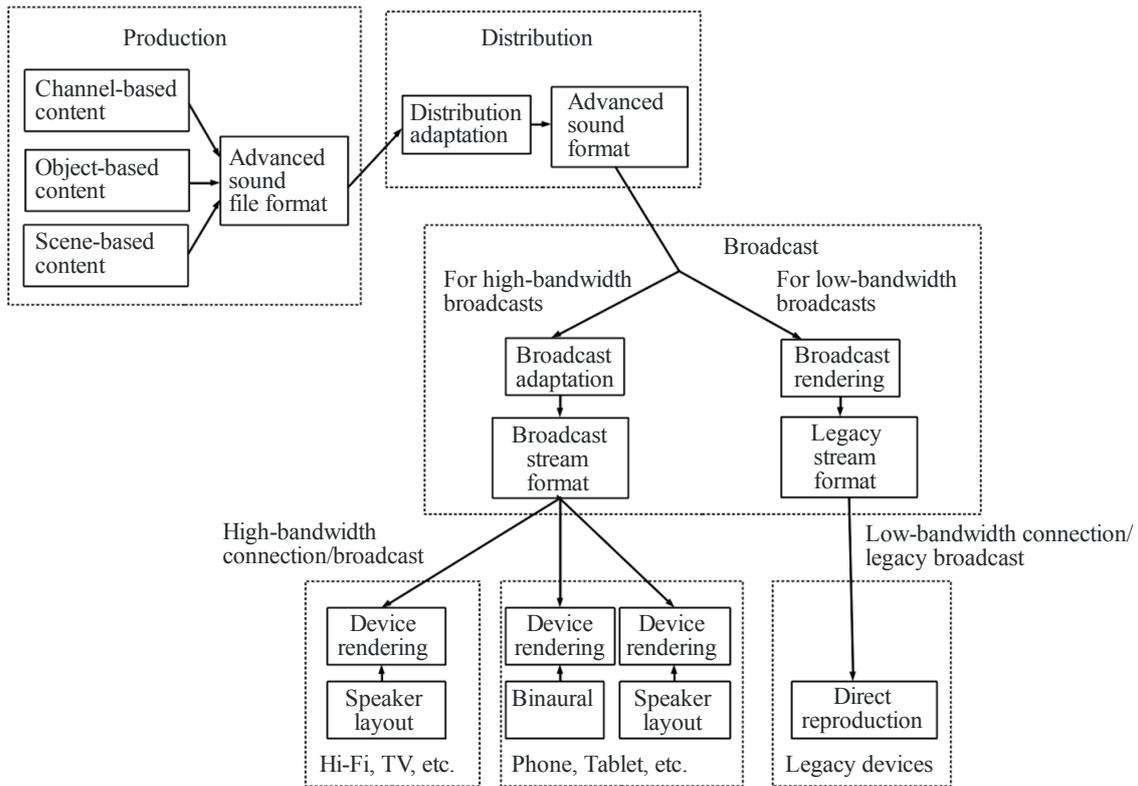
BS.2051-0 1

Figure 2 shows how an example of a typical broadcast chain could look for an advanced sound system. The production can use any type of audio source and content, but it should be fully annotated with the correct metadata to describe the audio signals, and this is stored in a file format that can support this metadata.

The distribution stage will adapt the representations from production into a more compact representation which will retain or generate new metadata to allow further rendering. The distribution file or stream is then passed to the broadcast stage which will render it towards a particular broadcast format. Higher bandwidth broadcasts will allow to deliver and rendering many object and channels, whereas low bandwidth broadcasts may have to render down to more traditional stereo formats. The broadcast format should retain as much of the metadata as required for the receiver end.

Each receiver device will have its own renderer designed for the loudspeaker layouts possible for that device. For example, a Hi-fi will need something very flexible to allow multiple loudspeakers to have different placements, whereas a TV will have its internal loudspeakers fixed in known positions. Future content delivery will potentially receive a representation, which will maintain full flexibility for user interaction or personalization.

FIGURE 2



BS.2051-02

Attachment 2 to Annex 1 (informative)

Audio – Visual Spatial Alignment

As described in Recommendation ITU-R BS.775, image displays for television viewing vary in size, and are frequently narrower than the defined angular spacing between the main left and right loudspeakers (e.g. M+030 and M-030). As a result of this inconsistent relationship between sound and picture display width, audio and visual images cannot be reliably aligned.

Object-based audio systems can overcome this issue by describing the object location relative to the screen. Using metadata that describes the screen location, an object renderer can appropriately render screen-referenced objects to the available loudspeakers such that the audio and associated visual elements are spatially aligned.

A similar capability in the horizontal dimension can be achieved with a pure channel-based audio program if screen-based audio elements are authored and distributed using a screen channel pair (Lsc and Rsc). In principle, the screen channels are played back using a loudspeaker pair at the left and right edge of the screen (M+SC and M-SC). In practice, the screen channel pair (Lsc and Rsc) can be

rendered using existing loudspeakers (e.g. M+030, M+000, and M-030): if the screen is large and spans the space between M+030 and M-030, the Lsc and Rsc channels can be summed into the L and R channels for playback; if the screen is smaller, as is often the case in domestic use, the Lsc and Rsc channels can be panned appropriately (depending on screen size) between the L and C, and C and R channels, respectively, and then routed to the appropriate loudspeakers. While a domestic system would often not include screen loudspeakers, a production environment may choose to include screen loudspeakers for screen channel playback during content creation and validation.

Union internationale des télécommunications

UIT-R

Secteur des Radiocommunications de l'UIT

Recommandation UIT-R BS.1116-3
(02/2015)

Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio

Série BS
Service de radiodiffusion sonore



Union
internationale des
télécommunications

Avant-propos

Le rôle du Secteur des radiocommunications est d'assurer l'utilisation rationnelle, équitable, efficace et économique du spectre radioélectrique par tous les services de radiocommunication, y compris les services par satellite, et de procéder à des études pour toutes les gammes de fréquences, à partir desquelles les Recommandations seront élaborées et adoptées.

Les fonctions réglementaires et politiques du Secteur des radiocommunications sont remplies par les Conférences mondiales et régionales des radiocommunications et par les Assemblées des radiocommunications assistées par les Commissions d'études.

Politique en matière de droits de propriété intellectuelle (IPR)

La politique de l'UIT-R en matière de droits de propriété intellectuelle est décrite dans la «Politique commune de l'UIT-T, l'UIT-R, l'ISO et la CEI en matière de brevets», dont il est question dans l'Annexe 1 de la Résolution UIT-R 1. Les formulaires que les titulaires de brevets doivent utiliser pour soumettre les déclarations de brevet et d'octroi de licence sont accessibles à l'adresse <http://www.itu.int/ITU-R/go/patents/fr>, où l'on trouvera également les Lignes directrices pour la mise en oeuvre de la politique commune en matière de brevets de l'UIT-T, l'UIT-R, l'ISO et la CEI et la base de données en matière de brevets de l'UIT-R.

Séries des Recommandations UIT-R

(Egalement disponible en ligne: <http://www.itu.int/publ/R-REC/fr>)

Séries	Titre
BO	Diffusion par satellite
BR	Enregistrement pour la production, l'archivage et la diffusion; films pour la télévision
BS	Service de radiodiffusion sonore
BT	Service de radiodiffusion télévisuelle
F	Service fixe
M	Services mobile, de radiorepérage et d'amateur y compris les services par satellite associés
P	Propagation des ondes radioélectriques
RA	Radio astronomie
RS	Systèmes de télédétection
S	Service fixe par satellite
SA	Applications spatiales et météorologie
SF	Partage des fréquences et coordination entre les systèmes du service fixe par satellite et du service fixe
SM	Gestion du spectre
SNG	Reportage d'actualités par satellite
TF	Emissions de fréquences étalon et de signaux horaires
V	Vocabulaire et sujets associés

Note: Cette Recommandation UIT-R a été approuvée en anglais aux termes de la procédure détaillée dans la Résolution UIT-R 1.

Publication électronique
Genève, 2016

© UIT 2016

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, par quelque procédé que ce soit, sans l'accord écrit préalable de l'UIT.

RECOMMANDATION UIT-R BS.1116-3*, **

Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio

(Question UIT-R 62/6)

(1994-1997-2014-2015)

Domaine d'application

La présente Recommandation est destinée à être utilisée pour évaluer des systèmes qui introduisent des dégradations si infimes qu'elles ne sont pas décelables sans contrôle rigoureux des conditions d'expérimentation et sans analyse statistique appropriée. Si elle était appliquée à des systèmes qui introduisent des dégradations relativement importantes et aisément décelables, elle entraînerait une dépense excessive de temps et d'effort et pourrait donner des résultats moins fiables que ceux obtenus avec un essai plus simple. La présente Recommandation constitue la référence de base pour les autres Recommandations, qui, elles, peuvent introduire des conditions particulières supplémentaires ou des spécifications plus souples que celles définies dans la présente Recommandation.

Mots clés

Qualité audio; dégradations faibles; évaluation subjective; essais d'écoute; codage audio; signaux audio de haute qualité; local d'écoute.

L'Assemblée des radiocommunications de l'UIT,

considérant

- a) que les Recommandations UIT-R BT.500, UIT-R BS.562, UIT-R BT.710 et UIT-R BT.811 ont défini des méthodes d'évaluation subjective de la qualité des systèmes audio et vidéo;
- b) que les essais d'écoute subjectifs permettent d'évaluer le degré de gêne qu'inflige à l'auditeur toute dégradation que subit le signal utile au cours de sa transmission de son origine jusqu'à l'auditeur;
- c) que les méthodes objectives classiques ne conviennent pas forcément pour évaluer les schémas de codage audio perfectionnés et que l'on élabore actuellement des méthodes d'évaluation objectives perceptuelles pour vérifier la qualité de son des systèmes sonores;
- d) qu'il est important d'utiliser des méthodes normalisées pour l'échange, la compatibilité et l'évaluation correcte des résultats des essais;
- e) que la mise en place de nouveaux systèmes audio numériques perfectionnés qui tirent parti des propriétés psychoacoustiques exige, notamment dans le cas des dégradations faibles, que les méthodes d'évaluation subjective progressent;
- f) que l'introduction de systèmes à son stéréophonique multivoies pouvant comprendre jusqu'à 3/2 voies définis dans la Recommandation UIT-R BS.775 et de systèmes sonores évolués

* Cette Recommandation doit être portée à l'attention du Groupe ad hoc audio ISO/MPEG (Organisation internationale de normalisation/ Moving Picture Experts Group).

** La Commission d'études 6 des radiocommunications a apporté des modifications de forme à cette Recommandation en juillet 2015 conformément à la Résolution UIT R 1.

définis dans la Recommandation UIT-R BS.2051, avec ou sans image associée, exige de nouvelles méthodes d'évaluation subjective y compris les conditions expérimentales,

recommande

1 que les procédures d'essai, d'évaluation et de présentation que décrit l'Annexe 1 servent à l'évaluation subjective des dégradations faibles des systèmes audio, y compris les systèmes sonores multivoies (avec ou sans image),

recommande en outre

1 de poursuivre les études sur les caractéristique des locaux d'écoute et des dispositifs de reproduction pour les systèmes sonores évolués et de mettre à jour la présente Recommandation une fois ces études achevées.

Annexe 1

1 Généralités

1.1 Table des matières

La présente Annexe comprend 11 parties qui indiquent en détail les conditions applicables aux divers aspects des essais:

- 1 Généralités
- 2 Conception des expériences
- 3 Choix des groupes d'auditeurs
- 4 Méthode d'essai
- 5 Caractéristiques
- 6 Eléments de programme
- 7 Dispositifs de reproduction
- 8 Conditions d'écoute
- 9 Analyse statistique
- 10 Présentation des résultats des analyses statistiques
- 11 Contenu des rapports d'essai.

Par ailleurs, des directives concernant le choix des auditeurs experts figurent dans des pièces jointes qui contiennent en outre un exemple des consignes données aux participants.

Un certain nombre de mots d'usage courant sont utilisés ici dans un sens technique. On trouvera un glossaire de ces termes dans la Pièce jointe 4.

2 Conception des expériences

Si l'on veut rassembler des informations fiables dans un domaine d'intérêt scientifique, on dispose d'un grand nombre de stratégies de recherche différentes. Pour l'évaluation subjective des dégradations faibles dans les systèmes audio, il faut recourir aux méthodes expérimentales les plus

strictes. Les expériences subjectives se caractérisent tout d'abord par une maîtrise réelle des conditions expérimentales et ensuite par les données quantitatives fournies par des observateurs humains.

