AUDIO ENGINEERING SOCIETY PRESENTS ...

Audio Engineering Society Presents

AUDIO METERING

MEASUREMENTS, STANDARDS, AND PRACTICE

Eddy B. Brixen

A Focal Press Book

ROUTLEDGE

Audio Metering

In this comprehensive guide, Brixen takes the reader through the complex and confusing aspects of audio metering, imparting the knowledge and skills needed to utilize optional signal levels and produce high-quality audio.

Covering all aspects of this fundamental subject, *Audio Metering: Measurements, Standards, and Practice* begins with the basics, such as audio definitions and digital techniques, and works up to more complex topics like hearing and psychoacoustics.

This revised and expanded third edition includes:

- Updated information on loudness metering, covering both existing and new standards.
- Definitions of terms such as LKFS, LUFS, gating, LRA.
- Explanations of signal types and musical sounds and structures.
- Further details on immersive audio.
- Skills needed for both small-room acoustics and large auditorium sound design without loss of sound quality.
- Descriptions of measurement signals and systems for audio and acoustic sound.
- A chapter on listening tests from small setups to large-scale comparisons of PA/SR-systems.

Packed full of valuable information with a wide range of practical applications, this is the essential reference guide to audio metering for technicians, engineers, and tonmeisters, as well as sound designers working with acoustics, electroacoustics, broadcast, studio recording, sound art, archiving, audio forensics, and theatrical and live-audio setups.

Eddy B. Brixen is an audio consultant and lecturer based in Denmark.

Audio Engineering Society Presents

Recording Orchestra and Other Classical Music Ensembles
Richard King

Recording Studio Design
Philip Newell

Modern Recording Techniques
David Miles Huber

Immersive Sound
The Art and Science of Binaural and Multi-Channel Audio
Edited by Agnieszka Roginska and Paul Geluso

Hack Audio
An Introduction to Computer Programming and Digital Signal Processing in MATLAB
Eric Tarr

Loudspeakers
For Music Recording and Reproduction
Philip Newell and Keith Holland

Beep to Boom
The Development of Advanced Runtime Sound Systems for Games and Extended Reality
Simon N Goodwin

Intelligent Music Production
Brecht De Man, Ryan Stables, and Joshua D. Reiss

Women in Audio
Leslie Gaston-Bird

Audio Metering
Measurements, Standards, and Practice
Eddy B. Brixen

For more information about this series, please visit: www.routledge.com/Audio-Engineering-Society-Presents/book-series/AES

Audio Metering

Measurements, Standards, and Practice

Third edition

Eddy B. Brixen

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Third edition published 2020
by Routledge
52 Vanderbilt Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

First edition published by the Broadcast Publishing 2001
Second edition published by Focal Press 2011

Library of Congress Cataloging-in-Publication Data
A catalog record for this book has been requested

ISBN: 978-1-138-90912-0 (hbk)
ISBN: 978-1-138-90911-3 (pbk)
ISBN: 978-1-31569-415-3 (ebk)

Typeset in Times NR
by Apex CoVantage, LLC

Contents

Preface

What is the dynamic range, how loud is the program, why is it overloaded, and what is dBFS? These are the eternal questions known by everyone who works with the practical aspects of audio.

This book was written to give everybody with interest in audio an explanation of the conditions that determine the answers to these questions.

In the book, you will find descriptions of fundamental acoustics, electronics, and psychoacoustic concepts. Many topics related to digital audio technology are also covered, and information can also be obtained here on the majority of the tools that apply to describing the magnitude of sound.

This book is the third edition of *Audio Metering*. The update is particularly concerned with loudness measures and metering. However, the basic chapters at the beginning of the book have been expanded, and new chapters on AoIP and listening tests are now part of the book.

HOW SHOULD THIS BOOK BE USED?

You may use *Audio Metering* as a reference book. The beginning contains a table of contents, and the end of the book provides an extensive glossary and an index.

Reading the book from cover to cover, however, is possible and may be a good choice in some instances, because the subject matter of the book is organized so that the most basic material is placed at the beginning while the more generally descriptive material is found toward the end of the book.

Enjoy!

Eddy B. Brixen
Smørum, Denmark, November 2019

Thanks to Lene and to my publisher for the extensive patience they have exhibited during the preparation of this third edition of *Audio Metering*.

CHAPTER 1

Acoustic Sound

CHAPTER OUTLINE

"Sound" is derived from French (*son*) and from Latin (*sonus*). It refers to vibrations in the air that can be heard by people. "Audio" is derived from Latin also. It comes from the verb *audire* ("to hear") and refers to things that are related to hearing.

In English, we tend to use the terms "sound" and "audio" indiscriminately. Sound is a pressure disturbance in the air that can be perceived by hearing. However, the sound does not necessarily have to be audible. Infrasound and ultrasound, which are below and above the normal range of human hearing, respectively, are examples of inaudible sounds.

Through recent years the term "audio" has been redefined by (audio) professionals. By this newer definition, audio is an electrical (or digital) representation of sound. A transducer such as a microphone converts sound to audio; audio is converted into sound by a transducer such as a loudspeaker. The audio itself cannot be perceived by hearing, and thus is not audible.

When discussing acoustical topics in this book, we will be dealing with "sound," and that is where we will begin.

WHAT IS SOUND?

Sound is normally understood to mean elastic molecular oscillations in the air or other media such as water, iron, or concrete. These oscillations result in pressure variations that are of such a magnitude that they can be sensed by human hearing.

However, sound can also be converted to, for example, variations in the electrical current in a conductor, in grooves on the surface of a vinyl record, or a sequence of numeric values. We call these forms intermediate formats – or audio – because we later convert them into acoustic (derived via Latin "acousticus" from the Greek word "akoustós" ("heard, audible) sound."

SPEED OF SOUND

Sound propagates by an oscillating solid body setting the particles next to it in motion, and those next to them, and so on (see Figure 1.1). The sound thus spreads with a certain propagation velocity. This is called the speed of sound, which varies depending on the medium.

Normally we specify the speed of sound (c) in a medium in meters per second **[m/s]** or feet per second **[ft/s]**. In the air, the speed of sound is dependent on the temperature:

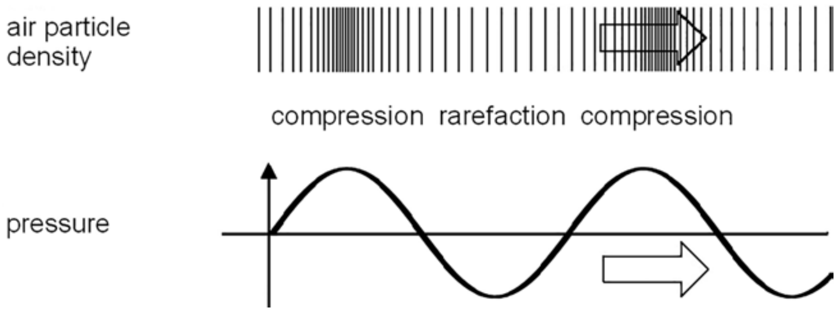

Figure 1.1 Sound (in the air) can be defined as variations or disturbance in the air pressure.

In the air, at 0° C or 32° F, the speed of sound is 331.4 m/s or 1086 ft/s (1193 km/h or 740.5 mph).

In the air, at 20° C or 68° F, the speed of sound is 343.54 m/s or 1126 ft/s (1237 km/h or 767.7 mph).

However, as a rule, 340 m/s or 1130 ft/s is used as approximations for the speed of sound for general purposes. Thus, in applied audio technology, one often encounters the following values:

34 cm/ms (Sound travels about 34 cm each millisecond.)

3 ms/m (Sound takes roughly 3 ms to travel 1 meter.)

or

1 ft/ms (Sound travels about 1 ft each millisecond.)

The speed of sound in solids (see Table 1.1), in general, depends on its elasticity, density, and a measure of deformation characteristics (Poisson's ratio).

In practical audio design and recording, a knowledge of the speed of sound is essential when designing and aligning PA systems with – or without – delay, or, for instance, when recording immersive audio. The speed of sound in water or ice also comes into play when doing underwater sound recordings and underwater sound design.

FREQUENCY

Frequency **(f)** is a measure of the number of oscillations or cycles per second and is specified in **Hertz** (abbreviated as **Hz**).

Table 1.1 Speed of sound in various media.

Medium	m/s	f/s
Air (at sea level), 0° C/32° F	331.4	1,086
Air (at sea level), 20° C/68° F	343.5	1,126
Air (3,048 m/10,000 ft above sea level), 5.1° C / 41.2° F	328.4	1,077
Water, pure, 0° C/32° F (not frozen*)	1,403	4,603
Water, pure, 20° C/68° F	1,481	4,858
Frozen water, pure, ice (bubble-free), 0° C/32° F	3,839	12,595
Frozen water, pure, ice (bubble-free), –20° C/–4° F*	3,894	12,776
bubbles in the ice will affect the speed	≈3,800–4,000	≈12,500–13,100
Seawater, just below surface, 0° C/32° F, 3.5% salt	1,407	4,616
Seawater, just below surface, 20° C/68° F, 3.5% salt	1,530	5,019
Glass	≈5,000–5,300	≈16,000–18,000
Concrete	≈3,000–3,400	≈10,000–11,000
Steel	≈5,000–6,000	≈16,000–20,000
Wood	≈3,300–5,000	≈11,000–16,000
Rubber	≈1,600–1,800	≈5,250–5,900

$$f = \frac{1}{T}[Hz]$$

where

f = frequency [Hz]

T = period [s]

A frequency of 1 Hz = 1 cycle per second. A frequency of 1000 Hz = 1000 cycles per second (1000 Hz is normally expressed as 1 kilohertz, abbreviated as 1 kHz).

The nominal audible frequency range of human hearing comprises frequencies from 20 Hz to 20 kHz. This range thus is called the audio frequency range.

WAVELENGTH

The wavelength is the distance a single oscillation takes to complete in a given medium. The wavelength is thus dependent on the frequency and the speed of sound in the medium concerned.

It is expressed like this:

$$\lambda = \frac{c}{f}$$

where

λ = wavelength [m]

c = speed of sound [m/s]

f = frequency [Hz]

or

λ = wavelength [ft]

c = speed of sound [ft/s]

f = frequency [Hz]

This relationship shows that the audio spectrum in the air contains wavelengths ranging from approximately 17 m at 20 Hz down to 17 mm at 20 kHz or from approximately 55 ft at 20 Hz down to 0.67 in at 20 kHz (see Table 1.2). Thus, the wavelength is in the same order as the dimensions of the rooms and equipment applied for recording and listening. It influences everything from the acoustics of rooms to the directional patterns of loudspeakers, as well as the frequency response and directionality of microphones.

If the concept of wavelength is a little difficult to comprehend, it helps to imagine the sound radiating from a sound source as pressure variations generated by the source being carried away on a conveyor at a given fixed speed (the speed of sound) (see Figure 1.2). If one "freezes" this image the distance between two maxima is the wavelength of that frequency. If the frequency is increased, one period of that frequency is finished in less time, and the traveled distance gets shorter, hence a shorter wavelength.

Table 1.2 Wavelengths in Air at 20° C / 68° F. In this calculation, the speed of sound is 344 m/s or 1126 ft/s.

Frequency			Wavelength	
20 Hz	⇒	17.20 m	or	56.3 ft
100 Hz	⇒	3.44 m	or	11.3 ft
200 Hz	⇒	1.72 m	or	5.65 ft
344 Hz	⇒	1.00 m	or	3.273 ft
1000 Hz	⇒	34.40 cm	or	1.126 ft
1126 Hz	⇒	30.20 cm	or	1.000 ft
2000 Hz	⇒	17.20 cm	or	0.563 ft
10,000 Hz	⇒	3.44 cm	or	0.113 ft
20,000 Hz	⇒	1.72 cm	or	0.056 ft

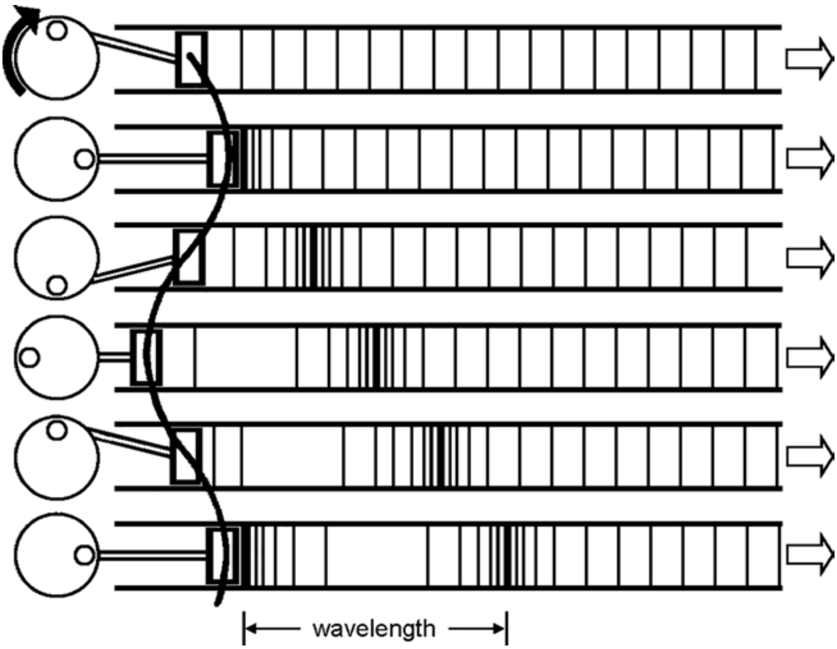

Figure 1.2 A sound source regarded as a piston. Pressure variations generated at the piston are transported away at a constant speed. Hence, the wavelength of a given frequency is the distance between two identical points in adjacent cycles.

SOUND PRESSURE

Air surrounds us. We call the static pressure of air the barometric pressure. Sound pressure is defined as the difference between the instantaneous pressure and static pressure.

The pressure is measured in pascals (abbreviated Pa**)**. One pascal is defined as 1 N/m^2 (newton per square meter).

The sound pressure can be regarded as a modulation of the static pressure. The greater the variations of sound pressure (+ and −), the stronger the perception of the sound.

However, the limitation of the sound pressure of a soundwave is reached when the negative part reaches 0 Pa. If the barometric pressure is 1000 hPa (or 100 kPa), the maximum modulation of the soundwave then is ±1000 hPa. Beyond that pressure, the air begins to distort the soundwave.

The weakest audible sound at 1 kHz has a sound pressure of approximately 20 µPa (20 micro-pascal = $20 \cdot 10^{-6}$ Pa), whereas the ear's threshold of pain lies at a sound pressure of around 20 Pa. The weakest audible sound and the threshold of pain thus differ by a factor of one million.

The sound pressure *level* (SPL) is expressed in dB with a reference of 20 μPa. Thus the reference sound pressure level is 0 dB re 20 μPa, and the level which creates pain in the ear is 120 dB re 20 μPa. The loudest sound pressure level in air without distortion is approximately 194 dB re 20 μPa (see Chapter 6: The dB Concept).

CONVERSION RELATIONSHIPS

The pascal (Pa) is a unit of the SI system (International System of Units). Other units are still seen in practice, such as on older data sheets for classic microphones, where the bar unit is used (see Table 1.3).

SOUND POWER

Sound is a form of energy, hence the concept of sound power.

Sound power is the sound energy transferred during a period divided by the period concerned. In a traveling plane wave with a sound pressure of 20 μPa, the power that passes through an area of 1 m^2 placed perpendicular to the direction of travel is 1 pW (one picowatt $=10^{-12}$ watt).

This value is used as a reference when specifying a sound power level in dB. For a sound pressure of 20 Pa, the power is 1 W.

Sound power, for example from a loudspeaker or a machine, is measured in practice either in a reverberation chamber (a room with highly reflective surfaces) or by performing a large number of sound pressure measurements around the object, normally on a hemi-

Table 1.3 Conversion between Pa, Bar, and Atmosphere (Atm).

Unit	Equivalent to:
1 Pa	1 N/m^2
	10 μbar
	$7.5006 \cdot 10^{-3}$ mm Hg
	$9.869 \cdot 10^{-6}$ atm
Bar	10^5 N/m^2
	10^5 Pa
	750.06 mm Hg
	0.98692 atm
1 atm	$1.01325 \cdot 10^5$ N/m^2
	$1.01325 \cdot 10^5$ Pa
	1.01325 bar
	760 mm Hg

sphere; these measurements subsequently are used to calculate the total radiated power. Unlike sound pressure, sound power is neither dependent on room nor distance. Where sound pressure is a property of the field at a point in space, the sound power is a property of the sound source, equal to the total power emitted by that source in all directions.

One purpose of measuring the sound power is to determine the efficiency of loudspeakers: What is the radiated acoustic power from a loudspeaker when a given electric power is fed into it? The ratio between the two indicates the efficiency. It is normally expressed as a percentage.

From a perspective of noise at the workplace, the sound power of machines or other equipment is the basis for calculating the sound pressure level, for instance, at operators' positions.

SOUND INTENSITY

Sound intensity is an expression of power per unit of area. For a flow of sound energy that is propagating in a specific direction, the intensity is the power transferred through an area perpendicular to the direction of travel, divided by the area.

For a traveling plane wave in the free field with a sound pressure of 20 µPa, this intensity is 1 pW/m^2.

Sound intensity has a directional component. In practice, the sound intensity from a sound source is measured using a sound intensity probe consisting of two transducers (pressure microphones) at a well-defined distance. By looking at the phase of the radiated sound, among other things, the direction of the sound can be determined.

SOUND FIELDS

SPHERICAL SOUND FIELD

When sound radiates from a point source, the intensity decreases with distance (see Figure 1.3). It can be compared with a balloon when being filled with air: the bigger the diameter, the thinner the rubber wall at a given area on the circumference. The balloon consists of a certain amount of rubber that has to cover a still larger volume. There is a quadratic relation between the radius and the area. Doubling the radius enlarges the area by four.

The intensity falls with distance according to:

$$\boldsymbol{I} = \frac{1}{\boldsymbol{d}^2}$$

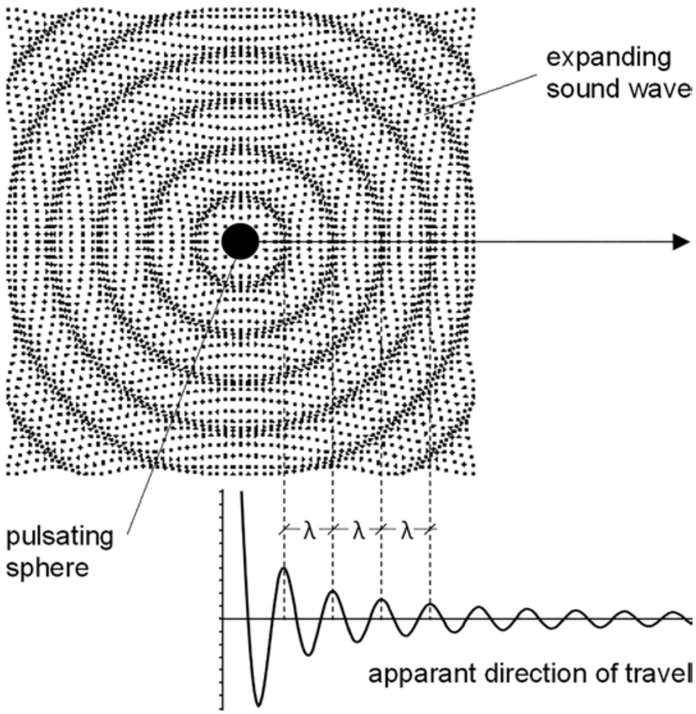

Figure 1.3 Sound radiating from a point source.

where

I = sound intensity

d = distance

This is called the inverse square law.

Therefore, the sound pressure falls according to:

$$\boldsymbol{P} = \frac{1}{\boldsymbol{d}}$$

where

P = sound pressure

d = distance

Thus, in the free field, doubling the distance halves the sound pressure. In dB, the sound level drops 6 dB when doubling the distance (see Figure 1.4).

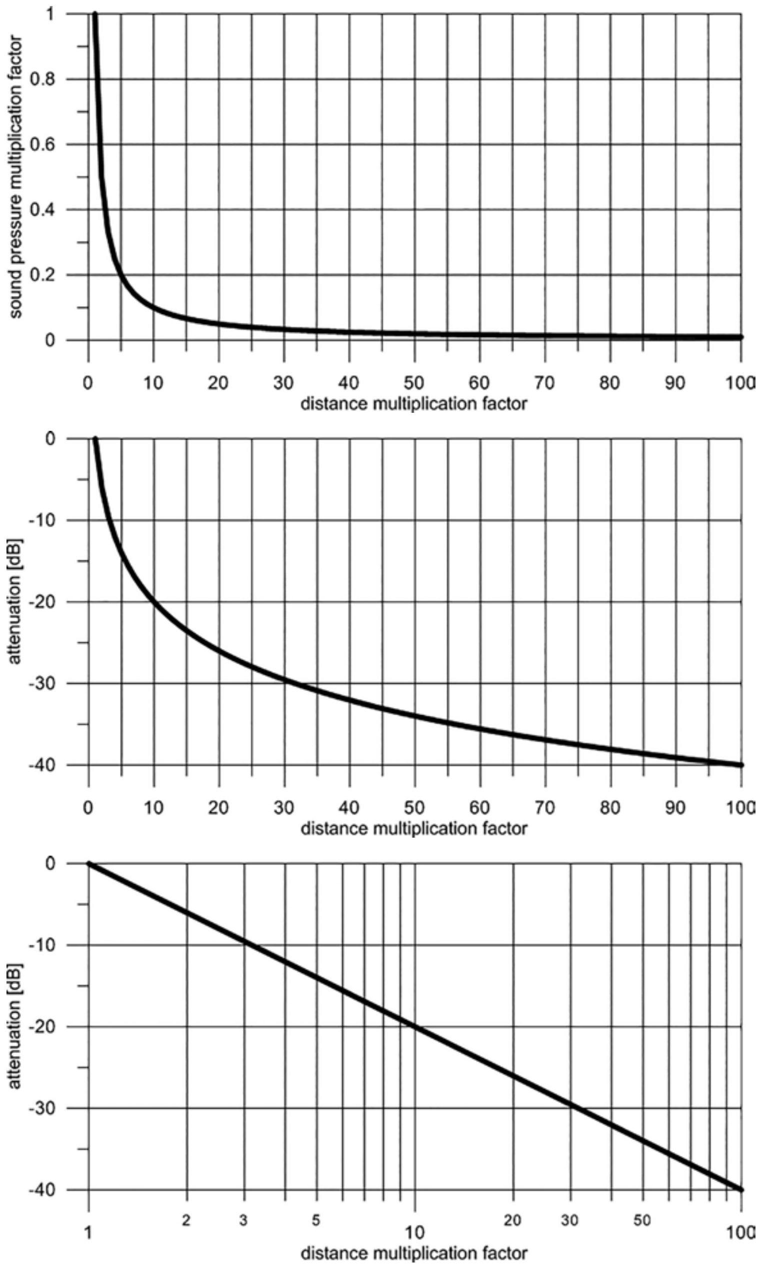

Figure 1.4 Point source: Three different ways to express the relation between sound pressure and distance.

A: Sound pressure (expressed as a multiplication factor) vs. distance (expressed as a multiplication factor). Example: Moving ten times the initial distance (meters, feet, yards, etc.) away from the sound source changes the sound pressure by a factor of 0.1.

B: Attenuation of sound pressure (expressed in dB) vs. distance (expressed as a multiplication factor). Example: Moving ten times the initial distance away from the sound source reduces the sound pressure by 20 dB.

C: Sound pressure (expressed in dB) vs. distance (expressed as a multiplication factor on a logarithmic scale). The values are the same as in B. However, the curve here is a straight line.

CYLINDRICAL SOUND FIELD

An (infinite) line source generates a cylindrical sound field.

The intensity falls with distance according to:

$$\boldsymbol{I} = \frac{1}{\boldsymbol{d}}$$

where

$\boldsymbol{I}$ = sound pressure

$\boldsymbol{d}$ = distance

Hence the sound pressure falls according to:

$$\boldsymbol{P} = \frac{1}{\sqrt{\boldsymbol{d}}}$$

where

$\boldsymbol{I}$ = sound pressure

$\boldsymbol{d}$ = distance

Thus, doubling the distance reduces the sound pressure by $\sqrt{2}$. In dB, the sound level drops 3 dB when doubling the distance.

Some loudspeaker designs are pronounced line sources, typically formed by an array of speaker units. In practical audio engineering, these loudspeakers may act as line sources within a limited distance and frequency range. In the far field, the practical (finite) line array is always considered to be a point source.

PLANE SOUND FIELD

In a plane sound field, the sound does not attenuate as the intensity of the sound field is kept constant in the direction of the propagation. In other words, the attenuation is 0 dB. However, infinite plane sound sources do not exist. In real life, an approximation of the plane sound field can be experienced far away from a point source where a limited sector of a spherical sound field can be regarded as a plane field. Another approximation exists very close to a large loudspeaker membrane or a vibrating wall. (For instance, if a microphone is placed 1 cm from a 12-inch loudspeaker unit and then moved 1 cm further away it does not change the sound level.)

CHAPTER 2

From Acoustic Sound to Electrical Signals

CHAPTER OUTLINE

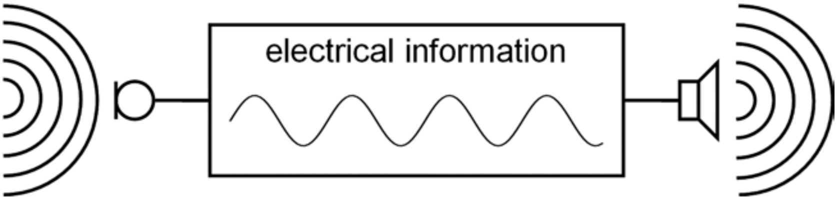

Figure 2.1 From acoustic sound to electrical signal and back to the acoustic sound.

In order to be able to measure, manipulate, or describe sound, we generally convert it from an acoustic to an electrical signal.

Sound exists in purely acoustical terms as pressure variations primarily in the air. By using an appropriate transducer, such as a microphone, we transform these pressure variations into variations in current or voltage (see Figure 2.1).

If we wish to record the sound, we then convert these electrical variations into another form. In the past, these variations were widely used to leave magnetic traces on audio tapes or physical variations in the surface of vinyl records. Today we mostly convert the sound into sequences of numerical values, as is the case with digital technology. Computer technology can subsequently be used to save or process these values, which presumably will end up as acoustic information again at a later stage.

ELECTRICAL SIGNALS

When a microphone is used to convert acoustic information to electrical information, we say that the electrical signal is analogous to the acoustic signal. In other words, the waveforms that describe the pressure and voltage or current variations resemble each other.

In the electrical world, we have a circuit in which a current of electrons flows. A potential difference drives the electrons in one direction or the other. It is the size of this potential difference, the voltage, which we most often regarded as the magnitude of the signal.

SPEED

The movement of the electrons occurs at speed close to the speed of light, which, in a vacuum, is $2.99792458 \cdot 10^8$ m/s [$9.8293 \cdot 10^8$ ft/s]. In practice, we use a rounded value of $3*10^8$ m/s, or 300,000 km/s [186,000 miles/s].

Electrical signals move at a much higher speed in a wire than acoustic signals propagate in air.

WAVELENGTH

To determine the wavelength of the electrical signal in cables and transmission lines, we use the same expression that we use for acoustic sound:

$$\lambda = \frac{c}{f}$$

where

$\boldsymbol{\lambda}$ = wavelength [m]

$\boldsymbol{c}$ = speed [m/s]

$\boldsymbol{f}$ = frequency [Hz]

or

$\boldsymbol{\lambda}$ = wavelength [feet]

$\boldsymbol{c}$ = speed [feet/s]

$\boldsymbol{f}$ = frequency [Hz]

In the audio frequency range (20 Hz–20 kHz) we thus have wavelengths ranging from 15,000 km down to 15 km or from 9,000 miles down to 9 miles. As opposed to an acoustic sound, which propagates in the air, electrical audio signals will very seldom have wavelengths that are of the same physical size as the applied circuits. In practice and for the most part, this only occurs when we are using cables with lengths in kilometers or miles.

If, however, we are working in the high-frequency spectrum, such as with video or radio frequencies, or with the transmission of digital signals, then the wavelengths will very quickly turn out to be comparable to the sizes of the physical circuits in which the current is flowing. This fact is the reason why we have to consider how we connect antennas, digital audio, and video. AF-signals, on the contrary, are more forgiving about impedance matching, cable length, and the like.

CHAPTER 3

Digital Representation

CHAPTER OUTLINE

Digital audio technology takes the starting point in the transformation of an analog signal to numerical values at an appropriate rate. After this conversion, computers can be applied for the processing, transmission, and storage (see Figure 3.1).

Advantages of digital audio technology include error correction, which allows for the execution of copying and transmission in a lossless manner – further, calculations in the digital domain substitute what otherwise would imply physical, electronic circuitry.

ANTI-ALIASING

Before the conversion of an analog signal, it is necessary to determine a well-defined upper cutoff frequency (f_u), which implies a low-pass filter. This filtering is called anti-aliasing; the term “alias” means a changed identity – or a “false version” of the original. The necessity of the filtering is due to the sampling process itself. The analog signal must not contain frequencies that are higher than half of the sampling frequency (a frequency also called the Nyquist frequency). If

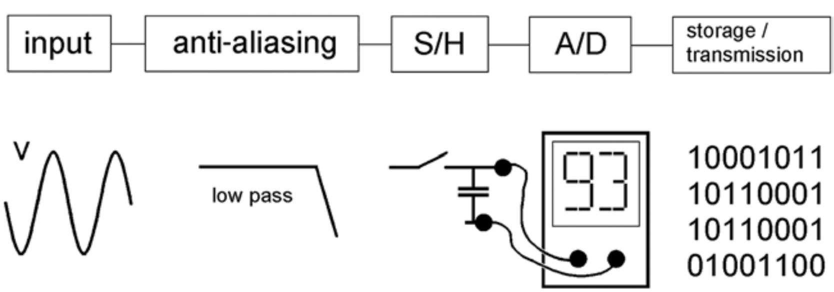

Figure 3.1 Principles for the digitizing of analog audio. The signal is low-pass filtered before sampling (anti-aliasing). The sampling (S/H) is carried out at a given rate that determines the frequency range of the digitized audio. Then the size of each sample is determined (quantization). The number of bits available per sample determines the precision of the quantized audio sample.

the sampling frequency is lower than twice the highest input frequency, then the reconstructed audio signal contains frequency components that were not present in the original. The filter ensures that the signal does not contain any aliasing frequencies after reconstruction.

SAMPLING

After the low-pass filtering, sampling is performed. Sampling is measurements of the instantaneous value of the signal. The rate of measurements per second is called the sampling rate or sampling frequency (f_s).

For a comparison, think of a movie camera that can record moving pictures by taking a single picture 24 times per second. One could then say that the camera's sampling frequency is 24 Hz. Now and then, you can also observe visual alias frequencies on film, such as the wheels of the stagecoach turning backward while the horses and the carriage are moving forward because the wheels are rotating faster than half the sampling frequency or the sampling frequency (number of pictures per second) is not high enough (see Figure 3.2).

Sampling frequencies of 32 kHz, 44.1 kHz, and 48 kHz have long been the standard for quality audio for things like CD or broadcast audio tracks. However, higher sampling frequencies of 88.2 kHz, 96 kHz, 176.4 kHz, and 192 kHz widely apply to

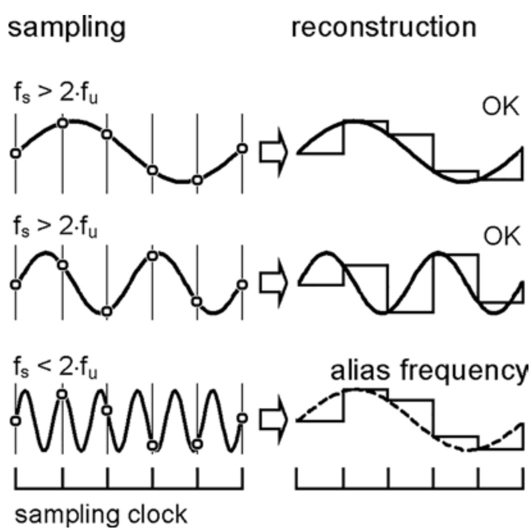

Figure 3.2 If the sampling frequency is not at least twice the highest audio frequency, the reconstructed signal is not in accordance with the input.

production environments, and high-quality audio delivery formats (see oversampling below). If the final delivery format is specified to have a lower sampling rate, downsampling is performed (sample rate conversion).

Sound clips for computer games, audio in communication systems, and other similar types of audio historically have used very low sampling frequencies, down to 8 kHz or even less. (However, computer games have become quite sophisticated, providing good quality and immersive audio to lift the storytelling to yet another level.)

For each sample, the instantaneous value of the analog signal is retained for as long as the analog-to-digital converter (also called an A-D converter or ADC**)** needs to perform its conversion. In the early converters, this was performed by a "hold" circuit. This circuit fundamentally was a capacitor that was charged/discharged to the instantaneous value of the signal at the time of sampling. The reading of the analog signal in modern converters occurs so quickly that the hold function is no longer needed. However, the understanding of the sampling process is easier when keeping a "virtual capacitor" in mind.

OVERSAMPLING

Oversampling describes the process when performing sampling at a frequency that is several times higher than the requisite minimum. Oversampling makes it easier to implement anti-aliasing and reconstruction filters, as the necessity for extremely steep filter slopes is reduced. The drawback, of course, is the increased amount of data. However, oversampling may provide a higher quality of the digitized audio, although it is at the cost of latency.

QUANTIZATION

Now comes the part of the process that determines the digital "number." This process is called quantization. The word comes from Latin (*quantitas* = size). During quantization, the size of the individual samples is converted to numbers. This transformation, or conversion, is not always ideal, however.

The scale applied for this purpose of comparison has a finite resolution that is determined by the number of bits. The word "bit" is a contraction of the words "binary digit," which refers to a digit in the binary number system. With quantization, it is the number of bits that determines the precision of the value read. Each time there is one more bit available, the resolution doubles, and hence the error in measurement is halved. This error is primarily referred to as quantization noise. In practice, this means that the signal-to-noise ratio (s/n) is improved by approximately 6 dB for each extra bit that is available for the comparator scale. To be more precise, the s/n equals $6.02 \times n + 1.761$, where n is the number of bits.

BINARY VALUES

The value ascribed to the quantization is not a decimal number but a binary number instead. The binary number system uses the number 2 as its base number. Thus, only two numbers are available, namely 0 and 1. These values are easy to create and detect in electrical or optical terms. For example, voltage present = 1; no voltage present = 0; the current run in one direction = 1; the current run in the opposite direction = 0; there is light =1; there is no light = 0.

With only one binary digit, or one bit, available, we thus only have two values, namely 0 and 1. With two bits available, we have four possible combinations, namely 00 (zero, zero), 01 (zero, one), 10 (one, zero), and 11 (one, one). The number of steps on the scale equals 2 to the power of the number of bits. In practice, for parallel quantization, between 8 and 32 bits are used in the quantization of analog signals. CD-quality audio corresponds to 16 bits per sample (= 2^{16} = 65,536 possible values). There are only a finite number of values available for the determination of the magnitude. Thus, the actual analog value at the moment of sampling is represented by the nearest value on the scale (see Figure 3.3).

With linear quantization (equal distance between the quantization steps), a resolution of only a few bits results in extreme distortion of the original signal.

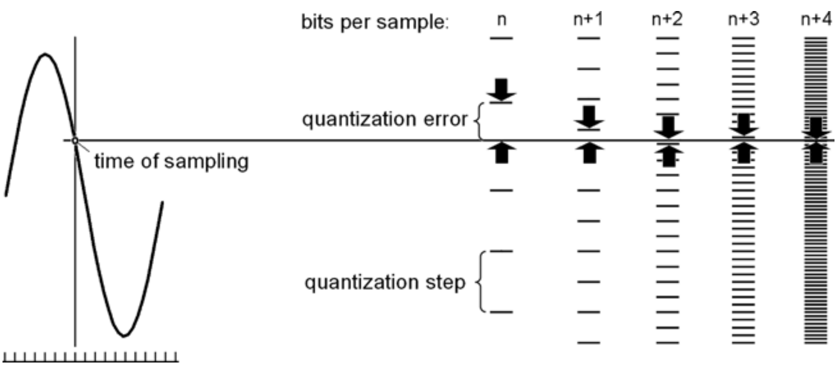

Figure 3.3 With quantization, it is the number of bits that determine the precision of the value read. Each time there is one more bit available, the resolution of the scale doubles, and the error in measurement is halved. In practice, this means that the signal-to-noise ratio improves by approximately 6 dB for each additional bit available.

Table 3.1 Dynamic range of digital formats.

Applications	Sampling Rate [kHz]	Bit Depth	Dynamic Range $dBFS_{rms}$
Audio CD	44.1	16	93
DVD Video	48	16	93
DVD Audio	96–192	24	141
Blu-ray Video	48–96	24	141
Digital Cinema	48–96	24	141
Pure Audio Blu-ray	96–384	24	141

When resolved with additional bits (or by the introduction of dither, see below), this distortion gradually becomes perceived as broadband noise. As mentioned above, the signal-to-noise ratio is estimated to be about 6 dB per bit (see Table 3.1).