Il faut définir et organiser avec soin les expériences pour assurer qu'aucun facteur non contrôlé n'affecte les essais d'écoute, ce qui créerait des ambiguïtés. Si, par exemple, la séquence d'éléments sonores utilisée est la même pour tous les participants à un essai d'écoute, on ne peut avoir la certitude que les jugements qu'ils ont portés ne tiennent pas davantage à la séquence qu'aux divers niveaux de dégradation présentés. Il faut donc que les conditions d'essai soient telles qu'elles ne fassent apparaître que les effets des facteurs indépendants.

Au cas où il serait probable que les dégradations éventuelles et d'autres caractéristiques soient distribuées de façon homogène tout au long de l'essai d'écoute, il faudrait que la présentation des conditions d'essai soit rendue tout à fait aléatoire.

S'il peut y avoir non-homogénéité, il faudra en tenir compte dans la façon de présenter les conditions d'essai. Par exemple, si les éléments à évaluer sont de difficulté variable, il faudra présenter les stimulus dans un ordre aléatoire, au cours d'une séance ou d'une séance à l'autre.

De même, on organisera les essais d'écoute de façon que les participants ne soient pas fatigués au point que leurs appréciations soient moins précises. Sauf si la correspondance entre son et image est importante, il est préférable que l'évaluation des systèmes audio s'effectue sans les images associées.

Il est très important d'inclure des contrôles appropriés. Il s'agit généralement d'insérer des éléments audio non dégradés sans que les participants puissent les déceler. Les différences entre l'appréciation de ces stimulus de contrôle et de ceux qui peuvent être dégradés permettent de conclure que les notes sont bien des évaluations des dégradations.

On reviendra par la suite sur ces considérations. Il ne faut pas oublier que la conception et la réalisation des expériences mais aussi l'analyse statistique sont des questions complexes et qu'une Recommandation comme celle-ci ne saurait donner que des directives très générales. Il est recommandé de faire appel à des spécialistes de la conception des expériences et des statistiques avant de commencer à organiser des essais d'écoute.

3 Choix des groupes d'auditeurs

3.1 Auditeurs experts

Il est essentiel que les données issues des essais d'écoute servant à évaluer les dégradations faibles dans les systèmes audio proviennent exclusivement de participants ayant reçu une formation pour détecter de tels défauts. La présence d'auditeurs experts est d'autant plus nécessaire que la qualité des systèmes étudiés est élevée.

3.2 Critères de choix des participants

Le résultat des essais subjectifs effectués avec des groupes d'auditeurs sélectionnés qui jugent des systèmes audio présentant de faibles dégradations ne se prête pas avant tout à une extrapolation au public en général. L'objectif visé consiste en principe à rechercher si un groupe d'auditeurs experts est capable, dans certaines conditions, de déceler des dégradations relativement subtiles et à fournir aussi une estimation quantitative des dégradations introduites. Cette procédure d'essai doit être très élaborée de façon à mettre en évidence les problèmes susceptibles de se poser à long terme dans les diverses conditions qui apparaissent en réalité dès lors qu'un système a été proposé à l'utilisateur.

Il faut parfois introduire une technique de rejet avant ou après l'essai réel (pré ou postsélection) ou encore recourir à l'une et à l'autre. L'élimination consiste ici en un processus où l'on néglige tous les jugements portés par tel ou tel participant.

Tout type de technique de rejet qui ne serait ni analysé ni mis en œuvre avec soin risque de fausser les résultats. Lorsque des données ont été éliminées, il est donc très important que le rapport relatif à l'essai décrive clairement le critère utilisé afin que le lecteur puisse se forger sa propre opinion.

3.2.1 Présélection des participants

Les procédures de présélection reposent sur des méthodes telles que les essais audiométriques, le choix des participants en fonction de leur expérience acquise et de leur compétence lors d'essais précédents, et l'élimination de certains d'eux d'après l'analyse statistique des essais préliminaires. La présélection peut aussi se fonder sur la procédure de formation.

La principale raison en faveur de l'introduction d'une technique de présélection est le souci d'avoir des essais d'écoute efficaces, encore qu'il faille tenir compte du risque d'une limitation trop importante de la validité des résultats.

3.2.2 Postsélection des participants

On peut, dans l'ensemble, distinguer au moins deux catégories de méthodes de postsélection; l'une est fondée sur les incohérences par rapport au résultat moyen, l'autre sur la faculté du participant d'effectuer les identifications correctes. La première n'est jamais légitime. Quand un essai d'écoute subjectif est effectué avec la méthode d'essai recommandée ici, l'information nécessaire à la seconde méthode de postsélection est automatiquement disponible. L'Appendice 1 décrit une méthode statistique suggérée à cet effet.

Les méthodes servent essentiellement à éliminer les participants qui ne peuvent faire les distinctions appropriées. L'application d'une méthode de postsélection peut aider à dégager les tendances dans les résultats d'essai. Il faut toutefois être prudent en raison des différences de sensibilité des participants aux divers défauts.

3.3 Effectif d'un groupe d'écoute

On peut prévoir l'effectif approprié d'un groupe d'écoute à condition de pouvoir estimer la variance et de connaître la résolution requise de l'expérience.

Lorsque les conditions d'un essai d'écoute sont bien maîtrisées tant du point de vue technique que psychologique, l'expérience a montré que les données fournies par 20 participants suffisent souvent pour tirer de l'essai des conclusions appropriées. Si l'analyse est effectuée au fur et à mesure des essais, il n'est pas nécessaire de tenir compte d'un nombre plus important de participants dès que l'on a obtenu un degré de précision statistique suffisant pour tirer les conclusions appropriées.

Si l'on prévoit que certains des systèmes à l'essai seront quasiment transparents, il faudra retenir un plus grand nombre de participants pour avoir la certitude qu'ils seront assez nombreux à passer avec succès le test de postsélection.

Si, pour une raison quelconque, on ne maîtrise pas bien les conditions de l'expérience, il faudra davantage de participants pour arriver à la résolution voulue.

L'effectif du groupe d'écoute ne dépend pas uniquement de la résolution désirée. Le résultat des expériences du type considéré dans la présente Recommandation n'est, en pratique, valable que pour le groupe d'auditeurs experts ayant participé à l'essai. En augmentant l'effectif du groupe, on peut donc estimer que le résultat serait valable pour un groupe plus général d'auditeurs experts et serait alors jugé plus convaincant. On peut aussi avoir besoin d'augmenter l'effectif pour qu'il y ait plus de chances que la sensibilité des divers participants aux différents défauts varie.

4 Méthode d'essai

Pour effectuer des évaluations subjectives dans le cas de systèmes qui présentent de faibles dégradations, il faut choisir une méthode appropriée. On estime que la méthode de «doublement aveugle à triple stimulus et référence dissimulée» est particulièrement sensible, stable et permet de détecter avec précision les faibles dégradations. Il faudra donc l'utiliser pour ce type d'essai.

Sous la forme préférée et la plus sensible de cette méthode, il n'y a qu'un participant à la fois et il est libre de choisir un stimulus parmi trois («A», «B», «C»). La référence connue est toujours présente comme stimulus «A». La référence dissimulée et l'objet de l'essai sont disponibles simultanément mais sont attribués au hasard des essais aux stimulus «B» et «C».

Le participant est prié d'évaluer la dégradation de «B» comparé à «A» et de «C» comparé à «A» selon l'échelle continue de dégradation à 5 notes. L'un des stimulus, «B» ou «C», ne devrait pas se distinguer du stimulus «A»; l'autre peut présenter une dégradation. Toute différence décelée entre la référence et les autres stimulus doit être interprétée comme une dégradation.

Dans cette méthode préférée, dès que le participant a fini de noter l'essai, il doit être possible de passer à l'essai suivant. On peut répéter l'extrait jusqu'à ce que le participant ait fourni son évaluation. C'est donc lui qui définit le rythme de la procédure d'essai.

L'échelle d'évaluation sera traitée de façon continue compte tenu des «repères» de l'échelle de dégradation à 5 notes (voir le Tableau 1) tirée de la Recommandation UIT-R BS.1284.

TABLEAU 1

Dégradation	Note
Imperceptible	5,0
Perceptible mais non gênant	4,0
Légèrement gênant	3,0
Gênant	2,0
Très gênant	1,0

NOTE 1 – On a montré que l'utilisation de points de repère intermédiaires prédéfinis pouvait introduire des distorsions [Poulton, 1992]. Il est possible d'utiliser l'échelle de chiffres sans la description des points de repère; dans ce cas, l'orientation voulue de l'échelle doit être indiquée. Cette façon de procéder peut permettre de surmonter les problèmes de traduction rencontrés lors de la comparaison d'essais effectués dans des langues différentes.

Si on n'utilise pas des points de repère intermédiaires, il est indispensable que les résultats de chaque participant soient normalisés par rapport à un écart moyen type. On peut appliquer l'équation suivante pour normaliser les résultats tout en conservant l'échelle initiale:

$$Z_i = \frac{(x_i - x_{si})}{s_{si}} \cdot s_s + x_s$$

où:

Z_i : résultat normalisé

x_i : note donnée par le résultat du participant i

x_{si} : note moyenne du participant i pendant la séance s

x_s : note moyenne de tous les participants pendant la séance s

s_s : écart type pour tous les participants pendant la séance s

s_{si} : écart type pour le participant i pendant la séance s .

L'utilisation d'échelles sans points de repère intermédiaires empêche également toute interprétation des résultats en termes absolus.

Il est recommandé d'utiliser cette échelle avec une précision de la première décimale.

La méthode d'essai comprend deux parties: la phase de familiarisation ou d'entraînement et la phase de notation.

4.1 Phase de familiarisation ou d'entraînement

Avant de passer à la notation proprement dite, il faut laisser aux participants le temps de bien se familiariser avec le dispositif d'essai, l'environnement, la méthode et les échelles de notation et leur utilisation. Ils doivent aussi avoir une bonne connaissance des défauts à étudier. Pour les essais les plus délicats, il faut leur présenter auparavant tous les extraits qu'ils auront à noter pendant les séances suivantes consacrées à la notation. Au cours de la familiarisation ou de l'entraînement, il est préférable que les participants soient groupés (par trois, par exemple) pour qu'ils puissent communiquer aisément et échanger leurs impressions sur les défauts qu'ils détectent.

La Pièce jointe 3, «Exemple d'instructions destinées aux auditeurs», donne, à titre d'exemple, un jeu d'instructions avec une description de la technique de «doublement aveugle à triple stimulus et référence dissimulée» de présentation du stimulus. Une familiarisation bien menée peut convertir des participants initialement peu doués en des experts pour les besoins de l'essai. A l'issue de la familiarisation, les participants doivent avoir acquis une connaissance précise de l'échelle qu'ils auront à utiliser dans la phase de notation qui succédera à la familiarisation ou à l'entraînement.

4.2 Phase de notation

Au début de la première séance réelle de notation de la journée, il faut présenter oralement à chaque participant les instructions pour les essais, complétés de préférence par des documents écrits. On peut effectuer à titre d'exemple plusieurs comparaisons juste avant le début des présentations pour la notation.

Comme la mémoire auditive à long et moyen terme n'est pas fiable, la procédure d'essai ne devra faire appel qu'à la mémoire à court terme. On y parvient au mieux avec une méthode de commutation quasi instantanée (voir la Note 1) associée au système à triple stimulus que décrit la Pièce jointe 3. Ce type de commutation exige une coordination temporelle très précise des stimulus.

NOTE 1 – Une commutation rigoureusement instantanée peut créer des défauts si les signaux des stimulus successifs ne sont pas les mêmes. On préférera une commutation quasi instantanée avec, par exemple, un délai de 40 ms en cas de disparition/changement/réapparition.

Dans les évaluations les plus critiques, un seul participant intervient à la fois. Ce n'est qu'ainsi qu'il peut, en toute liberté, passer d'un stimulus à l'autre dans la méthode à trois stimulus. Il est essentiel qu'il ait cette faculté pour explorer librement toutes les comparaisons entre les stimulus de chaque essai.

Il est préférable qu'il puisse effectuer les commutations entre stimuli sans information visuelle, les yeux fermés s'il le veut, pour mieux se concentrer et être moins distrait. Il ne faut pas que la commutation introduise le moindre bruit parasite (comme des «clics»), car ce genre de parasite risque de perturber gravement le processus d'évaluation.

Une séance de notation ne devrait pas durer plus de 20 à 30 min mais si, comme cela a été préconisé, le rythme des essais est laissé à la discrétion des participants, la durée variera de façon arbitraire de l'un à l'autre. D'après l'expérience acquise, il semble qu'il ne faut pas prévoir plus de 10 à 15 essais

par séance pour obtenir la durée voulue. La fatigue des participants est de nature à affecter sérieusement la validité des jugements portés. Pour l'éviter, il sera prévu pour chaque participant et entre les séances successives, un repos de durée égale ou inférieure à celle de la séance.