A-D

The principal components in the A-D converter are one or more comparators, which compare the instantaneous values of the individual samples with a built-in voltage reference. After the comparison, the comparator's output indicates the value 0 (or "low") if the signal's instantaneous value is less than the reference. If the signal's instantaneous value is equal to or greater than the reference, then the output of the comparator is 1 (or "high").

For serial (sequential) quantization, the comparator first determines the most significant bit, and then the next bit, and so on, until the least significant bit is determined. For a parallel conversion, a comparator is required for each level determined, which for n bits corresponds to 2^{n-1}. If, for example, there are eight bits available for the total signal, this corresponds to a resolution of 256 levels, represented by numerical values in the range 0–255. Written in binary format, this corresponds to the numbers from 00000000 to 11111111.

A form of encoding applied, called two's complement, lets the first digit specify the polarity of the signal. If the number is 0, then a positive voltage value has been sampled. If the number is 1, then a negative voltage value has been sampled. Generating code sample-by-sample is called pulse-code modulation or PCM, which relates to digital audio applying linear quantization, for instance, on a CD. A particular version of PCM is DXD, with a sampling rate of 352.8 kHz and a resolution of 24 bits.

DELTA-SIGMA AND BITSTREAM

One particular and very common conversion technique is ΔΣ- (Delta-Sigma). Here, one single bit is determined at a time but at a very high clock frequency, typically a sampling frequency in the MHz range. The sampling frequency is at least the number of bits per sample (e.g., 16) times faster than a corresponding conventional converter. A frequency corresponding to four times oversampling is commonly used. For a 16-bit system, it then becomes a clock frequency or sampling frequency of 2.8224 MHz. The converter may be constructed in such a way that an integrator sums the values coming out of a comparator. The built-in value of the integrator is compared with the signal at the time of sampling. It is determined whether the signal is now larger or smaller than the summed value. If the value is larger, the output is set to “1.” If the value is smaller, the value is set to “0.”

If an error occurs (say, reading 0 instead of 1) at the moment of comparison, the error never exceeds the smallest interval (in contrast to other converters, where the error can cause large level deviation). Since the comparator does not compare with predetermined levels but with the summed value, some noise is generated along with the signal. This noise can act as dither (see later discussion), but the noise can also be unfortunate, as it can form unwanted tones, so-called idle tones.

When the Delta-Sigma converter signal is to be stored or transmitted, the bitstream can be converted to normal PCM encoding (see later discussion), such as it is on a CD or, for instance, a broadcast server for playout. Decimation is used for this purpose.

Direct Stream Digital (DSD), used on the Super Audio Compact Disc (SACD), uses a Delta-Sigma bitstream which is applied directly to the disc without first being formed into sample values with a certain number of bits and ends up with a sampling frequency of 2.8224 MHz.

PDM

A PDM (Pulse Density Modulation) bitstream is encoded using Delta-Sigma modulation. A one-bit quantizer generates either a 1 or 0 depending on the amplitude of the analog signal (see Figure 3.4). A 1 or 0 corresponds to a full positive or a full negative signal, respectively. Because analog signals are rarely fully positive or fully negative, a quantization error occurs. It is the difference between the 1 or 0, and the actual amplitude it represents. This error is used to feed back negatively in the Delta-Sigma process loop. This has the effect of averaging out the quantization error.

PDM is applied in Micro Electro Mechanical System (MEMS)-microphones and many digital power amplifiers.

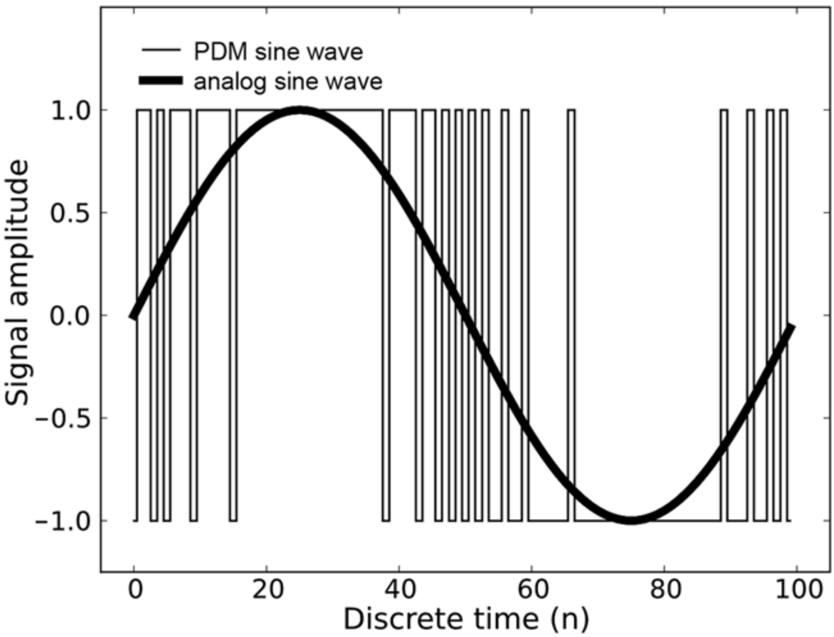

Figure 3.4 Principle of PDM, pulse-density modulation.

FLOATING POINT

In digital audio, full scale indicates the maximum signal you can pass. So the top of the dynamic range of the digitized signal is described by the max full-scale level, and the bottom is defined by the finite number of bits provided for this quantization. However, applying the floating-point technique provides an opportunity to record safely, even if the signal is out of scale.

By the floating-point technique, the sampled value is converted to a number (the mantissa) and an exponent. The files can be scaled, providing an effective dynamic range that is huge. The fundamental signal is preserved, even when scaling up or down. It is a feature found in most DAWs (Digital Audio Workstations) for recording. As a default format, any recorded audio is converted to – and stored as – 32-bit floating-point data.

In the DAW, it leaves you with the possibility of recording a heavily overloaded signal. When you open the file, it is all there as long as you keep it in the floating format. However, if you store in a non-floating format, then any overloaded signal, of course, is clipped at the full-scale level.

DITHERING AND NOISE SHAPING

Dither means trembling. The purpose of dither is to reduce distortion at low levels.

An analog signal may be digitized in a 24-bit converter. Each sample, therefore, consists of 24 bits. Subsequently, the signal must be transferred to, for instance, a playout system that uses only 16 bits. The simple way to reduce to 16 bits is to remove the last 8 least significant bits. It is called truncation." To benefit from the values below the 16th bit, rounding can be an option. But the addition of dither is better. Dither is like random noise that is added to the signal at a level just below the truncation level.

By adding random noise, the next significant bit (in this case bit no. 17) becomes significant in that, along with the random noise, bit 16 has an average value that is more accurate than achieved by simple truncation. The disadvantage is that the signal-to-noise ratio is impaired. However, this is improved by introducing noise shaping; shaping the noise spectrum, so it leaves the main energy of the noise in a frequency range that is less audible but still reduces distortion. Typically, the noise is moved to the 10–20 kHz region. The result: Less distortion and the added noise is moved out of the audible range (unless played with extraordinary gain). For example, when using a PA system, the premise of noise-shaped dither is a noise floor close to the threshold of hearing. This premise is no longer valid.

There are various methods for this with different types of noise. Different brands have special names. Bit mapping is one of them. The Delta-Sigma converter, as mentioned, will generate a kind of dither by oversampling itself. The problem may be that the noise generated can have a frequency distribution that is not always quite appropriate.

D-A

In the conversion of PCM from digital-to-analog, the objective is to produce a signal that is proportional to the value contained in the digital numerical information. Each bit represents a voltage source. The most significant bit (MSB) converts into the largest voltage; the next most significant bit converts into half of that voltage, and so on until the least significant bit (LSB) is included. By summing all the voltage steps, and by holding each summed value until the next sample takes over provides a continuous signal. The signal created is then smoothed out by applying a low-pass filter (see Figure 3.5).

The D-A conversion is, in principle, quite straightforward; however, it can be difficult to control in the real world, where, for example, $2^{16} = 65{,}536$ different levels are available for 16-bit PCM. There can certainly be differences in the quality of A-D

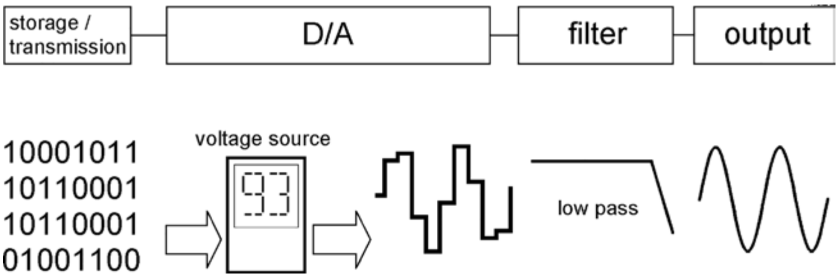

Figure 3.5 During digital-to-analog conversion, the stored numbers are converted back to analog values. Essentially, the numbers are read into the programmable power supply so that they recreate the corresponding voltage steps. The low-pass filter smoothes out the signal by removing the harmonic overtones (caused by the steps) lying above the desired frequency spectrum.

converters in practice. Poor converters can have a DC offset and poor linearity in their dynamics. Methods exist, however, to reduce these problems.

The conversion of the PDM-stream, in general, is simple, as it just needs low-pass filtering for the creation of the analog signal.

SAVING BITS

The number of bits per sample and the sampling rate are factors for the quality of digital audio. In both cases, the higher the better. However, for some purposes, like sound transmission over the Internet, it may not be possible to transfer the number of bits per second required for linear high-quality audio within a reasonable amount of time.

Therefore, we may introduce compromises, like lowering the number of bits spent per second. This method is called bit reduction," or bit companding (a mixture of the words compressing and expanding). Fundamentally, there are several different methodologies available.

LOSSLESS PACKING

One principle of reducing the number of bits does not throw any information away.

One system is FLAC, Free Lossless Audio Codec, which is very popular due to its fast decoding. As it is a nonproprietary format, several codecs are available. By this method, the stored data are reduced to approximately half the size. Other systems are available but not that widely used.

Another methodology is MLP, Meridian Lossless Packing (aka Packed PCM, PPCM). MLP is similar to zipping a data file. By packing the information, it takes up less space, but the contents are still intact. Dolby HD, Blu-ray, and HD DVD employ MLP.

DTS-HD Master Audio (DTS-HD MA) is a combined lossless/lossy audio codec created by DTS.

LOWER F_S AND FEWER BITS PER SAMPLE

The simplest method is to use a lower sampling frequency and fewer bits per sample; however, this results in deterioration of quality.

NONLINEAR QUANTIZATION

A method applied for many years is nonlinear quantization.” **S**pecifically, the A-law (telephony in Europe) and μ-law (mu-law, telephony in the US) methods are the most widely used variants. These require only 8 bits per sample but effectively give 12 bits of resolution obtained by implementing fine resolution at low levels and an increasingly more coarse resolution as levels get higher (aka logPCM). This methodology often applies in communications, such as for voice mail systems and the like; however, the quality is not good enough for music.

PERCEPTUAL CODING

The dominating methodology is based on psychoacoustics and thus called perceptual coding. It makes use of the fact that the ear does not necessarily hear everything in a complex spectrum. Strong parts of the spectrum may mask weaker parts. The principle is then to discard what is estimated not to be audible. (Read more about masking in Chapter 7: The Ear, Hearing, and Level Perception.)

For perceptual coding of the audio, spectrum analysis is first performed. As one single sample by itself does not contain any information about the frequency at that sampling point, a greater amount of samples are collected, typically 1024. By dividing the audio spectrum into a number of frequency bands, calculations then are performed across these bands. This analysis finds whether audio in the neighboring bands is masking precisely this band according to the perceptual model. The data in masked bands, more or less, are thrown away; the quantizing noise is allowed to rise in the masked region by using a lower resolution in these bands. Also, multiple channels can share the information they have in common. Depending on the algorithms applied, the content eventually is reduced to a few percent of the original size.

One of the drawbacks of all these methodologies is that it takes time to compress the bitstream, and it takes time to expand it again. Time delays in most cases are in the range of 5–20 ms, even though, in some systems, the latency may exceed these values in the transmissions, solely due to the complexity of the algorithms. When performing any signal processing on perceptually coded audio, another problem can arise. The thresholds that might have kept the artifacts at an audible minimum now may change and influence the sound quality perceived. Finally, artifacts could appear in excess of what was estimated when the codec was designed.

CODECS AND APPLICATIONS

There is an overwhelming number of bit reduction algorithms available. Some are initiated by standards organizations, while others are proprietary company standards. The different methods are in general optimized for various applications like download and storage for personal playback devices, Internet media, VoIP, video embedded audio, digital broadcast, and so on. Often new algorithms are based on older versions and may or may not be backward compatible. Perceptual coding is an area of constant development. So the following compressed overview in Table 3.2 can be regarded as a snapshot providing information on a few currently widely used algorithms.

Table 3.2 Some popular formats for perceptually coded audio.

Codec	bit rates kbps	sample rates kHz	filename extension
MP3 (MPEG 1, Layer 3)	32–320	32, 44.1, 48	mp3
MP3 (MPEG 1, Layer 3)	8–160	16, 22.05, 24	mp3
AAC (Advanced Audio Coding)	Variable and dependent on no. of channels (up to 48 channels) High-quality stereo@ 128 kbps High-quality 5.1 @ 320 kbps	8, up to 96	m4a, m4b, mop, m4p, m4r, 3GP, 3gp mp4, aac
HE-AAC v2 (High-Efficiency AAC)	(Up to 48 channels) Good quality music @ 24 kbps High-quality stereo @ 128 kbps.	-	aac, mp4, m4a ("a" indicates audio only)

HOW MUCH SPACE DOES (LINEAR) DIGITAL AUDIO TAKE UP?

When calculating the size of any digital information handled by computers, one must be aware that it is all based on bytes [B], which each contain 8 bits. Thus, the number of bits per sample is calculated as an integer multiplied by the number 8 (1×8, 2×8, 3×8, and so on). The number of bits per sample of linear PCM is either 8 (1 byte), 16 (2 bytes), 24 (3 bytes), or 32 (4 bytes). For internal processing, 64 bits or more may apply.

Because these numbers get large, the use of prefixes gets very handy. The prefix units are defined in the SI system, which uses “k” (kilo), “M” (Mega), “G” (Giga), “T” (Tera), and so on. However, while using the same prefix names, it is the binary definition we apply as soon as we describe file sizes; this is rather confusing!

Here is how to calculate file sizes as they appear on your computer:

1 B = 8 bits

1 kB = 1024 B = 8192 bits

1 MB = 1024 kB = 8,388,608 bits ($\approx 8.39 \times 10^6$ bits)

1 GB = 1024 MB $\approx 8.59 \times 10^9$ bits

1 TB = 1024 GB $\approx 8.8 \times 10^{12}$ bits

Example:

How much storage capacity is needed for a 1-hour stereo recording in 44.1 kHz/16 bit?

The total number of bits is calculated as follows:

Sampling frequency x no. of bits per sample × no. of audio channels × the duration of the recording (in seconds):

[1 hour = (60 min. × 60 seconds) = 3600 seconds]

44,100 (samples per second) × 16 (bits per sample) × 2 (channels) × 3600 (seconds) = $5.08 \cdot 10^9$ *bits*

Number of bytes: $5.08 \times 10^9 / 8 = 6.35 \times 10^8$ *B*

Number of kB: $6.35 \times 10^8 / 1024 = 6.20 \times 10^5$ *kB*

Number of MB: $6.20 \times 10^5 / 1024 = 605.6$ *MB*

This amount of data is in the range of the storage capacity of a CD. For this example, additional data, such as the file header and table of contents, is not taken into consideration.

Prefixes for binary-based numbers have existed for many years. Some manufacturers use binary prefixes when specifying their hard drives. Table 3.3 compares the *decimal*

Table 3.3 In this table, the decimal-based prefixes are compared to binary-based prefixes. Unfortunately, it is most common to use decimal prefixes like they were binary.

Decimal				**Binary**					
Value		**SI**		**Value**		**IEC**		**JEDEC**	
factor		**symbol**	**name**	**factor**		**symbol**	**name**	**symbol**	**name**
1000	10^3	k	kilo	1024	2^{10}	Ki	kibi	K	kilo
1000^2	10^6	M	mega	1024^2	2^{20}	Mi	mebi	M	mega
1000^3	10^9	G	giga	1024^3	2^{30}	Gi	gibi	G	giga
1000^4	10^{12}	T	tera	1024^4	2^{40}	Ti	tebi	-	-
1000^5	10^{15}	P	peta	1024^5	2^{50}	Pi	pebi	-	-
1000^6	10^{18}	E	exa	1024^6	2^{60}	Ei	exbi	-	-
1000^7	10^{21}	Z	zetta	1024^7	2^{70}	Zi	zebi	-	-
1000^8	10^{24}	Y	yotta	1024^8	2^{80}	Yi	yobi	-	-

prefixes to the *binary* prefixes, as defined by the IEC (International Electrotechnical Commission) [1] or the JEDEC (Joint Electron Device Engineering Council).

In the example mentioned earlier, the correct calculation yields this result: *605.6 MiB*.

Note also that the Ki uses capital letter K. Sometimes you may see K (without the "i") also meaning Ki.

FURTHER READING

Greenebaum, Ken; Ronen Barzel, eds.: *Audio anecdotes II: Tools, tips, and techniques for digital audio*. A.K. Peters, Wellesley, MA (2004). ISBN 1-56881-214-0.

Pohlmann, Ken, ed.: *Advanced digital audio*. Sams, Carmel, IN (1991) ISBN 0-672-22768-1.

Pohlmann, Ken, ed.: *Principles of digital audio*, 6th ed. McGraw-Hill, New York (2011). ISBN 978-0-07-166346-5.

Rumsey, Francis; Watkinson, John: *The digital interface handbook*, 3rd ed. Focal Press, Oxford (2003). ISBN 0240519094.

Watkinson, John: *The art of digital audio*, 3rd ed. Focal Press, Oxford (2001). ISBN 0-240-51587-0.

Watkinson, John: *The art of sound reproduction*. Focal Press, Oxford (1998). ISBN 0-240-51512-9.

Watkinson, John: *Compression in video & audio*. Focal Press, Oxford (1995) ISBN 0-240-51394-0. [Want to know all about MUSICAM, Dolby AC, JPEG and MPEG? This is your book.]

Watkinson, John: *An introduction to digital audio*. Focal Press, Oxford (1995) ISBN 0-240-51378-9. [Less rigorous than the one above, but no less valuable. An excellent first book on digital audio.]

Watkinson, John: *The MPEG handbook*. Focal Press. ISBN-13: 978-0240805788; ISBN-10: 024080578X.

REFERENCE

[1] IEC 60027-2, Second edition, 2000–11: Letter symbols to be used in electrical technology - Part 2: Telecommunications and electronics.

CHAPTER 4

Signal Types

CHAPTER OUTLINE

A sound is a pressure that varies over time. Represented in electrical form, it is voltage or current that varies over time. This means that every sound – regardless of how many frequencies it might contain – can be described by a signal's variation in time or its waveform.

PURE TONES

Pure tones are characterized by a sinusoidal waveform (see Figure 4.1). These are periodic signals containing one and only one frequency. In practice, pure tones nearly always occur as test signals only.

COMPLEX TONES

Complex tones are periodic signals that consist of a fundamental frequency with associated harmonic overtones. These signals are, among other things, characteristic of musical instruments. In musical contexts, overtones are referred to as harmonics or partials. When one refers to the frequency of a tone, what is actually referred to is the frequency of the fundamental frequency or the lowest partial.

The frequency of the harmonic overtones in this case is integer multiples of the frequency of the fundamental tone. If the frequency of the fundamental tone is 100 Hz, then the frequency of the second harmonic (second overtone) is 200 Hz, the third harmonic 300 Hz, the fourth harmonic 400 Hz, and so on.

Besides frequency and magnitude, we also use phase to describe a complex tone. The phase component of a signal expresses how much a sinewave is delayed compared to a reference sinewave with the same frequency. The phase is only relevant if it is relative to a reference (see Figure 4.2).

The phase is described by an angle (0°–360°), as we also here refer to the unit-circle which we know from trigonometry. If two – same frequency – sine waves are "in phase" it means that there is no delay between them. Or we can say that the phase

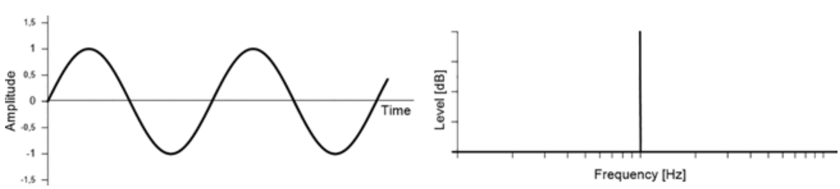

Figure 4.1 A pure (sinusoidal) tone contains one and only one frequency.

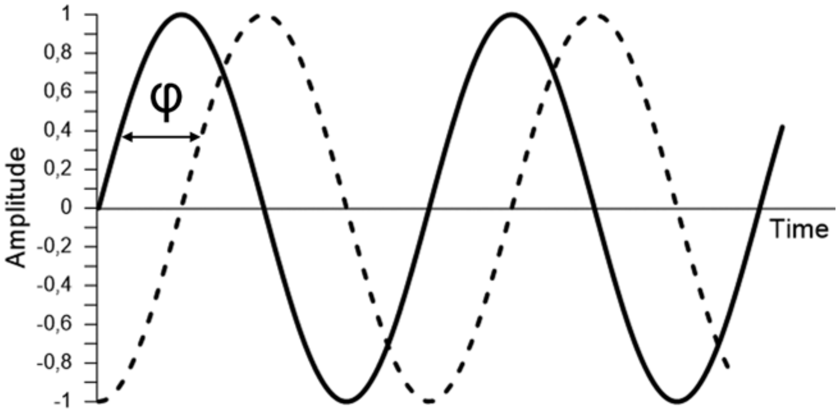

Figure 4.2 The phase is always relative to a reference.

angle = 0°. If the two sines are "out of phase," the phase angle is ≠ 0°, or the one is delayed compared to the other.

Sometimes the term "out of phase" is (wrongly) understood as two oppositely phased signals, one signal being an inverted version of the other. However, there is no delay between them. The phase angle is said to be 180°, or one signal has the inverted polarity of the other. On the input section of many mixing consoles, we can find a small button with the Greek letter "Φ" (phi). By pushing this button, the input signal is inverted – the polarity is swapped.

In general, when analyzing audio signals (or vibration signals) the frequency and the magnitude is called the "real part," and the phase is called the "imaginary part."

SPECIAL WAVEFORMS

Among the periodic waveforms, there are characteristic waveforms such as sinusoid, sawtooth, square, triangular, and pulse train (or pulse string) waveforms.

The sawtooth wave contains the fundamental frequency and all harmonics in a specific proportion corresponding to the inverse of the number of the harmonic (i.e., the second harmonic has 1/2 × amplitude of the fundamental, the third harmonic has 1/3 × amplitude of the fundamental, and so on.). Further, all harmonics are added in phase; at waveform zero-crossings of the complex tone, all the individual harmonic tones are crossing zero as well (see Figure 4.3).

The square wave contains the fundamental frequency and all the odd-numbered harmonics in the same declining proportion as the sawtooth wave. Further, as with the sawtooth wave, all harmonics are in phase.

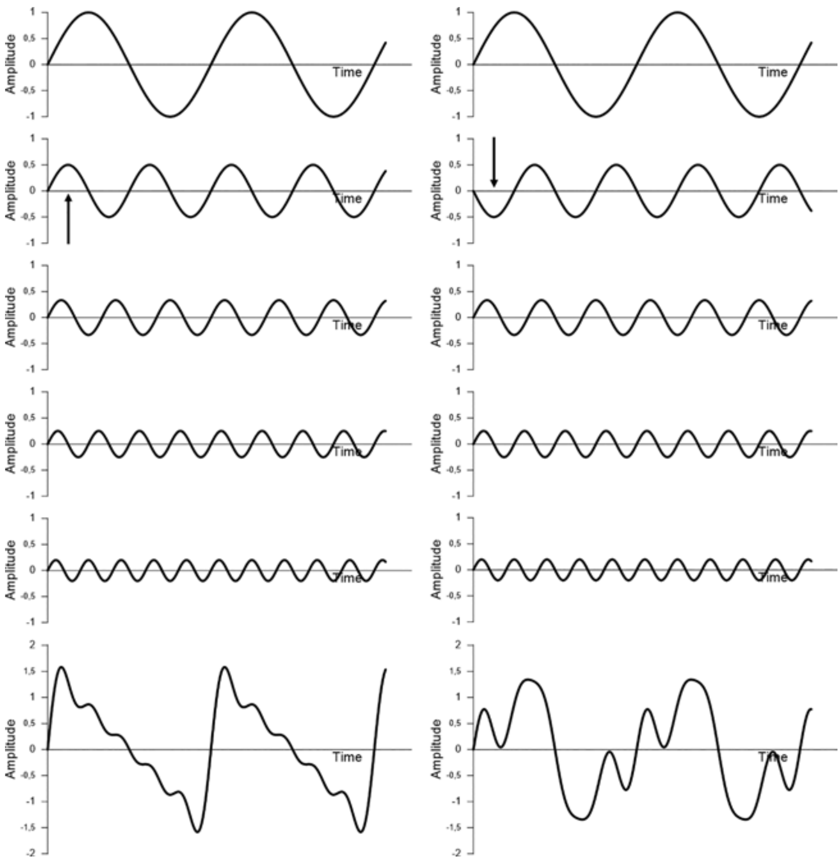

Figure 4.3 To the left, five pure tones are shown corresponding to a fundamental tone, the associated second harmonic overtone (with an amplitude = 1/2 × the fundamental tone), a third harmonic overtone (with an amplitude = 1/3 × the fundamental tone), a fourth harmonic overtone (amplitude = 1/4 × the fundamental tone), and a fifth harmonic overtone (amplitude = 1/5 × the fundamental tone). At the bottom, these five pure tones are added. The waveform approximates a sawtooth waveform. Note that all the curves here start at 0°. If the phase angles are different, then the resultant waveform would also have a different appearance, as shown to the right: A phase shift of 180° is introduced to the second harmonic. Now the resultant waveform is different despite the fact that frequencies and their magnitudes are identical.

The square wave is known from, among other places, electroacoustic equipment that is overloaded and thus "clip" the peaks (positive and negative) of the waveform. Clipping is one way in which harmonic distortion can occur. The magnitudes of the harmonic overtones created are stated with reference to the fundamental frequency and quantified as a percentage of the fundamental (i.e., total harmonic distortion (THD) equals *xx %*).

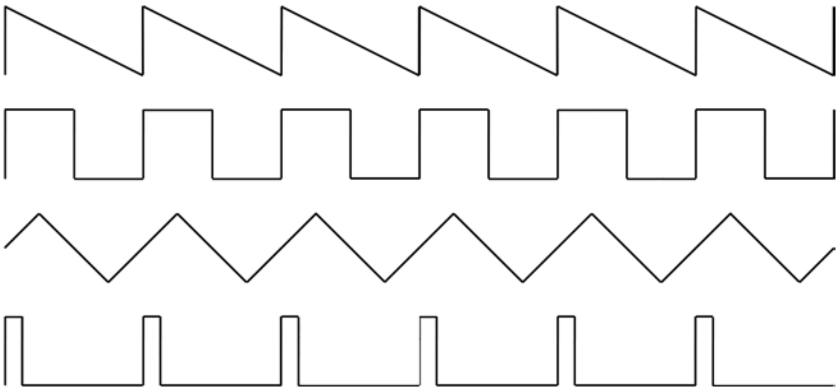

Figure 4.4 Special periodic, continuous waveforms: sawtooth, square, triangular, and pulse train.

The triangular wave contains the fundamental frequency and all the odd-numbered harmonics. However, the polarity of every second of the harmonics present (3rd harmonic, 7th harmonic, 11th harmonic, and so on) is inverted. The declining of the level of the harmonics corresponds to the inverse square of the harmonic number (i.e., 1/9 × 3rd harmonic, 1/25 × 5th harmonic, 1/49 × 7th harmonic, and so on) (see Figure 4.4).

NOISE SIGNALS

Those random signals for which no sensation of tone occurs are called noise. These are characterized by containing all frequencies in a given range. Some well-defined electrical (and acoustical) noise signals belong to this group (see Figure 4.5). Typically, these signals apply for audio tests:

White noise is a signal that contains all frequencies in the audible range, exhibiting constant energy per Hz bandwidth.

Pink noise is a signal that contains all frequencies in the audible range, exhibiting constant energy per octave band (or any fraction of octave band, for instance, 1/3 octave). The amplitude follows the function 1/f, which corresponds to it diminishing by 3 dB per octave or 10 dB per decade. The crest factor (see Chapter 5: How Large Is an Audio Signal?) of a standard pink noise signal for measurements is 4 (12 dB).

Under certain circumstances only a fraction of the pink noise spectrum applies. For instance, pink noise in the range 500 Hz–2 kHz applies to in-room alignment of monitoring levels.

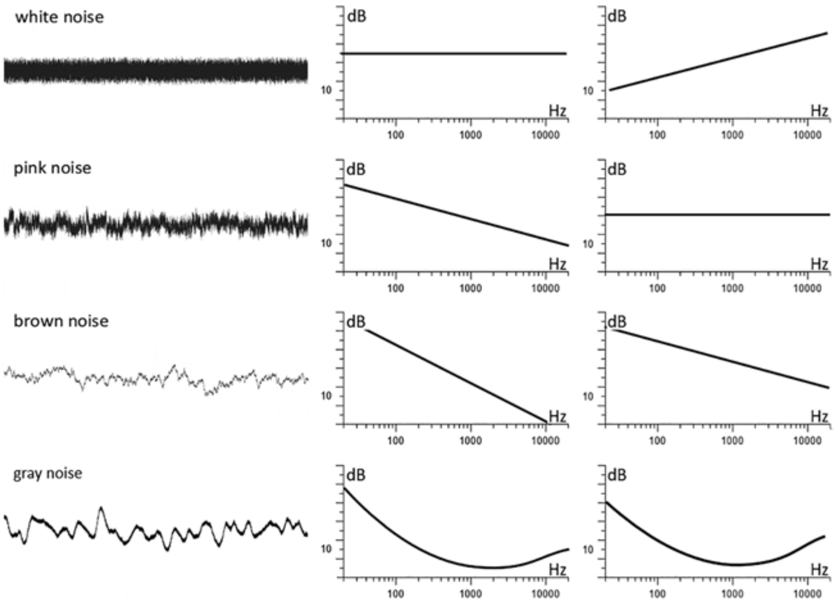

Figure 4.5 Left column: Waveforms of white, pink, brown, and gray noise (1-second excerpts). Middle column: FFT spectra (constant bandwidth of filters). Right column: 1/3-octave spectra (constant relative bandwidth of filters).

Brown noise is a signal that contains all frequencies within the audible range, exhibiting an amplitude vs. frequency that diminishes by $1/f^2$, corresponding to 6 dB per octave or 20 dB per decade.

Gray noise is a signal that also contains random noise. However, the noise has a frequency weighting or "shaping" that relates to the equal loudness curves. The frequency weighting corresponds to an inverse A-weighting (see Chapter 9: Frequency Weighting, and Filters).

Please note that different noise generators do not necessarily produce identical signals despite the fact that the signals are named the same. For instance, pink noise is widely applied for measurements; however, various sources of pink noise may exhibit differences regarding bandwidth and crest factor (see Chapter 5: How Large Is an Audio Signal?).

M-noise," introduced by Meyer Sound, is a mathematically derived test signal that emulates the dynamic characteristics of music noise. It is similar to pink noise below 500 Hz but has a gradually increasing crest factor above this frequency (from 4 (12 dB) to 10 (20 dB)). M-noise is designed especially for testing the maximum linear output of loudspeaker systems for sound reinforcement and public address systems (SR/PA).

Real noise signals such as traffic noise or ventilation noise can also contain audible tones in addition to broadband noise. The primary content often only occupies a limited part of the frequency spectrum. For example, ventilation noise contains a primary content of low frequencies; compressed air noise has a primary content at higher frequencies. Noise signals can also be a part of the sound of musical instruments, such as the "resin sound" of the strings or the air noise of various wind instruments.

THE VOICE AS A SOUND SOURCE

While language can be something that groups of people have in common, the sound and character of the voice is unique from person to person. Our familiarity with speech, regarded as an acoustic signal, allows us to have a good reference as to how it should sound.

VOICE LEVEL

The level of the voice can vary from subdued to shouting depending on the vocal effort. The level is, of course, individual from person to person and thus difficult to assign a number to. The values in Table 4.1 represent the long-term (unweighted) sound level of speech of an adult male.

The level of speech is normally found by measurements averaging several words or sentences. Longer sequences are preferred (like 1 minute or more). Pauses do not count. Depending on the purpose, both linear (unweighted) and A-weighted levels are applied.

The ability to understand speech is optimum when the level of the speech corresponds to normal speech at a distance of 1 meter, in other words, a sound pressure

Table 4.1 Average speech level (unweighted) as a function of listening/recording distance.

	Speech level [dB re 20 μPa]			
Listening distance [m]	**Normal**	**Raised**	**Loud**	**Shout**
0.25	70	76	82	88
0.5	65	71	77	83
1.0	58	64	70	76
1.5	55	61	67	73
2.0	52	58	64	70
3.0	50	56	62	68
5.0	45	51	57	63

level of approximately 58–60 dB re 20 μPa. ("re" means "with reference to." Here, it is with reference to the weakest audible sound pressure, around 1 kHz.) (See Chapter 6: The dB Concept.)

When we speak, and background noise gets higher, we tend to raise the level (and the pitch) of the voice. We do this involuntarily to enhance the audibility of the voice. This special behavior is called the **Lombard effect** or the Lombard reflex (named after the French otolaryngologist Étienne Lombard, who discovered the phenomenon in 1909).

The sound pressure levels in singing are somewhat different. The level often depends on the effort and training. Further, it may also depend on the note that the singer is aiming for. Singing a scale, from the lowest possible note to the highest, exhibits an increase in the level of approximately 12–13 dB/octave unless the singer pays special attention to keeping the level as constant as possible. It should also be mentioned that extremely loud and trained singers – measured at the lip plane – produce peak levels above 150 dB re 20 μPa.

THE SPECTRUM OF SPEECH

The spectrum of speech covers a large part of the total audio frequency range. Speech (Western languages) consists of vowels and consonant sounds. The vocal cords generate the voiced sounds of speech. The resonances of the vocal and the nasal cavities they pass shape the spectrum. The voiced sounds of speech are recognized as complex tones defined by a fundamental frequency and several harmonics. A whisper does not contain voiced sounds; however, the cavities that contribute to the formation of the different vowels still act on the passing flow of air. This acoustical filtering of the airflow is how some characteristics of vowels may occur in a whisper.

In general, the mean value of the fundamental frequency – also called the pitch or f_0 – is in the range of 110–130 Hz for men. In the same way, the fundamental frequency is approximately one octave higher at 200–230 Hz for women. In both cases, values outside these ranges occur. For children, f_0 lies at around 300 Hz.

Different derived values are defined in relation to f_0. The mean value of f_0 is the average frequency measured for normal conversation. The baseline f_0 (f_{base}) is the lower frequency at which the voice returns from higher frequency excursion. It is like finding the keynote of a melody. Alternatively, the baseline can be defined as the frequency from which any utterance starts. A definition given by Jonas Lindh suggests that the baseline is the frequency below which 7.64% of the f_0 values fall. The individual deviation from the base frequency is higher in tone languages (like Chinese) than in European languages like English or German. It should be mentioned that the mood of the speaker, the vocal effort, the conversation partner, the background noise, and much more affects the pitch/f_0 (see Figure 4.6).

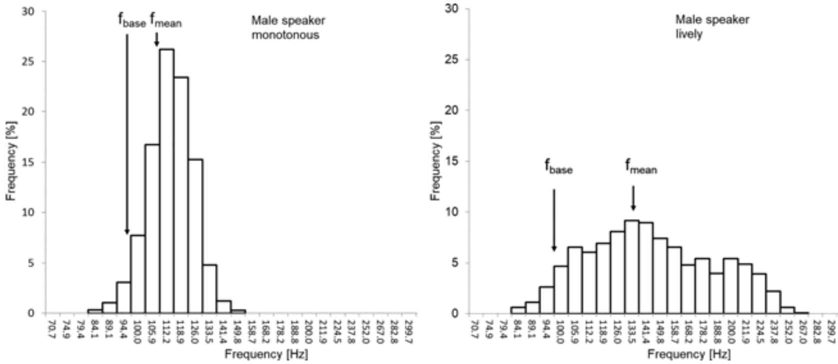

Figure 4.6 Two histograms illustrating the f_0-distribution of a given same speaker utterance. The first (left) is a very monotone voice; the second (right) is very lively. The f_{base} and the mean values are indicated. Note that the mean values are different, but the f_{base} is approximately the same.

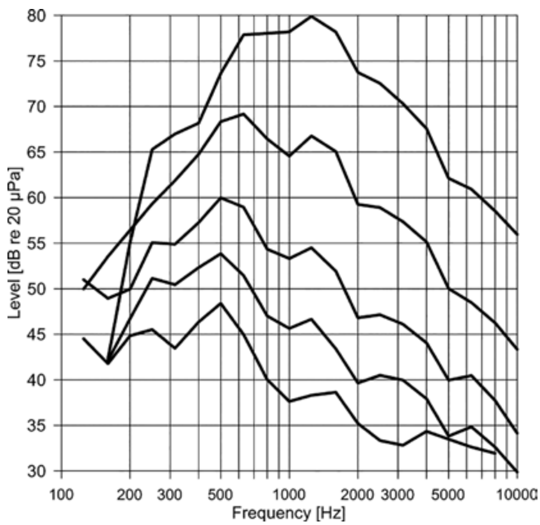

Figure 4.7 Male voices: average 1/3-octave spectra for various levels of speech. Note that the energy moves toward higher frequencies as the sound level of the voice is gradually increased.

The consonants are formed by air blockages and noise sounds created by the passage of air in the throat and mouth, and particularly the tongue and lips. In terms of frequency, the consonants predominantly lie above 500 Hz.

At a normal vocal intensity, the energy of the vowels normally diminishes rapidly above approximately 1 kHz. Note, however, that the emphasis in the speech spectrum shifts one to two octaves toward higher frequencies when raising the level of the voice (see Figures 4.7 and 4.8). One should also note that it is not possible to increase the

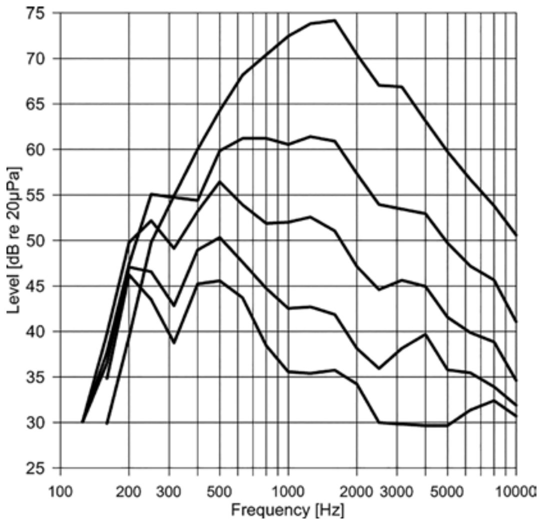

Figure 4.8 Female voices: average 1/3-octave spectra for various levels of speech.

sound level of consonants to the same extent as of vowels. In practice, this means that the intelligibility of speech is not increased by shouting in comparison to using a normal voice (assuming the background noise is not significant). It is also worth knowing when recording speech that the speech spectrum to some degree changes with distance and recording angle.

DIRECTIVITY OF THE VOICE

The radiation pattern of the voice depends on the position of the voice cords, the vocal, and the nasal tract and the size of the head. By and large, this gives us a directivity, that very much is like a small sphere-shaped loudspeaker box: lower frequencies propagate more or less in all directions, and the high frequencies (higher harmonics of the voice and consonants) are pretty much directional. Figure 4.9 is a diagram showing the directional pattern.

FORMANTS

If one listens to two or more people who are speaking or singing the same vowel at the same pitch (f_0), the correct vowel is presumably recognized in all cases, although there is a difference in terms of timbre.

The characteristics of the individual vowels are formed by the resonances of the vocal and the nasal tracts (see Figure 4.10). The cavities act like acoustic filtering of

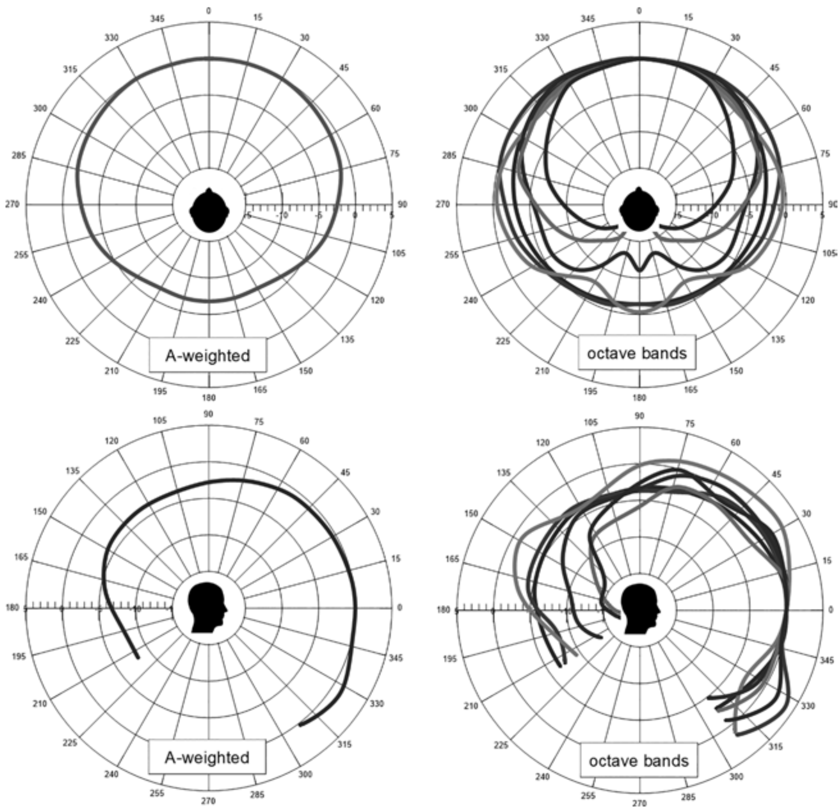

Figure 4.9 Directivity of the voice.

the glottis spectrum, the spectrum generated by the vocal cords. The result is several formants (i.e., frequency ranges that are particularly prominent) that provide the sound of the vowels, more or less regardless of the pitch of the voice (see Figure 4.11). However, each person produces speech with characteristics due to the individual anatomical differences. These differences are, along with other phenomena such as voice quality and intonation, part of what makes it possible to differentiate one voice from another.

CREST FACTOR

The **crest factor** expresses the ratio between the peak level and the RMS level (see Chapter 5: How Large Is an Audio Signal?). The consonants can have relatively strong peaks but limited energy content when viewed over a longer time. You can convince yourself of this by comparing the amplitude reading on a fast meter like TP (true peak)

Figure 4.10 The position of the tongue is shown here for four vowels.

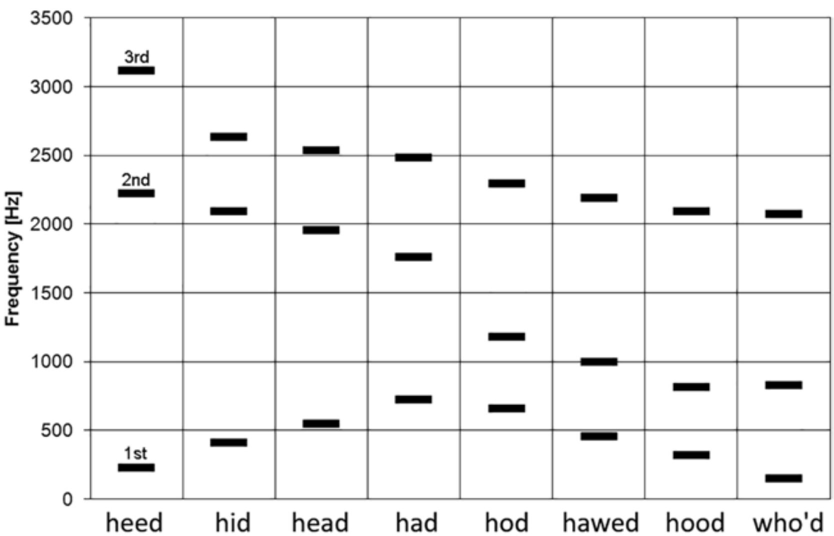

Figure 4.11 Placement of the formants for different vowels in English. The formants are marked from the bottom (lowest frequency) as 1st formant, 2nd formant, and 3rd formant.

or PPM (peak program meter) and a slower meter like VU (volume indicator) or LU (loudness unit), respectively. The crest factor is typically 10–15 (20–23 dB); this has significance when a voice is recorded or reproduced in electroacoustic systems. Give peaks a chance! The survival of the consonants is essential for speech intelligibility.

MUSICAL INSTRUMENTS

The sound of a musical instrument is characterized by parameters such as pitch, tone, timbre, range of harmonics, overtones, attack, decay, and formants.

TONE

The tone of a musical instrument essentially consists of a fundamental frequency and several harmonics and overtones/partials. Characteristic noise sounds add to this base.

The magnitude of the fundamental frequency may be very small compared to the rest of the spectrum. However, due to ears' ability to define the fundamental frequency based upon the interval between harmonics, in general, the pitch perception is correct. It is important to distinguish between the definition of harmonics and overtones or partials when looking at the acoustics of musical instruments. Harmonics are always a mathematical multiple of the fundamental frequency. However, overtones are not always necessarily harmonic; in this case, they are called inharmonic overtones.

Depending on the type of instrument, the tone is characterized by the attack and by decay (see Figure 4.12). Each partial may have different attack and decay patterns. Percussion instruments do not normally exhibit a sustained tone; however, by the stroke, a resonance is excited.

The noise attached as a part of the characteristic sound is common for almost all musical instruments. Examples of this are the sound of the air leakage on the mouthpiece, the sound of resin on the bowed string, and the sound of different instruments' moving mechanical parts.

Like the voice, many musical instruments also have formants, which relates to the physics of the instruments. For instance, a bassoon consists of a tube of a given

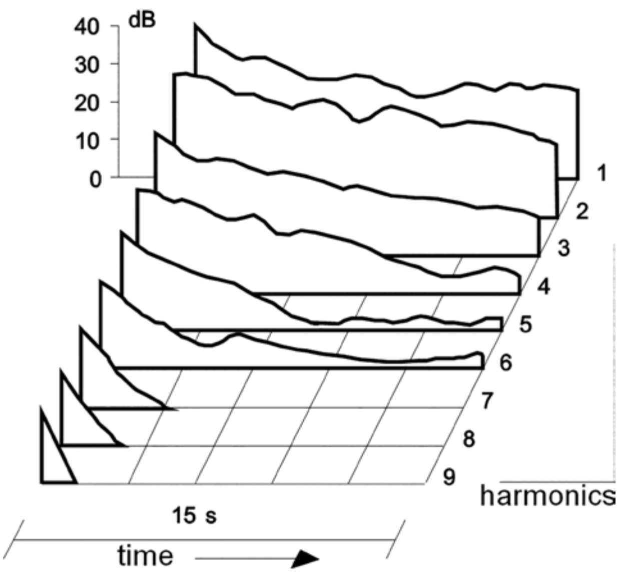

Figure 4.12 The buildup, attack, and decay of individual harmonics (piano).

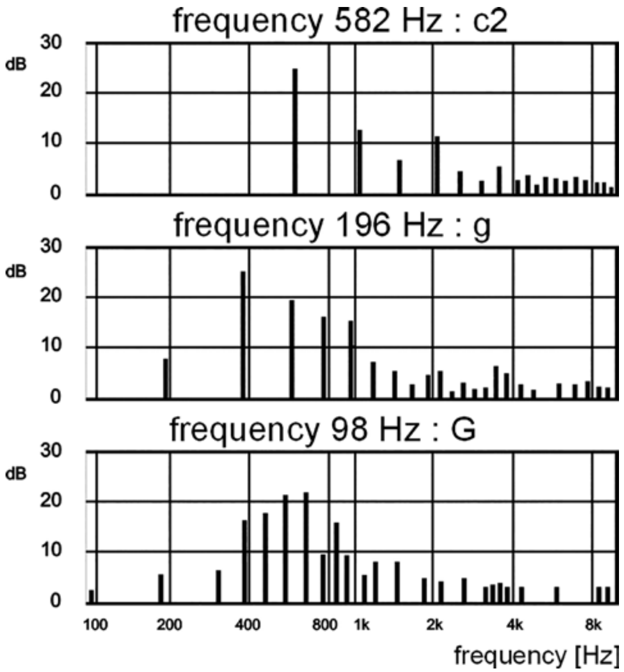

Figure 4.13 The spectra of three different notes on the bassoon. At the lowest note, the fundamental frequency is almost not present. Notice the formant range, 300–600 Hz.

length. This "pipe" provides a natural emphasis on a given frequency range, in this case, around 300–600 Hz (see Figure 4.13). No matter which note played, the harmonics within this frequency band are always the strongest. The specific formants are regarded as the typical range of the instrument when filtered in a complex downmix to avoid "mud" in the mix but still keep the characteristics of the individual instruments. In addition to the characteristics of the instrument itself, obviously the musicians' personal style and playing techniques also play a role.

ACOUSTIC MEASURES OF MUSICAL INSTRUMENTS

Table 4.2 contains some acoustical characteristics for a selection of musical instruments.

Table 4.2 Acoustical properties of musical instruments.

Instrument	Tonal range		Frequency content up to:	Typical max SPL, 3 m	SPL dB re 20μPa	Dynamic range	Attack time
	Lower	Upper	Harmonics/Noise				
Strings							
Violin	g (196 Hz)	g^4 (3136 Hz)	⇒ 10 kHz / 16 kHz	95 dB	At the ear: up to 109 dB	30 dB	5–30 ms
Viola	c (131 Hz)	c^3 (1047 Hz)	⇒ 10 kHz / 16 kHz			30 dB	30–50 ms
Cello	C (65.4 Hz)	c^3 (1047 Hz)	⇒ 8 kHz / 16 kHz			35 dB	Plucked string: 5–30 ms Struck string: 30–50 ms
Double bass	E_1 (41.2 Hz)	g (196 Hz)	⇒ 7.5 kHz / 10 kHz	95 dB		35 dB	Up to 110 ms Hard: up to 110 ms Soft: up to 450 ms
Brass							
Trumpet	e (165 Hz)	d^3 (1175 Hz)	⇒ 15 kHz / 15 kHz	106 dB	0.5 m from bell piece: normal forte: 108 dB extreme forte: 128 dB	Lower tonal range: 30 dB Higher tonal range: 10 dB	20–40 ms
French horn	H_1 (61.7 Hz)	f^2 (698 Hz)	⇒ 10 kHz / 10 kHz	95 dB		Middle tonal range: 40 dB Higher tonal range: 20 dB	20–40 ms In low tonal range: up to 80 ms

(*Continued*)

Table 4.2 (Continued)

Instrument	Tonal range		Frequency content up to:	Typical max SPL, 3 m	SPL dB re 20μPa	Dynamic range	Attack time
Trombone	E (82.4 Hz)	c^2 (523 Hz)	⇒ 7 kHz / 10 kHz	104 dB		Middle tonal range: 45 dB	20–40 ms
Tuba	Eb1 (39 Hz)	g^1 (392 Hz)	⇒ 4 kHz / 7.5 kHz	96 dB		Middle tonal range: 40 dB	20–40 ms
Woodwind							Generally 20–60 ms
Piccolo flute	d^2 (587 Hz)	c^5 (4186 Hz)	⇒ 10 kHz / 15 kHz	102 dB			
Flute	c^1 (262 Hz)	c^4 (2093 Hz)	⇒ 6 k Hz / 15 kHz	96 dB		35 dB	Flute, low register: Up to 180 ms
Oboe	c^1 (262 Hz)	d^3 (1175 Hz)	⇒ 15 kHz / 15 kHz	90 dB		30 dB	Obo, low register: Up to 120 ms
Clarinet	d (147 Hz)	e^3 (1319 Hz)	⇒ 10 kHz / 15 kHz	92 dB		50 dB	
Bassoon	Bb (58.3 Hz)	c^2 (523 Hz)	⇒ 8 kHz / 12 kHz	90 dB		35 dB	
Contra bassoon	Bb_2 (29.2 Hz)	c^1 (262 Hz)		92 dB			
Soprano sax	ab (208 Hz)	eb^3 (1245 Hz)	⇒ 12 kHz / 16 kHz	98 dB			
Alto sax	db (139 Hz)	ab^2 (831 Hz)		98 dB	At the center of the bell piece up to 130 dB		
Tenor sax	Ab (104 Hz)	eb^2 (622 Hz)		98 dB	At the center of the bell piece up to 130 dB	25 dB	
Baritone sax	Db (69.4 Hz)	ab^1 (415 Hz)		98 dB			

Bass sax	Ab_1 (51.9 Hz)	db^1 (278 Hz)		96 dB			
Acoustic guitar (nylon strings)	E (82.4 Hz)	c^3 (1047 Hz)		88 dB			
Acoustic guitar (steel strings)	E (82.4 Hz)	c^3 (1047 Hz)		92 dB			
Grand piano 3 strings: c^5 – F 2 strings: F – F_1	A_2 (27.5 Hz)	c^5 (4186 Hz)		Lid in 45° position: 100 dB			
Percussion	Tuning	Frequency range with the highest level					
Timpani 80 cm	Eb (77.8 Hz) – G (98 Hz)		60 Hz – 200 Hz	110 dB		70 dB	10–18 ms
75 cm	F (87.3 Hz) – c (131 Hz)						
65 cm	A (110 Hz) – f (175 Hz)						
60 cm	c (131 Hz) – ab (208 Hz)						
Bass drum			50 Hz – 400 Hz	115 dB	With pedal, 20 cm from drumhead: 120–128 dB Inside, close to drumhead	80 dB dB	8–12 ms
Snare drum			80 Hz – 4 kHz	108 dB		55 dB	5–8 ms
Cymbal			100 Hz – 5 kHz	105 dB		65 dB	2–4 ms
Sizzling cymbal			100 Hz – 10 kHz	80 dB			
Triangle			2 kHz – 12 kHz				

REFERENCES AND FURTHER READING

Allen, G.D.: Acoustic level and vocal effort as cues for the loudness of speech. *Journal of the Acoustical Society of America* Vol. 49, No. 6B, 1831–1841 (1971).

Brixen, Eddy B.: *Spectral degradation of speech captured by miniature microphones mounted on persons' heads and chests*. AES Convention 100, Copenhagen, Denmark (1996). Preprint 4284.

Brixen, Eddy B.: *Near field registration of the human voice: Spectral changes due to positions*. AES Convention 104, Amsterdam, The Netherlands (1998). Preprint 4728.

Brixen, Eddy B.: *A study of f0 as a function of vocal effort*. Proceedings of the 54th AES International Conference, London (2014).

Brixen, Eddy B.; Sadolin, Cathrine; Kjelin, Henrik: *Acoustical characteristics of vocal modes in singing*. AES Convention 134, Rome, Italy (2013). Preprint 8897.

Chu, W.T.; Warnock, A.C.C.: *Detailed directivity of sound fields around human talkers*. Institute for Research in Construction, National Research Council of Canada, Ottawa (2002).

Coleman, Robert F.; Markham, Ira W.: Normal variations in habitual pitch. *Journal of Voice* Vol. 5, No. 2, 173–177, Raven Press, Ltd., New York (1991).

Dickreiter, M.: *Tonmeister technology*. Temmer Enterprises, Inc., New York (1989). ISBN:3-7776-0388-0.

Fletcher, H.: *Speech and hearing in communication*. Van Nostrand, New York (1953).

Lindh, Jonas: *Fundamental frequency and the alternative baseline in forensic speaker identification*. Proceedings, IAFPA, the College of St. Mark & St. John, Plymouth (2007).

Meyer, Jürgen: *Acoustics and musical performance*. Frankfurt am Main (1972). ISBN:3-920-11256-3.

Olson, H.F.: *Musical engineering*. McGraw-Hill Book Company, Inc, New York (1952).

Pawera, N.: *Microphones*. Arsis Baedeker & Langs Verlag GmbH, Dachau (1981). ISBN 3-88714-007-9.

Pearsons, Karl S.; Bennett, Ricarda L.; Fidell, Sanford: *Speech levels in various noise environments*. Environmental Health Effects Research Series. EPA-600/1-77-025 (1977).

Pulkki, Ville; Karjalainen, Matti: *Communications acoustics*. John Wiley & Sons, Ltd. (2015). ISBN:978-1-118-86654-2.

Smits, Bertrand: *Measurement of the change in voice power level with speaker receiver distance in different environments* (Master thesis) Technical University of Denmark (2009).

Traumnüller, H.; Eriksson, A.: *The frequency range of the voice fundamental in speech of male and female adults. Manuscript*. Stockholms Universitet, Stockholm (1995).

Wilkens, H.; Bartel, H-H.: Wiedererkennbarkeit der Originallautstärke eines Sprechers bei elektroakustischer Wiedergabe. *Acoustica* Vol. 37, 45–49 (1977).

CHAPTER 5

How Large Is an Audio Signal?

CHAPTER OUTLINE

When describing the magnitude of audio signals, comparable methodologies of measurement must be agreed to or there is a risk that differently measured values will not be comparable.

ACOUSTICAL SIGNALS

As a rule, acoustic signals must be measured using transducers such as microphones. The output is an electrical signal that is analogous to the acoustic signal. Thus, the measurement is performed on the electrical signal.

ELECTRICAL SIGNALS

Amplitude values of the waveform can describe an electrical signal for the voltage or current. Alternately, the electrical signal can be defined by the electrical energy it contains, for example, the electrical power that is fed into a given load or during a given period.

When we study the signals' amplitude over time (waveform), we have several terms that apply.

PEAK VALUES

The peak value describes the instantaneous maximum amplitude value within one period of the signal concerned (see Figure 5.1). The peak value can also be the maximum value that is ascertained during any period under consideration, even though the usual designation here involves the maximum value within one period. Usually, the peak value measurement is used with symmetric signals (i.e., signals that deviate equally far from 0 in both a positive and negative direction).

For a sinusoidal wave:

$$\begin{aligned}\text{Peak value} &= \text{Maximum value}\\ &= \sqrt{2}\ \text{RMS value}\\ &= \frac{\pi}{2}\ \text{Average value}\end{aligned}$$

(See the following two sections for discussions of RMS value and average value.) In practical audio engineering, particularly insofar as it concerns circuits for the recording and transmission of audio signals, it is the peak-to-peak value that is significant. This value describes, so to speak, how much "space" (voltage range) the signal "takes

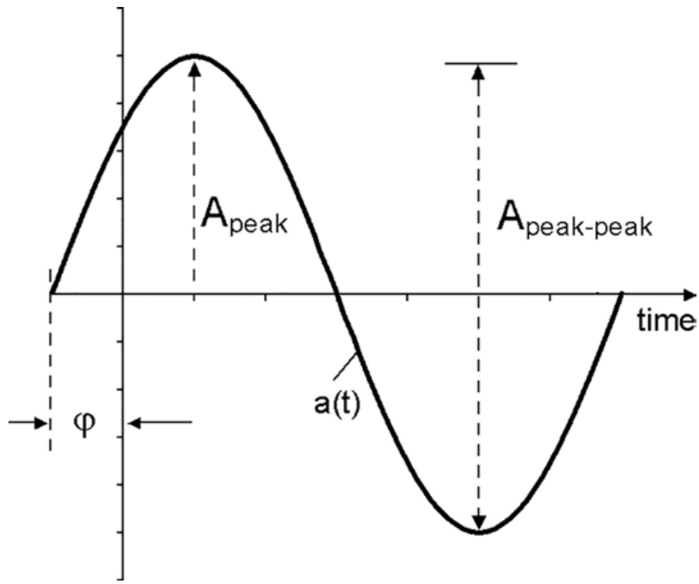

Figure 5.1 Sinewave where the peak value (Up), peak-to-peak value (Up-p) and the phase (ϕ) are indicated.

up" in the electrical circuit. Also, the peak-to-peak value indicates how much a membrane must move in a purely physical sense from one extreme to the other to record or reproduce the sound.

For a sinusoidal wave:

$$\begin{aligned}\text{Peak to Peak value} &= 2 \times \text{Maximum value}\\ &= 2.828 \times \text{RMS value}\end{aligned}$$

In relation to the recording of audio, it is important to know that possible phase shifts, which can arise, for example, in different forms of pre-emphasis or correction networks, can make a signal asymmetric (see Figure 5.2).

Also, there are many acoustic signals that contain an asymmetric wave. The voice, for example, or musical instruments, particularly percussion instruments, where the impact can be highly asymmetric.

AVERAGE VALUE

The average value is based on the average of the numerical values of the amplitude over a period. (Numerical value here means the indicated value without regard to its sign.)

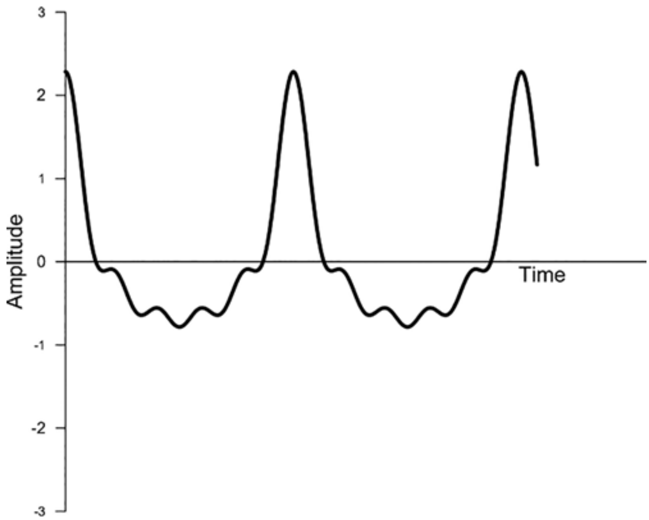

Figure 5.2 Example of a signal that is asymmetric around 0. It contains a fundamental tone and subsequently, four harmonic overtones (symmetric sinusoidal tones). Each of the harmonics has a 90° phase shift in relation to the fundamental frequency. The length of this curve corresponds to two periods of the fundamental frequency.

The average value is calculated according to the following expression:

$$A_{Average} = \frac{1}{T}\int_0^T \left| a \right| dt$$

where

T = period

a = amplitude

For a sinusoidal wave:

$$\begin{aligned}\text{Average value} &= \frac{2\times\text{Maximum value}}{\pi}\\ &= 0.9\times\text{RMS value}\end{aligned}$$

RMS VALUE

The RMS value is based on the energy that is contained in a given signal. RMS stands for root mean square and, in essence, is the square root of the square of the average value (see Figure 5.3 and 5.4).

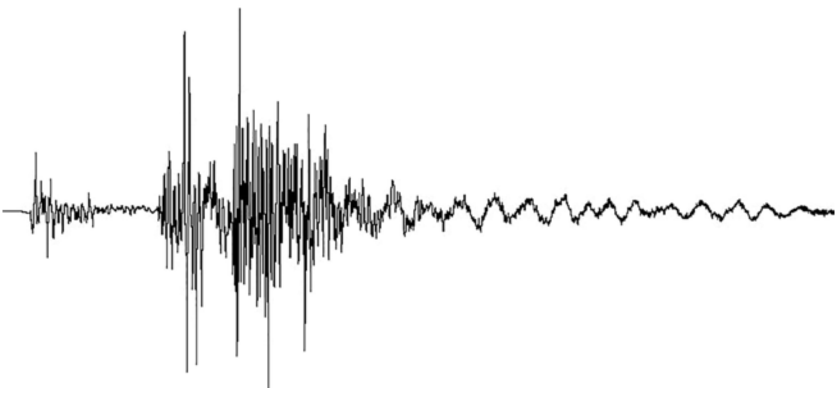

Figure 5.3 Sound impulse with a high peak value, but a low RMS value.

$$A_{RMS} = \sqrt{\frac{1}{T}\int_0^T a^2(t)dt}$$

where

T = period

a = amplitude as a function of time

For a sinusoidal wave:

$$\begin{aligned}\text{RMS value} &= \frac{\text{Maximum value}}{\sqrt{2}}\\ &= 1.11\times\text{Average value}\\ &= 0.707\times\text{Maximum value}\end{aligned}$$

For a square waveform:

$$\text{RMS value} = \text{Maximum value}$$

For a triangular wave:

$$\begin{aligned}\text{RMS value} &= \frac{\text{Maximum value}}{\sqrt{3}}\\ &= 0.576\times\text{Maximum value}\end{aligned}$$

For a half-wave:

$$\begin{aligned}\text{RMS value} &= \frac{\text{Maximum value}}{2\times\sqrt{2}}\\ &= 0.354\times\text{Maximum value}\end{aligned}$$

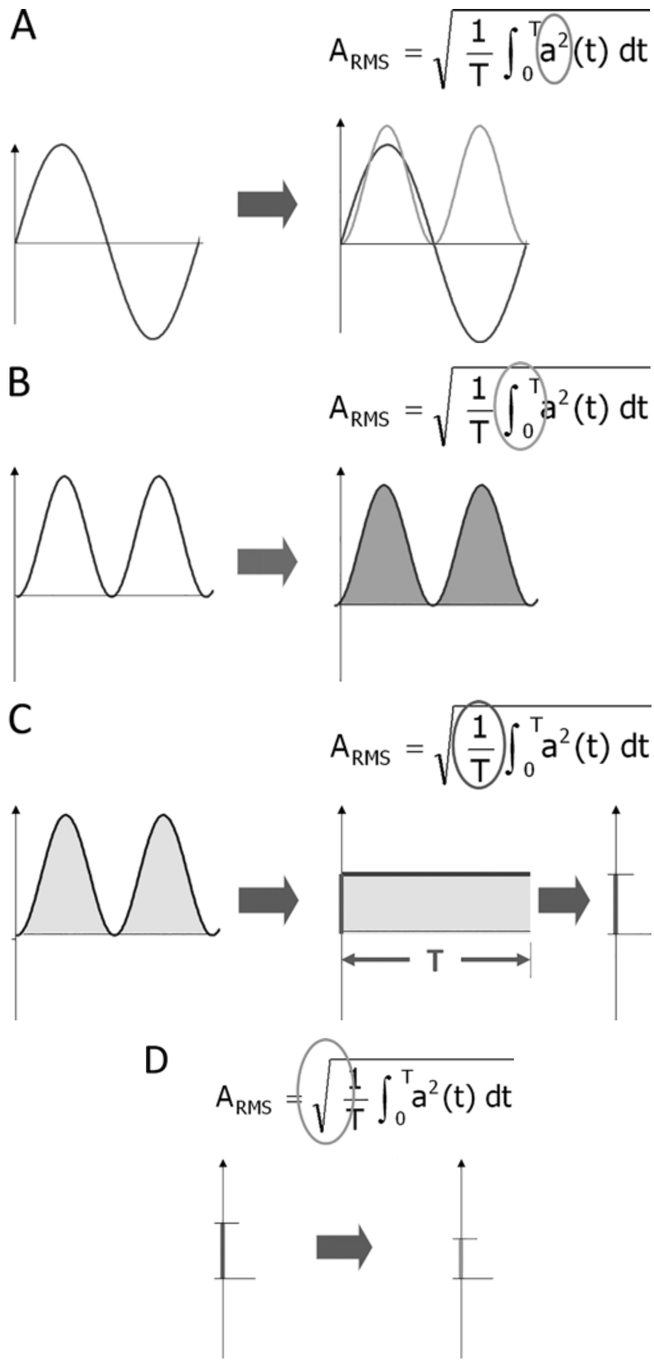

Figure 5.4 This strip explains what happens to the signal when finding its RMS value. A: All values are squared (provides positive figures). B: By integrating the area under the curve, the area is found. C: The area is divided by the time of the period. D: Take the square root, and we have a result: RMS!

CREST FACTOR

The crest factor expresses the relation between the peak value (maximum value) and the RMS value. It is a magnitude that is important to know when recording because the maximum value is an expression of the signal's amplitude, whereas the RMS value typically relates to what the level meter is showing.

$$\text{Crest factor} = \frac{\text{Maximum value}}{\text{RMS value}}$$

For a sinusoidal waveform (see Figure 5.5):

$$\text{Crest factor} = \sqrt{2} = 1.414$$

For a square waveform:

$$\text{Crest factor} = 1$$

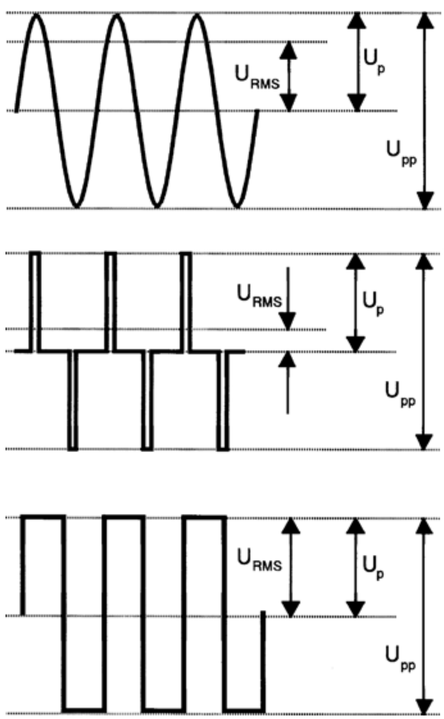

Figure 5.5 The relation between peak, peak-to-peak, and RMS values for different waveforms.

For a triangular waveform:

$$\text{Crest factor} = \sqrt{3} = 1.73$$

For pink noise:

$$\text{Crest factor} \approx 4$$

For (uncompressed) speech:

$$\text{Crest factor} \approx 10$$

FORM FACTOR

The form factor expresses the relation between the signal's RMS value and the average value. The form factor relates to the signal's waveform in that a small form factor is an indication of a flat waveform, and a large form factor is an indication of a waveform containing peaks.

$$\text{Form factor } |\xi| = \frac{\text{RMS value}}{\text{Average value}}$$

For a sinusoidal waveform:

$$\text{Form factor} = \frac{\pi}{2\sqrt{2}} = 1.11$$

For a square waveform:

$$\text{Form factor} = \frac{\pi}{\sqrt{3}} = 1.15$$

CHAPTER 6

The dB Concept

CHAPTER OUTLINE

When sound levels have to be specified, it is practical to do so in a manner that corresponds to how the ear perceives them. Human hearing is approximately logarithmic with respect to the perception of both level and frequency. Logarithmic increments mean that there is a constant ratio between the individual steps on the scale. For example, if 1 W is first imparted to a loudspeaker and then followed by 2 W, this will be experienced as a certain interval. To reach a similar interval, not 3 W, but, rather, 4 W must be added – and 8 W must be added for the next step. Here, there is a ratio of 2 between the individual steps.

In the calculation of dB, the base 10 logarithm is used. The base 10 logarithm of a number is that number to which 10 must be raised, to give the number itself. Hence the logarithm of 100 = 2, because $10^2 = 100$. Originally, the logarithm of the ratio between two power measurements was used (the Bel unit). Later, a tenth of this became preferred (the deciBel unit, abbreviated as dB).

The advantages of the dB unit are that large ratios can be described with the use of a maximum of 3 digits (and then the first being "1"), and changes in a sequence will be the same throughout, expressed in terms of dB, regardless of whether the unit of measurement concerned is the volt, ampere, watt, and so on.

We will first look at the dB value for the ratio between measured power. Power is measured in watts, so this transformation will apply to everything that can be measured in watts, in the form of acoustic as well as electrical power (this transformation will thus not apply to amplitude values such as sound pressure, current, or voltage).

dB; POWER RATIO

The ratio between the two measurements of power is expressed in the following manner:

$$10 \cdot \log \frac{P_1}{P_0}$$

where

P_0= the reference (i.e., the value against which the comparison is being made)

P_1= the relevant value to be specified

Example: I have a new amplifier. The old one was rated at 100 W, and the new one at 200 W. What is the difference in dB?

10 · log (200/100) [dB]

⇓

10 · log 2 [dB]

⇓

$$\mathbf{10 \cdot 0.3\ [dB]}$$
$$\Downarrow$$
$$\mathbf{3\ [dB]}$$

The answer is that there is 3 dB more power in the new amplifier as compared to the old one.

dB; AMPLITUDE RATIO

The ratio between two amplitude values (for example, sound pressure, current, or voltage) has a square relationship to power and is thus expressed in the following manner:

$$10 \cdot \log \frac{(a_1)^2}{(a_0)^2}$$

where

a_0 = the reference (i.e., the value to be compared against)

a_1 = the value to be specified

This expression can be reduced as the squaring is changed to multiplication when the logarithm is extracted:

$$20 \cdot \log \frac{a_1}{a_0}$$

Example: The input sensitivity of my amplifier is 0.775 V. The signal out of my mixer is only 0.3 V. What is the ratio expressed in dB?

$$\mathbf{20 \cdot \log (0.3/0.775)\ [dB]}$$
$$\Downarrow$$
$$\mathbf{20 \cdot \log 0.39\ [dB]}$$
$$\Downarrow$$
$$\mathbf{20 \cdot (-0.41)\ [dB]}$$
$$\Downarrow$$
$$\mathbf{-\ 8.2\ [dB]}$$

The signal is −8.2 dB in relation to the input sensitivity. In other words, the signal is 8.2 dB below the level needed for full modulation of the amplifier.

Note that the dB value is positive when the ratio is above 1 and negative when the ratio is less than 1.