5 Caractéristiques

Les caractéristiques spécifiques des évaluations des systèmes sonores évolués et des systèmes monophoniques, stéréophoniques à deux voies et stéréophoniques multivoies (c'est-à-dire jusqu'à 3/2 voies) sont énumérées ci-après. On préfère évaluer, dans chaque cas, la caractéristique «qualité audio de base». Les expérimentateurs peuvent choisir de définir et d'évaluer d'autres caractéristiques.

Si on demande aux participants d'essayer d'évaluer plus d'une caractéristique au cours de chaque essai, il se peut qu'ils aient du mal à répondre. S'ils se sentent perdus ou débordés du fait qu'ils tentent de répondre à plusieurs questions à propos d'un stimulus donné, il peut en résulter des notes peu fiables pour toutes ces questions.

5.1 Système monophonique

Qualité audio de base

- Cette caractéristique unique et globale sert à estimer toute différence décelée entre la référence et l'objet de l'essai.

5.2 Système stéréophonique à deux voies

Qualité audio de base

- Cette caractéristique unique et globale sert à estimer toute différence décelée entre la référence et l'objet de l'essai.

La caractéristique supplémentaire qui suit peut aussi être intéressante:

Qualité de l'image stéréophonique

- Cette caractéristique correspond aux différences entre la référence et l'objet de l'essai du point de vue de l'emplacement des images sonores, de l'impression de profondeur et de présence de l'événement audio.

Bien que des études aient montré que la qualité de l'image stéréophonique peut se dégrader, la question de savoir s'il est justifié de noter séparément la qualité de l'image stéréophonique et la qualité audio de base n'a pas fait l'objet de recherches suffisantes.

NOTE 1 – Jusqu'en 1993, pour la plupart des études d'évaluation subjective des dégradations faibles, on s'est servi uniquement de la caractéristique qualité audio de base. La caractéristique qualité de l'image stéréophonique était donc implicitement ou explicitement incluse dans la qualité audio de base en tant que caractéristique globale dans ces études.

5.3 Système stéréophonique multivoies

Qualité audio de base

- Cette caractéristique unique et globale sert à estimer toute différence décelée entre la référence et l'objet de l'essai.

La caractéristique supplémentaire qui suit peut aussi être intéressante:

Qualité frontale de l'image

- Cette caractéristique correspond à la localisation des sources sonores frontales. Elle comprend la qualité de l'image stéréophonique et les pertes de définition.

Impression de qualité ambiophonique

- Cette caractéristique correspond à l'impression d'espace, à l'ambiance ou aux effets spéciaux bidirectionnels d'immersion.

5.4 Systèmes sonores évolués*Qualité audio de base*

- On utilise cette caractéristique unique et globale pour évaluer toutes les différences décelées entre la référence et l'objet de l'essai. L'examen de cette caractéristique pour les systèmes sonores évolués devrait également porter sur toutes les caractéristiques décrites pour les systèmes multivoies.

En outre, les caractéristiques ci-après peuvent présenter un intérêt:

Qualité du timbre – Il a été établi que cette caractéristique est très importante

- La caractéristique de qualité du timbre peut être décrite par deux ensembles de propriétés:
Le premier ensemble de propriétés relatives au timbre concerne la *couleur sonore*, par exemple l'éclat, le timbre musical, la coloration, la clarté, la dureté, l'égalisation ou la richesse.
Le second ensemble de propriétés relatives au timbre concerne *l'homogénéité sonore*, par exemple la stabilité, la pureté, le réalisme, la fidélité et les nuances. Ces propriétés peuvent permettre de décrire le timbre sonore, mais également d'autres caractéristiques sonores.

Qualité de localisation

- Cette caractéristique correspond à la localisation de toutes les sources sonores directionnelles. Elle comprend la qualité de l'image stéréophonique et les pertes de définition. Cette caractéristique peut être divisée en *qualité de localisation horizontale*, en *qualité de localisation verticale* et en *qualité de localisation distante*. Si les essais sont accompagnés d'images, ces caractéristiques peuvent également être divisées en *qualité de localisation sur l'écran* et en *qualité de localisation autour de l'auditeur*.

Qualité de l'environnement – Cette caractéristique est une extension de la caractéristique qualité ambiophonique

- Cette caractéristique correspond à l'impression d'espace, à l'enveloppement, à l'ambiance, à la diffusivité ou aux effets spatiaux directionnels d'immersion. Elle peut être divisée en *qualité de l'environnement horizontale*, en *qualité de l'environnement verticale* et en *qualité de l'environnement distante*.

6 Eléments de programme

Il ne faut employer que des éléments critiques pour révéler les différences entre les systèmes soumis aux essais. Les éléments critiques sont ceux qui mettent à l'épreuve ces systèmes. Il n'existe pas d'éléments de programme universels «convenables» pour évaluer tous les systèmes dans toutes les conditions. Il faut donc trouver, pour chaque système à tester dans chaque expérience, un élément de programme critique. La recherche de l'élément approprié prend en général du temps; toutefois, faute d'avoir trouvé les éléments critiques pour chaque système, l'expérience ne saura révéler les différences entre systèmes et ne sera pas probante.

Il faut, avant qu'une conclusion «nulle» soit tenue pour valable, s'assurer de façon empirique ou statistique que, si on ne parvient pas à trouver des différences entre systèmes, ce n'est pas parce que l'expérience ne les décèle pas en raison d'un mauvais choix des éléments audio ou encore parce qu'elle est mal organisée. Dans le cas extrême où l'on constate que tous les systèmes ou une majorité d'entre eux sont parfaitement transparents, il peut être nécessaire de prévoir des essais spéciaux avec des notes faibles ou moyennes afin de vérifier expressément la compétence des participants (voir la Pièce jointe 1).

Il doit être établi, à partir de recherches préalables par exemple, que ces repères peuvent être détectés par des auditeurs très expérimentés, mais non pas par des auditeurs inexpérimentés. Ces repères sont introduits en tant qu'éléments d'essai afin de vérifier non seulement la compétence des participants mais encore la sensibilité de tous les autres aspects des conditions expérimentales.

Si ces repères, qu'ils soient intégrés de manière non prévisible dans le contexte d'éléments apparemment transparents ou qu'ils soient intégrés dans un essai distinct, sont correctement identifiés par tous les auditeurs dans le cadre d'une méthode d'essai type (voir le § 3 de la présente Annexe) avec application des considérations statistiques indiquées dans la Pièce jointe 1, on peut alors conclure que la compétence des auditeurs est acceptable et qu'il n'y a pas de problème de sensibilité en ce qui concerne les autres aspects des conditions expérimentales. Dans ce cas alors, les résultats de transparence apparente établis par ces auditeurs sont une preuve de «vraie transparence» pour les éléments ou les systèmes pour lesquels ces auditeurs n'ont pu différencier la version codée de celle non codée.

Par ailleurs, si ces repères ne peuvent être correctement identifiés par n'importe quel auditeur, on peut alors en conclure soit que ces auditeurs ne sont pas assez compétents, soit qu'il y a des défauts de sensibilité dans le cadre expérimental même, soit que ces deux facteurs sont associés. Dans ce cas, la transparence apparente des systèmes ne peut être interprétée de manière adéquate et il faudra renouveler l'expérience avec de nouveaux auditeurs pour remplacer ceux qui n'ont pu réussir cet essai supplémentaire et apporter toute autre modification susceptible d'accroître la sensibilité expérimentale.

Dans la recherche d'éléments critiques, il faut accepter tout stimulus qui pourrait se présenter dans un programme radiodiffusé. On ne tiendra pas compte des signaux synthétisés spécialement conçus pour perturber le fonctionnement d'un système. Le contenu artistique ou intellectuel d'une séquence de programme ne sera ni intéressant, ni déplaisant, ni fatigant au point de détourner le participant de la recherche des dégradations. Il sera tenu compte de la probabilité d'apparition de chaque type d'élément de programme dans les diffusions réelles. Il faut toutefois se souvenir que la nature des éléments de programme peut changer en fonction de l'évolution des styles musicaux et de la mode. A l'avenir, des modèles perceptuels objectifs pourraient aider à choisir les éléments critiques.

Lorsqu'on sélectionne les éléments de programme, il est important de définir avec précision les caractéristiques à évaluer. Il faut confier cette tâche à des participants expérimentés ayant une connaissance de base des dégradations prévues. Ils doivent disposer au départ d'une vaste gamme d'éléments que l'on peut élargir au moyen d'enregistrements spécialisés.

En vue de la préparation de bandes d'essai pour comparaison subjective, il faut qu'un groupe de participants expérimentés règle le niveau sonore de chaque extrait avant de l'enregistrer sur le support des essais. On pourra ensuite se servir de ce support pour régler de façon fixe le gain de tous les éléments de programme.

Pour toutes les séquences d'essai, le groupe de participants expérimentés devra donc se réunir et convenir de façon unanime des niveaux sonores relatifs de chaque extrait. En outre, les experts devront parvenir à un accord sur le niveau de pression acoustique absolu reproduit par rapport au niveau d'alignement pour l'ensemble d'une séquence.

Au début de chaque enregistrement, il y aura une salve de tonalités (par exemple, 1 kHz, 300 ms, -18 dBFS) (FS: full scale – échelle totale) au niveau du signal d'alignement de façon qu'on

puisse régler le niveau d'alignement à la sortie sur le niveau d'alignement à l'entrée exigé par le canal de reproduction (voir le § 8.4.1). Pour les éléments d'essai en enregistrement numérique, le niveau d'alignement doit correspondre à -18 dB par rapport au niveau de codage maximal possible du système numérique [UER, 1992]. Il faut contrôler le signal radiophonique de façon que les crêtes d'amplitude ne dépassent que rarement l'amplitude de crête maximale permise pour le signal (voir la Recommandation UIT-R BS.645; signal sinusoïdal à 9 dB au-dessus du niveau d'alignement). A noter que, dans ces conditions, un indicateur de crête indiquera des niveaux qui ne dépassent pas le niveau de signal maximal autorisé. La salve de tonalités peut aussi être utile pour l'alignement dans le temps de la référence et les stimulus d'essai.

Le nombre d'extraits que l'on peut inclure dans un essai est variable: il doit être égal pour chaque objet d'essai. Un nombre raisonnable serait de 1,5 fois le nombre d'objets, pourvu qu'il y ait au moins 5 extraits. En général, les extraits audio durent de 10 à 25 s. Etant donné la difficulté de la tâche, il faut que le ou les objets soient disponibles. La sélection ne sera réussie que si on a défini un horaire approprié.

Pour l'évaluation des systèmes monophoniques et stéréophoniques, il serait intéressant que les extraits proviennent de sources aisément accessibles pour que l'on puisse vérifier directement les bandes préparées en les comparant aux originaux, si besoin est. Le disque compact SQAM est un exemple de source de ce genre. Il est toutefois plus important d'utiliser des extraits vraiment critiques même s'ils proviennent de sources d'accès plus difficile.

On vérifiera le fonctionnement d'un système multivoies en reproduction stéréophonique à deux voies au moyen d'un mixage réducteur de référence. Bien que dans certains cas, on puisse estimer que le recours à un mixage réducteur fixe est restrictif, c'est certainement à long terme l'option la plus raisonnable pour les radiodiffuseurs. Les formules du mixage réducteur de référence sont les suivantes (voir la Recommandation UIT-R BS.775):

$$L_0 = 1,00L + 0,71C + 0,71L_s$$

$$R_0 = 1,00R + 0,71C + 0,71R_s$$

Lorsque les essais portent sur un système sonore évolué, le rapport d'essai devrait contenir les équations utilisées pour le mixage réducteur permettant de passer d'un système sonore évolué à un système à deux voies ou multivoies ou une description du processus de restitution, le cas échéant.

La présélection d'extraits pour essai adaptés à l'évaluation critique du fonctionnement du mixage réducteur de référence à deux voies devrait être fondée sur la reproduction des éléments de programme à mixage réducteur à deux voies.

7 Dispositifs de reproduction

7.1 Généralités

On choisira les haut-parleurs et les casques de contrôle de référence de manière qu'ils reproduisent au mieux les signaux radiophoniques et les autres signaux d'essai; ainsi, quel que soit le type de reproduction, ils doivent donner un son neutre et pouvoir servir aux évaluations tant en monophonie qu'avec les systèmes sonores stéréophoniques à deux voies ou plus.

Certains défauts sont plus perceptibles avec une reproduction au casque et d'autres par haut-parleurs. Il faudra donc déterminer au moyen d'essais subjectifs préliminaires quel est le mode de reproduction approprié.