FROM dB TO POWER OR AMPLITUDE RATIO

One can calculate backward from dB to power measurement ratios as follows:

$$\textit{Power ratio} = 10^{(x\ [dB]\ /10)}$$

Similarly, calculate backward from dB to amplitude ratios as follows:

$$\textit{Amplitude ratio} = 10^{(x\ [dB]/20)}$$

Example: The sensitivity of a microphone in a catalog is stated as −56 dB re 1 V. What does that correspond to in terms of voltage?

$$10^{(-56/20)}$$
$$\Downarrow$$
$$10^{(-2.8)}$$
$$\Downarrow$$
$$0.00158 \text{ (or } 1.58 \cdot 10^{-3}\text{)}$$

The result must be multiplied by the reference, which in this case is 1 V.

The answer is thus $1.58 \cdot 10^{-3} \cdot 1$ *[V], which is* $1.58 \cdot 10^{-3}$ *V or 1.58 mV*

CONVERSION TABLE

Table 6.1 shows the conversion between power ratios and dB as well as amplitude ratios and dB.

REFERENCE VALUES

In acoustics and electronics, specific reference values are applied as a basis for the specified ratio in dB. In certain cases, the reference is stated directly, such as in the specification of sound pressure level: 60 dB re 20 μPa. This means that the sound pressure is 60 dB over $20 \cdot 10^{-6}$ pascal, which equates to a sound pressure of 0.02 pascal.

Instead of specifying the reference in full, in some situations one or more letters can be appended after the "dB" to indicate the reference. The following are some examples of this usage:

dBμ: Reference relative to 1 μW (microwatt).

dBd: Antenna gain in relation to a half-wave dipole.

Table 6.1 Conversion between power ratios and amplitude ratios.

dB	Power ratio	Amplitude ratio	dB	Power ratio	Amplitude ratio
0	1	1	0	1	1
1	1.259	1.122	–1	0.794	0.891
2	1.585	1.259	–2	0.631	0.794
3	1.995	1.413	–3	0.501	0.708
4	2.512	1.585	–4	0.398	0.631
5	3.162	1.778	–5	0.316	0.562
6	3.981	1.995	–6	0.251	0.501
7	5.012	2.239	–7	0.200	0.447
8	6.310	2.512	–8	0.158	0.398
9	7.943	2.818	–9	0.126	0.355
10	10	3.162	–10	0.100	0.316
12	15.85	3.981	–12	0.063	0.251
14	25.12	5.012	–14	0.040	0.200
15	31.62	5.623	–15	0.032	0.178
20	100	10	–20	0.010	0.100
26	398.1	19.95	–26	0.003	0.050
30	1000	31.62	–30	0.001	0.032
40	10^{4}	100	–40	10^{-4}	10^{-2}
50	10^{5}	316.2	–50	10^{-5}	$3162 \cdot 10^{-3}$
60	10^{6}	1000	–60	10^{-6}	10^{-3}
70	10^{7}	$3.162 \cdot 10^{3}$	–70	10^{-7}	$3.162 \cdot 10^{-4}$
80	10^{8}	10^{4}	–80	10^{-8}	10^{-4}
90	10^{9}	$3.162 \cdot 10^{4}$	–90	10^{-9}	$3.162 \cdot 10^{-5}$
100	10^{10}	10^{5}	–100	10^{-10}	10^{-5}
110	10^{11}	$3.162 \cdot 10^{5}$	–110	10^{-11}	$3.162 \cdot 10^{-6}$
120	10^{12}	10^{6}	–120	10^{-12}	10^{-6}

dBFS: For digital equipment: reference to “Full Scale.” 0 dBFS is the RMS of a sinewave.

dBi: Antenna gain in relation to an omnidirectional antenna.

dBk: Reference relative to 1 kW (kilowatt).

dBm: Reference relative to 1 mW (milliwatt) imparted at 600 ohms (often used incorrectly for the reference relative to 0.775 V RMS).

dBm0: Absolute power level (re 1 mW/600 ohm) at a point of 0 relative level.

dBm0p: Absolute psophometric (frequency-weighted) power level re 1 mW/600 ohm at a point of 0 relative level.

dBm0ps: Absolute psophometric (frequency-weighted) power level re 1 mW/600 ohm for a point of 0 relative level for program transmission (sound).

dBm0s: Absolute power level re 1 mW/600 ohm for a point of 0 relative level for program transmission (sound).

dBov: dB overload; not a standard, but similar to dBFS.

dBp: Reference relative to 1 pW (picowatt).

dBq0ps: Absolute and weighted voltage level re 0.775 V for a point of 0 relative level for program transmission (sound).

dBq0s: Absolute and unweighted voltage level re 0.775 V for a point of 0 relative level for program transmission (sound).

dBr: Relative specification in relation to an arbitrary reference, which must then be stated.

dBrn: Relative specification in relation to noise floor; used in telecommunications.

dBrs: Relative specification in relation to sound program level.

dB SPL: Sound Pressure Level, with the reference 20 μPA.

dBTP: (True Peak) Reference relative to the true peak of a signal.

dBu: Reference relative to 0.775 V RMS (in general the effective voltage that must lie across a resistance of 600 ohms in order for a power of 1 mW to be imparted); dBu is sometimes used incorrectly for the reference relative to 1 μW.

dBu0: Absolute voltage level re 0.775 V for a point of 0 relative level.

dBu0s: Absolute voltage level re 0.775 V for a point of 0 relative level for program transmission (audio).

dBuv: American, reference relative to 1 μV (should be avoided).

dBuw: American, reference relative to 1 μW (should be avoided).

dBv: Often used in American data sheets for dBu (should be avoided).

dBV: Reference relative to 1 V (volt).

dB W: Reference relative to 1 W (watt).

WEIGHTED MEASUREMENTS

dB(A), dB(B), dB(C), dB(D), dB(G), and dB(Z): These designations do not refer to a specific level but rather to the fact that the signal has been measured with an inserted frequency weighting filter (see Chapter 9: Frequency Weighting and Filters).

ADDITION OF dB

To find the total magnitude of two acoustic sound levels specified in dB, the starting point will normally be that the two sound sources are un-correlated (i.e., either the two signals do not resemble each other, or they were recorded in a diffuse sound field). In practice, it is only in the near field between two loudspeakers that are reproducing the same sound where this condition would not be fulfilled or electrically, when the two signals are identical and in phase.

The total sound level is calculated according to the following expression:

$$L_{p1} + L_{p2} = 10 \cdot \log\left(10^{\frac{L_{p1}}{10}} + 10^{\frac{L_{p2}}{10}}\right)$$

Example

What is 84 dB + 90 dB?

$10 \cdot \log(10^{(84/10)} + 10^{(90/10)}$ [dB]

$\Downarrow$

$10 \cdot \log(0.25 \cdot 10^9 + 10^9)$ [dB]

$\Downarrow$

$10 \cdot \log(1.251 \cdot 10^9)$ [dB]

$\Downarrow$

$10 \cdot 9.1$ [dB]

$\Downarrow$

91 [dB]

NOMOGRAM (ADDITION)

One can avoid the work of performing the calculation by using a nomogram instead (see Figure 6.1).

The difference is found between two sound levels (from the previously mentioned example: 90 dB −84 dB = 6 dB). This number is located on the horizontal axis. Then the vertical line is followed to the curve, after which you move horizontally to the left where the value 1.0 dB can be read. This value is then added to the highest level, which was 90 dB. The result is then (90 + 1.0) = 91.0 dB.

As the nomogram shows, two equally strong sound sources together would comprise a level that is 3 dB louder than the individual source. In practice, if the difference between the two sources is more than 10 dB, then the weakest one can be ignored.

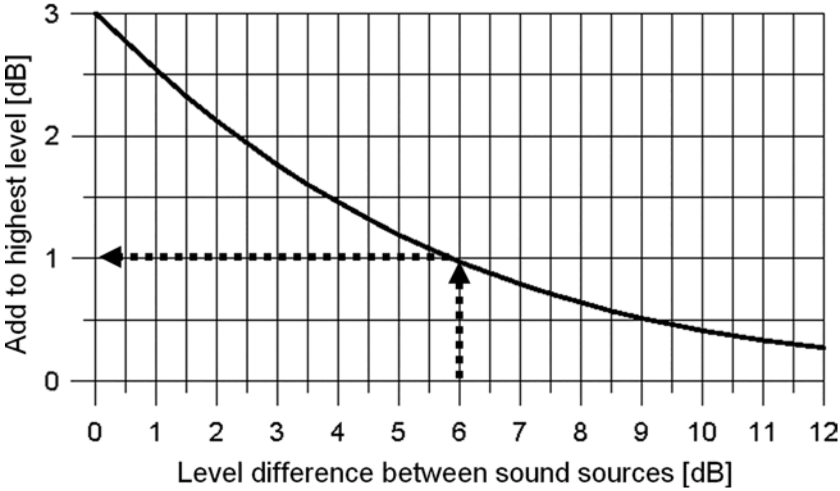

Figure 6.1 Nomogram for the addition of sound levels. The addition is performed on an energy basis and applies for uncorrelated sound sources (the sounds from the sources are different from each other) or sound sources in diffuse sound fields.

SUBTRACTION OF dB

Subtracting one sound level from another is related to the calculation of uncorrelated acoustic sound sources. This is calculated in the following way:

$$L_{p1} - L_{p2} = 10 \cdot \log\left(10^{\frac{L_{p1}}{10}} - 10^{\frac{L_{p2}}{10}}\right)[dB]$$

Example:

The noise level L_{p1} in a server room is 60 dB re 20 µPa. What is the level when one server having a noise level L_{p2} of 57 dB re 20 µPa is removed from the room?

$$10 \cdot \log\left(10^{\frac{60}{10}} - 10^{\frac{57}{10}}\right)[dB]$$

$$\Downarrow$$

$$10 \cdot \log\left(10^{6} - 10^{5.7}\right)[dB]$$

$$\Downarrow$$

$$10 \cdot \log\left(0.5 \cdot 10^{6}\right)[dB]$$

$$\Downarrow$$

$$57[dB]$$

Thus the result is that the noise level is 57 dB re 20 μPa after removing one noise source.

NOMOGRAM (SUBTRACTION)

As with addition, subtraction can be performed using a nomogram (see Figure 6.2).

The difference between the two sound levels is found (from the previously mentioned example: 60 dB −57 dB = 3 dB). This number is located on the horizontal axis. Then the vertical line is followed to the curve, after which you move horizontally to the left where the value −3 dB can be read. This (negative) value is then added to the highest level, which was 60 dB. The result is then (60 + (−3)) = 57 dB.

NOTES

- A change of 1 dB is only just audible.
- A change of 3 dB is a clearly audible change.
- A change of 10 dB corresponds to a subjective doubling of the perceived sound pressure level.
- The addition of electrical and acoustic signals is covered later in Chapter 22 on the summation of audio signals.

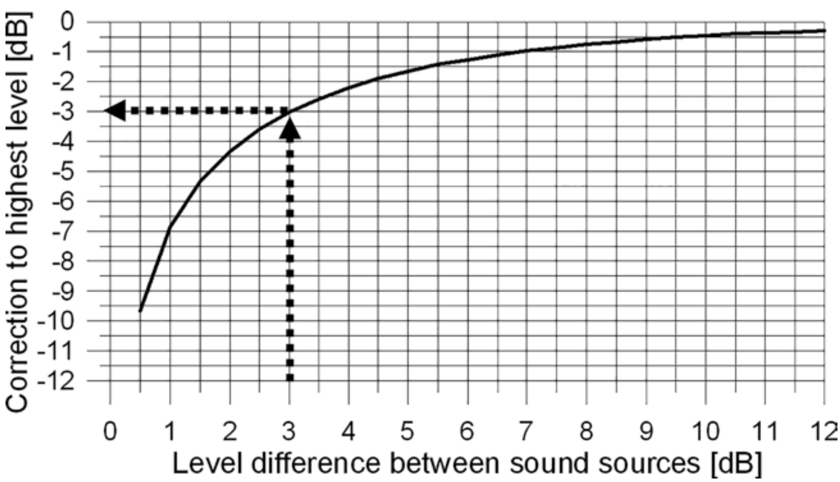

Figure 6.2 Nomogram for the subtraction of sound levels. The subtraction is performed on an energy basis and applies for uncorrelated sound sources (the sounds from the sources are different from each other) or sound sources in diffuse sound fields.

OTHER UNITS IN AUDIO

VU: Understood as Volume Unit; used in audio metering. The change of 1 VU is a change of 1 dB. 0 VU has different definitions; however, +4 dBm is commonly used (see Chapter 12: The Standard Volume Indicator (VU Meter).

NEPER (Np): Used by the telephone companies.

$1\ \mathrm{Np} = (20 \cdot \log e)\ \mathrm{dB} \sim 8.686\ \mathrm{dB}$.

where e is a constant: 2.71828182845904523536 02

$1\ \mathrm{dB} = (0.05 \cdot \ln 10)\ \mathrm{Np} \sim 0.1151\ \mathrm{Np}$

LK (or L_k): Loudness unit. Expresses the loudness of the audio program, applying frequency weighting (K-weighting) and by power summing of all active channels into one single measure. Changing the level by 1 LK is a change of 1 dB. This unit was introduced and standardized by the ITU, primarily for digital television; later applied in general audio production. Equivalent to LU.

LKFS: Loudness, K-weighted, relative to Full Scale. Measured with the equipment specified by ITU-R BS.1770. Equivalent to LUFS.

LU: Loudness Unit. (Equivalent to LK). Expresses the loudness of the program, based on the electrical signal using frequency weighting (K-weighting) and by power summing all active channels to one single measure. Changing the level by 1 LU is a change of 1 dB. This unit is introduced and standardized by the EBU.

LUFS: Loudness, K-weighted, with reference to Full Scale. Equivalent to LKFS. LUFS is preferred by EBU for being compatible with ISO conventions.

PHON: Loudness unit, acoustic measure; equal to the sound pressure in dB re 20 μPa of an equally loud 1 kHz tone.

SONE: Measure of subjective loudness. One sone corresponds to 40 phons. A doubling/halving of the sone value corresponds to 10 phons (for example 2 sones equate to 50 phons).

BARK: A scale applied in psychoacoustics for subjective measurement of loudness. (Named after Heinrich Barkhausen, German scientist, electronic engineering). The scale ranges from 1 to 24 and relates to the critical frequency bands of hearing.

MEL: The mel [abbreviated *m*] scale is a perceptual scale of pitch. The scale is derived from subjective listening, each step on the scale is assessed to be equal in distance. A tone of 2n mel is judged to have the double pitch as one of 1n mel.

The human pitch perception is rather linear up to around – or slightly above – 500 Hz. Above this frequency, increasingly larger intervals are judged to produce equal pitch increments. Since introduced around 1937, different researchers have presented various versions of the scale. In the earliest versions, the frequency 131 Hz (C0) was

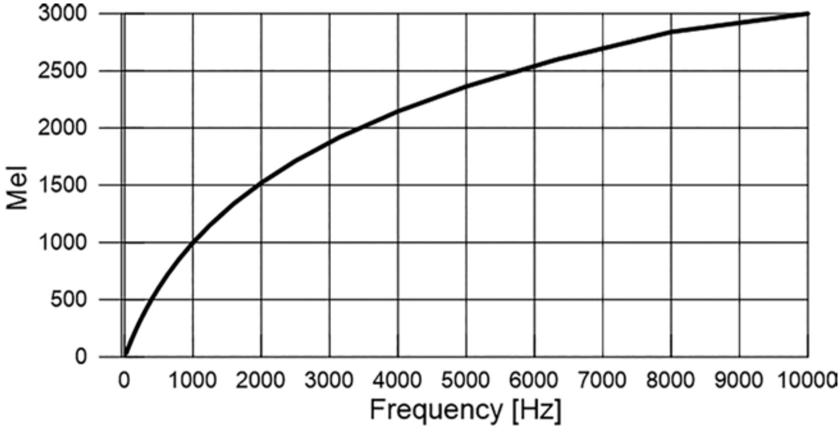

Figure 6.3 The relation between frequency and perceived pitch, Mel. The curve attempt to express the best fit to experimental data.

defined as an anchor and got the value 131 mels. Today most accepted scales assign a perceptual pitch of 1000 mels to a 1000 Hz tone despite the fact that the curve is not very accurate below 1000 Hz.

There are various formulas that seek to express a best fit to the experimental data. This is one of them:

$$m = 2595 \cdot \ log\left(1 + \frac{f}{700}\right)$$

where f is the frequency (see Figure 6.3).

CHAPTER 7

The Ear, Hearing, and Level Perception

CHAPTER OUTLINE

Hearing is one of man's senses. Many would say the most important. How the sense of hearing works is still the subject of research.

This chapter is only a brief review of the ear's function. However, it covers important elements from psychoacoustics (the field which encompasses human perception of sound) [1–4].

In anatomical terms, the ear is divided up into three parts, namely, the outer ear, middle ear, and inner ear (see Figure 7.1).

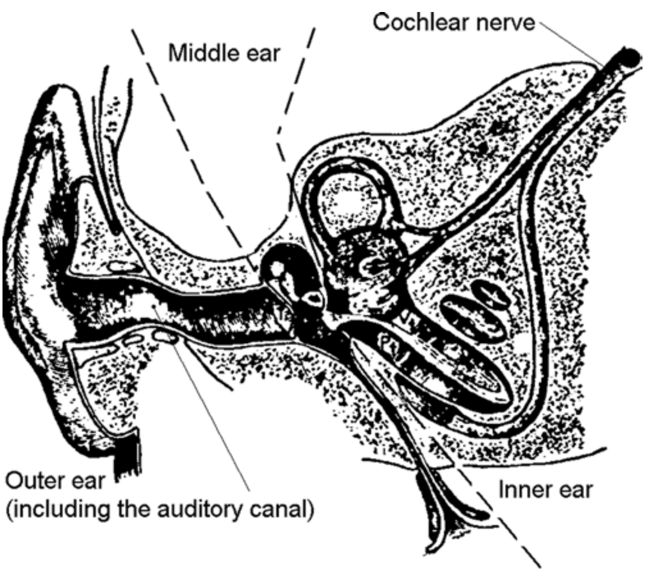

Figure 7.1 A cross section of the human ear is shown here: The outer ear with the pinna and auditory canal. The eardrum comprises the transition to the middle ear with the three ossicles (the hammer, anvil, and stirrup), which via the oval window connects to the inner ear with the cochlea.

THE OUTER EAR

The outer ear consists of the pinna, the auditory canal, and the eardrum. The pinna is important for directional perception in determining whether a sound is coming from the front or the rear. The auditory canal has dimensions that cause it to work as an acoustic resonator, which influences the ear's frequency response. The eardrum has a diameter of approximately 9 mm (0.35") and a thickness of approximately 0.1 mm (0.004").

THE MIDDLE EAR

Three small bones, called ossicles, are contained in the middle ear. These are named the hammer, anvil, and stirrup, respectively. These three bones function like a system of levers which ensure that the sound pressure variations in the outer ear transfer to the inner ear. The system could be called a pressure transformer in that the pressure increases by approximately a factor of 15 between the eardrum and the inner ear.

The hammer, which sits on the eardrum, is approximately 8 mm (0.32") long and weighs approximately 25 mg (0.0009 ounce). The stirrup, which connects to the inner ear, is approximately 3 mm (0.12") and weighs approximately 3 mg (0.0001 ounce).

The middle ear is filled with air and connects with the nasal cavity via the Eustachian tube, which is a narrow passage that opens when a person yawns or swallows. This passage allows for the equalization of any possible pressure differences between the surrounding environment and the middle ear.

THE INNER EAR

The inner ear is a fluid-filled organ, which is well-protected with its placement in the cranial tissue. The inner ear is an organ that never grows and for which there is no form of regeneration of destroyed cells.

The inner ear has a snail-shaped cavity, which is called the cochlea. With its slightly more than 2.5 turns, the cochlea has a volume of approximately 100 mm^3. The cavity is divided into two parts by the basilar membrane, save for a small hole (the helicotrema) at the top of the cochlea.

The width of the basilar membrane varies from approximately 0.1 mm (0.0039") to 0.5 mm (0.02"). The footplate of the stirrup connects to the oval window, one of the cochlea's two openings. The round window, which is the other opening, ensures pressure equalization during the influence of the oval window.

The organ of Corti sits on the basilar membrane and is equipped with thousands of hair cells placed in rows, approximately 3,500 inner hair cells and approximately

12,000 outer hair cells. A gelatinous membrane, the tectorial membrane, sits above these rows. When the ear is affected by a sound wave, the stirrup's movements converts to a movement of the fluid in the inner ear. This movement sets the basilar membrane in motion, which in turn causes a shearing motion between the organ of Corti and the tectorial membrane. The tectorial membrane activates the hair cells, releasing electrical impulses sent to the brain via the auditory nerve, which consists of approximately 25,000 nerve fibers.

The outer hair cells predominantly amplify low-level sound, whereas the inner hair cells mostly generate and deliver the electrical information to the brain. However, the inner hair cells are afferent delivering feedback in the hearing system.

The upper cutoff frequency for the transmission of signals to the brain is approximately 1 kHz.

THE SENSITIVITY OF THE EAR

The sensitivity of the ear varies with the frequency. Due to the shape of the auditory canal, the sensitivity is the greatest in the range around 2–4 kHz.

This frequency dependence is not constant but rather changes with the loudness of the sound (which is described in more detail later). By and large, one can say that the ability to hear bass increases relative to the midrange frequencies, when the sound pressure level increases. Humans will thus perceive differences in the sound of an acoustic image depending on the sound pressure level at which it is reproduced. Hence it is always important for audio engineers to perform a listening test of any program material at the contemplated level.

HEARING LOSS

Age-related hearing loss affects the highest frequencies in the audible spectrum and begins at around 25–30 years of age.

Noise-related hearing loss can arise from individual events such as explosions, screeching loudspeaker systems, and so on. However, exposure to larger doses of noise is the main reason for permanent hearing loss, like extensive use of loud-playing headphones, loud concerts, and loud rehearsal rooms. In these cases, the hair cells in the inner ear fall over or become deformed, and no regeneration takes place.

Before the actual noise-related hearing loss sets in, it can be ascertained that the ear's ability to differentiate between frequencies has deteriorated.

Normally it has been the opinion that short temporary noise-induced hearing loss does not affect the hearing. However, new research [5, 6] has shown that even the short temporary noise-induced hearing loss actually generates cochlear nerve degeneration.

Tinnitus is a particularly annoying form of hearing disorder. In its mild form, which many musicians are familiar with, a ring or a hiss arises in the ears as a consequence of the influence of loud sounds. This phenomenon will normally diminish after a few hours. However, it may never disappear. A person with tinnitus will always hear a sound that can be perceived as a ringing, hissing, screeching, or roaring. This unfortunate situation is a sufficient reason to take care of your hearing.

THE EAR AS A FREQUENCY ANALYZER

Under the influence of sound, the basilar membrane undergoes fluctuations of a certain size. These fluctuations are approximately 10^{-10} mm ($4 * 10^{-12}$") at the threshold of hearing. Depending on the frequency that is affecting the ear, the fluctuation will have a maximum at a certain distance from the oval window. The position of this maximum is of significance to the ear's ability to determine the frequency of the sound since the nerve fibers of the organ of Corti also function as selective filters with a certain bandwidth. For tones in the middle-frequency range above 200 Hz, the ear can differentiate frequency variations of less than 1%.

When the ear has to differentiate between the frequency components in complex sounds, the filter function works as a series of parallel filters, each with a certain bandwidth. Until recently, it was supposed that these frequency bands, which are referred to as "critical bands," had a bandwidth of approximately 20% of the center frequency (almost comparable with the bandwidth in a 1/3 octave filter). More recent research has shown, however, that this bandwidth is closer to 10%.

LOUDNESS AS A FUNCTION OF FREQUENCY

As described previously, human hearing is frequency dependent. Moreover, this frequency dependence varies with loudness. The extent of this relationship is expressed in the "equal loudness" curves (see Figure 7.2). These curves were created through the testing of a large number of people with normal hearing.

The basis is pure tones in a free, frontal field, which is to say that the test subjects in practice have found themselves in an anechoic chamber with a loudspeaker in front of them. The threshold value is then determined for the subject (i.e., the sound pressure precisely sufficient for a tone at a given frequency to be heard). Also, as

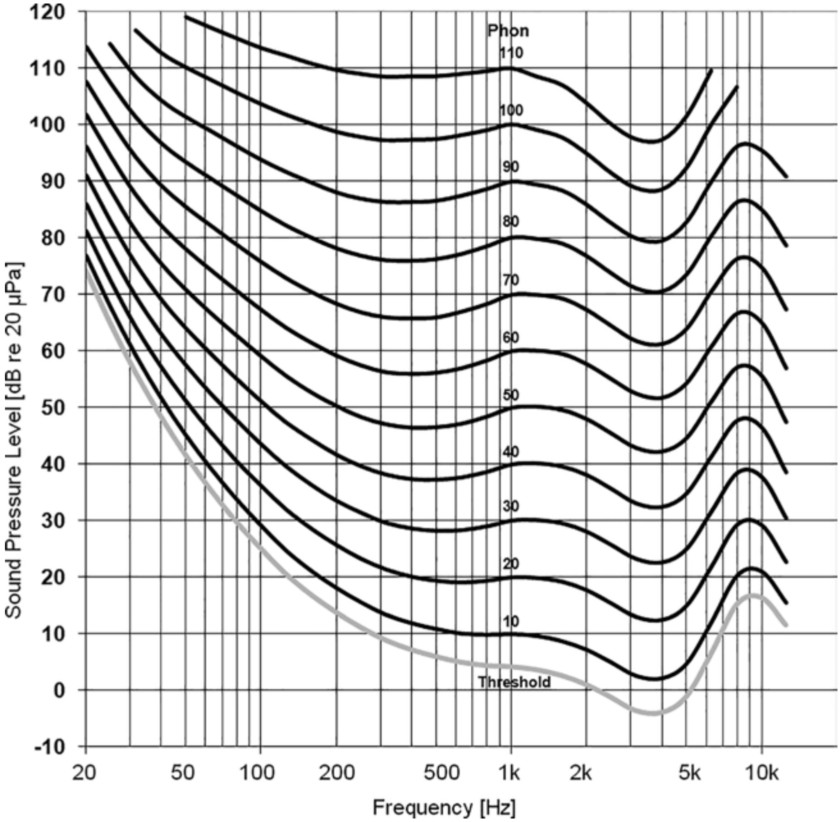

Figure 7.2 The equal loudness curves as they are standardized in ISO 3746. The curves express the sound pressure level at which pure tones must be represented in the free field to perceive constant level regardless of the frequency. The lower curve describes the MAF (Minimum Audible Field) threshold value.

a function of frequency, the required sound pressure for constant loudness is also registered.

This is done by comparing the tone concerned with a tone at 1000 Hz and a known sound pressure level. The curves thus express constant loudness – all points on a curve sound equally loud. The unit for loudness is the phon.

Example

The loudness for a pure tone with a frequency of 1 kHz and sound pressure of 50 dB re 20 µPa is 50 phons.

The loudness of a pure tone with a frequency of 100 Hz and sound pressure of 50 dB re 20 µPa is approximately 40 phons.

LOUDNESS AS A FUNCTION OF THE SOUND FIELD

The direction of a sound is also significant to its loudness. A frequency dependency occurs again. This can be of significance in the assessment of acoustic images in a loudspeaker arrangement for the reproduction of surround sound. The timbre can appear to be different depending on whether the sound comes from the center speaker, from the right or left front, or right or left rear speaker.

It is thus clear that the sound in a center speaker cannot be substituted by a phantom image created by the left and right front speakers. The relationship between a frontal direct sound field and a diffuse sound field, where all sound directions are equally probable, is shown on the curve in Figure 7.3.

LOUDNESS AS A FUNCTION OF THE DURATION OF A SIGNAL

The duration of a signal is as important to the loudness of the signal as is the frequency. This is one of the key points when it comes to the dynamic range: How large is the signal and how loud does it sound?

Studies have shown that signals (tones) with duration down to approximately 200 ms sound just as loud as an equivalent constant signal at the same sound pressure level.

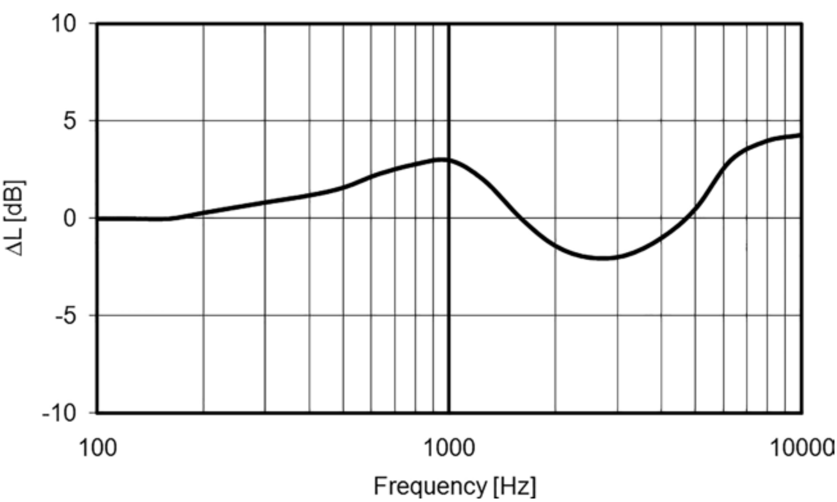

Figure 7.3 The curve shows the difference between the perception of a direct, frontal sound field and a diffuse sound field.

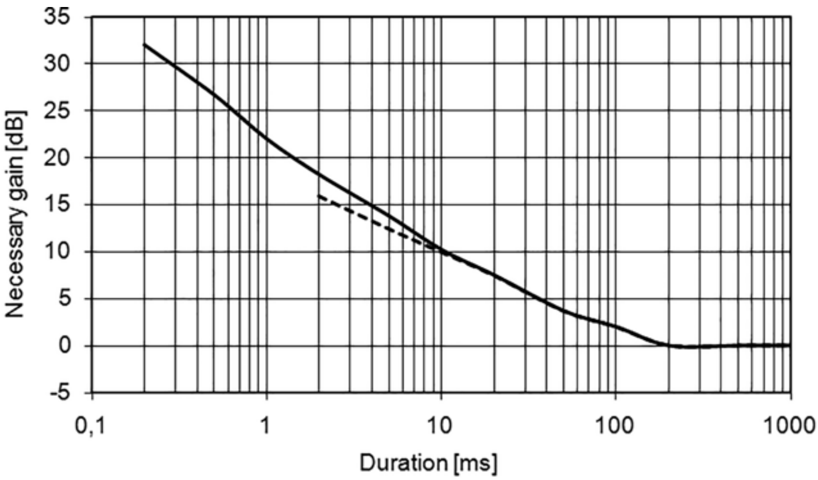

Figure 7.4 The curves describe the required increase in the level of impulses to maintain the same loudness as a constant signal. The fully connected line applies to broadband noise. The dotted line applies to pure tones.

If the duration of the signal is less than 200 ms, the perceived loudness is reduced (see Figure 7.4).

The curve shows the necessary increase in level to maintain constant loudness. For example, a signal of 10 ms duration must be increased by 10 dB to sound just as loud as an equivalent constant signal. Standardized instruments such as a PPM can show the full level for signals with a duration of 10 ms. A PPM does not take into consideration how loud the signal sounds, but rather its physical magnitude. A VU meter has an integration time of 300 ms; an LU meter has a short-term integration time of 400 ms. Hence these instruments far better illustrate loudness, or volume, especially as far as it concerns the duration of the signal.

Also, part of the assessment of the duration of a signal is the fact that the perceived loudness will fall during extended exposure due to the effects of fatigue.

MASKING

If the ear is exposed to a sound in a limited part of the frequency spectrum (such as a pure tone or a narrow band) then this sound, depending on the level, to a certain extent will mask or "hide" sounds of a similar or slightly higher frequency and lower level, even though this level also lies above the threshold of hearing. If the ear is exposed to two pure tones, which are very close to each other in frequency, then these tones will mask each other to a greater extent than if these tones were further apart from each

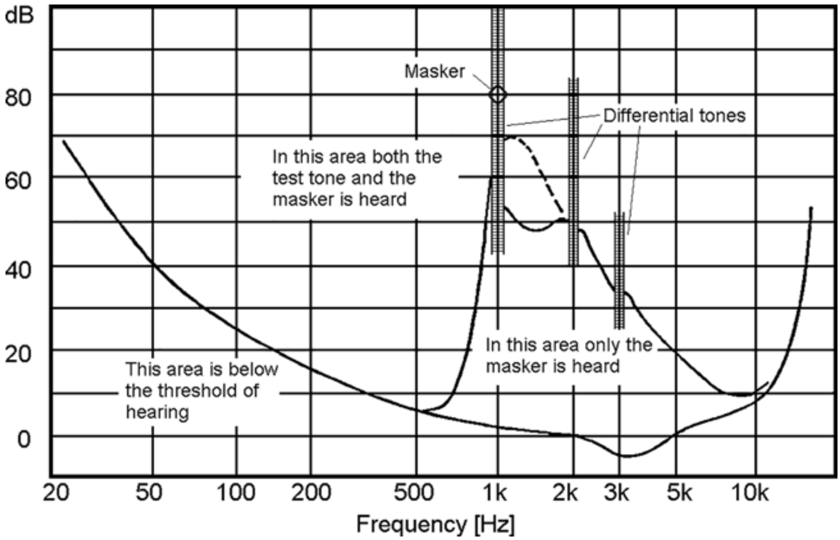

Figure 7.5 The masking curve for a pure tone at 1 kHz and a sound pressure level of 80 dB re 20 μPa.

other. Two tones that are close to each other often are perceived as one tone that varies in level (differential tones).

These phenomena were studied through experiments that are similar to those applied for measuring the sensitivity of the ear (i.e., by the test subjects being asked to adjust the level of individual tones of different frequencies until these are just barely audible). This experiment differs simply in that the ear is constantly subjected to a fixed tone at 1 kHz with a sound pressure level of 80 dB (see Figure 7.5). The results show that a tone of 1.5 kHz in this situation must have a level of approximately 45 dB above the threshold value in order barely to be audible simultaneously with the disrupting tone at 1 kHz. This phenomenon does not exist only at 1 kHz. Similar results can be obtained in the other part of the frequency spectrum, just as noise confined to a narrow bandwidth can have approximately the same effect as pure tones.

FORWARD AND BACKWARD MASKING

Masking does not occur only with simultaneous tones but can also occur based on sounds that start or stop split seconds before or after the masking or dominant sound.

Forward masking refers to the phenomenon whereby a tone can be masked by a sound that has stopped up to 20–30 ms before the tone concerned starts. It appears as if hair cells that have just been stimulated are not nearly as sensitive as cells that have rested for a longer period.

Backward masking refers to a tone being able to be masked by a sound with a higher level that begins up to 10 ms after the tone has begun. This effect does not originate in the inner ear, but rather in the brain. It appears that the brain's processing of the weaker tone is set aside when a sound impulse with a greater intensity appears.

The masking ability of the ear is utilized in connection with noise reduction and in particular with "intelligent" bit reduction, more precisely called "perceptual coding."

AUDIBILITY OF PEAKS AND DIPS

When changes are made to the frequency response in an electroacoustic system, tops (i.e., local rises in frequency response) are more audible than corresponding dips.

Experiments show that a gain of 10 dB in a narrow band around 3.2 kHz in an otherwise neutral frequency response would be noticeable by all listeners with normal hearing (see Figure 7.6).

Correspondingly, a drop in the same frequency band of 10 dB is only noticeable by approximately 10% of the listeners. A drop of 20 dB would only be noticed by approximately 40%.

This knowledge is useful in particular when adjusting loudspeaker systems to minimize problems with acoustic feedback at specific frequencies. Or for removing disturbing tonal noise from program material. If one by accident has recorded a test

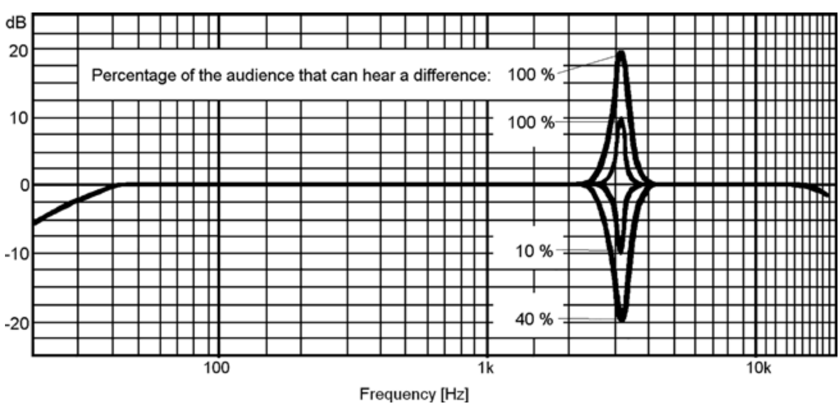

Figure 7.6 Based on psychoacoustic experiments, it has been demonstrated that everyone can hear gaining in a narrow frequency range, whereas only 10% can hear a corresponding attenuation of 10 dB. If the drop is 20 dB, only 40% can hear the change.

tone (pure tone) on top of the program material, a notch filter can remove the tone, as this filter has the property only attenuating a very limited frequency range. The filtering appears to be inaudible.

THE LAW OF THE FIRST WAVEFRONT

The ability of the hearing to localize a sound source in a closed space is affected by the reflections in the space. However, experience show that even in a room with many reflections – and hence a long reverberation time – it is possible to localize a single sound source even if the sound energy of the direct sound is much smaller than the energy of the reverberant sound field at the position of the listener. This phenomenon is called "The law of the first wavefront" and first described as such by Lothar Cremer in 1948.

A study carried out by Helmut Haas in the 1950s has become important background information in the understanding of how the hearing system perceives the combination of direct and delayed sounds. In this experiment, subjects were seated in an acoustically damped room in front of two loudspeakers (LS1 and LS2) like a stereo setup with a listening angle of 80 degrees (see Figure 7.7). A speech signal (approxi-

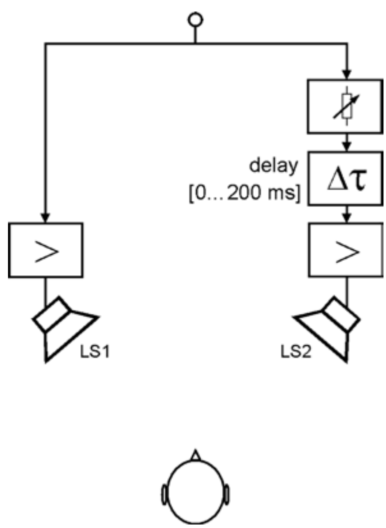

Figure 7.7 This is the basic setup for Haas's experiment. The same signal is reproduced by loudspeakers LS1 and LS2; however, the level and delay of LS2 can be controlled. By adding various delays to the LS2 signal, the task of the listener is to define what is the level difference between LS1 and LS2 when 1: Only LS1 is audible, 2: Only LS2 is audible, and 3: When the sound is localized to a position right between LS1 and LS2.

mately 5 syllables per second) was reproduced by both LS1 and LS2, however with a possibility of changing both the level and the delay of the sound reproduced by LS2. The level of LS1 was approximately fixed at 50 dB re 20 μPa.

When changing level and delay of LS2 the subjects' task was to determine either of the three: LS1 was inaudible, LS2 was inaudible, or the auditory event was localized in a direction right between the two loudspeakers LS1 and LS2. In Figure 7.8, the main results are shown. In electroacoustic applications, the lower curve is the most interesting. It shows that if the sound from LS2 arrives between approximately 5 and 35 ms later than the first arrived sound (from LS1), then the level of LS2 can be up to 5 dB louder than LS1 without being audible. This is often referred to as "the Haas effect."

If the delay exceeds 32–50 ms (depending on the type of signal reproduced) the delayed sound is perceived as an echo. It is easier for the ear to detect echo effects on impulsive/percussive sounds than more sustained sounds.

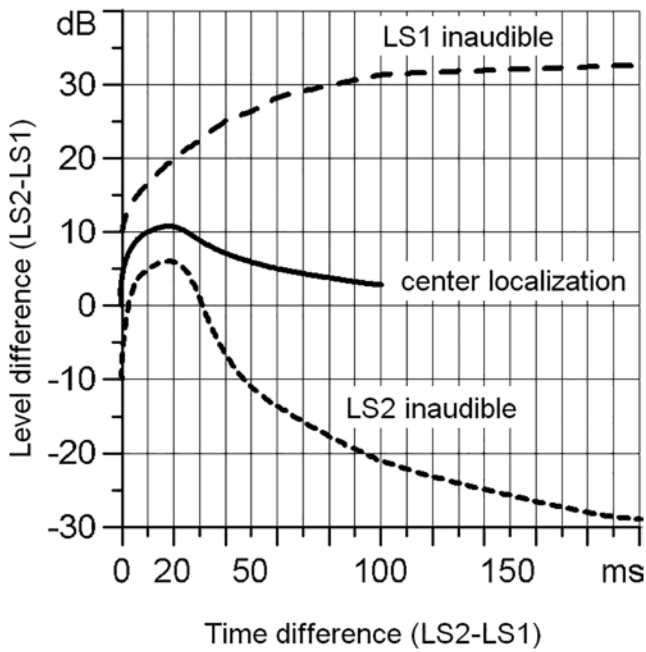

Figure 7.8 These curves represent the result of the experiment, the level difference between LS1 and LS2 (LS2-LS1): The upper curve shows that at zero delay LS2 must be 10 dB louder than LS1 if LS1 must not be audible. However, increasing the delay, the level difference also has to be increased, still, LS2 being even louder. The middle curve shows that LS2 must be louder to keep center localization when adding a delay. The bottom curve is the most interesting: When LS2 is delayed 15–20 ms, the level of LS2 can be up to 5 dB louder than LS1 without being audible.

REFERENCES

[1] Juhl Petersen, O.: *Menneskets lydopfattelse*. Laboratoriet for Akustik, DtH, Denmark (1974).

[2] Blauert, Jens: *Spatial hearing*. The MIT Press (1997). ISBN 13-978-262-02413-6.

[3] Evans, E.F.: *Basic physiology of the hearing mechanism*. Proceedings of the AES 12th International Conference (1993).

[4] Zwicker, E.; Feldtkeller, R.: *Das Ohr als Nachrichtenempfänger*. S Hirl Verlag, Stuttgart (1967).

[5] Kujawa, S.; Liberman, C.: Adding insult to injury: Cochlear nerve degeneration after “temporary” noise-induced hearing loss. *The Journal of Neuroscience* Vol. 29, No. 45 (November 11, 2009).

[6] Viana, L. et al.: Cochlear neuropathy in human presbycusis: Confocal analysis of hidden hearing loss in post-mortem tissue. *Hearing Research* Vol. 327, 78–88 (September 2015).

CHAPTER 8

Time Weighting

CHAPTER OUTLINE

Most instruments that display levels create the value displayed through an RMS averaging of the signal over time. In practice, two different methods are used to create the averaged signal, namely, linear and exponential averaging. The term time weighting refers to the time window applied for the creation of the value displayed.

LINEAR AVERAGING

With linear (time) averaging, an average level is created as a simple RMS value of the signal within a given period, such as 2 seconds. The weighting for linear averaging is shown in Figure 8.1.

Instruments for acoustic measurements typically have a selection of averaging times to choose from: 70 ms, 250 ms and 2 s corresponding to the terms I (impulse, not widely used anymore), F (fast), and S (slow), respectively.

For measurements with program level meters (PPM) for audio recording and transmission, the averaging time windows or averaging times applied are typically 0.1 ms (fast), 10 ms (standard PPM), 300 ms (VU meter), 400 ms (LU meter, momentary), and 3s (LU meter, short-term).

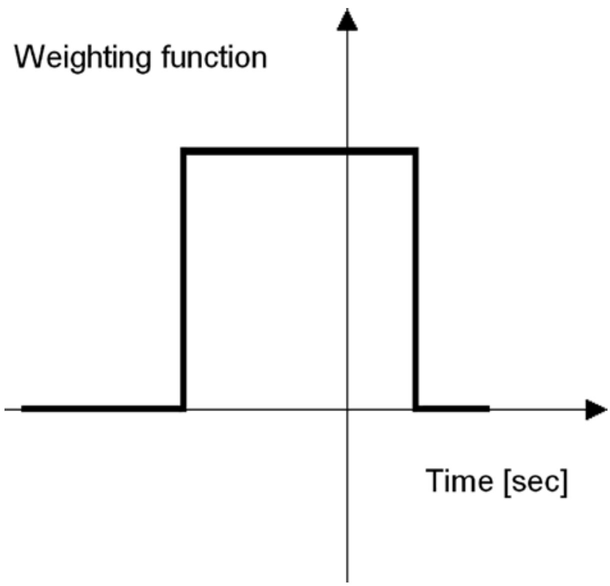

Figure 8.1 Principle of linear time weighting: the signal that occurs within the time window is averaged.

EXPONENTIAL AVERAGING

With exponential averaging, the weighting is an exponential function, which, as opposed to linear averaging, does not extend over a fixed period. The RMS signal created will thus "remember the past," but in such a manner that events that lie far away in the past will have lesser weight than events that have just occurred.

Exponential weighting is shown in Figure 8.2. With exponential averaging, the concept of a time constant is used rather than a period for averaging. The time constant is a measure of how fast the exponential function "dies" out, or more precisely, it specifies the time before the exponential function is reduced to 69% of its beginning value.

In connection with the acoustic measurement of sound, there normally is an option to select time constants of 125 ms (fast) and 1 s (slow), as well as an option for impulse weighting.

IMPULSE

Impulse weighting differentiates itself from fast and slow weighting by the fact that it contains a peak detector. It is thus the peak value rather than the RMS value that is decisive for the impulse value displayed when the instrument or meter is in I, or impulse, mode. As the name suggests, this form of weighting is intended for the measurement of

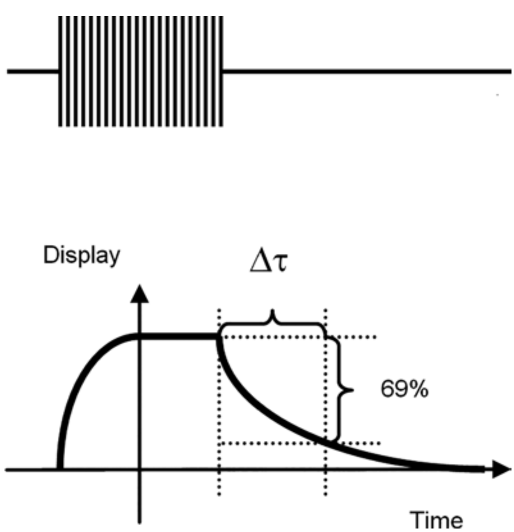

Figure 8.2 Principle of exponential averaging.

impulse noise. The time constant is short (35 ms), and a peak detector with an ensuing long time constant (1.5 s) will ensure that impulses will take a long time to "die" out. This form of level display is applied for the measurement of environmental noise, such as when measuring noise from shooting ranges. However, it must be mentioned that the term "I" is gradually finding its way out of the standards. Instead, most standards now refer to "peak" measurements, which are displayed with a hold function that leaves the indicator at the max peak detected during the measurement. Depending on national legislation, this value is measured using A- or C-frequency weighting.

PEAK

In many countries, the measurement of impulse "I" at the workplace has been superseded by the measurement of the peak level. In this case, a frequency network (A- or C-weighting) may or may not be used. The peak measurement is supported by a hold function that maintains the highest peak level measured until the next reset.

TRUE PEAK

In some digitally based level-reading meters (like loudness meters), we find the expression "true peak." Normally oversampling is applied to search for any possible exceeding value between any two samples of the basic sampling rate.

Please note that it is not possible to measure the true peak of a digital signal if it has passed through an asynchronous sample rate converter.

THE TIME FACTOR IN PROGRAM LEVEL METERS

In program level meters, time weighting and integration time are defined in a slightly different manner when compared to equipment for acoustic measurements.

The integration time is defined in the standards as the time it takes for a signal to reach close to full amplitude. To be more precise, the IEC standard (IEC 60268–10) as an example states that the integration time is the time it takes a 5 kHz sinusoidal burst at reference level to reach a display 2 dB below that reference level. There must be a silent interval between each tone burst large enough that the instrument can settle back to its minimum.

To be able to see what the level was when it was there, a fallback time can also be defined. The fallback time is the time that elapses from a continuous signal stops

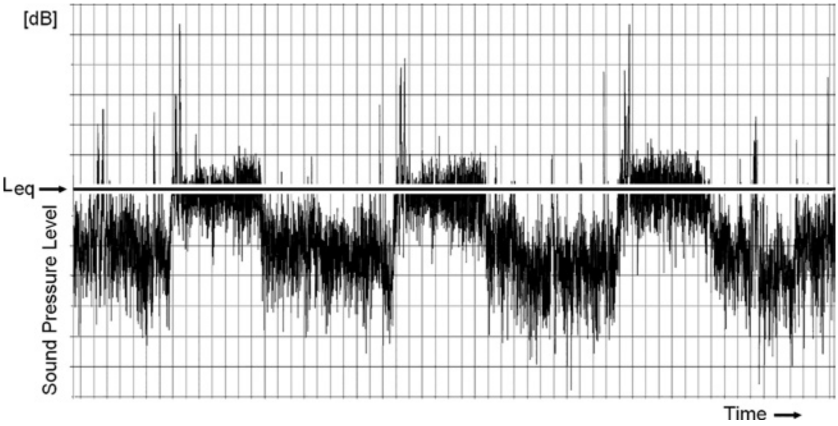

Figure 8.3 L_{eq} of a time-varying sound pressure level.

until the display reaches a specific lower point on the scale. Typically dB/s specifies this feature.

EQUIVALENT LEVEL, L_{EQ}

L_{eq} expresses the energy equivalent level. The L_{eq} value is an averaged value of (on an energy basis) the level over a longer interval of time (minutes to hours) (see Figure 8.3). This concept is used in connection with acoustic measurements of sound; however, L_{eq} is also applied to electrical measurements for the calculation of program loudness. In many cases, the value is calculated from a frequency-weighted signal, such as IEC A (acoustical measurements) or ITU K-weighting (loudness, program material):

$$L_{eq}(w) = 10 \log_{10}\left(\frac{1}{T}\int_{t_1}^{t_2} a_w^2(t)dt \right)$$

where

w = frequency weighting, if applied, e.g., IEC A or ITU k

$a_w(t)$ = the amplitude waveform of the signal (weighted)

$t_2 - t_1$ = integration time

CHAPTER 9

Frequency Weighting and Filters

CHAPTER OUTLINE

Frequency weighting involves emphasizing/suppressing certain parts of the spectrum of an audio signal. There are three reasons for using frequency weighting:

- To take the human perception of sound into account when measuring sound levels (acoustic) or the internal noise of devices and systems (electrical).
- To estimate the risk of hearing loss from sound exposure.
- For the pre-emphasis of audio signals, to obtain an optimization of the frequency response or signal-to-noise ratio.

FREQUENCY WEIGHTING EMULATING THE RESPONSE OF THE EAR

The hearing system exhibits a kind of frequency weighting. Thus, it may be appropriate to apply a similar weighting for the measurement of sound. This technique applies mainly to the measurement of noise. It can be acoustic noise or electrically generated internal noise in devices or systems.

Typically, the applied weighting filters suppress lower frequencies and slightly emphasize the midrange just above 1 kHz. Thus, low frequencies do not have as large an influence on the result of the measurement as do those frequencies that lie between, say, 1 and 4 kHz.

In practice, the weighting filter is inserted into the electrical part of the measurement sequence, either in form of a physical filter or a computational calculation. However, as a rule, filters are applied in measuring devices, such as sound level meters or audio analyzers.

IEC A/ANSI A

A-weighting applies, in particular, to the measurement of acoustic noise at the workplace and in the environment. However, it also applies to electrical measurements. It is developed from the inverse of the 30 phon equal loudness curve.

A-weighting is specified in IEC 60 651, but references are also made in other standards like ANSI S1.4–1981 (see Figure 9.1).

Measurements performed with A-filters are often described with an "A" in parentheses, for example, xx dB(A) or with an "A" as an index, L_A = xx dB.

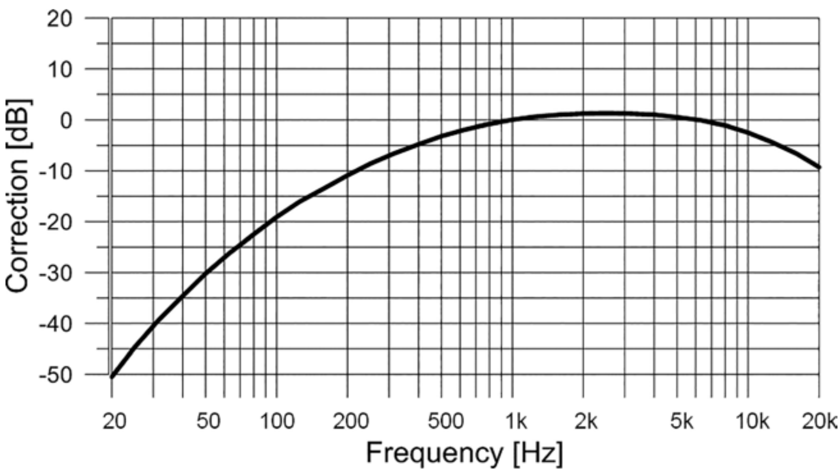

Figure 9.1 IEC A-weighting curve.

IEC B, C, AND D

The B, C, and D weightings belong to the same "family" as A-weighting (see Figure 9.2). Originally the idea was that different weightings should be used depending on the level of sound. A-weighting should be used at the lowest sound pressure levels since the ear does not hear that much bass at these levels. At higher levels, a change should be made first to the B- and later to C-weighting.

Today, the original B-weighting standard is by and large not used at all. However, a part of it has been implemented for loudness measures of program material.

The C-weighting curve is "flat," but with limited bandwidth, with −3 dB corners of 31.5 Hz and 8 kHz, respectively. The C-weighting often applies to the measurement of the maximum sound pressure levels of PA systems and cinema loudspeaker arrays, and also in some cases for the measurement of monitoring levels.

D-weighting applies to the measurement of noise from aircraft.

IEC Z

Z-weighting (IEC 61.672–1) is for use with acoustic measurements and is no weighting (Z for zero-weighting). However, the standard defines a flat frequency response of 10Hz to 20 kHz ± 1.5dB. This response replaces the older "linear" or "unweighted" responses, as these did not define the frequency range over which the meter would be linear.

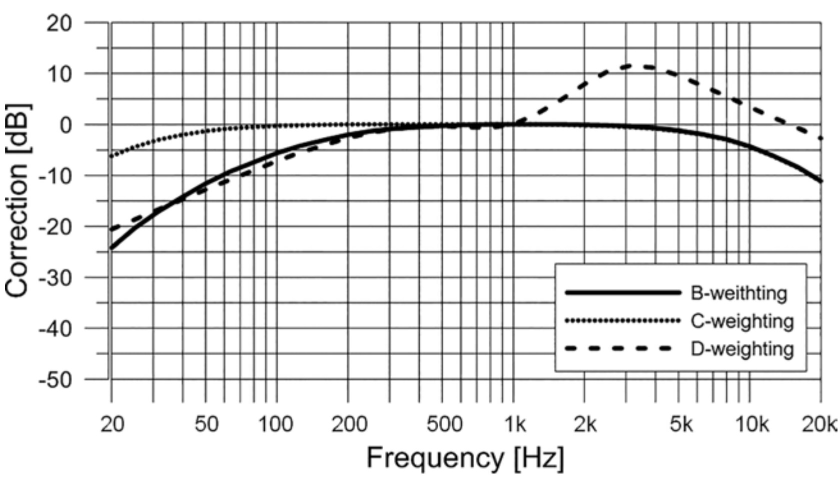

Figure 9.2 Weighting curves B, C, and D.

ISO/ANSI G

G-weighting is for measuring infrasound (i.e., acoustic sound below the frequency range which is normally regarded as the audible range). Hence the G-weighting is defined within the range of 1 Hz to 20 Hz (see Figure 9.3). The curve is defined in the two standards ISO 7196:1995 and ANSI S1.42–2001 (R2011).

RLB – REVISED LOW-FREQUENCY B-WEIGHTING (ITU-R BS.1770)

Work by the ITU has led to the conclusion that it is possible to determine the loudness of a program by a relatively simple algorithm involving a weighting curve. This weighting is basically a high-pass filter with approximately the properties of the low-frequency part of the IEC B-weighting curve. Hence the name: RLB – Revised Low-frequency B-weighting (see Figure 9.4).

K-WEIGHTING

K-weighting is a further development of the RLB curve. At an early stage, this curve was called RL2B, but now it is known as K-weighting. This curve includes a high-frequency shelving curve to emulate the presence of a listener's head. Find more about this curve in Chapter 12: Determination of Loudness.

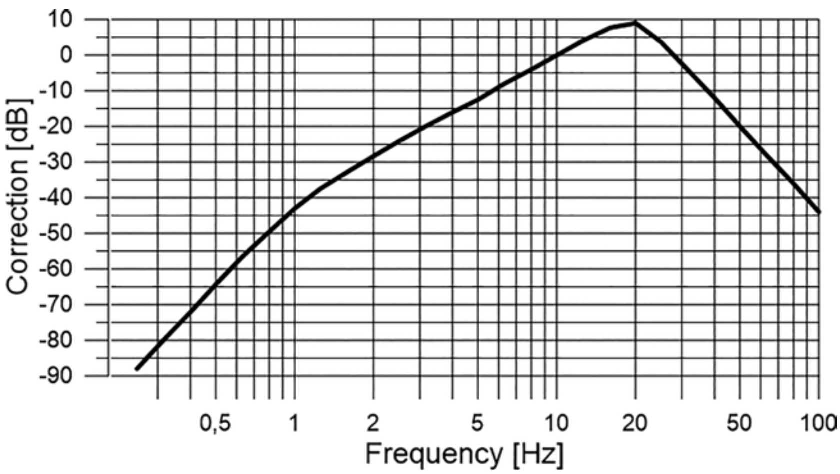

Figure 9.3 G-weighting curve ISO 7196:1995 and ANSI S1.42–2001 (R2011).

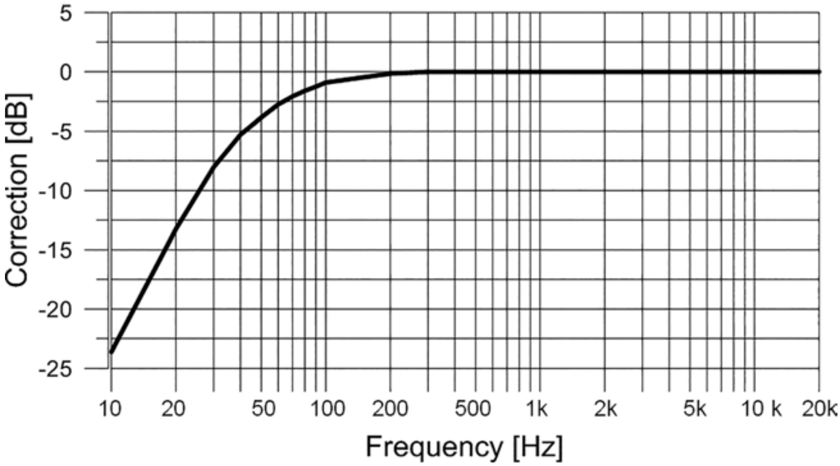

Figure 9.4 RLB weighting curve (ITU 1770).

ITU-R BS.468-4 (CCIR)

This weighting curve (known in the past as CCIR 468, see Figure 9.5) applies to the measurement of noise in electroacoustic devices, predominantly broadcast equipment. As opposed to most measuring systems, RMS is not used, but rather a quasi-peak detector (here, the meaning of quasi is "the same as").

The ITU have published different documents on the same weighting curve.

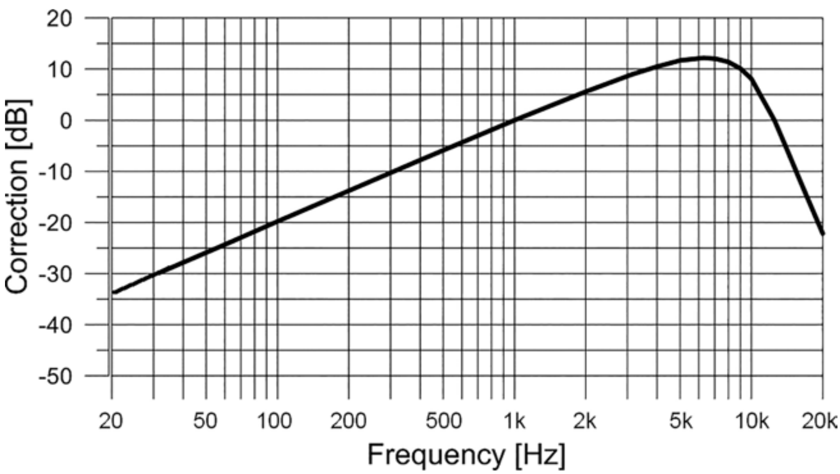

Figure 9.5 ITU-R BS.468–1 weighting curve (formerly known as "CCIR").

The ITU-R 468–2 exhibits unity gain at 2 kHz, which has been adopted by Dolby Labs for the CCIR/ARM.

ITU-R BS.468–4 is the most resent version, which is very similar to the first implementation in ITU-R BS.468–1.

CCIR/ARM/ISO

In 1979, Dolby Laboratories presented a simplified procedure for the analysis of signal-to-noise ratio in electroacoustic equipment. The purpose was to establish weighted measurements utilizing inexpensive equipment. Also, it had to provide a value that was easily understandable and which could be accepted by the industry in general.

The result became "CCIR/ARM," which uses CCIR weighting (now ITU-R BS.468) but with a 5.6 dB offset (see Figure 9.6). This offset makes the curve cross 0 dB (unity gain) at 2 kHz instead of 1 kHz. The value is read using a simple millivolt-meter that shows the average value. "ARM" stands for Average Responding Meter.

This weighting was subsequently applied to Dolby's system for measuring sound levels in cinema films. In this case, the average is not used, but an approximated L_{eq} (equivalent level) is used instead. The result is specified as L_{eq}(m), where the "m" is an abbreviation for "movie."

The application has been standardized in ISO 21727(en): Method of measurement of perceived loudness of short-duration motion-picture audio material.

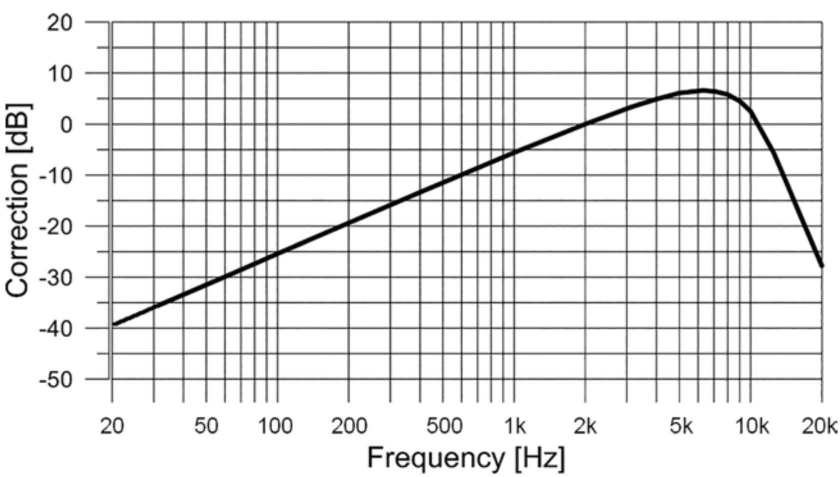

Figure 9.6 The CCIR/ARM curve crosses the zero-line (unity gain) at 2 kHz.

OTHER WEIGHTINGS

Other frequency weightings exist, of which some are only of limited use, and some are losing popularity. The first-mentioned group includes, for example, psophometric curves, which apply to frequency weighting, especially in the measurement of telephones and telephone components. Others include weightings for the measurement of loudspeakers. The assumption is that the frequency content does not have a flat, but has a shaped, spectrum in average music.

Among the soon-to-be-forgotten weighting curves are those used in the measurement of rumble and the like in gramophones (those things used to play vinyl records).

EMPHASIS

Pre-emphasis and de-emphasis are used in several contexts, particularly where analog signals have to be recorded or transmitted. The reason for the use of emphasis is to attain a good frequency response or to attain an improved signal-to-noise ratio or a combination of both.

On transmission lines, this typically only involves a high-frequency lift. Much attention must be paid to the measurement of the signal output from an FM transmitter or satellite equipment, as that output must be measured with pre-emphasis included, even if the output is passing through a limiter. With pre-emphasis, there is a risk of burning out a transmitter or knocking out the satellite link due to overamplification. Even though a limiter controls the output, if not connected after the pre-emphasis, it cannot save the situation.

With digital signals, it was earlier also possible to get mixed up in something that can later be quite difficult to solve. The AES3 standard (formerly AES/EBU) allows for the use of different pre-emphasis types. When combining signals in a mixer, emphasizing some of the signals but not others, the result can be mixed signals that subsequently get wrongly processed. This problem is not as common today, but it arose in the old days of the Sony PCM1610, F1, and so on when recordings could be pre-emphasized. So the problem is only relevant to those working with early digital recordings.

THE µS CONCEPT

The µs concept refers to the time constant that an RC element must have to attain the relevant frequency characteristic (see Figure 9.7).

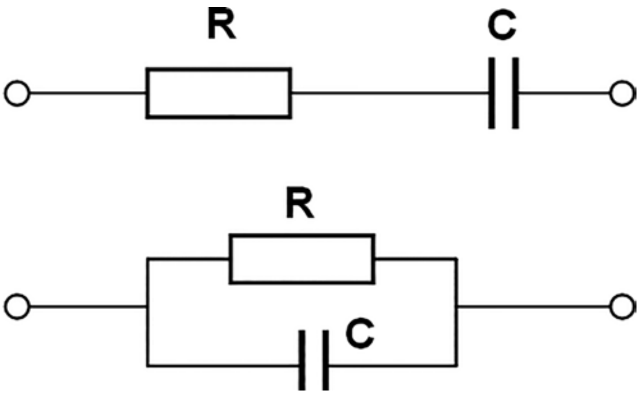

Figure 9.7 RC circuits.

The time constant is expressed as follows:

$$\tau = \boldsymbol{R \cdot C}$$

where

τ = time constant [s]

R = resistance [Ω]

C = capacitance [F]

or

$$\boldsymbol{\tau = 1 / 2 \cdot \pi \cdot f_0}$$

where:

τ = time constant [s]

f_0 = corner frequency of the emphasis circuit (the 3 dB point) [Hz].

This filter is of first-order (i.e., the slope of the inclined part of the curve is 6 dB/octave).

A number of standardized time constant based pre-emphasis techniques are applied, some in transmission techniques and others in tape recording (see Table 9.1). Some of the most frequently used time constants are listed here along with the corresponding corner frequencies:

50/75 µS

In particular, 50 µs or 75 µs apply to FM transmission either involving a typical radio broadcast or transmission systems for wireless microphones (see Figure 9.8).

Table 9.1 Time constants of 1st order RC-circuits and the corresponding corner frequencies.

Time constant [µs]	Frequency [Hz]
17.5	9094.57 ≈ 9100
35	4547.57 ≈ 4550
50	3183.10 ≈ 3180
70	2273.64 ≈ 2275
75	2122.01 ≈ 2120
90	1768.39 ≈ 1770
120	1326.29 ≈ 1325
318	505.00 ≈ 500
3180	50.50 ≈ 50

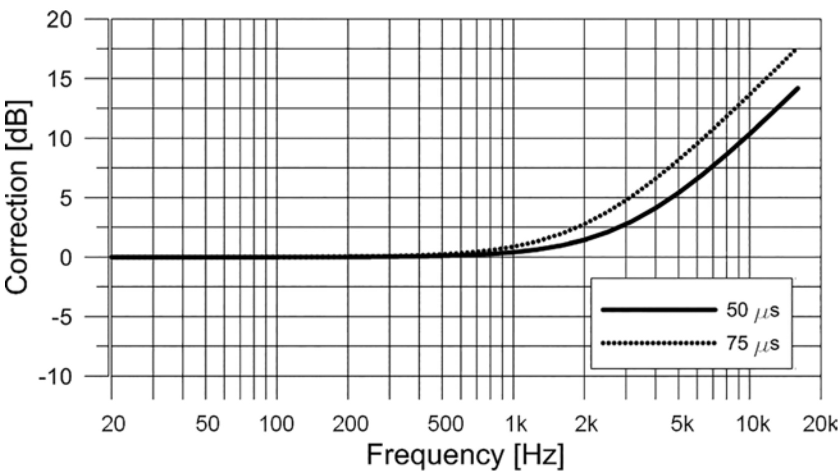

Figure 9.8 Emphasis defined by the time constants 50 µs and 75 µs.

The advantage is an improved signal-to-noise ratio, especially at higher frequencies. The drawback is that the curve rises endlessly toward higher frequencies. One then becomes dependent on frequency-limiting a signal during FM broadcasts among other things. In Europe, broadcasters employ a 50-µs time constant for FM transmission, whereas American stations use 75 µs.

50/15 µS

The pre-emphasis curve employed by consumer digital formats such as CD and DAT rises at high frequencies in the same way as the 50 µs curve for FM radio. However, due to the second time constant (corner frequency @ 10.6 kHz) the gain levels off toward the top of the audio band, reaching around 10 dB by 20 kHz.

THE EMPHASIS IN TAPE RECORDERS

In tape recorders, different time constants are used in the frequency-shaping networks of the recording and playback circuits (see Table 9.2). Even though analog tape recorders are not very common, it is still essential to keep track of the time constants for handling archived material. Mixing the standards can displace the tonal balance of a recording significantly.

Two time constants are used in certain standardized applications since the correction is performed at both ends of the spectrum. The correction at low frequencies (3180 μs) is intended to lift the signal above hum components.

Look up some of the standardized time constants in Table 9.2.

Table 9.2 Emphasis for tape recording. The 3180 μs correction is introduced to raise the signal above hum and the short time constants have the purpose of compensating for losses.

	Standards	Tape speed [cm/2]	Time constants [μs]	
Reel-to-reel	NAB	38	3180	50
	NAB	19	3180	50
	NAB	9.5	3180	90
	NAB	4.75	3180	90
	IEC 1	76	–	35
	IEC 2	76	–	17.5
	IEC 1	38	–	35
	IEC 2	38	3180	50
	IEC 1	19	–	70
	IEC 2	19	–	50
	IEC	9.5	3180	90
Cassette tape	IEC I	4.75	3180	120
	IEC II	4.75	3180	70
	IEC III	4.75	3180	70
	IEC IV	4.75	3180	70

J.17

J.17 emphasis got its name after the ITU standard described this curve. It is a weighting that applies to the pre-emphasis of signals that are either transmitted or stored. Some

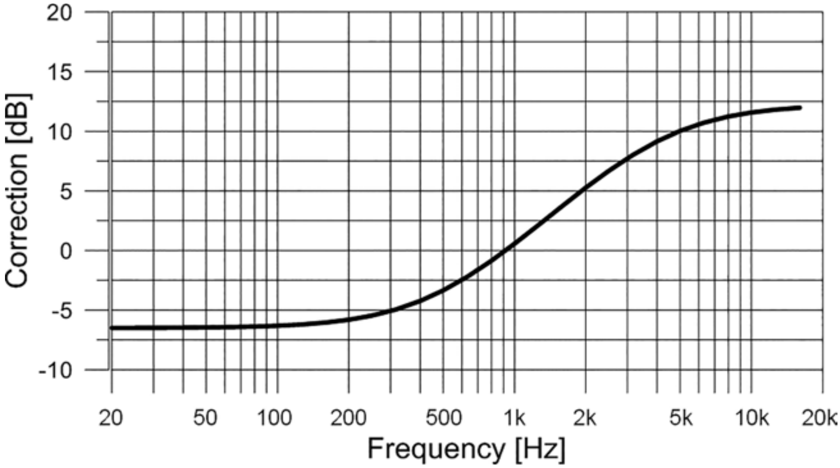

Figure 9.9 ITU J.17 emphasis.

examples of application are the pre-emphasis of NICAM stereo sound for TV and the audio channels of satellite transmissions.

J.17 pre-emphasis influences not only high frequencies but also reduces low frequencies, both around 10 dB relative to its unity gain at 2 kHz (see Figure 9.9).

The advantage of this pre-emphasis in comparison with 50 μs or 75 μs is that the rise flattens out toward higher frequencies. The level is thus less sensitive to the selection of the upper-frequency bandwidth.

RIAA

RIAA is an abbreviation for the Recording Industry Association of America. The RIAA emphasis applies to the recording of gramophone records. The purpose is in part to give a correct frequency response and in part to “economize” the needle’s movement in the groove (see Figure 9.10).

If one were to cut a record with the same amplitude throughout the whole spectrum, the physical displacement of the groove would be greater when a lower frequency is being recorded. The pickup would have a difficult time tracking, and there would be less room for content on the record. Therefore, the recording curve falls at a slope of 6 dB/octave toward the lower frequencies from a transition frequency of 500 Hz (318 μs). However, the curve flattens out below 50 Hz (3180 μs). Above a transition frequency of approximately 2.1 kHz (75 μs), the recording curve rises by 6 dB/octave. A constant amplitude could have been chosen at higher frequencies, but the choice was made to introduce this rise. The reasons include avoiding dust

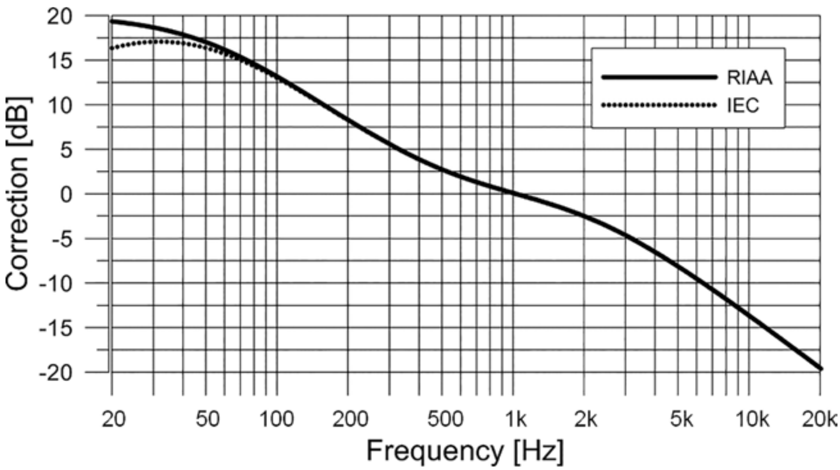

Figure 9.10 The RIAA and the RIAA/IEC curves for gramophone records (with a minor correction in the low bass range in the RIAA/IEC curve).

particles and the like, which have a disproportionately large and destructive effect on the acoustic image. In practice, it is difficult to see from the curve that the interval between 500 Hz and 2 kHz has a constant amplitude.