Dans le cas particulier des défauts qui affectent les caractéristiques de l'image sonore stéréophonique, on choisira la reproduction par haut-parleurs.

Pour évaluer les systèmes sonores stéréophoniques à deux voies, il peut être nécessaire d'utiliser à la fois des haut-parleurs stéréophoniques et des casques. Pour évaluer les systèmes sonores monophoniques, on peut utiliser un haut-parleur central ou des casques.

Si on a le choix entre les haut-parleurs ou les casques pour des essais isolés ou groupés, on pourra trouver la corrélation entre l'importance auditive d'un phénomène et l'appareil utilisé mais le nombre effectif des participants en sera diminué. En revanche, si les participants peuvent passer librement des haut-parleurs aux casques, il ne sera pas possible de trouver la corrélation entre l'usage d'un appareil et l'importance auditive du phénomène.

Pour évaluer les systèmes sonores multivoies et les systèmes sonores évolués avec ou sans image associée, il faut recourir aux haut-parleurs si l'on veut évaluer ce qui se passe pour tous les canaux de reproduction employés simultanément.

Quoiqu'il en soit, il faut que tous les haut-parleurs soient acoustiquement homogènes dans la bande de fréquences utile, de façon que les différences de timbre entre eux soient aussi faibles que possible.

7.2 Haut-parleur de contrôle de référence

7.2.1 Généralités

Le «haut-parleur de contrôle de référence» désigne un équipement d'écoute de haute qualité pour studio qui comprend un ensemble intégré de haut-parleurs dans une enceinte de dimensions spécifiques et une égalisation spéciale, d'excellents amplificateurs de puissance et les réseaux de transition appropriés.

Les caractéristiques électroacoustiques doivent répondre aux exigences minimales suivantes mesurées en espace libre. Les niveaux sonores absolus correspondent à une distance de mesure de 1 m par rapport au centre acoustique, sauf mention contraire.

7.2.2 Exigences électroacoustiques

7.2.2.1 Courbe amplitude fréquence

Pour la présélection des haut-parleurs, la courbe de réponse en fréquence dans la gamme 40 Hz-16 kHz, mesurée sur l'axe principal (azimut = 0°) au moyen de bruit rose dans des bandes d'un tiers d'octave, doit de préférence tenir dans une bande de tolérance de 4 dB. Les courbes de réponse en fréquence mesurées pour des azimuts de ± 10° ne doivent pas s'écarter de plus de 3 dB de la réponse sur l'axe et de plus de 4 dB pour des azimuts de ± 30° (uniquement dans le plan horizontal).

Les réponses en fréquence des divers haut-parleurs doivent se correspondre. Dans la gamme de fréquences comprises au moins entre 250 Hz et 2 kHz, il est préférable qu'elles ne s'écartent pas les unes des autres de plus de 1,0 dB.

NOTE 1 – La courbe de réponse du local utilisé, mentionnée au § 8.3.4, décrit la caractéristique de fréquence dans le champ sonore du local d'écoute.

7.2.2.2 Indice de directivité

L'indice de directivité C , mesuré pour les fréquences comprises entre 500 Hz et 10 kHz avec un bruit de largeur de bande d'un tiers d'octave, doit satisfaire à l'équation suivante:

$$6 \text{ dB} \leq C \leq 12 \text{ dB}$$

L'indice de directivité doit augmenter régulièrement avec la fréquence.

7.2.2.3 Distorsion non linéaire

Un signal d'entrée à tension constante qui produit un niveau de pression acoustique moyen de 90 dB est appliqué au haut-parleur. Avec cette pression, aucune composante de distorsion harmonique, dans la gamme des fréquences fondamentales de 40 Hz à 16 kHz, ne doit dépasser les valeurs suivantes:

-30 dB (3%)	pour $f < 250$ Hz
-40 dB (1%)	pour $f \geq 250$ Hz

7.2.2.4 Restitution des transitoires

Mesuré à l'oscilloscope, le temps de descente jusqu'au niveau à $1/e$ (soit environ 0,37) du niveau de départ (sur l'axe principal seulement) doit être:

$$t_s < 5 / f$$

où f est la fréquence.

Autrement dit, le temps de descente d'une salve de tonalité sinusoïdale ne doit pas dépasser 5 fois la période de l'onde sinusoïdale correspondante.

7.2.2.5 Temps de propagation

La différence de temps de propagation entre les voies d'un système stéréophonique ou multivoies ne doit pas dépasser 100 μ s.

NOTE 1 – Cela n'inclut pas le temps de propagation entre le haut-parleur et la position d'écoute.

Dans le cas d'un système avec image associée, le temps de propagation global du haut-parleur de contrôle de référence par rapport au ou aux systèmes étudiés ne doit pas dépasser les limites imposées dans la Recommandation UIT-R BS.775.

7.2.2.6 Dynamique

Le niveau de fonctionnement sonore maximal qu'un haut-parleur peut produire pendant au moins 10 min sans dégâts d'ordre thermique ou matériel et sans surcharge des circuits actifs, mesuré avec un signal de bruit imitant un programme (conformément à la Publication 268-1c de la Commission électrotechnique internationale (CEI)) doit être:

$$L_{\text{eff max}} > 108 \text{ dB}$$

mesuré au moyen d'un sonomètre à réponse uniforme et en position valeur quadratique moyenne (lent).

Le niveau équivalent de bruit acoustique émis par un seul haut-parleur de contrôle de référence et son amplificateur associé, à une distance de 1 m de son centre acoustique (voir la Note 1) doit être:

$$L_{\text{bruit}} < 10 \text{ dBA}$$

NOTE 1 – Le centre acoustique est le point de référence pour les mesures. Il correspond d'habitude au centre géométrique de la surface qui rayonne les fréquences les plus élevées du haut-parleur. Il devra être indiqué par le constructeur.

7.3 Ecouteurs de contrôle de référence

7.3.1 Généralités

Il s'agit d'un matériel d'écoute de haute qualité pour studio, égalisé en fonction de la réponse en champ diffus.

7.3.2 Exigences électroacoustiques

7.3.2.1 Réponse en fréquence

La réponse en fréquence en champ diffus des casques de contrôle pour studio est précisée dans la Recommandation UIT-R BS.708.

7.3.2.2 Temps de propagation

La différence de temps de propagation entre les voies d'un système stéréophonique ne doit pas dépasser 20 μ s.

Dans le cas d'un système avec image associée, le temps de propagation global des casques de contrôle de référence combinés au ou aux systèmes étudiés ne doit pas dépasser les limites imposées dans la Recommandation UIT-R BS.775.

8 Conditions d'écoute

8.1 Généralités

L'expression «conditions d'écoute» désigne les exigences acoustiques complexes imposées à un champ sonore de référence que subit un auditeur dans un local d'écoute, au point d'écoute de référence, en présence de sons reproduits par des haut-parleurs. Ces conditions d'écoute recouvrent:

- les propriétés acoustiques du local d'écoute,
- la disposition des haut-parleurs dans le local d'écoute,
- l'emplacement du point ou de la zone d'écoute de référence,

qui produisent les caractéristiques du champ sonore résultant en ce point ou en cette zone.

Comme l'état actuel de la technique ne permet pas encore de fournir une description du champ sonore de référence au moyen des seuls paramètres acoustiques, on indiquera certaines exigences géométriques et acoustiques imposées au local d'écoute de référence afin de garantir que les conditions d'écoute décrites sont viables.

8.2 Local d'écoute de référence

8.2.1 Généralités

Dans le cas de la reproduction par haut-parleurs, les essais subjectifs doivent satisfaire aux exigences suivantes. Les exigences minimales imposées aux locaux d'écoute sont décrites ci-dessous.

Dans le cas de la reproduction au casque uniquement, le local d'écoute doit satisfaire au moins aux exigences concernant le niveau de bruit ambiant.

8.2.2 Propriétés géométriques

On trouvera ci-après les dimensions nettes qui conviennent à un local d'écoute de référence. Si elles ne peuvent être obtenues, il faut au moins satisfaire aux conditions imposées ci-dessous au champ sonore et à la disposition des haut-parleurs.

8.2.2.1 Dimensions du local (surface au sol)

- Pour une reproduction monophonique ou stéréophonique à deux voies: 20-60 m².
- Pour une reproduction avec un système stéréophonique multivoie ou un système sonore évolué: 30-70 m².

NOTE 1 – Avec les plus faibles valeurs ci-dessus, le nombre d'auditeurs qu'il est possible de recevoir ensemble sera limité.

NOTE 2 – Il est nécessaire de poursuivre les études pour déterminer les caractéristiques optimales du local d'écoute pour les systèmes sonores évolués. La dimension du local, sa forme, ses proportions et ses propriétés acoustiques devraient être indiquées dans le rapport d'essai.

8.2.2.2 Forme du local

Le local doit être symétrique par rapport au plan vertical médiateur de la base stéréo. La surface au sol aura de préférence la forme d'un rectangle ou d'un trapèze.

8.2.2.3 Proportions du local

Les rapports des dimensions du local devront satisfaire aux relations suivantes pour que la distribution des tonalités propres à basse fréquence du local soit raisonnablement uniforme:

$$1,1 l / h \leq L / h \leq 4,5 l / h - 4$$

où:

L : longueur

l : largeur

h : hauteur.

On respectera en outre les conditions $L / h < 3$ et $l / h < 3$.

8.2.3 Propriétés acoustiques du local

8.2.3.1 Temps de réverbération

La valeur moyenne du temps de réverbération, T_m , dans la gamme de fréquences comprises entre 200 Hz et 4 kHz, doit être:

$$T_m = 0,25 (V / V_0)^{1/3} \quad \text{s}$$

où:

V : volume du local

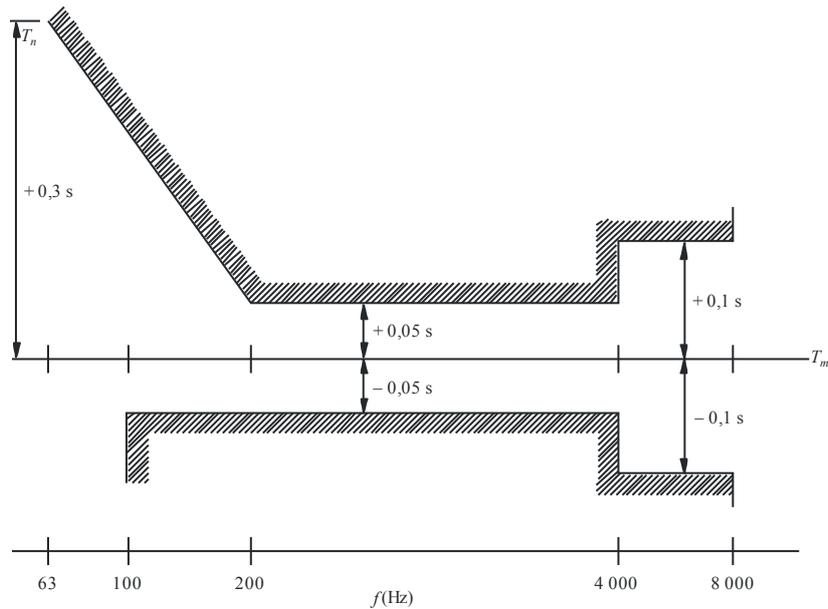
V_0 : volume de référence de 100 m³.

La Figure 1 présente les tolérances valables pour T_m dans la gamme comprise entre 63 Hz (voir la Note 1) et 8 kHz.

NOTE 1 – Il est difficile de mesurer de faibles temps de réverbération aux basses fréquences.

FIGURE 1

Limites des tolérances pour le temps de réverbération, rapportées à sa valeur moyenne, T_m



BS.1116-01

8.3 Caractéristiques du champ sonore de référence

8.3.1 Généralités

Les caractéristiques du champ sonore dans la zone d'écoute exercent une influence déterminante sur la perception subjective et sur l'évaluation de la qualité des phénomènes audibles et de leur reproductibilité à d'autres emplacements ou dans d'autres locaux. Ces caractéristiques résultent de l'interaction du ou des haut-parleurs et du local d'écoute et correspondent au dispositif d'écoute utilisé (voir le § 8.5).

On peut actuellement décrire les caractéristiques suivantes.

8.3.2 Son direct

8.3.2.1 Réponse en fréquence du haut-parleur de contrôle

La courbe de réponse du ou des haut-parleurs mesurée en espace libre doit répondre aux exigences exposées au § 7.2.2.

8.3.3 Son réfléchi

8.3.3.1 Réflexions rapides

Les réflexions rapides dues aux parois du local d'écoute, qui arrivent dans la zone d'écoute moins de 15 ms après le son direct, doivent être affaiblies dans la gamme 1-8 kHz d'au moins 10 dB par rapport au son direct.