In 1976, IEC issued a correction to the original curve, becoming known as the RIAA/IEC curve. A new time constant that modifies only the extreme bass was introduced. However, this correction has never achieved significant success; the original RIAA curve is still the most widely used.

If working with the world of sound archives, it is recommended to look up a sufficient number of equalization data for the disk recording. More than 100 years of the invention have left us with many systems.

FILTERS IN AUDIO

Electrical filters are used everywhere in audio. As a rule, they affect the frequency balance of the signal. In addition to the concepts already mentioned, some of the fundamental general definitions are described here (see Figure 9.11).

CUTOFF FREQUENCY

The cutoff frequency is the frequency at which a filter begins to attenuate. Commonly defined by the frequency where attenuation is 3 dB. (The “−3 dB” is the power cut down by a factor of 2).

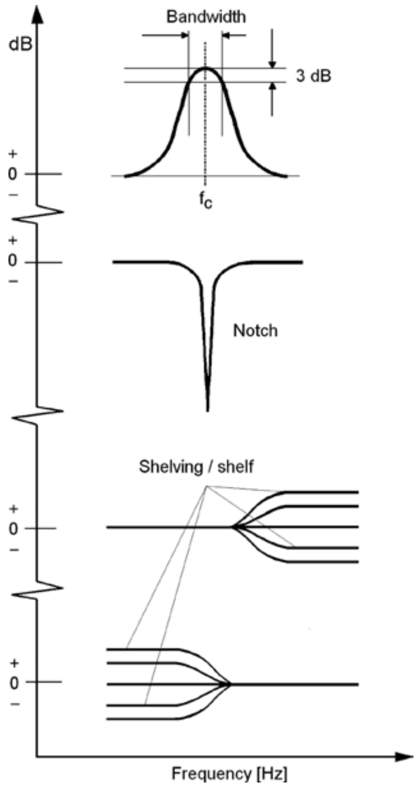

Figure 9.11 Upper curve: Band-pass filter. The bandwidth is defined (in Hz) by the points on the curve where the attenuation is 3 dB. Middle curve: The notch filter is a "narrow" version of a band-stop filter. Lower curve: The shelving filter is characterized by the curve being flat in its active portion.

BANDWIDTH

Bandwidth is the frequency span between the −3 dB cutoff points on a response curve (i.e., $f_{upper} - f_{lower}$ [Hz]). Expressed either in absolute terms in Hz or relative terms in octaves (often in 1/1 octave, 1/3 octave or fractions of an octave expressed with decimals, for example, 0.1 octaves) or in percentages:

$$b = \frac{(f_u - f_l) \times 100}{f_c} [\%]$$

where

f_u = upper cutoff frequency [Hz]

f_l = lower cutoff frequency [Hz]

f_c = center frequency [Hz]

1/1 octave ~ 70%, 1/3 octave ~ 22%. See also Q-Factor described below.

CROSSOVER FREQUENCY

In a crossover filter for a loudspeaker, this is the frequency or frequencies that determine what portion of the frequency spectrum distributed to the individual loudspeaker unit.

ATTENUATION

Attenuation is the magnitude by which a signal's strength is reduced. The attenuation is expressed in dB or the reduction factor.

GAIN

The gain represents the magnitude by which a signal's strength is increased. The gain is expressed in dB or the gain factor.

SLOPE

The slope is the response curve outside the actual passband, defined by dB per octave (or dB per decade).

CORNER FREQUENCY/TURNOVER FREQUENCY

These terms have the same meaning as the cutoff frequency.

Q FACTOR

A filter's Q factor is an expression of the relative bandwidth.

$$Q = \frac{f_{res}}{b}$$

where

f_{res} = resonance/center frequency [Hz]

b = bandwidth [Hz]

PASSBAND

That span of a filter's frequency response that lets the signal pass without attenuation.

STOPBAND

That span of a filter's frequency response that attenuates/rejects the signal.

FILTER ORDER

The easy way to comprehend the order of a filter is by comparing the number of frequency-dependent components in that filter. One condenser (and a resistor): first order. One coil and a condenser: second order. One condenser, one coil, one condenser: third order. For each order "*n*," the slope of that filter is $n \times 6$ dB per octave (or n × 10 per decade). Thus, a third-order filter has a slope of 18 dB/octave (or 30 dB/decade). See Figure 9.12.

RIPPLE

When designing filters, especially with steep slopes, ripples may occur in the transition area around the cutoff frequency.

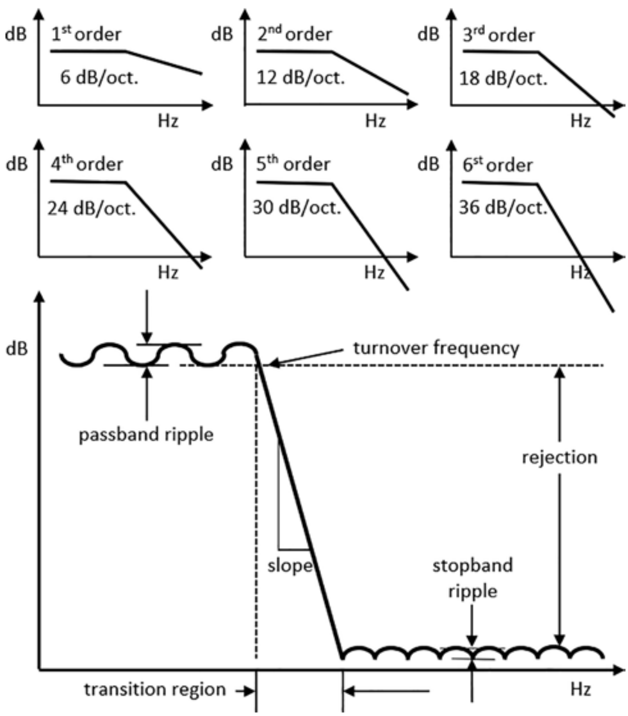

Figure 9.12 Upper: the order of a filter is related to the number of frequency-dependent components of that filter. Lower: filter is exhibiting a ripple in the transition area.

FILTER TYPES

The following gives a short description of the most commonly used filter types or equalizers.

ADAPTIVE FILTER A filter that can adapt itself to a given spectrum, for example, a noise spectrum, to be subsequently used to remove the noise concerned. Typically based on digital signal processing.

BELL The characteristics of the band-pass filter curve, having the shape of a bell.

BAND-PASS FILTER Allows a particular frequency range to pass but attenuates frequencies outside of it. This designation however, often describes filters that provide both amplification and attenuation.

BAND-STOP FILTER Suppresses a particular frequency range but allows frequencies outside of it to pass.

CONSTANT Q/CONSTANT BANDWIDTH These have more or less become type designations for graphic equalizers, where the individual (band-pass) filters have the same bandwidth regardless of the magnitude of the amplification (or attenuation).

GRAPHIC EQUALIZER Consists of several band-pass filters, each of which can amplify or attenuate their part of the frequency range and which cover the entire frequency range collectively. The division into bands often follows the international standards for 1/1 or 1/3 octave, and possibly 1/2 octave. The "graphic" part of the term refers to the sliders or controls on the front panel of the device that provide an indication of the frequency response that is presumed attained.

HIGH-PASS FILTER Allows high frequencies to pass but suppresses low frequencies. Also called LO-cut or bass-cut filters.

LOW-PASS FILTER Allows low frequencies to pass but suppresses the higher frequencies. The term HI-cut filter also applies.

NOTCH FILTER Suppresses a narrow range of frequencies. The bandwidth is a few percents.

OCTAVE FILTER A band-pass filter with a passband of precisely an octave. An octave filter is typically designated by the center frequency for the range concerned. The bandwidth is approximately 70%.

In sound technology, ten standard octaves are used, as a rule, as set in the ISO standard (see Chapter 27: Spectrum Analyzer, ***Preferred Frequencies,***

Table 27.1). The center frequency is the geometric average of the upper and lower cutoff frequencies:

$$f_c = \sqrt{f_u \times f_1}$$

The upper and lower boundaries are determined in the following manner:

$$f_u = 2 \times f_l \ ; \ f_l = \frac{f_c}{\sqrt{2}} \ ; \ f_u = f_c \times \sqrt{2}$$

1/3 octave filter The 1/3 octave filter, like the octave filter, has a constant relative bandwidth. There are also standardized center frequencies for these. The bandwidth is 22%.

Parametric equalizer A filter bank where the following parameters can be controlled: gain/attenuation, bandwidth/Q, and turnover frequency.

Peak filter A filter that amplifies a very narrow band.

Presence A filter that amplifies in the frequency range around 2–5 kHz. This filter can add "closeness" or "presence" to acoustic images, particularly to voices.

Shelving filter This type of filter is often applied in sound mixing or sound processing devices as a high- or low-pass filter. It is characterized by the active frequency range having a constant gain or attenuation. In particular filters, the gain or attenuation can be maintained, whereas the transition frequency can be varied.

Sweeping filter See **Parametric equalizer** (in this section).

FILTERS AND PHASE PROPERTIES

In this chapter, we only have looked into the frequency properties of filters. However, filters are more complicated than that. Filtering also exhibits a phase-dependent response. Whether dealing with analog or digital filters, the phase response should be considered. The phase response and time delay are two sides of the same coin. Thus, a phase shift at a given frequency also leads to a delay at that frequency. In (digital) filter design, it is possible to create filters without phase shift. However, the majority of filters has it, including digital designs. Some unique designs got the name after their inventors or the particular function that describes them. Here are some of the filters which apply in audio engineering:

Bessel This filter exhibits a limited phase shift.

Butterworth This filter exhibits a flat amplitude response of the passband and attenuation of exactly 3 dB at $\omega/\omega_0 = 1$.

Chebychev This filter has a steeper initial part of the stopband's slope but introduces ripples in the transition region of the passband.

Elliptical This filter provides a steep cutoff and a narrow transition width. However, it introduces ripples both in the passband and in the stopband.

To illustrate the effect of a given phase response, Figure 9.13 shows an example. A 1kHz low-cut (filter: Bessel, sixth order) of a 200 Hz square wave. The fundamental frequency is attenuated. The reduction of the RMS value amounts to approximately 10 dB. However, due to the phase response and the "displacement" in time of the harmonics, the peak level of the resulting waveform exhibits an increment of 5 dB! Repeating: *Reducing the frequency content may increase the peak level!* (That is why we need headroom when processing audio).

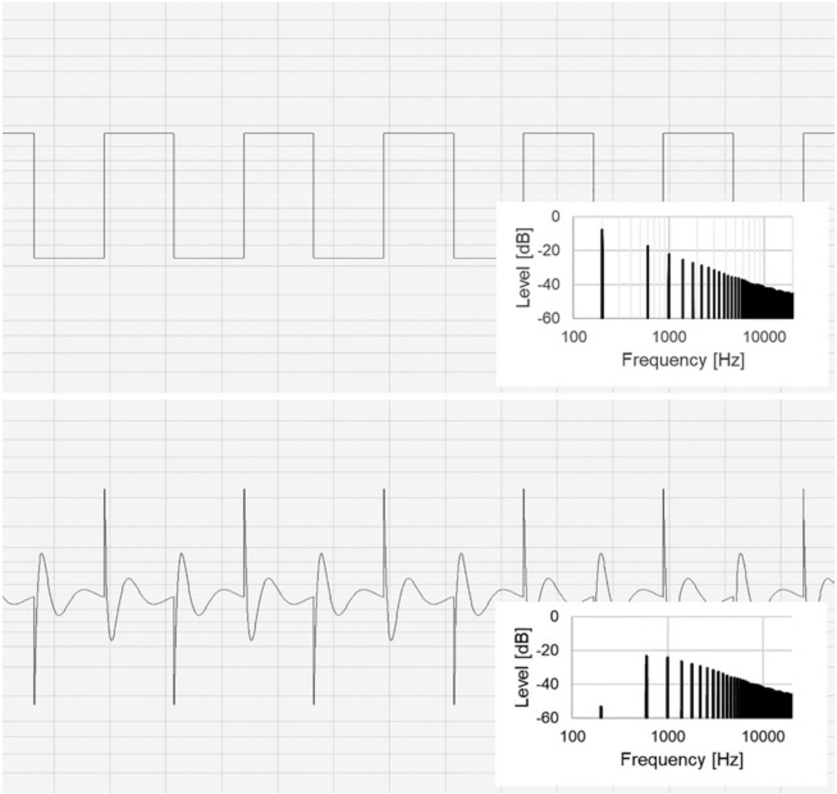

Figure 9.13 The initial signal is a 200 Hz square wave. A low-cut at 1000 Hz is introduced. The filter is a Bessel 6th order. Due to the phase response of the filter, the result shows reduced RMS value but increased peak values.

CHAPTER 10

Determination of Loudness

CHAPTER OUTLINE

Loudness is a measure of the perceived magnitude of a sound. In some applications, we are interested in measuring acoustic sound in a manner that expresses precisely the loudness experienced. Depending on the application, there are several methods for doing this.

The basis of loudness measurement is primarily the “equal loudness” curve. Each point (and thus frequency) on the same curve sounds equally loud, even though the sound pressure varies with the frequency. The curves express constant loudness regarding constant tones, and this loudness is expressed in phon. When we leave pure tones and move toward complex and composite sounds, which consist of many frequencies of different duration, it is necessary to compute data more comprehensively.

ZWICKER’S METHOD

Eberhard Zwicker described the most important method for the measurement and computation of loudness of noise. That method is internationally standardized.

The basis of this method is a frequency division into the so-called critical bands. As input data for the model, 1/3-octave spectra are used. However, frequencies under 280 Hz are added on an energy basis. Also, the character of the sound field must be determined. Here we distinguish between sound recorded in a direct, frontal sound field (the sound comes from only one direction, the front) or in a diffuse sound field (where all sound directions are equally possible). Loudness can then be computed manually by using one of ten charts as described in Figure 10.1. Alternatively, several instruments and software packages can measure or calculate loudness directly and provide the result in sone or phon.

The ISO 532–1975 describes the principles. This standard specifies two methods for determining a single numerical loudness value of a given sound. The first method is based on physical measurements, such as spectrum analysis in octave bands. The second method, referred to as ISO 532B, is based on spectrum analysis in 1/3 octave bands.

Zwicker’s articles that formed the basis for the standard are published in *Acustica*, No. 10, p. 303 (1960) and in the *Journal of the Acoustical Society of America*, No. 33, p. 248 (1961).

PROCEDURE

Manual determination of loudness by Zwicker’s method is performed in three steps.

First, the proper chart is selected. Do we measure a direct frontal sound field or a diffuse sound field? Then the chart is selected where the highest occurring value in the 1/3-octave spectrum can be plotted within the scaling. The 1/3-octave values from 45 to 90 Hz are added to a single value. The same applies for the 90–180 Hz values and the 180–280 Hz values. These values are then transferred to the measurement chart.

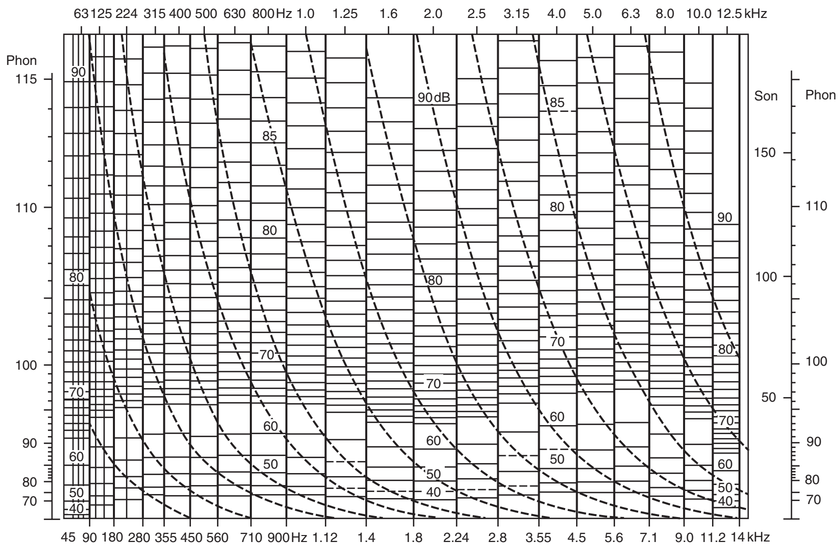

Figure 10.1 One of the ten charts used for the determination of loudness. This particular chart is called C8 and is used to determine loudness in a diffuse sound field for sound pressure in the 60–90 dB range.

The next step is connecting the plotted values from left to right. A following higher frequency band is connected with a vertical line if the level in this next band is higher compared to the previous one. If, however, the next band has a lower value, the line follows a parallel to the slanted lines (see Figure 10.2).

The third step consists of drawing a transverse line such that there is exactly as much area inside the curve above the line as outside the curve under the line. Extend this transverse line out to the scale on the sides of the measurement charts. The point of the intersection then indicates the loudness value either in phon or in sone.

As mentioned earlier, the method was standardized by ISO in 1975. The latest revision is from 2012. You can find all the charts in the standard.

CORRECTIONS TO ZWICKER'S METHOD

Some researchers revised Zwicker's method. These corrections are not currently a standard, however. Find a description in the following article:

Moore, B.C.J.; Peters, R.W.; Glasberg, B.R.: A revision of Zwicker's loudness model. *Acta Acustica* Vol. 82, 335–345 (1996).

In practice, there are only insignificant differences between the original and revised versions of Zwicker's method.

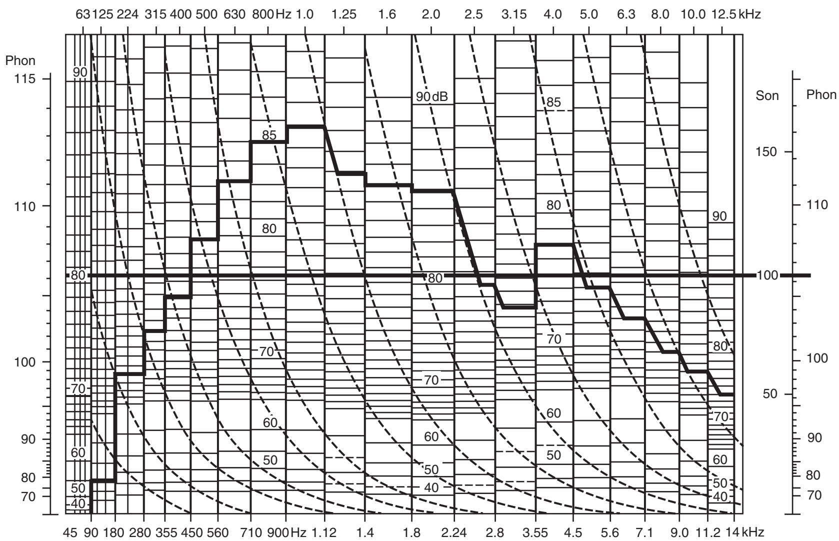

Figure 10.2 The measured 1/3-octave values are plotted in the measurement chart. Draw lines between the plotted values from left to right. For a rising level, use vertical lines. For a falling level, follow the dotted lines. Finally, draw a horizontal line that provides identical areas inside the curve above the line and outside the curve below the line. Extend the line to the scale on the right side. The point where this line intersects the scale indicates the result.

STEVENS'S METHOD

S.S. Stevens designed a somewhat simpler method. This method applies 1/1-octave frequency analysis. It is not as precise as Zwicker's method but also included in ISO 532–1975.

Each octave band is converted to a loudness index based on a measuring chart similar to the example shown in Figure 10.3. It requires differentiating between a direct, frontal sound field and a diffuse sound field. Two different charts are available for the calculation.

The various loudness indices are added using the following expression:

$$S_t = S_m + 0.3\left(\Sigma S - S_m\right)$$

where

S_m = the highest loudness index occurring in the measurement concerned

ΣS = the sum of the loudness indices in all bands

S_t = the total loudness expressed in sones

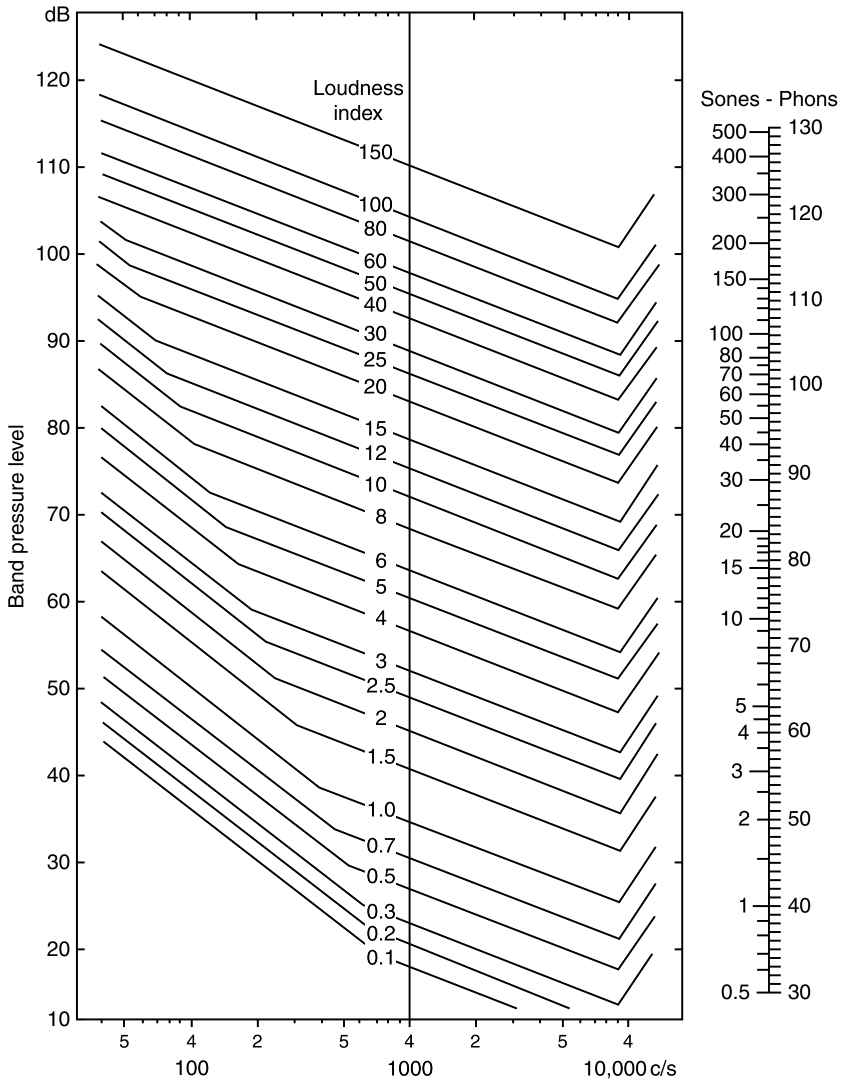

Figure 10.3 Measurement chart used in the manual calculation of loudness as per Stevens's model. This chart applies to diffuse sound field calculations.

DOLBY® $L_{EQ(M)}$ MOVIE LOUDNESS/ANNOYANCE

Dolby's method, which was mentioned earlier when discussing frequency weighting, claims to measure loudness. However, the value obtained is rather "annoyance." It is a method that was developed to measure the sound level of movies, particularly trailers and advertisements.

The determination of a film's sound level is not measured acoustically but as an electrical measurement. The method assumes that the soundtracks are played in a theater with a calibrated sound system. The audio in each track is frequency-weighted and filtered applying CCIR 468–2 weighting (level-corrected to 0 dB at 2 kHz). All tracks are extracted individually and then summed on an energy basis to a single value. An average value is calculated over the total length of the film. The final value derived at in the measurement is designated $L_{eq}(m)$.

This measure is standardized by ISO in 2004 and revised in 2016 for motion-picture audio material to be presented in cinemas with X-curve alignment of B-chain.

ITU-R BS.1770–4 – LOUDNESS OF BROADCAST PROGRAMS

In broadcasting, it is a problem that viewers watching television programs often complain about audio loudness jumps at every commercial break. Television commercials have unfortunately been infamous for their high compression and loud playout. Mixed with other program material, the level has been assessed as very uneven. The problem is not solved with standard metering and the traditional level setting alone.

Much effort to address this issue was made in standards organizations as well as in private companies around the years 2000–2005. Most of these interested parties gathered in an ITU-initiated workgroup formed to address the problem. Large-scale tests were carried out to find a practical solution for the measurement and the control of loudness, which in this context is a *perceptual* property of an audio signal when it is reproduced acoustically and listened to.

In these tests, many old and new algorithms and meters were tested against real program material. Fortunately, the outcome of the ITU work presented in 2006 was a surprisingly simple solution. The starting point was program material in mono. The recommendation, however, has been updated several times, and the most recent version is 1770–4 from 2015 to include, for example, measurement of NHK's 22.2 audio format (22 audio main audio channels).

The basis for the ITU recommendation is a weighting filter – the RLB curve (see Chapter 9: Frequency Weighting and Filters) combined with a L_{eq} measurement. The RLB (Revised Low-frequency B-curve) proved to be the best solution for continuous mono signals. As broadcast programs may contain not only mono but also two-channel mono, stereo, 5.1 surround, and other formats, further studies were made to provide one single loudness value regardless of the number of channels.

The presence of a human head in the sound field influences the spectrum. The acoustic effect of the head, modeled as a rigid sphere, is accounted for by adding pre-filtering to compensate the (electric) measurement of an acoustic phenomenon by raising the levels above 2 kHz by 4 dB (see Figure 10.4).

Humans' awareness of sounds coming from the rear is also taken into account. This is done by adding approximately 1.5 dB (a factor of 1.41) to the left surround and right surround channels, respectively, before the summation. Figure 10.5 shows the complete processing.

The additional weighting combined with the original RLB curve received a new name: k-weighting. In this respect, the scale indicates loudness with reference to a full-scale maximum. The loudness level thus became LKFS, meaning: level,

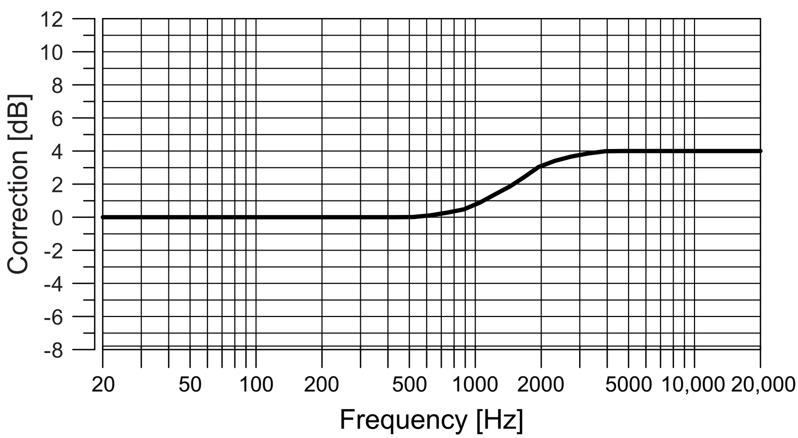

Figure 10.4 The ITU 1770 pre-filter added to all channels before summation to compensate for the acoustic effect of the head.

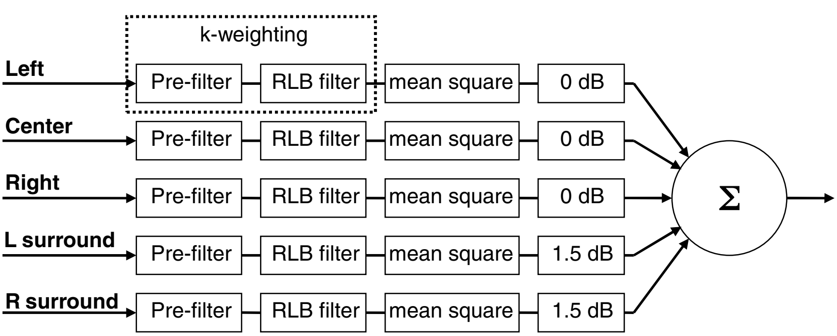

Figure 10.5 Block diagram of the multichannel (5.1) loudness algorithm (ITU 1770). The ".1-channel" is excluded from the calculation.

k-weighted, with reference to full scale. Going by ISO conventions, EBU instead use LUFS, meaning Loudness Units referenced to Full Scale. A unit of LUFS or LKFS is equivalent to a dB.

The recommendation "ITU-R BS.1770 – Algorithms to measure audio programme loudness and true-peak audio level" has been a breakthrough for loudness quantification in the field of broadcast. It provides the possibility of defining a target level for the instantaneous loudness as well as of the full length of the program based on an adequate loudness measure. (It is however noted in the recommendation that in general, it is not suitable for estimating the loudness level of pure tones.)

The standard was followingly considered by the European Broadcast Union which led to an introduction of gating as a part of the measurement. The gating was introduced to prevent silent sequences from unintendedly lowering the loudness measure of the complete program (see further background under "EBU R-128 - Loudness normalization.")

ITU adopted the gating technique and wrote it into the updated version of the recommendation of ITU-R BS.1770–2.

EBU R-128 – LOUDNESS NORMALIZATION

The European Broadcast Union accepted the ITU-R BS.1770 loudness algorithm when first published. However, in the recommendation EBU R-128, the program loudness estimation was taken a bit further:

The ITU 1770 algorithm rather measures a long-term or continuous loudness and did not count for certain types of programs that may contain larger sequences with very low-level audio. An example of this type of program is a feature about wildlife with long views overlooking the savannah while the wind is silently whispering in the grass in between the narration. If the level of the total program is averaged to reach the target level, then the narrating part is gained too much and become too loud on air. Additionally, commercials could be produced "creatively," including silence to become louder in the non-silent parts.

As already mentioned, the EBU workgroup P/LOUD adopted the ITU algorithm for loudness estimation. One thing that was changed was substituting the named LKFS scaling by Loudness Unit relative to Full Scale (LUFS). However, also, the gating function was introduced. The gating has two steps. The first being an absolute level of -70 dBFS. This gating level is a pure detection of whether or not any signal is present.

The second function halts the calculation whenever the loudness level is 10 LU (at an early stage 8 LU) below the target program loudness level measured within a window of 400 ms with 75% overlapping (see Figure 10.6).

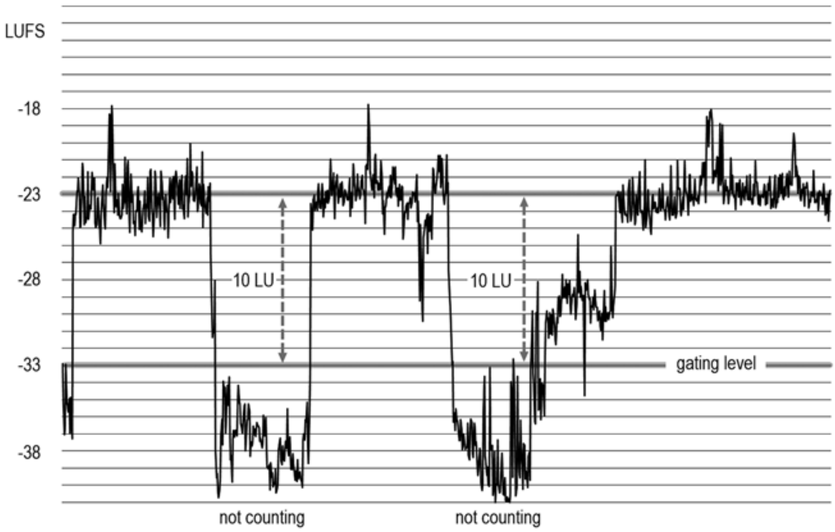

Figure 10.6 EBU-initiated gating function.

The practical reading includes long-term (full program, includes gating), short-term (3 seconds), and instantaneous (400 ms) loudness readings. Further, the EBU applied Loudness Range as a measure. (More about this in Chapter 14: Loudness Metering).

BIBLIOGRAPHY

ANSI S3.4–1980 (R1997). Procedures for the computation of loudness of noise.

EBU Recommendation R 128 (2014). Loudness normalisation and permitted maximum level of audio levels.

Fastl, H.; Zwicker, E.: *Psychoacoustics – facts and models*. Springer Verlag, Berlin Heidelberg (2007)

ISO 21727:2016 Cinematography – method of measurement of perceived loudness of motion-picture audio material.

ISO 532:1975 (reviewed 2012). Acoustics – method for calculating loudness level.

ITU-R BS.1770–4 (2015). Algorithms to measure audio programme loudness and true-peak audio level.

Moore, B.C.J.; Peters, R.W.; Glasberg, B.R.: A revision of Zwicker's loudness model. *Acta Acustica* Vol. 82, 335–345 (1996).

Skovenborg, E.; Nielsen, S.: *Evaluation of different loudness models with music and speech material*. AES 117th Convention (2004). Preprint 6234.

Soulodre, G.A.: *Evaluation of objective loudness meters*. AES 116th Convention (2004). Preprint 6161.

Soulodre, G.A.; Lavoie, M.C.: *Stereo and multichannel loudness perception and metering*. AES 119th Convention (2005). Preprint 6618.

Soulodre, G.A.; Lavoie, M.C.; Norcross, S.G.: *The subjective loudness of typical program material*. AES 115th Convention (2003). Preprint 5892.

Zwicker, E.: Ein Verfahren zur Berechnung der Lautstärke (A procedure for calculating loudness). *Acustica* Vol. 10, 304–308 (1960).

CHAPTER 11

Characteristics of Level Meters

CHAPTER OUTLINE

The use of analog metering in audio is by and large reserved for those who work with vintage gear. However, many of the related meter standards are still active. Also, many of the specifications may apply to digital metering.

The characteristics of various types of instruments mostly are defined in international standards. Some of the most important standards are **IEC 60268–10**, **IEC 60268–17**, and **IEC 60268–18**. In addition, recommendations may be provided by the **ITU** (International Telecommunications Union) or the **EBU** (European Broadcast Union). Alternatively, there can be national standards which apply, such as **ANSI** (American Standards Institute) or "**Technische Pflichtenhefte** / **Technische Richtlinie** der öffentlich-rechtlichen Rundfunkanstalten in der Bundesrepublik Deutschland" (Publication of technical requirements for public broadcasters in Germany).

DEFINITIONS AND REQUIREMENTS

In this chapter, a number of general definitions and specifications are reviewed together with examples of related specifications. The specifications here all refer to the IEC standards.

REFERENCE INDICATION

Analog Instruments: The instrument must have a marking on its scale that shows the maximum level allowed by the connected system. This marking is expressed by a percentage (e.g., 100%), dB (e.g., 0 dB), or Volume Units (e.g., 0 VU). The marking does not necessarily involve an absolute level.

Digital instruments: The reference indication corresponds to full scale (sine wave). The indication is normally marked 0 in terms of dB.

REFERENCE INPUT VOLTAGE

Analog instruments: This is the reference signal necessary to reach reference indication. The voltage is defined by the RMS value of a 1000 Hz sinusoidal signal.

IEC Type I: If nothing else is specified, the voltage is 1.55 V. Also the input voltage can be expressed in dB relative to 0.775 V. For example, +6 dB re 0.775 V or +6 dBu.

Volume indicators, VU (as per IEC 60268–17): The applied voltage for reaching the reference position is 1.288 V.

DIVISION OF THE SCALE

Analog instruments: This is the graphical display of the signal concerned. IEC Type I: −40 dB to +3 dB (minimum). IEC volume indicator: −20 to +3 VU

Digital instruments: The graphical display can either be an incremental dot or bar graph. Numerical values may indicate headroom or a reference if the input is analog.

Scale markings:

0 dB to −20 dB	numbers per 5 dB and minor ticks every 1 dB
−20 dB to −40 dB	numbers per 10 dB and no minor tics
−40 dB to −60 dB	numbers per 10 dB and a minor tic at −45 dB

AMPLITUDE FREQUENCY RESPONSE

Amplitude Frequency Response is within the effective frequency range, the deviation from ideal "flat" response expressed in dB.

Analog instruments: IEC PPM: 31.5 Hz – 16 kHz ± 1 dB

Digital instruments: IEC digital: 20 Hz – 20 kHz ± 0.5 dB

DYNAMIC RESPONSE

Dynamic Response expresses what the instrument displays when applying 5 kHz tone bursts of various durations: 100, 10, 5, 1.5 ms. The reading compares with reference to a display of a continuous 5 kHz sinusoidal tone (or >10 kHz for very short responses like 0.5 ms). The requirements are expressed by a curve or in a table.

DELAY TIME

Analog instruments: When applying a reference voltage, the delay time is the time it takes until the display reaches a reading 1 dB below the reference value.

IEC PPM: Less than 300 ms.

Digital instruments: Delay time is the time interval between the application of the reference input signal and the moment when the indicator passes a point 1 dB below reference indication. This delay relates to processing time.

IEC Digital: Less than 150 ms

INTEGRATION TIME

Integration time is the duration of a tone burst at a reference level that brings the indication to a given point below the reference value.

Analog volume indicator: 300 ms to reach 99% of the reference value

Analog and digital PPM instruments: 5 ms results in an indication 2 dB below the reference value

OVERSWING

Owerswing represents the maximum indication above a reference value that occurs when applying a 1000 Hz signal at the reference level.

Analog and digital instruments: ≤ 1 dB

RETURN TIME

Analog and digital instruments: The Return Time is the time it takes the indicated level to fall to a defined point below the reference value by the removal of an applied constant signal.

IEC PPM and digital: 1.7 s ± 0.3 s / 20 dB

REVERSIBILITY ERROR

Analog instruments: The difference in the indication when an applied asymmetric signal is phase-inverted.

IEC Type I: <1 dB

INPUT IMPEDANCE

Input Impedance is the internal impedance of the instrument in its entire active frequency range.

Analog, PPM: >10 kΩ

Analog, volume indicator: 7.5 kΩ

DISTORTION INTRODUCED BY THE PEAK PROGRAM LEVEL METER

The distortion represents the total harmonic distortion caused by the presence of the instrument.

IEC Type I: <0.1%

OVERLOAD CHARACTERISTIC

The expresses the maximum input signal (sinusoid) that the instrument can handle without it subsequently altering the specifications of the instrument.

IEC Type I: >20 dB for a 5 second signal

>10 dB for a continuous signal

SUPPLY VOLTAGE RANGE

Analog and digital instruments: The maximum allowable variation in the supply voltage for a given deviation in the indication.

The chapters that follow describes the relevant instruments in more detail.

REFERENCES

IEC 60268 Sound system equipment. Part 10: Peak programme level meters (ed. 2.0, 1991–03).

IEC 60268 Sound system equipment. Part 17: Standard volume indicators (ed. 1.0, 1990–10).

IEC 60268 Sound system equipment. Part 18: Peak programme level meters – digital audio peak level meter (ed. 1.0, 1995–11).

Technische Richtlinie Nr. 3/6 Aussteuerungsmesser für Produktion und Sendung (ed. 3, 2013–12).

CHAPTER 12

The Standard Volume Indicator (VU Meter)

CHAPTER OUTLINE

The Standard Volume Indicator (SVI) (also commonly known as a VU meter) was originally developed by the Bell System around 1940 and standardized in 1942. (American Standards Association (ASA), C16.5–1942: American Recommended Practice for Volume Measurements of Electrical Speech and Program Waves).

The SVI has ever since 1942 been the most used – and perhaps most misused and misunderstood – analog level meter in audio. The IEC issued the latest reaffirmation of the standard in 1990. The ANSI standard was withdrawn in 1999.

BASIC SPECIFICATIONS

The original Standard Volume Indicator consisted of a full-wave rectifier and a galvanometer (i.e., an electromechanical transducer). Hence, the meter is based on an average reading of the program signal, not peaks. According to the standard, the scale ranges from −20 VU to +3 VU (see Figure 12.1). The integration time, that is, the time it takes the deflection to reach from the bottom to the reference point of the scale (0 VU), corresponds to 300 ms ± 10%. The overshoot must be within 1–1.5%.

"VU" was originally denoted as a term designated to indicate volume. It was not meant to be a unit. However, later standards and practice have translated the VU as "Volume Unit."

ATTENUATOR

An important part of the Standard Volume Indicator is the attenuator circuitry connected to the meter. This passive electronic network determines the impedance and sensitivity of the complete meter. The reading of the program volume actually takes place on the attenuator – not the meter!

As can be seen from Table 12.1, the standard attenuator operates in steps of 1 dB. At 0 dB attenuation, the meter reaches the reference point on the scale (0 VU) when

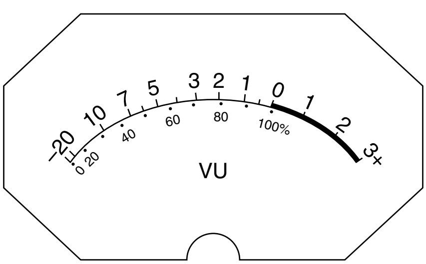

Figure 12.1 The VU-scale on the Standard Volume Indicator ranges from –20 to +3.

Table 12.1 Values for the Standard Volume Indicator attenuator.

Attenuation dB	Level VU	Arm A Ohm	Arm B Ohm
0	+4	0	Open
1	+5	224.3	33801
2	+6	447.1	16788
3	+7	666.9	11070
4	+8	882.5	8177
5	+9	1093	6415
6	+10	1296	5221
7	+11	1492	4352
8	+12	1679	3690
9	+13	1857	3166
10	+14	2026	2741
11	+15	2185	2388
12	+16	2334	2091
13	+17	2473	1838
14	+18	2603	1621
15	+19	2722	1432
16	+20	2833	1268
17	+21	2935	1124
18	+22	3028	997.8
19	+23	3113	886.3
20	+24	3191	787.8

connected to a source-level of +4 dBm. As the appendage “m” indicates, the reference is 1 mW. This reading corresponds to a voltage of 0.775 V across a load of 600 Ω; hence, the +4 dBm corresponds to 1.288 V.

IMPEDANCE

The input impedance of the attenuator (including the galvanometer) is 7500 Ω. A load resistor of 600 Ω can be connected across the input terminals (see Figure 12.2). So, when terminated with 600 Ω (impedance matching) and supplied by a constant signal of +4 dBm, the reading of the meter is 0 VU. The initial standard was not restricted to 600 Ω systems. If used with other impedances, the power level would change, and the calibration of the meter would be different from that of a 600 Ω system.

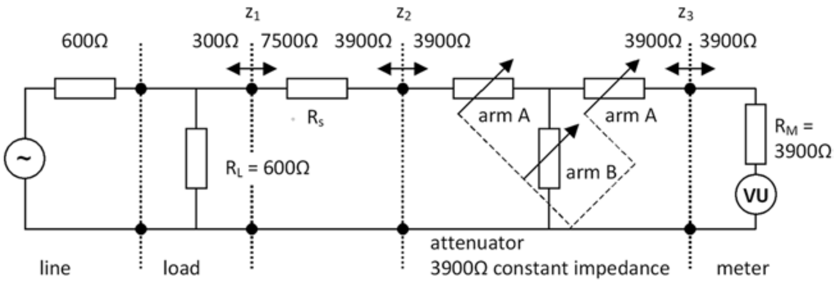

Figure 12.2 VU meter circuit is providing attenuation for an indication of reference levels above +4 dBm.

THE CONFUSION

The problem of understanding the reading of the SVI or VU meter is that the scale does not indicate VU.

Notice: *When*

- the line is terminated with 600 Ω.
- and the reading on the attenuator is "+4 VU" (the attenuation is 0 dB).
- and the reading of the meter is "0."

then this together indicates that the *volume level* of the program is "4."

If the program level is raised, then more attenuation is needed to align with the 0 VU marking on the meter; thus, the volume of the signal is higher.

The confusion is that a 0 VU reading on the scale of a meter, meeting the standard in a 600 Ω system, can indicate any volume level between +4 and +24 depending on the setting of the attenuator (if this attenuator exhibits a 0 to 20 dB attenuation).

THE SVI AND PEAKS

Because the SVI – or VU meter – is an average-reading instrument, it reacts relatively slowly, and often peaks in music signals will not be registered at all. In practice, the peaks can be 6–12 dB above the actual reading of the instrument. If program material with impulsive content is modulated to a display around 0 VU, overloading can quite easily occur on tapes, amplifiers, transmitters, and so on.

To prevent the overloading, a genuine VU meter is equipped with a so-called lead. This lead is an amplification circuit that can provide an increased reading on the meter. If this lead is not present, then correctly modulated impulsive type music would only show deflection at the bottom of the scale. Unfortunately, for years virtually no device equipped with VU instruments have included this lead.

THE "MODERN" VU METER

The understanding and use of the (remaining) VU meters today include only a few of the initial intentions. The scale (including the background color US Postcard Yellow) and the integration time of 300 ms are intact. The reference today is no longer power but voltage. So, the 0 VU deflection is reached for the voltage of 1.288 V. With most applications, the attenuator is completely forgotten, and the scale is translated to indicate the VU as a unit. Usually, the input impedance is high (i.e., $\geq 7.5\ \mathrm{k\Omega}$). While not included in any standard, a very practical LED overload indicator can be included with the VU meter.

CALIBRATION USING THE VU METER

The VU meter was widely used for the calibration of, for instance, tape-based media. The precise calibration values (magnetization levels) have changed over the years with the change in tape sensitivity. Hence, the VU reading in these cases is related to the media rather than to volume or voltage. The individual references have to be looked up (see Chapter 21: Standards and Practices).

REFERENCES AND FURTHER READING

American Standards Association (ASA), C16.5–1942: American recommended practice for volume measurements of electrical speech and program waves.

Hertz, B.: *New CCITT "Three-level test signal": The latest aspect in the VU–PPM debate*. AES 80th Convention (March 1986). Preprint 2357.

IEC 60268–17: Sound system equipment. Part 17: Standard volume indicators (ed1.0, 1990).

McKnight, J.G. (Jay): Some questions and answers on the standard volume indicator ("VU meter") (revised 2006). http://www.aes.org/aeshc/pdf/mcknight_qa-on-the-svi-6.pdf.

Pavel, E.A.; Gastell, A.; Bidlingmaier, M.: Über Vergleichende Messungen mit dem Volumenmesser und dem Spitzenwertmesser bei der Kontrolle von Rundfunkübertagungen. *FTZ, Heft* Vol. 4 (1955).

Wilms, H.A.: *0. VU- versus PPM-indicators. The end of a continuous misunderstanding*. AES 56th Convention (March 1977). Preprint 1221.

CHAPTER 13

Peak Program Meter – PPM

CHAPTER OUTLINE

THE PPM

The PPM (Peak Program Meter) is rather level oriented and thus works significantly faster than the VU meter. The instrument's time constant, the integration time, that is, the time it takes the reading to reach a point 2 dB below the reference indication, is 5 ms. In earlier standards, the integration time was defined as 10 ms. However, this is rather a question of definition. It still takes 10 ms to reach a reading 1 dB below the reference indication. In any event, the short integration time ensures that program peaks of short duration primarily can be read at (nearly) the correct level.

The fallback time (or return time) is relatively long for the eye to be able to register the indication: 1.5 sec. per 20 dB. The PPM meter is so fast (30 times faster than the VU meter) that peaks are often not under-read by much more than 3 dB. Thus, the analog PPM nowadays is often called QPPM (Quasi Peak Program Meter) due to the fact that it is still too slow to show the true peaks of the signal.

In a digital PPM, even continuous tones at frequencies that are a subdivision of the sampling frequency can also under-read.

This problem may happen even if the calculation of the value displayed includes all samples. The standard does not pay particular attention to this phenomenon, as the error only exists at very few frequencies. For real-world signals rather than pure tones, this particular type of under-read tends to smooth out.

However, in newer broadcast standards (like EBU R-128 or ITU 1770) peaks may reach a level 1 dB below full scale. To avoid any errors, it is recommended to measure "True Peak." The method involves (at least) four times oversampling. In this way, erroneous measurements due to short peaks and an unfortunate relationship between audio frequency and sampling frequency are avoided.

DIN SCALE

The DIN scale is characterized by the "0 dB" marking situated at the edge of the "red area," which is the typical designation for the overload region. Considering analog instruments, this modulation corresponds to a voltage of 1.55 V (+6 dBu). The scale ranges from −50 to +5 dB. Marking for the test level lies at −9 dB on the scale (see the "Test Level" section later). In addition to dB markings, the scale can be equipped with a percentage scale, where 0 dB = 100% (level), −6 dB = 50% and −20 dB = 10%.

The DIN scale has the advantage of the 0 dB marking positioned at the reference indication (i.e., at the edge of the "red" range). If choosing a different voltage reference, as in a digital recording, the scale can still be used.

Originally the DIN scale was standardized in DIN 45406. However, this has subsequently been replaced by the type I meter described in the IEC 60268–10 standard.

NORDIC SCALE

The Nordic broadcasters agreed that it would be practical to use a scale calibrated in dBu. This scale is called the "Nordic scale."

The edge of the red area lies at +6 dBu (where 0 dBu = 0.775 V). At a minimum, the scale covers a range from −36 to +9 dBu; however, implementations exist in which it runs from −42 to +12 dBu. The test mark lies at 0 dBu, 6 dB below the edge of the red area. Note that full modulation to the edge of the red area is attained at a voltage of 1.55 V for both the DIN scale and the Nordic scale.

BBC SCALE

The BBC developed a scale that ran from 0 to 7. The scale was designed in this way for reasons of being easy to read. The number "4" is located in the middle of the scale. At this level, the instrument's indicator is in a vertical position, and it relates to a voltage of 0.775 V (0 dBu).

The standard requires there being approximately 4 dB between each of the digits 1–7. The number "5" corresponds to 0 on the VU meter scale (for a constant tone and no attenuation added). Meters exist with the BBC scale that (outside the standard) use approximately 6 dB in the 1–2 and 2–3 intervals.

TEST LEVEL

The concept of a "test" or "test level" is used in a broadcasting context. What this refers to is one or more tones with a well-defined frequency, level, and duration.

What can seem a little strange is that two of the PPM instruments mentioned earlier provide a "Test" marking, but at two different positions on the scale in relation to the red area! The explanation is quite amusing in historical terms; however, the result can be a bit unmanageable in practice. The test tone concept originates from the era when AM radio transmitters came into use. Engineers wanted a test signal containing the same energy as "Gewöhnlisch Tanzmusik" (conventional dance music), which on the average would drive the transmitters at a 35% modulation. This 35% corresponds to 9 dB below the full signal (9 dB below 100% modulation). Hence the DIN scale's test point at −9 dB.

When the Nordic scale was developed, it was taken into consideration that the PPM meter, with its integration time of 10 ms, in some situations could indicate a value that was 3 dB too low. In other words, level peaks up to 3 dB above the level of +6 dBu can slip through even if the display never moves into the red area. As the test point lies 9 dB below the maximum, the result becomes 0 dBu! In the Nordic countries, the test level is thus 6 dB below the indicated full modulation.

When providing audio or video programs with test tones, as much information as possible should always be given in order to avoid misunderstandings about the test level. When the material is tape-based, the corresponding magnetic flux should be noted. In the digital domain, the levels must be observed with reference to full scale.

TEST LEVEL, DIGITAL METERS

In digital meters, test level marking is not always required unless the scaling conforms to the standard scales referring to analog audio. It is most common that the interface between the analog audio and the digital meter meets the specifications of IEC 60958–1 Digital audio interface – Part 1: General. However, meters connecting via ethernet-based networks are still more frequently used.

Levels reported in dBFS are always RMS Full-Scale level. 0 dBFS, is the level of a dc-free 997-Hz sine wave whose un-dithered positive peak value is positive digital full scale, leaving the code corresponding to negative digital full scale unused.

It is invalid to use dBFS for non-RMS levels. Because the definition of full scale is based on a sine wave, the level of signals with a crest factor other than that of a sine wave may exceed 0 dBFS; thus a Full-Scale square-wave reads +3 dBFS.

Occasionally confusion occurs when specifying or recording a test level in program material using software-based recording (DAWs): In some workstations, 0 dBFS may refer to either a sine wave or a square wave, the latter being the default value. As mentioned, the difference is 3 dB! Make sure that the sine wave reference applies.

EXTRA FUNCTIONALITY

The instruments commonly are equipped with additional practical functionality:

INTEGRATION TIME: FAST

A “Fast” option, which is normally 0.1 ms, may supplement the standardized integration time of PPM instruments of 5 ms (to reach a level 2 dB below that of a constant tone). This capability can be useful with both analog and digital recordings. However, in most digital meters this function is superseded by a “true-peak” indication.

PEAK HOLD

This function can show what the maximum level has been during a given period. A reset button typically releases it, or it can have a hold time of some seconds.

ADDITIONAL GAIN

Here, the signal is amplified before displaying the reading; this gain is typically 20 dB for PPM instruments. This function is applied when working with a large dynamic range or when assessing program material with varying levels.

PEAK INDICATION

Certain instruments are equipped with an additional LED indicating whether peak values exceed the maximum permitted level or the maximum of the scale. This function is beneficial if the time constant and level is known, which is seldom the case in non-professional meters. However, if a time constant of 5 ms (as a PPM instrument) or shorter is provided, then this can be a practical feature.

BIBLIOGRAPHY

IEC 60268 Sound System equipment. Part10 (1991): Peak programme level meters.

IEC 60268 Sound System equipment. Part18 (1995): Peak programme level meters – Digital audio peak level meter.

IEC 60958–1 (2014) Digital audio interface – Part 1: General.

CHAPTER 14

Loudness Metering

CHAPTER OUTLINE

Loudness is defined as the subjectively assessed level of sound. By naming an instrument as a loudness meter, the task of the instrument is more or less to emulate human hearing. The problem, however, is that the hearing and the assessment of sound to some degree is individual from person to person. Hence a true loudness meter is very difficult to define. Level, dynamic range, frequency range, the direction of the received sound, and the character of the sound itself will affect the perceived loudness.

Further, it is not possible to express the loudness of an electrical signal; it is the acoustical sound received and perceived by the listener that should be the basis for the measurement or calculation of the loudness. So, in reality the truth is impossible to reach. It is rather a question of defining a system that exhibits minimum error, a system that is satisfactory to most listeners.

For years there have been systems claiming to measure loudness, the Zwicker method being the most serious. This method is still valid regarding human assessment of noise. From the beginning of the 21st century, major attempts to define the optimum system for audio production and broadcast have been made. These efforts have led to recommendations such as the ITU-R BS.1770 [1], and the EBU R128 [2], [3], which even may develop further in the future. This chapter tries to provide an overview of the most important existing and new possibilities for loudness metering. Most attention is paid to the EBU-meter, as this is regarded as state of the art.

LOUDNESS METERS, FACTORY STANDARDS

Some loudness indicating systems have been in use for audio production and broadcast for many years. Since the Standard Volume Indicator or VU meter was developed and standardized, alternative ideas have found their way to audio production facilities.

DORROUGH LOUDNESS METER

The loudness meters from Dorrough Electronics have been widely used both in music production and broadcast since their introduction in the early 1970s. From that time, the meters were manufactured as stand-alone units for analog and digital audio or as plug-ins for DAWs.

Basically, the meters simultaneously display the instantaneous peak level of the signal and the average program level on the LED scale, which for basic use has a linear range of 40 dB. The difference between the readings expresses the "density" of the program. Each channel is measured individually.

The peak acquisition period is 10 µs to full scale, (measured with a 25 kHz sine input). The peak decay period is 180 ms, from full scale to all LEDs off. The loudness impression is derived from an average reading of the program signal. The time constant for the average reading is 600 ms.

Most systems in the line facilitate two-channel/stereo signals. Inter-channel phase relations can be monitored. The reading can be changed from LR to MS (Sum/Dif-modes). The Model 380D also encompasses 5.1 monitoring (see Figure 14.1).

Figure 14.1 The Dorrough Loudness Meters provides readings of the peak and the average program level simultaneously.

Special automatic zooming of scales for calibration is provided (Dorrough Window Expansion Mode). Also, various alarms can be set, such as for over/under level and phase errors.

DOLBY® MODEL 737 SOUNDTRACK LOUDNESS METER – $L_{EQ(M)}$

Dolby Labs is a major provider of audio equipment and systems primarily to the film industry and broadcast. Over the years, several Dolby inventions have become de facto standards. The Dolby metering systems have been developed to meet the demands of the industry.

Model 737 was developed for the measurement of Leq(m), primarily on film (see Chapter 10: Determination of Loudness). The unit has been discontinued. However, a large number of Model 737s are still in service, as the Leq(m) has become a standard for the measurement of commercials and trailers in many countries. It should be mentioned that what is actually measured is more "annoyance" than actual "loudness."

LM100 BROADCAST LOUDNESS METER

The Dolby® LM100 Broadcast Loudness Meter is developed as a tool for measuring the subjective loudness of dialogue within broadcast programming. The background for this is the fact that the dialogue can be regarded as the "anchor element" of a program; lining up dialogue levels provides better equality of the loudness within and between programs.

The LM100 employs the proprietary technology Dialogue Intelligence™ (DI) to measure the perceived loudness of dialogue in a complex program. However, DI's ability to reliably separate dialogue from other sources has been questioned [4], and early versions utilized Leq(A) for this measurement. However, the newer versions also have implemented the ITU-R BS.1770 algorithm as well. Further, the instrument can determine the unweighted peak and a range of other information about the signal. The unit can simultaneously display the incoming dialogue normalization (dialnorm; see Chapter 17: Dynamic Scales) value of a Dolby Digital program or any program within a Dolby E bitstream for direct comparison to the actual measured value (see Figure 14.2).

Figure 14.2 An example of the display readout on the Dolby® LM100.

A set of user-definable alarms and monitoring functions can inform an operator of input loss, signal clipping, over modulation (LM100-NTSC version), high or low signal levels, silence, and incorrectly set dialnorm values.

LOUDNESS METERS, INTERNATIONAL STANDARDS

With the Digital Age entering broadcast at the beginning of this century, the International Telecommunications Union realized that several problems in transmission needed solutions. Now a larger dynamic range was available. The transmission would cover formats from mono to surround. Downstream conversion to lower bit rates would create an alteration of peak levels. However, the most serious problem was the different levels perceived by the listener when zapping between the channels or when the same channel was changing between different contents: the loudness problem.

A workgroup initiated by the ITU facilitated psychoacoustic testing of different already available and new loudness meters and algorithms. The "best fit" to real audio mono samples was a Leq of the RLB-weighted signal. After this conclusion, the work was expanded to include the evaluation of program contents in stereo and surround as well. This involved a pre-filter to compensate for the presence of the head and a gain to surround channels due to the special awareness effect of sound coming from the rear. The total weighting of the signal (frequency and level) got the name K-weighting ("K" being an available symbol for this purpose). The algorithms were first defined in the Recommendation ITU-R BS.1770 from 2006.

THE ITU LOUDNESS METER

The advantage of the applied algorithm is that it is not owned by any private company, and anybody can manufacture loudness meters. The meter specifications are explained in the ITU-R BS.1771–1: Requirements for loudness and true-peak indication meters (2012).

The practical implementation of a meter led to the introduction of the Loudness Unit, LU. In this standard, the LU is a relative unit. The absolute loudness level at the reference indication can be defined elsewhere. The absolute level of the program loudness is defined by LKFS (in EBU the same unit is called LUFS), meaning the level of the K-weighted signal relative to full scale. So the reference indication, the target loudness (0 LU), must be stated in LKFS.

Also, the absolute peak level (true-peak) had to be considered. In digital systems, oversampling must be used. At four times oversampling, the worst case of under-read is in the range of 0.6 dB. At eight times oversampling, the worst case of under-read is in the range of 0.15 dB. For the standard, the four times oversampling was chosen for the true-peak reading.

DEFINITIONS

The loudness measurement encompasses three different settings for loudness reading.

Integrated Loudness or Program Loudness makes use of a relative measurement gate. The program is integrated over its complete length.

Momentary Loudness is defined as the ungated loudness in Recommendation ITU-R BS.1770 loudness algorithm when passed through a first-order IIR (infinite impulse response) low-pass filter with a 400 ms time constant.

Short-term Loudness is defined as an ungated measurement. The integration time is 3 seconds.

Loudness Unit (LU): The loudness unit is the scale unit of the loudness meter. The value of the program in loudness units represents the loss or gain (dB) that is required to bring the program to 0 LU (e.g., a program that reads −10 LU will require 10 dB of gain to bring that program up to a reading of 0 LU).

Further, the display was standardized:

Type I electronic display: Electronic display with resolution of one or more segments per loudness unit.

Type II electronic display: Electronic display with resolution of one segment per 3 loudness units.

The following features are required to fulfill the recommendation:

General requirements	The loudness display reading must not vary by more than 0.5 LU when the signal polarity is reversed.
Common requirements for program loudness displays	The loudness display shall be calibrated in LU. The loudness of a stereo or multichannel sound program shall be shown by a single display. (This does not prevent meters from also displaying individual channel loudnesses.)
Requirements for program loudness display – mechanical type	A mechanical loudness meter display shall have a nonlinearity of not more than 1% of full-scale deflection over its operating range (see Figure 14.3).
Display requirements – Optional peak level indicator on loudness meter	The threshold for overload indication shall be −2 dB re full-scale digital input. The overload indicator shall activate if the true-peak digital audio level exceeds the threshold (see Figure 14.4). Once the indicator light is activated, it shall remain activated for at least 150 ms after the signal has fallen below the threshold.

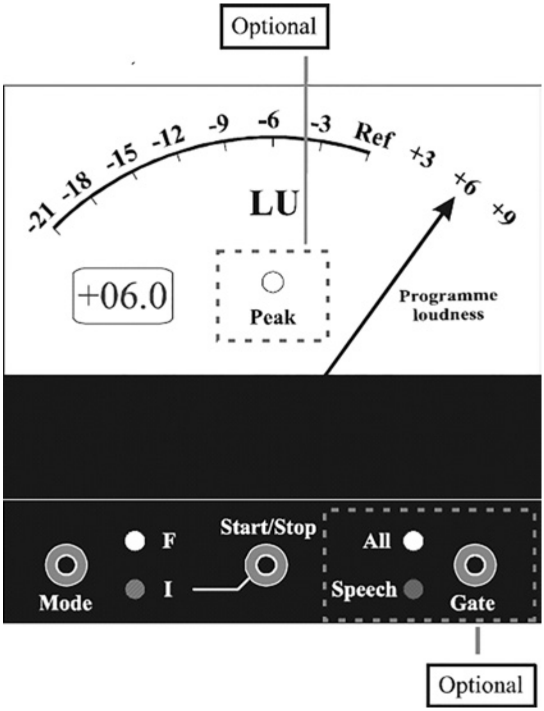

Figure 14.3 Example of program loudness display, the mechanical type given by ITU-R BS.1771.

In addition to these mandatory requirements, several optional facilities are proposed. One of these is that the loudness meter may have at least two operating modes: F mode (fast) and I mode (integrating). Further, the integrating mode may have a start/stop button or switch (see Figures 14.4 and 14.5).

Later further updates were adopted by the ITU 1770 recommendation.

EBU R128 LOUDNESS METERING

The European Broadcast Union wanted to implement the loudness measure already accepted by the ITU. Further to this, a gating technique was considered as well as a new descriptor called "Loudness Range." The new practice should, at the same time, replace the existing standard for PPM and exploit the dynamic range provided by digital systems.

The EBU R128 was first published in autumn 2010. This recommendation introduced three descriptors: Program Loudness, Loudness Range, and Maximum True

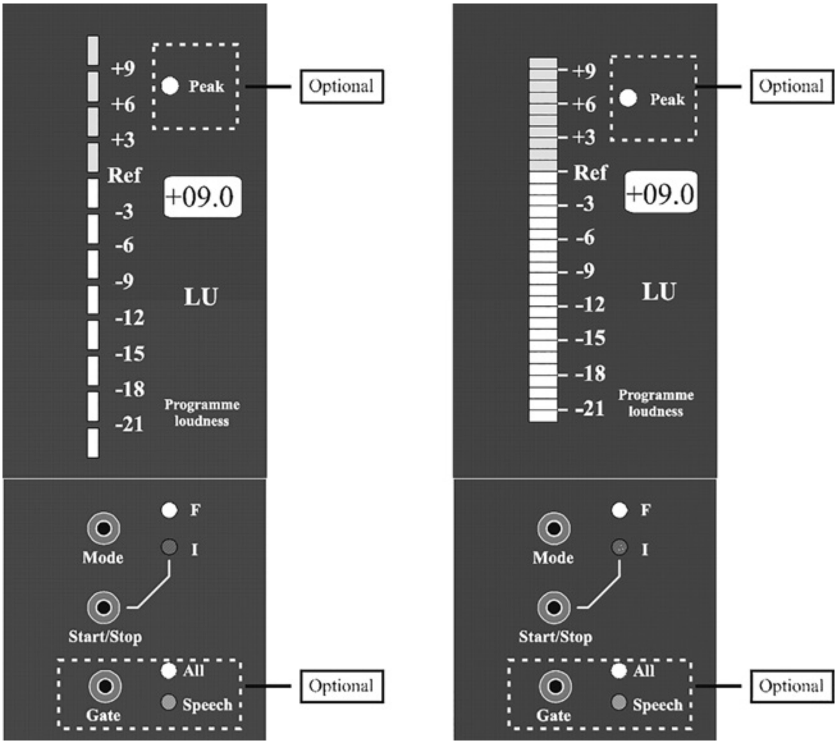

Figure 14.4 Example of program loudness level display, optoelectronic Type I (left) and Type II (right) given by ITU-R BS.1771 [5].

Peak Level. It also defined target levels as well as specifications for a meter to display the measures.

PROGRAM LOUDNESS

The Program Loudness is determined using the K-weighting (as per ITU-R BS.1770) and to average over the total length of the program. However, this includes a gating 10 LU below the target level. Whenever the program is below this gating level, the loudness calculation is paused (see Chapter 10: Determination of Loudness).

LOUDNESS RANGE (LRA)

The Loudness Range is defined in the EBU Technical Document 3342 [6]. Originally this descriptor was developed by TC Electronic (first named "Consistency"). LRA is defined as the difference between the estimates of the 10th and the 95th percentiles of the distribution. The lower percentile of 10%, can, for example, prevent the fade-out of

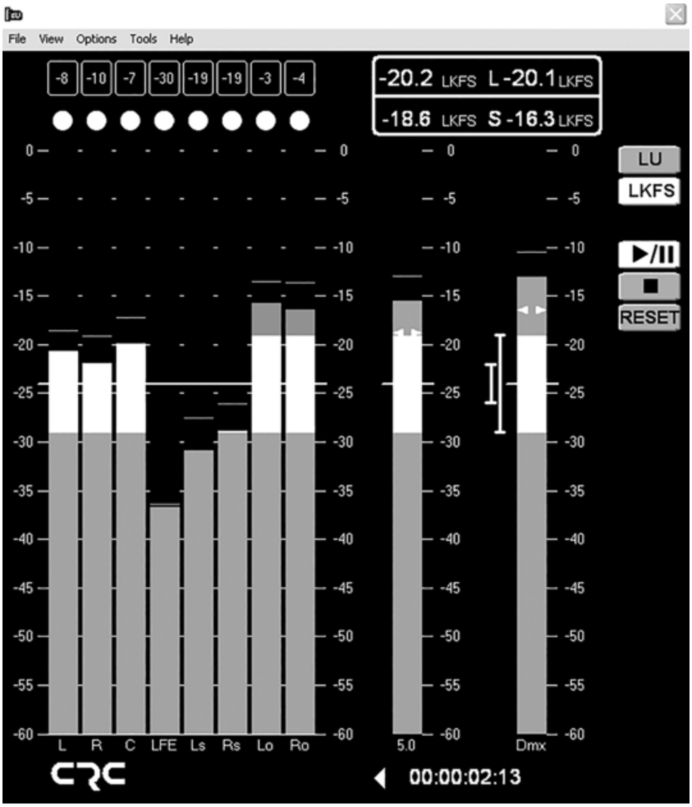

Figure 14.5 This display, of the ITU-R loudness meter developed at CRC, is the result of CRC's collaboration with CBC/Radio-Canada. It can be seen that the scale can display either LKFS or LU.

a music track from dominating Loudness Range. The upper percentile of 95% ensures that a single unusually loud sound, such as a gunshot in a movie, cannot by itself be responsible for a large Loudness Range.

The computation of the Loudness Range is based on the statistical distribution of measured loudness using a sliding 3-second analysis window for integration. An overlap between consecutive analysis windows is used to retain precision of the measurement of shorter programs. A minimum block overlap of 66% (i.e., minimum 2 s of overlap) between consecutive analysis windows is required. By doing this, a short but very loud event will not affect the Loudness Range of a longer segment. Similarly, the fade-out at the end of a music track, for example, will not increase Loudness Range noticeably. Specifically, the range of the distribution of loudness levels is determined by estimating the difference between a low and a high percentile of the distribution.

Loudness Range, furthermore, employs a gating method. The relative threshold is set to a level of −20 LU relative to the absolute-gated loudness level. Certain types of programs may be, overall, very consistent in loudness but have some sections with very low loudness, for example, only containing background noise (e.g., like atmos-sounds). If Loudness Range did not apply the gating, programs like that would (incorrectly) get quite a high Loudness Range measurement.

The purpose of the absolute-threshold gate is to make the conversion from the relative threshold to an absolute level robust against long periods of silence or low-level background noise. The absolute threshold is set to −70 LUFS because, generally, no relevant signals are found below this loudness level.

In EBU documents, LRA is a production guiding measure rather than a content qualification tool, and the optimum LRA is somewhat genre dependent.

MAXIMUM TRUE PEAK LEVEL

The maximum true-peak level of content is defined to −1dBTP measured with a meter compliant with both ITU-R BS.1770 and EBU Technical Document 3341 [7]. Hence four times oversampling is applied to ensure a correct reading. Delivery via lossy codecs may require the true-peak level to be limited to a lower value prior to encoding.

METER, EBU MODE

To ensure that the different descriptors are measured and reported correctly and not mixed up with other measures, any meter that measures, according to R128, must have an "EBU Mode." When set in this mode, it complies with EBU Tech. Doc. 3341.

The EBU Mode does not concern the graphical/UI details or the implementation of a meter.

THE THREE TIME SCALES

Regarding integration time scales and their terminology:

The shortest time scale is called "momentary," abbreviated "**M**."

The intermediate time scale is called "short-term," abbreviated "**S**."

The program- or segment-wise time scale is called "integrated," abbreviated "**I**."

The loudness meter shall be able to display the maximum value of the "momentary loudness." This maximum value is reset when the integrated loudness measurement is reset.

INTEGRATION – TIMES AND METHODS, METER BALLISTICS

In all cases in which the measurement is performed as specified in the ITU-R BS.1770, the measurement parameters for EBU Mode are:

The **momentary loudness** uses a sliding rectangular time window of length 0.4 s. The measurement is not gated.

The **short-term loudness** uses a sliding rectangular time window of length 3 s. The measurement is not gated. The update rate for "live meters" shall be at least 10 Hz.

The **integrated loudness** uses gating as described in ITU-R BS.1770. The update rate for live meters shall be at least 1 Hz.

The EBU Mode loudness meter shall at least provide functionality that enables the user to start/pause/continue the measurement of integrated loudness and Loudness Range simultaneously, that is, switch the meter between "running" and "stand-by" states, as well as reset the measurement of integrated loudness and Loudness Range simultaneously, regardless of whether the meter is in the "running" and "stand-by" state.

THE MEASUREMENT GATE

The "integrated loudness" shall be measured using the gating function as described earlier.

LOUDNESS RANGE (LRA) DESCRIPTOR

An EBU Mode meter shall be able to compute the Loudness Range, which is supplementary to the measure of overall loudness, that is, "integrated loudness." The computation of the Loudness Range is based on a measurement of loudness level. The term "Loudness Range" is abbreviated "LRA." LRA is measured in units of "LU." It is noted that 1 LU is equivalent to 1 dB.

UNITS

A **relative** measurement, such as relative to a reference level, or a range: L_K = xx.x LU

An **absolute** measurement, L_K = xx.x LUFS

The "L" in "L_K" indicates loudness level, and the "K" indicates the frequency weighting used.

TRUE PEAK MEASUREMENT

True peak is measured using four times oversampling.

SCALES AND RANGES

The display of an EBU mode meter may be simply numerical or an indication on a scale.

The scale used may either be an absolute scale, using the unit "**LUFS**," or alternatively the zero points may be mapped to some other value, such as the target loudness

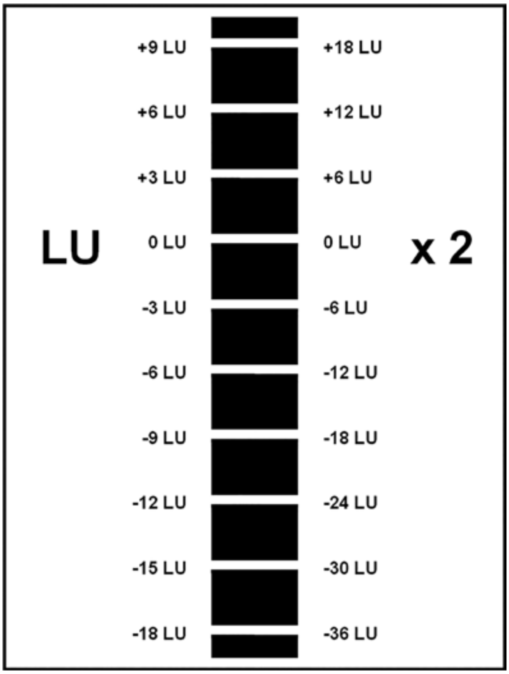

Figure 14.6 EBU Mode scales.

level (as in BS.1771). In the latter case, the unit shall be "**LU**," indicating a relative scale (see Figure 14.6). For an EBU Mode meter, the target loudness level shall be **−23 LUFS = 0 LU** (as defined in EBU R128). The EBU Mode meter shall offer both the relative and the absolute scale.

The location of the target/reference loudness level shall remain the same, regardless of whether an absolute or relative scale is displayed.