8.3.3.2 Traîne sonore

En plus des exigences spécifiques imposées aux réflexions rapides et à la réverbération (voir le § 8.2.3), il faut éliminer d'autres anomalies significatives du champ sonore, comme les échos flottants (flutter), la coloration sonore, etc.

8.3.3.3 Temps de réverbération

(Voir le § 8.2.3.1.)

8.3.3.4 Réponse impulsionnelle

La réponse impulsionnelle de chaque haut-parleur, mesurée à toutes les positions d'écoute des auditeurs, le local étant dans la configuration qui sera utilisée pour l'essai (meubles compris), devrait apparaître, dans le domaine temporel, dans le rapport d'essai. Cette caractéristique peut être utilisée pour aider à vérifier dans quelle mesure les haut-parleurs, associés aux caractéristiques acoustiques du local, sont conformes aux exigences en matière de réflexions rapides, de traîne sonore et de réverbération.

8.3.4 Champ sonore stationnaire

8.3.4.1 Courbe de réponse du local d'écoute

Les courbes de réponse du local d'écoute sont définies comme étant les courbes de réponse d'un tiers d'octave des niveaux de pression acoustique produite par chaque haut-parleur au point d'écoute de référence, avec un bruit rose dans la gamme de fréquences 50 Hz-16 kHz. Il faut que les courbes de réponse du local d'écoute mesurées soient comprises dans les limites de tolérances de la Fig. 2.

Les écarts entre les diverses courbes de réponse du local d'écoute pour chacun des haut-parleurs au point d'écoute de référence ne devraient pas dépasser, de préférence, la tolérance cible de 2 dB dans toute la gamme de fréquences; l'adaptation entre les haut-parleurs avant frontal (± 60 degrés en azimut) est très importante, notamment en ce qui concerne ceux qui se trouvent au milieu de la couche horizontale. Les courbes de réponse du local d'écoute mesurées devraient figurer dans le rapport d'essai. Il est possible de respecter ces valeurs en incluant une égalisation. Dans ce cas, le rapport d'essai doit indiquer qu'une égalisation a été utilisée et contenir des informations détaillées sur le type d'égalisation utilisée.

8.3.4.2 Bruit de fond

Il est préférable que le bruit de fond continu (dû à la climatisation, aux appareils internes ou à d'autres sources extérieures), mesuré dans la zone d'écoute à la hauteur nominale de l'oreille d'un auditeur assis ne dépasse pas la courbe NR 10 (voir les Fig. 3 et 4).

En aucune circonstance le bruit de fond ne dépassera la courbe NR 15.

Il ne faut pas que le bruit de fond apparaisse comme impulsif, cyclique ou sinusoïdal.

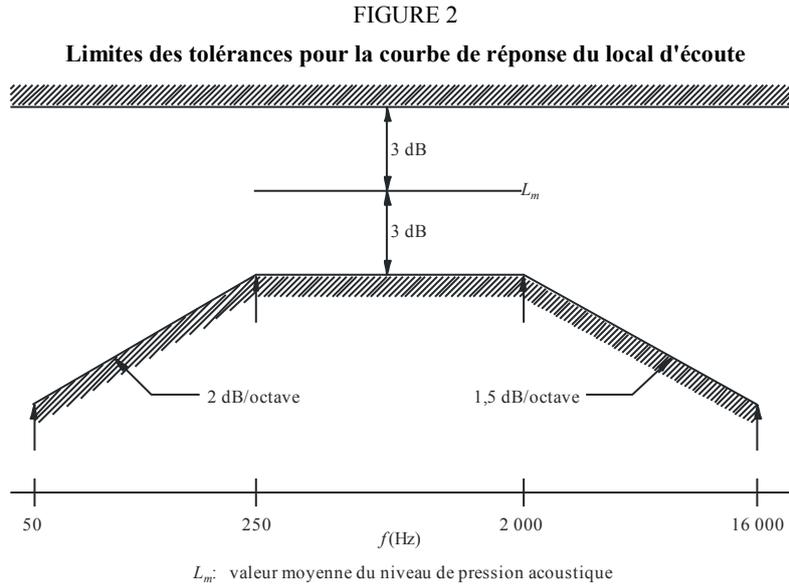
8.4 Niveau d'écoute

8.4.1 Ecoute sur haut-parleurs

8.4.1.1 Niveau de pression acoustique de travail (niveau d'écoute de référence)

On définit le niveau d'écoute de référence comme le niveau d'écoute préféré pour un signal de mesure donné, écouté au point de référence. Il caractérise le gain acoustique du canal de reproduction afin de garantir l'obtention du même niveau de pression acoustique pour un même extrait dans des locaux d'écoute différents.

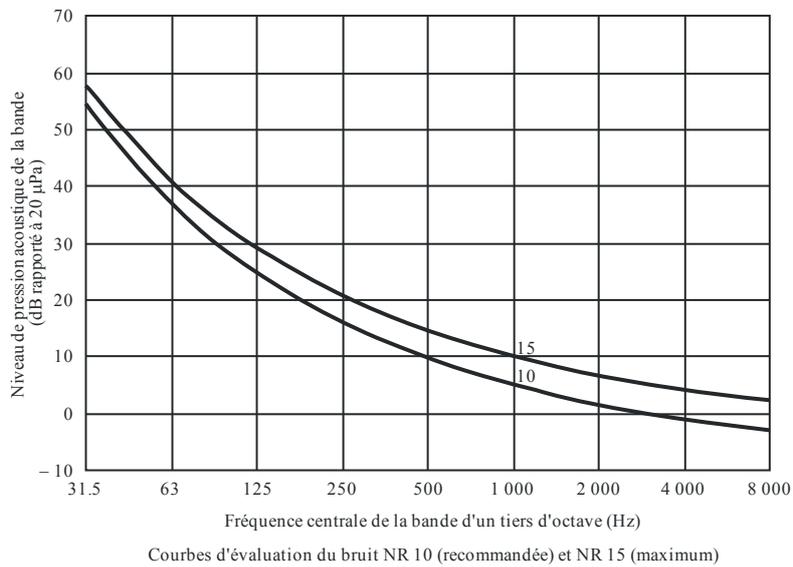
L'alignement des niveaux de chacun des haut-parleurs dans une disposition d'écoute donnée s'effectuera au moyen d'un bruit rose.



BS.1116-02

FIGURE 3

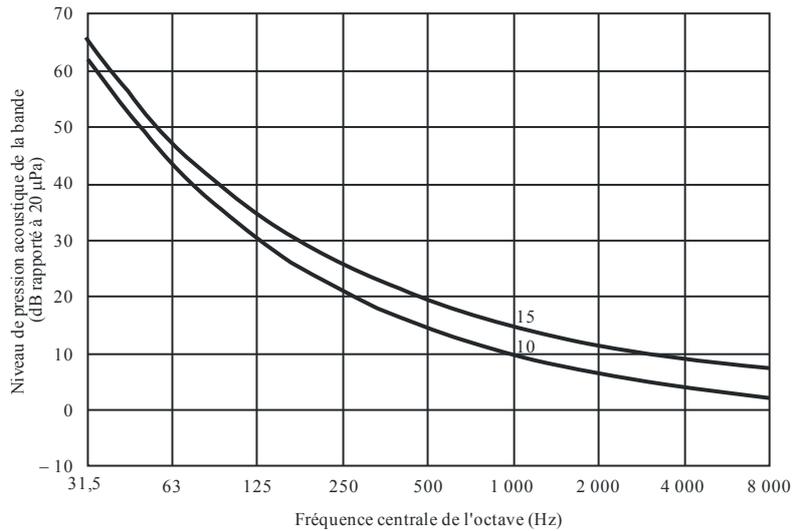
**Limites du bruit de fond par bande d'un tiers d'octave
(Courbes d'évaluation du bruit (NR) fondées sur les précédentes
courbes NR de l'ISO, Recommandation R1996 (1972))**



BS.1116-03

FIGURE 4

Limites du bruit de fond par octave
 (Courbes d'évaluation du bruit (NR) fondées sur les précédentes courbes NR de l'ISO, Recommandation R1996 (1972))



Courbes d'évaluation du bruit NR 10 (recommandée) et NR 15 (maximum)
 BS.1116-04

Pour un signal de mesure de tension efficace égale au «niveau du signal d'alignement» (0 dBµ0s, selon la Recommandation UIT-R BS.645; -18 dB au-dessous du niveau d'écrêtage d'un enregistrement numérique sur bande, selon [UER, 1992]), appliqué successivement à l'entrée de chaque canal de reproduction (c'est-à-dire un amplificateur de puissance et son haut-parleur associé), il faut régler le gain de l'amplificateur de façon à obtenir le niveau de pression acoustique de référence (CEI/pondération A, lent).

$$L_{réf} = 78 - \pm 0,25 \quad \text{dBA}$$

NOTE 1 – La mesure des paramètres acoustiques des systèmes sonores évolués peut-être beaucoup plus complexe que celle des systèmes audio multicanal antérieurs. Il faut prêter une attention particulière au choix du microphone de mesure et à son orientation lors des mesures .

(Au cours des séquences d'essai précédentes, on a noté que chaque auditeur peut préférer des niveaux absolus d'écoute différents. Bien que ce ne soit pas la meilleure solution, on ne peut pas toujours empêcher les participants de demander autant de souplesse. On ne sait pas encore si cela aura une influence sur l'audibilité de certains des défauts étudiés. Par conséquent, si les participants eux-mêmes règlent le gain du système, il faudra le mentionner dans le résultat des essais.)

8.4.2 Reproduction au casque

Il faut régler le niveau de façon à obtenir une intensité sonore égale au champ sonore de référence produit par des haut-parleurs. Pour apprécier cette égalité de niveau, il faut que le participant se trouve au point d'écoute de référence.

8.5 Dispositions d'écoute

8.5.1 Généralités

La disposition d'écoute décrit la position des haut-parleurs et les points d'écoute (zone d'écoute) dans le local d'écoute.

En principe, les essais d'écoute s'effectuent aux positions d'écoute de référence et à d'autres positions recommandées. Il faut toutefois évaluer ce qui se passe si l'écoute n'a pas lieu près du centre. C'est pourquoi on envisage les positions d'écoute du «cas le plus défavorable».

8.5.1.1 Hauteur et orientation des haut-parleurs de contrôle

La hauteur de tous les haut-parleurs dans le plan azimutal, mesurée au centre acoustique de chaque haut-parleur, doit être à la hauteur de l'oreille d'un auditeur assis. Les haut-parleurs seront orientés de telle manière que leurs axes de référence passent à la hauteur de l'oreille d'un auditeur assis à la position de référence. Si le système sonore évolué comprend des haut-parleurs placés à des hauteurs différentes, il est nécessaire d'indiquer et de décrire la position de tous les haut-parleurs dans les plans horizontaux et verticaux par rapport à la dimension du local et à la position d'écoute.

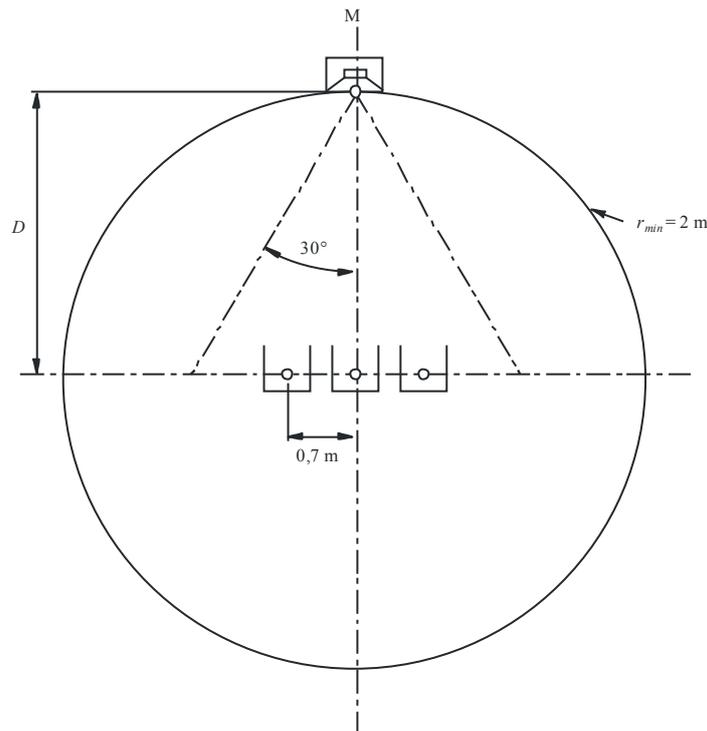
8.5.1.2 Distance jusqu'aux murs

Pour des haut-parleurs sur support indépendant, le centre acoustique du haut-parleur doit se situer à au moins 1 m des surfaces de réflexion d'enceinte. Si les dimensions du local ne permettent pas de respecter cette condition, les méthodes décrites dans la présente Recommandation pourraient tout de même être utilisées, mais le rapport d'essai doit alors indiquer que le critère de distance par rapport aux murs n'est pas respecté. Il faudrait alors contrôler les réflexions rapides d'une autre manière afin de respecter les exigences décrites au § 8.3.3.1 et la méthode utilisée pour ce faire devrait être indiquée dans le rapport d'essai.

8.5.2 Reproduction monophonique

Pour la reproduction des signaux monophoniques, on ne doit utiliser qu'un haut-parleur. La distance d'écoute est d'au minimum 2 m et toutes les positions d'écoute doivent se trouver dans un angle de $\pm 30^\circ$ par rapport à l'axe du haut-parleur (voir la Fig. 5).

FIGURE 5
Disposition de référence pour écoute avec haut-parleur M et zone d'écoute autorisée pour les systèmes sonores monophoniques



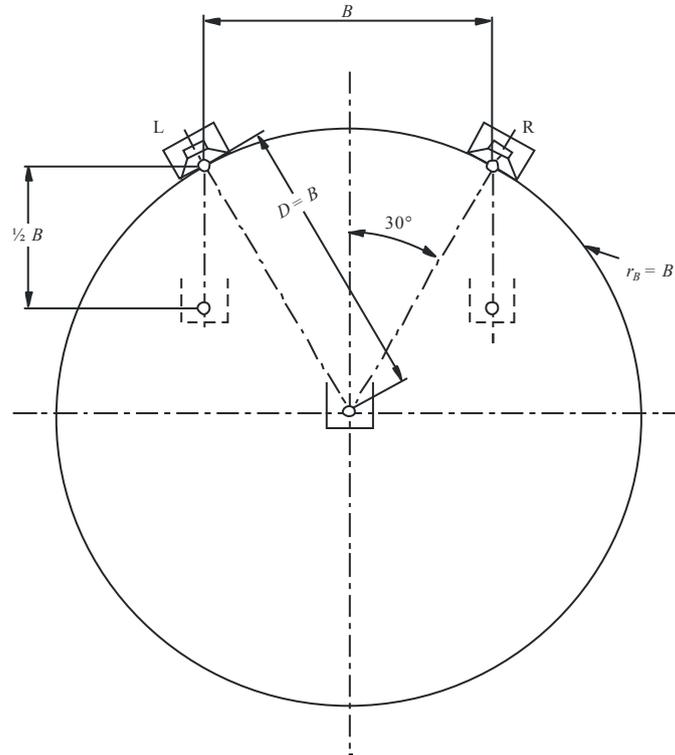
 Position d'écoute de référence

D : distance d'écoute

8.5.3 Reproduction stéréophonique à deux voies

FIGURE 6

Disposition pour essais d'écoute avec haut-parleurs L et R
Systèmes sonores stéréophoniques avec faibles dégradations



○ Position d'écoute de référence

○ Positions d'écoute les plus défavorables

B : largeur de la base des haut-parleurs
 D : distance d'écoute

BS.1116-06

8.5.3.1 Largeur de la base, B

La largeur de base B sera comprise de préférence entre 2 et 3 m. Dans des locaux de conception appropriée, elle peut atteindre 4 m.

8.5.3.2 Distance d'écoute, D (distance du haut-parleur à l'auditeur)

La distance d'écoute D sera comprise entre 2 et $1,7 B$ (m).

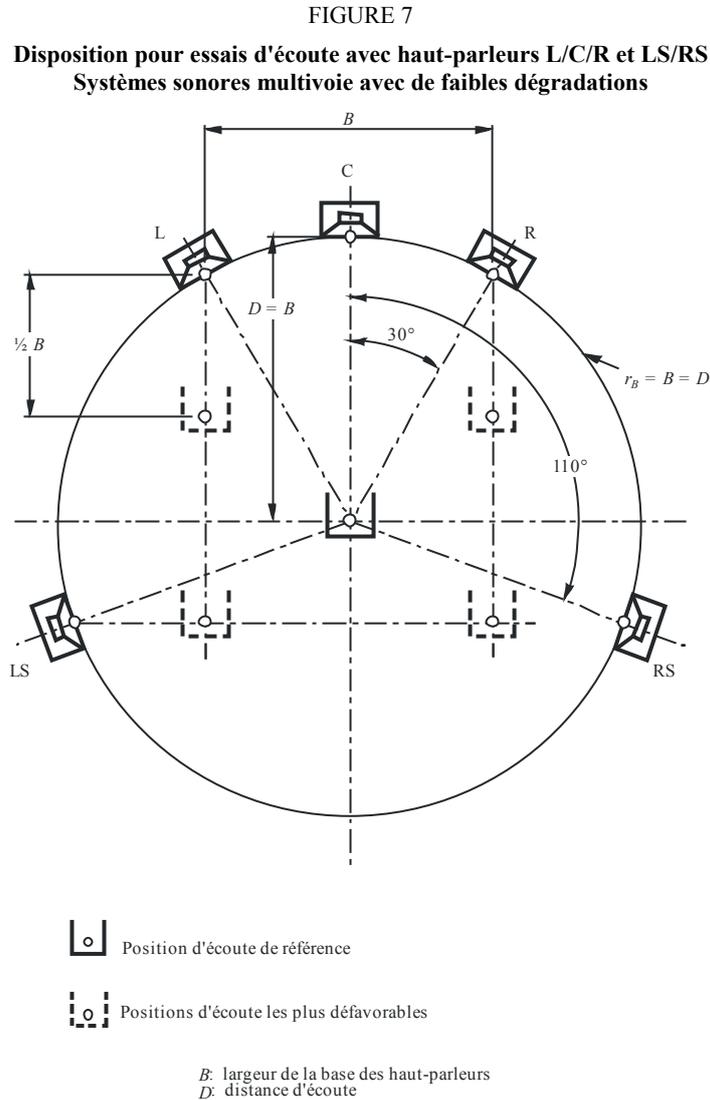
8.5.3.3 Positions d'écoute

Le point d'écoute dit de référence est défini par un angle d'écoute de 60° .

La zone d'écoute recommandée ne doit pas décrire un rayon supérieur à 70 cm autour du point d'écoute de référence. La Fig. 6 présente aussi d'autres positions d'écoute dans le «cas le plus défavorable».

8.5.4 Reproduction stéréophonique multivoies

La disposition d'écoute doit, en principe, correspondre à l'arrangement sonore multivoies 3/2 que spécifie la Recommandation UIT-R BS.775, Fig. 1: «Disposition de référence des haut-parleurs avec haut-parleurs L/C/R et LS/RS».



BS.1116-07

8.5.4.1 Largeur de la base, B

La largeur de base B sera comprise de préférence entre 2 et 3 m. Dans des locaux de conception appropriée, elle peut atteindre 5 m.

8.5.4.2 Distance d'écoute et angle de base

La distance d'écoute de référence sera B et l'angle de base de référence est donc égal à 60° .

8.5.4.3 Positions d'écoute

Le point d'écoute dit de référence est défini par un angle d'écoute de 60° comme ci-dessus. La Fig. 7 présente aussi d'autres positions d'écoute dans le «cas le plus défavorable».

8.5.5 Reproduction d'un système sonore évolué

Pour préciser les conditions d'essai, la position de tous les haut-parleurs (distances et angles) pendant les essais, ainsi que la disposition de ces haut-parleurs par rapport à la position d'écoute, doivent être décrites en détail dans le rapport d'essai. La forme et le contenu de cette description doivent correspondre aux dispositions des haut-parleurs et aux positions d'écoute définies dans la Recommandation UIT-R BS.775. Il faudra en outre indiquer et décrire la position de chaque haut-parleur dans le plan vertical pour les systèmes sonores évolués comprenant des haut-parleurs installés à des hauteurs différentes. La Recommandation UIT-R BS.2051 donne des informations pouvant être utiles dans ce contexte.

9 Analyse statistique

L'analyse statistique des résultats des essais vise principalement à déterminer avec précision la qualité moyenne de chacun des systèmes soumis aux essais et la validité des différences relevées entre ces appréciations de qualité moyenne. Pour mesurer cette validité, il faut connaître la variabilité ou la variance des résultats.

Si les essais ont été effectués conformément aux procédures décrites dans les autres parties de la présente Recommandation, il y a des chances pour que l'échelle soit régulière, c'est-à-dire que chaque échelon de l'échelle d'évaluation soit approximativement égal aux autres. Toutefois, aucune méthode statistique particulière n'est ni interdite ni recommandée pour obtenir cette propriété de l'échelle.

Pourvu que les hypothèses qui sous-tendent les statistiques paramétriques soient assez bien satisfaites, cette méthode se révèle la plus sensible et la plus efficace et elle est donc recommandée. Ce n'est que lorsqu'on constate que des propriétés importantes des données sont très différentes des hypothèses que sous-tend le modèle ANOVA (analyse de la variance – analysis of variance) qu'il convient d'envisager d'autres méthodes d'analyse (non paramétriques, par exemple). Il est précisément conseillé de commencer l'analyse en utilisant le modèle ANOVA. Par la suite, d'autres méthodes comme celle de l'essai t , de Neuman-Keuls, de Scheffé, etc., qui utilisent des estimations de la variance fournies par l'ANOVA peuvent servir à étudier plus en détail les secteurs où l'on observe les effets globaux significatifs révélés par ANOVA (le cas échéant).

Une hypothèse spécifique peut souvent être validée au moyen de plusieurs méthodes statistiques différentes. La légitimité d'une décision peut être renforcée si une hypothèse particulière apparaît convenir aussi pour une validation avec une autre méthode statistique. Il est donc suggéré d'appliquer une analyse des données supplémentaire (comme Wilcoxon, etc.).

Il est aussi important, à un moment donné, de considérer les aspects psychométriques qui influencent certainement le type de conclusions intéressantes que l'on peut obtenir avec une échelle non matérielle.

Il convient de noter qu'à moins que l'échelle d'évaluation ne s'avère linéaire, les comparaisons de notes différentes ne peuvent se faire que sur la base de l'ordre de classement.

10 Présentation des résultats des analyses statistiques

10.1 Généralités

Il faut une présentation avec laquelle le lecteur non averti aussi bien que l'expert puisse trouver les informations pertinentes. Tout lecteur veut d'abord voir le résultat global de l'expérience, sous forme graphique de préférence. Une telle présentation pourra être enrichie d'informations quantitatives plus détaillées mais les analyses numériques détaillées seront consignées dans des appendices.

10.2 Notes absolues

Grâce à une présentation des notes moyennes absolues, pour l'objet des essais et la référence dissimulée, on peut avoir un bon aperçu initial des données.

Il ne faut toutefois pas oublier que cette présentation ne saurait servir de base à une analyse statistique détaillée. En effet, quand on recourt à la méthode d'essai recommandée ici, le participant sait bien qu'une des sources de la paire à comparer est identique à la référence. Les observations ne sont donc pas indépendantes et l'analyse statistique de ces notes absolues ne conduira pas à une information significative, de sorte qu'il ne faut pas entreprendre cette analyse.

10.3 Notation des différences

La différence entre les notes données à la référence dissimulée et à l'objet de l'essai est un bon élément pour une analyse statistique. Une présentation graphique montre de combien on s'écarte vraiment de la transparence, ce qui est en principe du plus haut intérêt.

10.4 Niveau de signification et intervalle de confiance

Le rapport d'essai doit fournir au lecteur des renseignements sur la nature intrinsèquement statistique de toutes les données subjectives. Il convient d'indiquer les niveaux de signification ainsi que d'autres détails sur les méthodes statistiques et les résultats, comme l'inclusion des intervalles de confiance ou des intervalles d'erreur dans les graphiques, pour aider le lecteur à comprendre les conclusions.

Il n'existe pas naturellement de niveau de signification «correct». Toutefois, on retient en général la valeur de 0,05. On peut en principe utiliser des essais à une ou deux sorties selon les hypothèses vérifiées.

11 Contenu des rapports d'essai

Les rapports d'essai doivent exposer, aussi clairement que possible, les principes de l'étude, les méthodes mises en oeuvre et les conclusions obtenues. Il faut donner suffisamment de détails pour qu'une personne compétente puisse en principe reprendre l'étude pour en vérifier les résultats de façon empirique. Un lecteur qui connaît le sujet doit être capable de comprendre et de critiquer les points les plus importants de l'essai comme les raisons profondes de l'étude, les méthodes expérimentales et la réalisation ainsi que les analyses et les conclusions.

On s'attachera particulièrement aux aspects suivants:

- la spécification et la sélection des participants et des extraits;
- les précisions pratiques sur l'environnement et le matériel d'écoute, y compris les dimensions et les caractéristiques acoustiques du local, le type et l'emplacement des appareils de reproduction et la spécification de l'équipement électrique;
- le type de configuration des voies utilisé pour les essais (Recommandation UIT-R BS.775 ou Recommandation UIT-R BS.2051) et description;
si le système sonore faisant l'objet des essais n'est pas défini dans la Recommandation UIT-R BS.775, la position de chaque haut-parleur du système en question doit être présentée avec un niveau de détail équivalent à celui prévu dans la Recommandation UIT-R BS.775 afin de permettre une reproduction des essais. La position d'écoute de référence par rapport à la position des haut-parleurs associés au système sonore faisant l'objet des essais doit elle aussi être présentée en détail (voir les § 8.5.4 et 8.5.5);
- si les exigences relatives aux distances indiquées au § 8.5.1.2 sont respectées. Si ces distances ne sont pas respectées, ce point doit être noté;

- si les exigences relatives aux distances indiquées au § 8.5.1.2 ne sont pas respectées, les méthodes utilisées pour contrôler les réflexions rapides et respecter les exigences données au § 8.3.3.1 devraient être décrites;
- la réponse mesurée du local d'écoute pour tous les haut-parleurs. L'éventuelle application d'une égalisation doit être indiquée, de même que les méthodes utilisées pour ce faire;
- tout écart par rapport aux exigences acoustiques et physiques définies pour le local dans le présent document devrait être indiqué. Ces écarts peuvent concerner les mesures et les réponses acoustiques du local spécifiées au § 8.3, les critères de réponse comportementale de tous les haut-parleurs donnés au § 8.4 et toutes les exigences en matière de distance physiques données au § 8.5;
- la réponse impulsionnelle de chaque haut-parleur, mesurée à toutes les positions d'écoute de l'auditeur, le local étant dans la configuration qui sera utilisée pour l'essai (meubles compris), donnée dans le domaine temporel;
- la conception de l'expérience, l'entraînement, les instructions, les séquences de l'expérience, les procédures d'essai, la production des données;
- le traitement des données, y compris des précisions sur les statistiques déductives descriptives et analytiques;
- les bases précises de toutes les conclusions obtenues.

Pièce jointe 1 de l'Annexe 1

Aspects statistiques de la postsélection des participants

1 Evaluation de la compétence des auditeurs

La méthode de doublement aveugle à triple stimulus et référence dissimulée fournit deux notes pour chaque essai et permet de comparer ces deux notes directement, participant par participant, et d'étudier ces comparaisons sur tous les essais pour chaque participant. A chaque essai, on peut calculer la différence algébrique entre les deux notes, en soustrayant bien sûr toujours le même ordre. Supposons que nous soustrayions la note de la référence dissimulée de celle de l'objet de l'essai.

Si le participant n'a pas été capable, dans l'ensemble, de distinguer correctement la référence dissimulée de l'objet, la moyenne de toutes les différences de notes pour ce participant et cet essai d'écoute sera nulle ou proche de zéro car, en moyenne, les différences positives et négatives tendront à s'équilibrer. Si le participant a été capable, dans l'ensemble, de distinguer correctement la référence dissimulée de l'objet, la moyenne des différences s'écartera de zéro côté négatif, car elles seront pour la plupart négatives.

Les données ainsi obtenues sont soumises à un essai t unilatéral pour évaluer la probabilité que la moyenne de la distribution pour chaque participant soit égale à zéro. Si cette hypothèse de nullité est rejetée pour un participant donné, on peut alors en conclure que les données correspondant à ce participant proviennent d'une distribution à moyenne supérieure à zéro, côté négatif, pour un niveau de confiance donné. On peut alors conclure que chaque participant pour lequel cette hypothèse est vraie a donné la preuve qu'il ou qu'elle n'était pas tout simplement en train de deviner; on pourrait plus exactement dire de ces participants qu'ils ont fait preuve d'une compétence suffisante pour justifier l'inclusion de leurs données dans les analyses finales des résultats expérimentaux. Les données correspondant aux autres participants – ceux qui surtout devinaient – peuvent être rejetées pour toute analyse complémentaire au moyen de ce critère statistique.

Il convient de rappeler que les recommandations qui font l'objet de l'intégralité de la présente Recommandation n'ont trait exclusivement qu'aux faibles dégradations. S'il s'avère, pour une raison ou une autre, que de nombreuses dégradations «importantes» sont introduites dans un essai, et non uniquement des «faibles», la méthode de postsélection appliquée aveuglément, comme indiqué ci-dessus, peut aboutir à des conclusions erronées ou inappropriées. Une dégradation «importante» signifie ici une dégradation relativement facile à détecter, même par des auditeurs «non-experts». Il est évident que quelques dégradations véritablement «faibles» (difficiles à détecter), incluses dans un contexte dans lequel la plupart des dégradations sont «importantes» (faciles à détecter) auront peu d'incidence dans un essai t tel que celui décrit ci-dessus. En conséquence, les experts qui jugent correctement les éléments ayant de faibles dégradations peuvent être confondus dans les résultats globaux avec les non-experts qui jugent en «devinant» ces éléments. Cette situation tient au fait que dans des évaluations d'essai t , les résultats pour des éléments à faible dégradation peuvent se perdre dans le bruit statistique, étant donné que l'ampleur de t est la plus importante pour les éléments à dégradation importante.

Même dans le meilleur des essais portant sur de «faibles dégradations», on trouvera presque toujours inévitablement des dégradations importantes ou faciles à détecter, même si elles sont loin de constituer la majorité des dégradations. Cela étant, il est recommandé pour effectuer exclusivement des essais t de postsélection suffisamment rigoureux dans ce cas d'exclure systématiquement toutes les dégradations «faciles» à détecter ou importantes de la procédure d'essai t pour évaluer les compétences des auditeurs. Il peut s'agir de tous les éléments qui ont reçu de faibles notes moyennes de tous les participants, par exemple, des différences comprises entre $-2,0$ et $-4,0$. Pour ces éléments, la majorité des participants auront discerné avec justesse l'objet de la référence dissimulée et l'inclusion de ces éléments dans l'essai t rendra plus difficile l'évaluation de la compétence différentielle des participants au lieu de la faciliter. En laissant de côté les éléments à dégradation importante dans l'analyse de l'essai t , on exagérerait ou on surestimerait la compétence apparente des participants.

Le cas contraire, c'est-à-dire dans lequel il existe un trop grand nombre d'éléments «véritablement transparents», a été exposé au § 5 du corps de la présente Recommandation. Dans ce cas, ce sont les éléments apparemment transparents («trop difficiles») qui pourraient être omis dans les essais t de postsélection. Les éléments spéciaux introduits en raison de leur caractère critique avéré auraient davantage d'importance dans les essais t . En laissant les éléments apparemment transparents dans l'essai t , on sous-estimerait la compétence des sujets.

En règle générale, les éléments qui sont de manière permanente soit «trop difficiles» soit «trop faciles» ne permettent pas de distinguer les experts compétents des autres.

Le seul avantage que présente la bonne utilisation d'essais t de postsélection est que l'adéquation des participants pour une expérience donnée est évaluée au moyen de leurs résultats au cours de cette expérience. Lors d'une série d'expériences impliquant les mêmes participants dans des expériences différentes, il se peut que même si tous les participants passent le test de postsélection, certains d'entre eux conviennent parfaitement pour un sous-ensemble d'expériences mais pas pour toutes les

expériences, comme le montre la postsélection. Dans ce cas, les données correspondant à un participant particulier peuvent être acceptées ou rejetées selon le cas pour des résultats d'essai précis. On peut ainsi affirmer le concept «de compétence» mieux qu'en se fiant exclusivement à la présélection.

Il faut ici faire preuve de prudence. Un participant qui n'est pas suffisamment compétent ne peut fournir de bonnes données. En conséquence, il est tout à fait justifié de rejeter des données pour cause de compétence insuffisante, déterminée objectivement par une postsélection rigoureuse. Par ailleurs, il n'existe aucune garantie que les données provenant d'un participant qui a bien passé une postsélection d'essai t soient à coup sûr des données satisfaisantes. A titre d'exemple extrême, un participant peut très bien distinguer des objets des références dissimulées pour 100% des essais dans le cadre d'une expérience. Mais les données peuvent révéler qu'il ou qu'elle a donné une note de 1,0 à tous les objets soumis aux essais. En d'autres termes, l'ensemble total des données de ce participant peut correspondre à des valeurs de différence de $-4,0$ pour tous les essais.

Si l'on suppose que tous les autres participants dans la même expérience ont fait preuve d'une distribution «plus habituelle» des notes dans tous les essais, la structure de réponse très curieuse de ce participant particulier (valeurs de différence toutes de « $-4,0$ ») pourrait amener à la conclusion qu'il faut rejeter ces données. Toutefois, à l'exception peut-être d'un seul cas de toute évidence extrêmement déviant décrit ici à des fins d'exemple, il serait très difficile d'appliquer ces critères *a posteriori* en ce qui concerne l'acceptabilité des données. Cela reviendrait à délibérément modeler les données en fonction d'une préconception de l'expérimentateur plutôt qu'accepter la preuve empirique des résultats effectifs.

Ces méthodes *a posteriori* NE doivent PAS être employées. Tant que le nombre total de participants à une expérience est adéquat, même si les données d'un participant expert sont extrêmement déviantes, elles auront très peu d'influence négative sur l'ensemble total des données. Des résultats significatifs et reproductibles sont tout à fait habituels lorsque l'on procède à des expériences sensibles, même lorsque celles-ci comprennent des participants déviants mais experts. Une fois l'expérience menée à bien, en cas de doute sur la «valeur» des données, le seul recours consiste à refaire entièrement l'expérience, à l'aide d'un tout nouvel ensemble de participants en s'efforçant de corriger tout défaut éventuel dans les procédures expérimentales employées auparavant.

2 Evaluation complémentaire de la compétence des auditeurs

A mesure que la qualité des codecs à pertes fondés sur le mode perceptuel augmente, le nombre d'auditeurs assez compétents pour discerner les défauts de codage restants diminuera inévitablement. Un auditeur qui s'est montré suffisamment compétent lors d'un essai antérieur incluant des défauts assez «aisément audibles» peut ne pas être suffisamment compétent lors d'un essai dans lequel ces défauts plus aisément audibles n'ont pas été introduits. En outre, bien que le résultat t d'un auditeur puisse montrer qu'il est suffisamment expert pour l'expérience dans son ensemble, il peut ne pas être suffisamment expert pour faire des différences entre le signal de référence et un signal codé d'excellente qualité. Dans ce cas, les données du participant peuvent ajouter du «bruit statistique» aux données totales, ce qui masque les véritables différences perçues par d'autres participants.

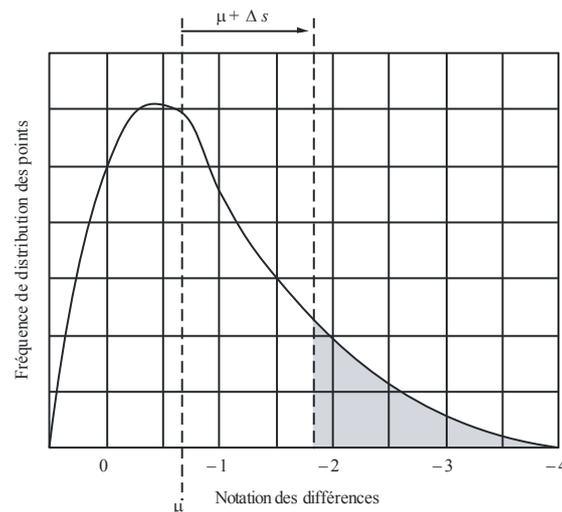
Pièce jointe 2 de l'Annexe 1

Évaluation du niveau de compétence des participants

Actuellement, toutes les données d'un participant au cours d'un essai donné servent à évaluer son résultat t . Les données de tous les participants ayant obtenu des résultats t suffisamment élevés sont alors prises en compte dans l'ANOVA.

FIGURE 8

Méthode permettant de ne pas tenir compte des points de données avant l'essai t



BS.1116-08

Dans la présente proposition, nous suggérons que les essais t soient refaits plusieurs fois pour les sous-ensembles de données de chaque participant. A chaque renouvellement de l'essai, le critère d'évaluation du niveau de compétence du participant deviendra plus exact.

Le niveau de compétence d'un participant sera réévalué et si ce dernier a fait preuve de compétence suffisante, les données qui lui correspondent seront alors incluses dans l'ANOVA suivante. En conséquence, pour chaque essai renouvelé, le critère de niveau de compétence suffisant augmentera et une ANOVA sera effectuée avec les données des autres participants. Les critères proposés d'évaluation de la compétence sont indiqués ci-après.

Le processus est illustré à la Fig. 8 pour un ensemble de données hypothétique. Tout d'abord, on calcule l'écart moyen type pour les données du participant; cet écart servira ensuite à déterminer les résultats z (voir la Note 1) correspondant aux données de ce participant. A partir de là, tous les points de données d'un participant situés au-delà d'un certain critère ($\mu + \Delta 1 s$) ne seront pas pris en compte et un nouvel essai t sera fait pour les points de données restants. Comme on le voit sur la Figure, les points de données situés au-delà de $\mu + \Delta 1 s$ (la zone ombrée) ne sont pas pris en compte et les autres points de données (la zone non ombrée) sont utilisés dans l'essai t suivant. Si, pour les points de données restants, il ressort de l'essai t que le participant a la compétence suffisante, toutes les données de ce participant seront incluses dans l'ANOVA suivante. Si l'essai t révèle que le participant n'a pas la compétence suffisante, les données qui lui correspondent seront éliminées entièrement de toutes les ANOVA suivantes. On répète ensuite ce processus en appliquant des critères de compétence encore plus stricts, $\mu + \Delta 2 s$. On répète ensuite le processus N fois avec les critères $\mu + \Delta i s$ où $i = 0, 1, \dots, N$. On étudie actuellement des valeurs appropriées de $\Delta i s$ et de N à partir de données

d'études précédentes menées à bien au CRC (Communications Research Center – Centre de recherches sur les communications, Canada).

NOTE 1 – Les résultats z représentent les résultats normalisés pour une distribution à zéro de moyenne et un écart type de 1. Le résultat est défini comme suit $z = \frac{x - \mu}{s}$ où x est le point de données, μ est la moyenne d'échantillon et s est l'écart type pour l'échantillon:

$$s = \sqrt{\frac{N \sum x^2 - (\sum x)^2}{N(N - 1)}}$$

Pièce jointe 3 de l'Annexe 1

Exemple d'instructions données aux participants

La terminologie de ces instructions n'est pas strictement conforme aux définitions du glossaire.

1 Familiarisation ou phase d'entraînement

La phase d'entraînement a pour but de permettre aux auditeurs d'identifier les distorsions et déformations possibles du système soumis aux essais et de se familiariser avec elles. Après l'entraînement, vous devez être capable de savoir «ce qu'il faut rechercher à l'écoute». On vous demandera, dans l'après-midi, de noter en aveugle le matériel audio que vous entendrez le matin. Au cours de la phase d'entraînement, vous vous familiariserez aussi avec la procédure d'essai.

Vous entendrez une version de référence (originale) et une version modifiée de chaque élément du matériel audio. Sur l'écran de contrôle vidéo, la version de référence sera identifiée par la lettre «A»; la version du signal modifiée et la «référence dissimulée» seront désignées par les lettres «B» et «C». Au cours de la présentation, vous pouvez passer librement de l'une à l'autre de ces trois lettres. Cela permettra d'établir une comparaison précise et détaillée entre «A», «B» et «C». Il faut donner des notes aux différences entre «A» et «B» et entre «A» et «C». Les séquences audio dureront en général de 10 à 25 s et seront répétées autant de fois que vous le voudrez. Pour l'entraînement, vous êtes libres d'utiliser les haut-parleurs, les casques ou les deux. Vous disposez de trois heures pour vous entraîner sur tous les éléments que vous noterez réellement dans l'après-midi au cours de la phase de notation en aveugle.

Au cours des essais de l'après-midi, on vous demandera de noter les présentations selon l'échelle du Tableau 2:

TABLEAU 2

Dégradation	Note
Imperceptible	5,0
Perceptible mais non gênant	4,0
Légèrement gênant	3,0
Gênant	2,0
Très gênant	1,0

Il faut expliquer au participant la signification de cette échelle et insister sur le fait qu'elle doit être considérée comme une échelle continue à échelons égaux avec des repères définis pour des valeurs spécifiques.

Comme chaque essai de l'après-midi comprend une référence dissimulée (c'est-à-dire une répétition exacte de la référence), il serait normal qu'il y ait pour chaque essai au moins une note 5,0 (mais une seulement (voir la Note 1)). Si vous trouvez que «B» ou «C» est meilleur que la référence, cela veut dire que vous trouvez une différence «perceptible mais non gênante» et vous attribuerez une note comprise entre 4,0 et 4,9 selon la différence décelée.

Bien que vous soyez obligé de vous interroger, au cours de la phase d'entraînement, sur la façon dont vous interprétez personnellement les dégradations audibles, à l'aide de l'échelle de notation, il est important que vous ne discutiez jamais de votre interprétation personnelle avec les autres participants.

NOTE 1 – L'objectif du changement recommandé est de forcer le participant à «mieux deviner» le matériel codé parmi les stimulus. Nous estimons que certains participants sont en fait capables de détecter des défauts très faibles, mais par prudence, ils donneront deux notes de 5,0 plutôt que prendre position. Le changement recommandé permettrait de résoudre ce problème.

2 Exemple de contenu d'une séance de formation

La formation principale, d'une durée de 3 h, doit être effectuée avec des groupes de quatre participants environ dans la matinée du premier jour. Les participants auront dû recevoir au préalable des instructions écrites.

La séance de formation doit comprendre les éléments suivants:

- brève introduction des raisons et des objectifs de l'essai,
- passage répété des extraits choisis pour l'essai pour se familiariser avec la présentation sonore et avec les éléments de programme à évaluer ultérieurement,
- brève explication des systèmes à l'essai et du type de dégradation et présentation orale des catégories de dégradation fixées par le groupe de présélection,
- démonstration des dégradations au moyen des éléments comportant les dégradations les plus importantes,
- explication de la caractéristique à évaluer,
- explication de l'échelle de dégradation à cinq notes,
- entraînement à la commutation et à la notation.

Les jours d'essai suivants, il convient de rappeler aux participants les points traités pendant la principale séance de formation, ce qui peut comprendre à nouveau l'écoute des éléments pour l'essai, avant de procéder aux essais formels.

3 Phase de notation en aveugle

L'essai en aveugle a pour but la notation des divers éléments audio entendus le matin au cours de la phase d'entraînement.

A chaque essai, vous entendrez trois versions d'un élément audio donné. Elles seront appelées «A», «B» et «C» sur l'écran de contrôle vidéo. «A» est toujours la version de référence (originale) avec laquelle «B» et «C» seront comparés et notés. L'une d'elles, «B» ou «C», est une version modifiée et l'autre la référence dissimulée (identique à la référence). On ne vous dit pas laquelle de «B» ou «C» est la version modifiée et laquelle est la référence dissimulée, d'où le terme «aveugle» pour cette phase de notation. A tout moment, vous pourrez passer librement à «A», «B» ou «C». On peut répéter

les séquences audio jusqu'à ce que vous soyez sûrs de votre évaluation. Vous pouvez à volonté passer à l'essai suivant quand vous êtes satisfaits de votre évaluation d'un essai.

Pour chaque essai, on vous demande de noter la différence éventuelle que vous remarquez entre «B» et «A», d'une part, et entre «C» et «A», d'autre part, au moyen de l'échelle à cinq notes du Tableau 3. Il faut donc donner deux notes pour chaque essai, une pour «B» et une pour «C». Il doit y avoir au moins une note 5,0 (mais une seulement (voir la Note 1 du § 1 de la présente Pièce jointe)) à chaque essai. Veuillez introduire vos notes dans l'ordinateur à la fin de chaque essai.

On peut aussi utiliser une fiche de notation au lieu d'introduire les notes dans un ordinateur.

Le Tableau 3 sera alors montré au participant et une copie sera disponible tout au long des séances de notation en aveugle.

Il faut expliquer au participant la signification de cette échelle et insister sur le fait qu'elle doit être considérée comme une échelle continue à échelons égaux avec des repères définis pour des valeurs spécifiques.

TABLEAU 3

Dégradation	Note
Imperceptible	5,0
Perceptible mais non gênant	4,0
Légèrement gênant	3,0
Gênant	2,0
Très gênant	1,0

Pièce jointe 4 de l'Annexe 1

Evaluation subjective: glossaire

Pour plus de clarté, on définit ici différents termes employés dans la présente Recommandation. La Fig. 9 illustre des correspondances entre certains de ces termes.

Caractéristique

Caractéristique perçue pour un événement auditif, conformément à une définition orale ou écrite.

Elément du programme

Extrait, traité par le système à l'essai.

Emplacement

Endroit où est effectué l'essai d'écoute. Il peut s'agir uniquement de l'emplacement géographique ou de la position du participant dans le local d'écoute. Peut être un des facteurs de l'essai.

Essai

Une partie de la séance qui va de la présentation d'un ensemble de stimulus à leur notation.

Essai doublement aveugle

Essai en aveugle où il n'y a aucune possibilité d'intervention non contrôlée entre l'expérimentateur et l'essai d'écoute.

Essai en aveugle

Essai où le participant ne dispose, à propos des essais, d'aucune autre information que les stimulus.

Extrait

Echantillon d'un passage de musique, de paroles ou autre convenant à l'évaluation des caractéristiques ou paramètres individuels de qualité sonore du système soumis aux essais.

Les extraits pour essai se présentent généralement sous forme d'enregistrements sonores (disque compact, R-DAT ou autres formats d'enregistrement ou de source).

Groupe d'écoute

Tout le groupe des participants qui fournissent les données d'un essai d'écoute.

Note

Expression numérique de la grandeur d'une caractéristique sur une échelle donnée.

Objet de l'essai

Système soumis aux essais, représenté par un certain nombre d'extraits que traite le système en question.

Participant

Personne qui évalue les stimulus d'un essai d'écoute.

Référence

Extrait pour essai, reproduit sans modification par l'objet de l'essai et servant de base de comparaison pour un essai avec dégradation.

Référence dissimulée

Référence non signalée au participant à l'essai.

Séance

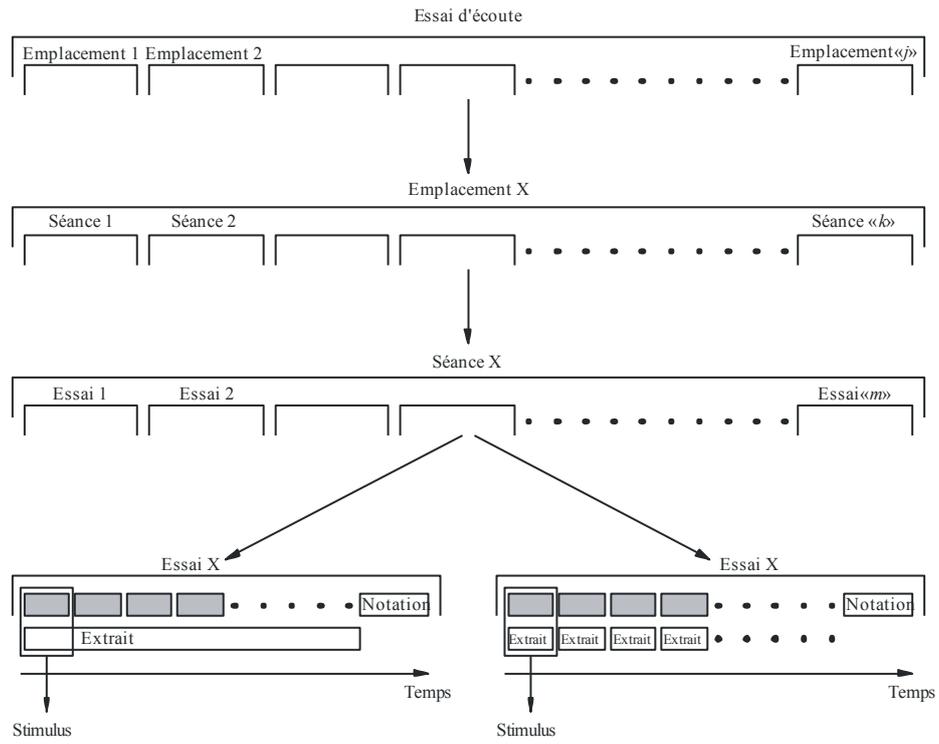
Ensemble des essais qu'un participant ou un groupe d'écoute doivent évaluer au cours d'une période continue.

Stimulus

Combinaison d'un objet d'essai ou de la référence dissimulée ou de la référence et d'un extrait ou d'une portion d'extrait.

FIGURE 9

Illustration des correspondances entre certains termes du glossaire



■ . . . Représente le choix qu'a fait le participant entre la référence, la référence dissimulée et l'objet de l'essai.

Les deux essais représentés montrent les extrémités sur une gamme de dispositions possibles.