An EBU Mode meter shall offer two scales, for when a scale is shown, selectable by the user:

range −18.0 LU to +9.0 LU (-41.0 LUFS to -14.0 LUFS), named "**EBU +9 scale**"

range −36.0 LU to +18.0 LU (-59.0 LUFS to -5.0 LUFS), named "**EBU +18 scale**"

The "EBU +9 scale" shall be used by default.

TC ELECTRONIC LM2N/LM6N

TC Electronic has taken an active role in the work of establishing standards for loudness measures, including those for metering and loudness normalization. The TC Loudness

Radar Meter was developed early in the process (2006) and has since become a widely used tool [8].

The LM2n/LM6n (two or six-channel Loudness Metering) are plug-ins for implementation with most DAWs as the software complies with the AAX, VST, and AU standards.

The main screen has a basic graphic circular shape. Some of the descriptors displayed by the meter's main screen are Short-Term Loudness (outer ring) and Loudness History (radar). Also, the true peak levels are displayed along with Loudness Range (LRA) and of course, Integrated Loudness (see Figure 14.7).

It can be normalized to both the ITU and the EBU recommendations. However, the reference can be set manually, which may especially apply to loudness levels of streaming services.

An additional measure is added to the gallery: Peak-to-Loudness Ratio (PLR) to estimate how compressed the program material is. In version LM6n, it is possible to generate log files.

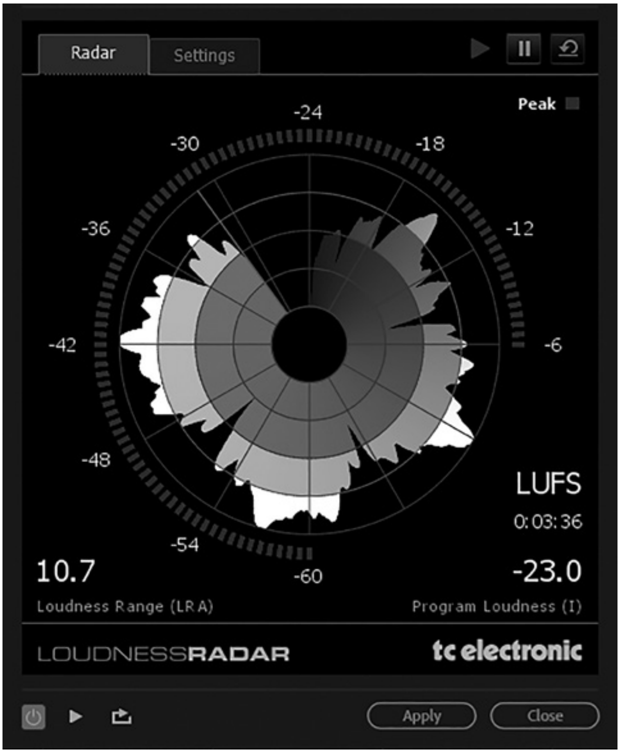

Figure 14.7 The characteristic radar display on LM2n from TC Electronic. The speed of the radar can be set.

TC Electronic also delivers the metering software along with various processing tools for Loudness Normalization.

As a standard, Adobe's Audition CC and Premiere Pro CC are delivered with the TC Electronic Loudness Radar Meter.

RTW TOUCHMONITOR TMR7

RTW is another company that is found among the pioneers of loudness metering. The company is a manufacturer of a very comprehensive monitor for broadcast, the TMR7 [9]. It is software-based but delivered as a stand-alone piece of hardware, either for built-in or tabletop placement. It is a four-channel meter that as a standard includes PPM, Stereo-Correlator, VSC Vectorscope, AES-status, and Gain Reduction. It is designed for connection via a network. A software packet for this meter transforms it to a complete loudness measurement solution that conforms to all relevant standards and recommendations.

The module expands the basic 4-channel PPM with Loudness measuring functions as described in current guidelines: EBU R128, ITU-R BS.1770–4/1771–1, ATSC A/85, ARIB, OP-59, AGCOM, CALM, LEQ(M), TASA, SAWA, and the LRA instrument for the graphical display of the Loudness Range.

A customer-specific mode allows the user to modify parameters. Besides, the module provides the SPL display mode with various weighting filters and integration times as well as reference level adjust to calculate an SPL value from an electrical signal.

RTW also manufactures less comprehensive models, like the TM3-Primus.

NUGEN VISLM2

Nugen is a company that has much experience in software-based metering. VisLM is one of the solutions [10]. The main screen displays the relevant values as described by the standards and recommendations. However, as a special feature, it also shows a histogram of the actual distribution of levels. The meter is highly configurable. It includes the Leq(m) for specific annoyance-measurements as described by TASA.

IZOTOPE INSIGHT 2

iZotope is a company that has introduced an overwhelming number of plug-ins for DAWs. Many of these algorithms are said to be built on artificial intelligence, thus providing many audio optimization tools. One plug-in is the Insight, which is a tool for metering and analyzing the audio [11]. Some of the measures include loudness. Almost all essential standards and recommendations are included. Also, the target values, not only from ITU, EBU, and AES [12], but also ATSC A/85, and the requirements for Netflix delivery.

REFERENCES

[1] ITU-R BS.1770-4 Algorithms to measure audio programme loudness and true-peak audio level (2015).

[2] EBU Technical Recommendation R128 Loudness normalisation and permitted maximum level of audio signals (2014).

[3] EBU Technical Recommendation R128 s1 Loudness parameters for short-form content (adverts, promos, etc.) (2014).

[4] Lund, Thomas; Skovenborg, Esben. *Level-normalization of feature films using loudness vs speech*. SMPTE Meeting Presentation (2014).

[5] ITU-R BS.1771-1 Requirements for loudness and true-peak indicating meters (2012).

[6] EBU Technical Document, 3342, Loudness range. A descriptor to supplement loudness normalization according to EBU R128. Version 3 (2016).

[7] EBU Tech. Doc., 3341, Loudness metering. 'EBU Mode' metering to supplement Loudness normalisation according to EBU Technical Document R128. Version 3 (2016).

[8] TC Electronic: Manual LM2n/LM6n loudness radar meters. http://cdn-downloads.tcelectronic.com/media/3641528/tc_electronic_lm2n_lm6n_manual_english.pdf.

[9] RTW: Manual, TouchMonitor TMR7. Software Version 7 (01 February 2019).

[10] Nugen Audio: Operation manual VisLM standardised loudness metering solution. https://nugenaudio.com/files/manuals/VisLM2%20Manual.pdf.

[11] Izotope: Insight help guide. http://help.izotope.com/docs/izotope-insight-help.pdf.

[12] AES Technical Document, AESTD1006.1.17–10. Loudness Guidelines for OTT and OVD Content (October 2017).

CHAPTER 15

Calibration of Level and Loudness Meters

CHAPTER OUTLINE

A level meter is a measuring instrument. However, if a measuring instrument is not calibrated, how does one know that the instrument is measuring correctly?

CONSTANT TONE

An analog instrument is calibrated first and foremost using a constant pure tone. In many cases, the frequency is chosen to be 1 kHz. However, that depends on the type of meter or instrument you want to calibrate. Sound level meters or any measuring device related to acoustic measurements can be calibrated using a continuous 1 kHz pure tone. The level must be constant. At 1 kHz, the weighting filters (like IEC A) are neutral. Many program meters accepts or require a constant 1 kHz tone as well for checking the reading. In loudness meters (ITU/EBU), one should be aware that 1 kHz lies on the slope of the pre-filter. Hence, very narrow tolerances are needed in the filter design.

If you are in possession of a high-precision digital RMS voltmeter, you can then measure the magnitude of the signal for the comparison.

TONE BURSTS

The integration time, or perhaps more correctly, the reaction time of an instrument is tested using a tone burst generator. The time it takes the meter to reach a given percentage of the reference level usually defines the calibration. A tone burst generator supplies a signal for a well-defined time interval, for example, 5 ms or 300 ms, and repeats after a longer fixed interval such as 1.5 s or 300 ms (see Figure 15.1).

Note that the frequency carried within the burst is 5 kHz in order to have a sufficient number of periods to comprise a tone burst. If the test frequency is too low or the burst does not contain an integer number of periods, there is a risk of making a wrong measurement.

IEC 60268–10 requires 10 kHz for very short impulses because at least five periods are required to form a suitable tone burst.

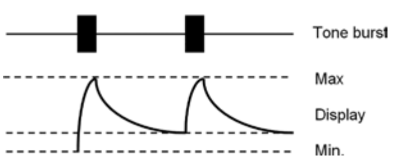

Figure 15.1 Testing an instrument's integration time using a tone burst. The individual standard specifies what the device should read at given time intervals.

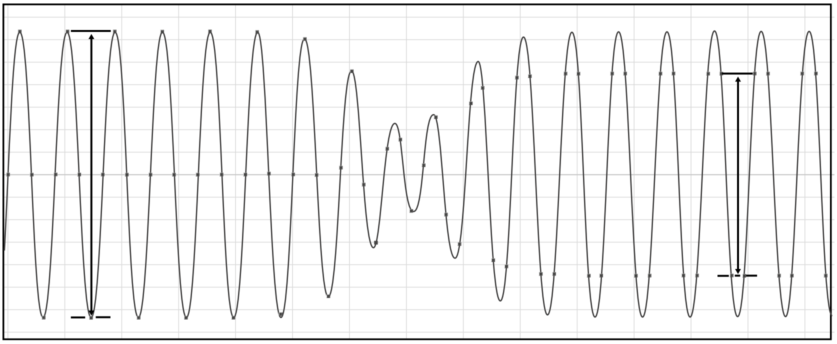

Figure 15.2 A crossfade between two 12 kHz sine waves, both generated to have the same level, and at a sampling frequency of 48 kHz, which provides four samples per period. On the left, the tone has its first sample at 0 degrees (the dots), and hence the others at 90, 180, and 270 degrees of the sine wave period. On the right, the first sample was made at 45 degrees and the following samples at 135, 225, and 315, respectively. The peak level of the first part reads a level 3 dB higher than that of the second part.

Test signals for the calibration of instrumentation exist on various CDs. However, if used for analog inputs or analog instruments, the analog output of the CD players must be checked. If using a DAW, it is possible to generate customized test tone sequences.

It is necessary to design the test signals for digital instruments in a way that does not compromise the sampling rate. For instance, if the test frequency is exactly one fourth of the sampling rate, there may be a problem (see Figure 15.2).

Further, it is possible to download audio files designed for calibration or measurement.

PROCEDURE

Initially, the analog instrument is calibrated with a constant sine wave at the alignment level (reference indication). Then the tone burst signal is supplied. Each of the bursts has the same level as the constant sine wave.

SVI OR VU

According to IEC 60268–17, a Standard Volume Indicator or VU instrument must reach 99% (±10 %) of the reference indication (the “0” marking) when supplying a sine wave with a length of 300 ms. If the length of the sine wave has a shorter duration, the reading must be lower. The fallback time should be identical to the reaction time, meaning the reading should fall to the bottom of the scale in 300 ms.

PPM (QPPM)

The reference indication of the IEC PPM (Quasi Peak Program Meter) instrument must be as follows according to IEC 60268–10:

Type I	1.55 V (+6 dB on Nordic scale)
Type IIa	1.94 V (6 on the BBC scale)
Type IIb	2.18 V (+9 on the IIb scale)

When using the instruments to check line levels, the accuracy of the scales must also be tested by supplying other well-defined levels.

The reaction time defines the dynamic response. The Type I PPM must reach a level 2 dB below the reference indication for a 5 ms sine wave. In the pause of 1.5 seconds, the instrument must manage to fall 20 dB. The reading of the meter should be in accordance with the following scheme (see Table 15.1):

Table 15.1 Dynamic response for IEC PPM instruments in normal mode. However, for a tone burst with a duration of 5 ms, the reading must reach a point 2 db from the reading that occurs at a constant sine wave.

Burst, duration	ms	Display dB	Tolerance dB
IEC Type I	10	–1	±0.50
	5 (ITU recommendation)	–2	±1.00
	3	*–4*	±1.00
	0.4	–15	±4.00
IEC Type IIa	100	6	±0.50
	10	5.5	±0.50
	5	5	±0.75
	1.5	3.75	±1.00
	0.5	1.75	±2.00
IEC Type IIb	100	+8	±0.50
	10	+6	±0.50
	5	+4	±0.75
	1.5	–1	±1.00
	0.5	–9	±2.00

LOUDNESS METER

Calibrating the loudness meter is a little tricky, especially if you want to check more parameters than just constant level. The meter reads both linear audio for the TP

(True Peak) and k-weighted audio for the loudness measurement, resulting in LUFS (Loudness Unit with reference to full scale). The loudness reading includes gating for the integrated loudness. Additionally, the loudness meters calculate the loudness range (LRA).

To perform a full calibration of a loudness meter, various types of test signals are needed. Fortunately, the working groups behind the ITU and the EBU standards have provided test signals that meet the task of a full calibration.

ITU-R BS.1770–1 COMPLIANCE MATERIAL

The ITU (International Telecommunication Union) has prepared compliance material for the loudness meter, according to ITU-R BS.1770–4, 2015. The content is described in ITU-R BS.2217–2, 2016. Due to the complexity of the loudness meter, several audio files are available along with the expected reading of the meter. You can find files for 1, 2, 6, 8, 10, 12, and 24 -channel meters. It is possible to download these files from the ITU website. Browse for "ITU-R BS.2217–2." In the pdf-file, there are links to the individual files.

The gating can be checked, both the absolute gate @-70 LKFS and the relative gate 10 LU below the target level.

The frequency response is checked by files containing sine waves from 25 Hz to 10 kHz.

The various signal also applies to the summing of from 2 up to 24 channels. Further, the pool of files comprises signals for a measured loudness level of −24 LKFS, which is the ITU standard and for −23 LKFS (or LUFS), which refer to the EBU standard.

Finally, a mapping of the channels is provided (Table 15.2).

LOUDNESS METER, EBU MODE

For a loudness meter in the EBU mode, these k-weighted readings are available: M (momentary loudness), S (short-term loudness) and I (integrated loudness). EBU has recommended a series of test signals for alignment and calibration. These signals are described in EBU Tech 3341 (2011) and more fully in EBU Tech 3342 (see Table 15.3).

The signals have been prepared by members of the PLOUD and are available from:

https://tech.ebu.ch/news/2016/03/new-ebu-test-signal-sets-the-lou

Table 15.2 Mapping of the channels.

No. of Channels		Channel ordering and channel weighting					
1 channel (0+1+0)	Channel ID, Label	1 (Mono)					
	Weighting	1.00 (±0.0 dB)					
2 channels (0+2+0)	Channel ID, Label	1 L	2 R				
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)				
6 channels (0+5+0)	Channel ID, Label	1 L	2 R	3 C	4 LFE	5 Ls	6 Rs
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	N/A*	1.41 (+1.5 dB)	1.41 (+1.5 dB)
8 channels (0+7+0)	Channel ID, Label	1 L	2 R	3 C	4 LFE	5 Lss	6 Rss
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	N/A*	1.41 (+1.5 dB)	1.41 (+1.5 dB)
	Channel ID, Label	7 Lrs	8 Rrs				
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)				
10 channels (4+5+0)	Channel ID, Label	1 L	2 R	3 C	4 LFE	5 Ls	6 Rs
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	N/A*	1.41 (+1.5 dB)	1.41 (+1.5 dB)
	Channel ID, Label	7 Tfl	8 Tfr	9 Tbr	10 Tbr		
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)		
12 channels (4+7+0)	Channel ID, Label	1 L	2 R	3 C	4 LFE	5 Lss	6 Rss
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	N/A*	1.41 (+1.5 dB)	1.41 (+1.5 dB)
	Channel ID, Label	7 Lrs	8 Rrs	9 Tfl	10 Tfr	11 Tbr	12 Tbr
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)
24 channels (9+10+3)	Channel ID, Label	1 FL	2 FR	3 FC	4 LFE1	5 BL	6 BR
	Weighting	1.41 (+1.5 dB)	1.41 (+1.5 dB)	1.00 (±0.0 dB)	N/A*	1.00 (±0.0 dB)	1.00 (±0.0 dB)

No. of Channels		Channel ordering and channel weighting					
	Channel ID, Label	7 FLc	8 FRc	9 BC	10 LFE2	11 SiL	12 SiR
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	N/A*	1.41 (+1.5 dB)	1.41 (+1.5 dB)
	Channel ID, Label	13 TpFL	14 TpFR	15 TpFC	16 TpC	17 TpBL	18 TpBR
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)
	Channel ID, Label	19 TpSiL	20 TpSiR	21 TpBC	22 BtFC	23 BtFL	24 BtFR
	Weighting	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)	1.00 (±0.0 dB)

* The LFE channels are not included in a Recommendation ITU-R BS.1770 measurement.

Table 15.3 Minimum requirements test signals used with the EBU loudness meter.

Test case	Test signal	Expected response and accepted tolerances
1	Stereo sine wave, 1000 Hz, −23.0 dBFS (per-channel peak level); signal applied in phase to both channels simultaneously; 20 s duration	M, S, I = −23.0 +/−0.1 LUFS M, S, I = 0.0 +1−0.1 LU
2	As #1 at −33.0 dBFS	M, S, I = −33.0 +/−0.1 LUFS M, S, I = −10.0 +/−0.1 LU
3	As #1, preceded by 20 s of −40 dBFS stereo sine wave, and followed by 20 s of −40 dBFS stereo sine wave	I = −23.0 +/−0.1 LUFS I = 0.0 +/−0.1 LU
4	As #3, preceded by 20 s of −75 dBFS stereo sine wave, and followed by 20 s of −75 dBFS stereo sine wave	I = −23.0 +/−0.1 LUFS I = 0.0 +/−0.1 LU
5	As #3, but with the levels of the 3 tones at −26 dBFS, −20 dBFS, and −26 dBFS, respectively	I = −23.0 +/−0.1 LUFS I = 0.0 +/−0.1 LU
6	5.0 channel sine wave, 1000 Hz, 20 s duration, with per-channel peak levels as follows: −28.0 dBFS in L and R −24.0 dBFS in C −30.0 dBFS in Ls and Rs	I = −23.0 +/−0.1 LUFS I = 0.0 +/−0.1 LU
7	Authentic program 1, stereo, narrow loudness range (NLR) program segment; similar in genre to a commercial/promo	I = −23.0 +/−0.1 LUFS I = 0.0 +/−0.1 LU
8	Authentic program 2, stereo, wide loudness range (WLR) program segment; similar in genre to a movie/drama	I = −23.0 +/−0.1 LUFS I = 0.0 +/−0.1 LU

The above-mentioned test signals were defined in the 2011 version of the EBU Tech 3341. Further “minimum requirements test signals” are defined in the 2015 version of the same document. The files are available from the EBU Technical website, synthesized at a sampling rate of 48 kHz.

REFERENCES

IEC 60268–17: Sound system equipment - Part 17: Standard volume indicators (1991).

IEC 60268–10 A: Sound system equipment - Part 10: Peak programme level meters (1991).

BS.2217–2: Compliance material for Recommendation ITU-R BS.1770 (10/2016).

EBU Tech 3341–2015 Loudness metering: ‘EBU Mode Metering to supplement EBU R 128 loudness normalization. Version 3 (2016).

CHAPTER 16

Relationships Between Scales

CHAPTER OUTLINE

When it comes to level meters, the abundance and different types of appearances of the scales are overwhelming. However, most of them share a common trait in that their scales use divisions that are based on units in dB. There can be a difference between the scales as far as the portion of the dynamic range that they cover. For certain purposes, it is important to be able to monitor the entire dynamic range. In other cases, it is only important to see what is occurring near full modulation. Some scales combine both properties.

As far as professional equipment is concerned, good and readable scales predominate. With consumer equipment, it is often the case that the aim is just to have something that moves, and here there is rarely any real possibility of performing a calibration.

COMMENTS ON THE SCALES

Several of the scales currently used are shown in the following Figure 16.1 together with their relationship. Some comments follow in connection with how the scales are shown.

VOLTS

This applies to analog equipment only. The scale in volts is included for the sake of comparison. The value of this voltage, is the RMS and not the peak value.

dBu

This applies predominantly to analog equipment. However, also digital equipment may use the term dBu if any analog connections are present, such as in converters. The unit dBu is an absolute magnitude with the reference 0.775 V. Most of the scales, therefore, express the voltage of the signal.

IEC I, NORDIC SCALE

The most important scale among professionals and broadcasters in the Nordic countries is the Nordic scale. It refers to analog values but may apply to digital equipment.

volts	dBu	IEC I Nordic	IEC IIa BBC	IEC IIb	DIN	SVI or VU			EBU AD/DA
						Direct reading	North Am., Australia	France	
12.28	24								
10.95	23								
9.76	22								
8.70	21								
7.75	20								
6.91	19								
6.16	18								0
5.49	17								–1
4.89	16								–2
4.36	15								–3
3.88	14								–4
3.46	13								–5
3.09	12	12	7	12					–6
2.75	11	11		11	5				–7
2.45	10	10		10	4				–8
2.18	9	9		9	3				–9
1.95	8	8	6	8	2				–10
1.74	7	7		7	1	+3			–11
1.55	6	6		6	0	+2			–12
1.38	5	5		5	–1	+1			–13
1.23	4	4	5	4	–2	0			–14
1.10	3	3		3	–3	–1	+3		–15
0.976	2	2		2	–4	–2	+2		–16
0.870	1	1		1	–5	–3	+1	+3	–17
0.775	0	test	4	test	–6	–4	0	+2	–18
0.691	–1	–1		–1	–7	–5	–1	+1	–19
0.616	–2	–2		–2	–8	–6	–2	0	–20
0.549	–3	–3		–3	test	–7	–3	–1	–21
0.489	–4	–4	3	–4	–10	–8	–4	–2	–22
0.436	–5	–5		–5	–11	–9	–5	–3	–23
0.388	–6	–6		–6	–12	–10	–6	–4	–24
0.346	–7	–7		–7	–13	–11	–7	–5	–25
0.309	–8	–8	2	–8	–14	–12	–8	–6	–26
0.275	–9	–9		–9	–15	–13	–9	–7	–27
0.245	–10	–10		–10	–16	–14	–10	–8	–28
0.218	–11	–11		–11	–17	–15	–11	–9	–29

Figure 16.1 Relation between common scales.

volts	dBu	IEC I Nordic	IEC IIa BBC	IEC IIb	DIN	SVI or VU			EBU AD/DA
0.195	–12	–12	1	–12	–18	–16	–12	–10	–30
0.174	–13	–13			–19	–17	–13	–11	–31
0.155	–14	–14			–20	–18	–14	–12	–32
0.138	–15	–15			–21	–19	–15	–13	–33
0.123	–16	–16			–22	–20	–16	–14	–34
0.109	–17	–17			–23		–17	–15	–35
98 m	–18	–18			–24		–18	–16	–36
87 m	–19	–19			–25		–19	–17	–37
78 m	–20	–20			–26		–20	–18	–38
69 m	–21	–21			–27			–19	–39
61 m	–22	–22			–28			–20	–40
55 m	–23	–23			–29				–41
49 m	–24	–24			–30				–42
44 m	–25	–25			–31				–43
39 m	–26	–26			–32				–44
35 m	–27	–27			–33				–45
31 m	–28	–28			–34				–46
28 m	–29	–29			–35				–47
⇓	⇓	⇓	⇓	⇓	⇓				⇓
–	–	–42	–∞	–∞	–50				–

Figure 16.1 (Continued)

IEC IIA, BBC

This scale also refers to the analog signal, but it is also implemented in digital equipment. Concerning the standard, there are 4 dB between each subdivision. However, there are some instruments with 6 dB between the lowest steps on the scale.

IEC IIB

In broadcast, this scale is used for the adjustment of analog transmission lines.

DIN

The broadcast standard for German-speaking countries is found in what formerly was known as the “Pflichtenheft 3/6” (see Figure 16.2). Now the title is: Technische

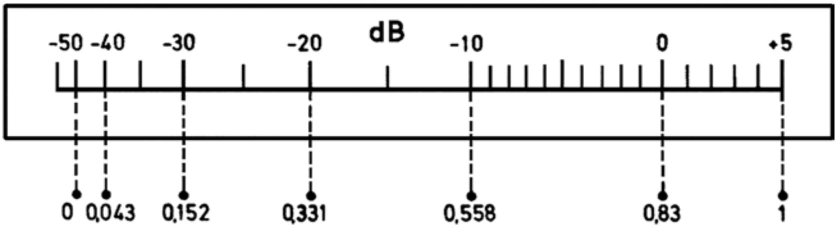

Figure 16.2 Scale subdivisions as originally described by IRT in Richtlinie Nr. 3/6 (2013).

Richtlinie der öffentlich-rechtlichen Rundfunkanstalten in der Bundesrepublik Deutschland, Richtlinie Nr. 3/6: Aussteurungsmesser für Produktion und Sendung (2013). This is issued by Institut für Rundfunktechnik.

However, the reference behind this is now the IEC standards, which is also the DIN standard regarding metering.

SVI OR VU

The VU-scale is used with very little respect for any standard. It must be mentioned that the scale originally does not have VU as a unit. The VU on the scale only indicates that this instrument is a Standard Volume Indicator. The reading of the instrument is found on the attenuator in front of the meter.

That said, we must acknowledge the different ways this instrument and its scale is used – and mostly without the attenuator applied.

There are three different scale offsets in Figure 16.1:

Direct reading: When 1.23 volts is applied, the reading on the scale is 0, if either no attenuator is used in front of the meter *or* the attenuator in front of the meter is set in a position where it does not attenuate. However, the volume is +4 (dBm) according to the standard.

North America/Australia: The reading of 0 is reached at 0.775 volts and no attenuation is applied.

France: The scale is further displaced so that 0 on the scale corresponds to 0.616 V.

M/S-SCALES

Some meters provide an M/S-scale, displaying the direct relation of the in-phase contents compared to the out-of-phase contents of left and right channel. This applies to both analog and digital equipment.

EBU VS. SMPTE

The last scale shows the relation to EBU R68, which gives the correlation between the analog signal level and its digital coding. Another standard exists, namely, SMPTE RP155. Unfortunately, they are not identical. The relationship between the two is shown in the following section. It shows that 0 dBFS corresponds to +18 dBu according to the EBU standard (see Figure 16.3).

Similarly, 0 dBFS is equal to +24 dBu, according to SMPTE. Hence, you must know the conversion factor to ensure that inexplicable jumps in the level of up to 6 dB do not occur. This happens, in particular, if you have a mixture of American and European sources.

	EBU R68			**SMPTE RP155**			
	dBFS	dBu	Volts	dBFS	dBu	Volts	
	0	18	6.16	0	24	12.28	**Max level**
	-1	17	5.49	-1	23	10.95	
	-2	16	4.89	-2	22	9.76	
	-3	15	4.36	-3	21	8.70	
	-4	14	3.88	-4	20	7.75	
	-5	13	3.46	-5	19	6.91	
	-6	12	3.09	-6	18	6.16	
	-7	11	2.75	-7	17	5.49	
	-8	10	2.45	-8	16	4.89	
	-9	9	2.18	-9	15	4.36	
	-10	8	1.95	-10	14	3.88	
	-11	7	1.74	-11	13	3.46	
	-12	6	1.55	-12	12	3.09	
	-13	5	1.38	-13	11	2.75	
	-14	4	1.23	-14	10	2.45	
	-15	3	1.09	-15	9	2.18	
	-16	2	0.976	-16	8	1.95	
	-17	1	0.870	-17	7	1.74	
Test	-18	0	0.775	-18	6	1.55	
	-19	-1	0.691	-19	5	1.38	
	-20	-2	0.616	-20	4	1.23	**0 VU**
	-21	-3	0.549	-21	3	1.09	
	-22	-4	0.489	-22	2	0.976	

Figure 16.3 Relationship between analog signal level and digital coding in converters as per the EBU and SMPTE standards, respectively.

-23	-5	0.436	-23	1	0.870
-24	-6	0.388	-24	0	0.775
-25	-7	0.346	-25	-1	0.691
-26	-8	0.309	-26	-2	0.616
-27	-9	0.275	-27	-3	0.549
-28	-10	0.245	-28	-4	0.489
-29	-11	0.218	-29	-5	0.436
-30	-12	0.195	-30	-6	0.388
-31	-13	0.174	-31	-7	0.346
-32	-14	0.155	-32	-8	0.309
-33	-15	0.138	-33	-9	0.275
-34	-16	0.123	-34	-10	0.245
-35	-17	0.109	-35	-11	0.218
-36	-18	0.098	-36	-12	0.195
-37	-19	0.087	-37	-13	0.174
-38	-20	0.078	-38	-14	0.155
-39	-21	0.069	-39	-15	0.138
-40	-22	0.062	-40	-16	0.123
-41	-23	0.055	-41	-17	0.109
-42	-24	0.049	-42	-18	0.098

Figure 16.3 (Continued)

LOUDNESS SCALES

The loudness scales are not comparable unless the algorithms forming the reading are identical. However, most scales are based on dB-sized units.

The scales implemented in meters that conform to the ITU or the EBU recommendations are basically the same.

REFERENCES

EBU Technical Recommendation R68-2000, Alignment level in digital audio production equipment and in digital audio recorders.

IEC 60268-10 A: Sound system equipment – part 10: Peak programme level meters (1991).

IEC 60268-17: Sound system equipment – part 17: Standard volume indicators (1991).

ITU-R Recommendation BS.645-2: Test signals and metering to be used on international soundprogramme connections.

ITU-R Recommendation BS.646-1: Source encoding for digital sound signals in broadcasting studios.

SMPTE RP 155 Reference levels for analog and digital audio systems (2014).

Technische Richtlinie der öffentlich-rechtlichen Rundfunkanstalten in der Bundesrepublik Deutschland, Richtlinie Nr. 3/6: Aussteurungsmesser für Produktion und Sendung (2013).

CHAPTER 17

Dynamic Scales

CHAPTER OUTLINE

It is a fact that the dynamic range can vary in different types of program material. Pop music can be very compressed and thus have a very limited dynamic range. Classical music and films mixed for the cinema have a large dynamic range (see Figure 17.1).

When recording, this means that the need for headroom will vary depending on the program material.

Also, in the listening situation, the demand for dynamic range is different, depending on where and under which conditions, you listen. If you listen in a car, then the dynamic range must be limited to perceive audio above the background noise. Or if the material is to be played back in the home, where you for some reason are not allowed to play loud, there is similarly a need for limited dynamic range.

Over the years, proposals have been presented by several groups in the pro audio field, leaving possibilities fully to exploit the dynamic range of the storage media or the transmission line.

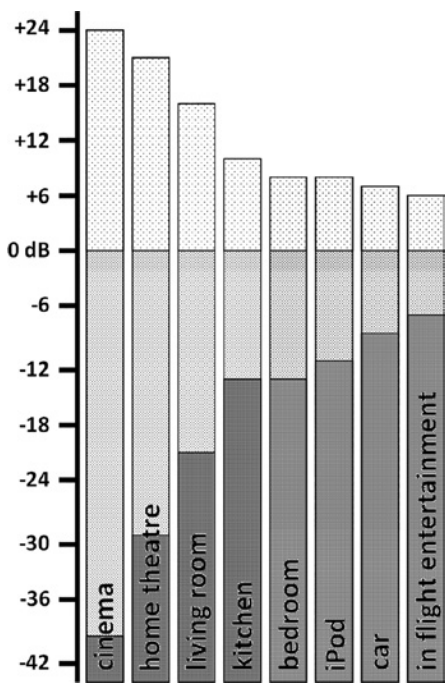

Figure 17.1 Dynamic ranges in various environments normalized with reference to dialogue level (Ref. Thomas Lund, TC Electronic).

The loudness standards from the ITU and the EBU, in general, handle the problem. However, other systems are in use, two of which are described here: an applicable factory standard from Dolby and one from the American producer Bob Katz.

DTV

In the United States, ATSC (Advanced Television Systems Committee) was introduced as a standard for DTV (Digital TV). In this system, it is possible to reproduce up to six channels of sound encoded in Dolby Digital.

Along with the digitized (and bit-reduced) sound, supplemental information is sent in the form of metadata. In popular terms, it can be said that the information that previously was written on the box containing the tape or the disk now is included with the signal itself.

What is interesting about the metadata is that, among other things, it can contain different options regarding level, dynamic range, number of channels, and so on, which can then be selected when listening to the program.

The bitstream that contains the digital sound is divided into frames and blocks. The metadata can either be attached to each frame or each block (for example, compression data). Other metadata applies to the entire program (such as information on level). It is the sender who decides which parameters will be associated with the program. However, the user can select from the options offered and customize the sound according to their requirements.

It must be mentioned that although it is efficient to use metadata, the functionality is never better than the data provided. On the one hand, it can be a problem to generate relevant data if it is not already delivered by the content provider along with the program. On the other hand, it is also a question whether the receiver can apply these metadata for the correct reproduction.

DIALNORM

As earlier mentioned in this book, one of the biggest problems in TV sound is the level differences between programs and level differences between stations. When a viewer browses through the channels, the levels encountered are quite different, particularly regarding speech. It is in a way quite natural because the level of dialogue in movies is mixed at a lower level and with more variation than the dialogue levels in an interview program for television.

One can, of course, decide to always run dialogue at a specific (low) level; however, that would mean a poor utilization of the system's dynamic range.

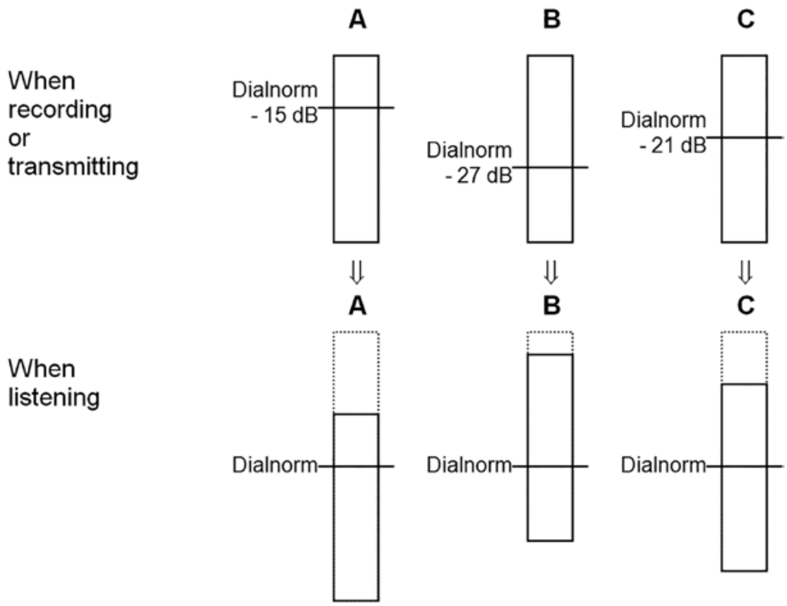

Figure 17.2 Dialnorm is used to adjust dialogue (or the program's average level) to the same listening level, regardless of whether the individual programs were recorded with different level settings, head rooms, and dynamic ranges. The first version of the standard stated that the dialogue level was determined by an A-weighted L_{eq}. Now the measurement/calculation is performed as per ITU 1770.

Instead, "Dialnorm" – an abbreviation of dialogue normalization – was introduced by Dolby. This technique lets the program material get recorded in an optimum manner, exploiting the available dynamic range. Then the dialogue level is specified separately as an anchor of the program. All the program audio subsequently can be shifted in level so that the volume or loudness of the dialogue is uniform from program to program (see Figure 17.2).

The level at which the dialogue is reproduced is also determined by the selection of the dynamic range at the receiver. If the program is reproduced uncompressed (setting: "lin"), then the dialogue level is around −31 dBFS. If the program is reproduced with high compression (setting: "RF"), then the dialogue level is around −20 dBFS, 11 dB higher.

In the U.S., the A/85:2013 ATSC "Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television" states how to achieve the right values regarding the loudness and dialnorm settings in the transmission of digital television.

In general, the program loudness, as measured according to ITU 1770, is the value sent along with the program. This value can be corrected if the content provider has another impression of the dialogue level. However, a content delivery specification should specify the Target Loudness for all content. The content provider should indicate the actual average loudness with the deliverable. Essentially, the recommendation tells the providers to deliver at the right loudness (−24 LKFS) and report this number (24) with the program. If in doubt, then the content supplier should measure and correct according to ITU 1770, even though there are possibilities for correction at later stages in the chain, such as at the AC-3 encoder.

It should be mentioned that interstitials (short-form content) always are measured as integrated loudness rather than dialogue.

THE K SYSTEM

In an attempt to better utilize the dynamic range when recording and to have an indication of the modulation and subjective loudness, a proposal has been presented for a meter system by Bob Katz, an American whose experience includes many years of recording and mastering in the American music industry.

The system – named the K System after himself – builds on three scales, namely, K-20 with 20 dB headroom above 0 dB, K-14, and K-12 with 14 and 12 dB headroom, respectively. On the scales, the color green is used below 0 dB, yellow between 0 and +4 dB, and red for above +4 dB.

Each scale can be used with three different time/frequency weightings, named RMS, LEQA, and Zwicker.

Thus K-X*)/RMS is used with a flat frequency response of 20 Hz–20 kHz ±0.1 dB.

K-X*)/LEQA uses A-weighting (IEC A) and an integration time of 3 seconds.

K-X*)/Zwicker uses, as the name suggests, Zwicker's model for loudness.

For calibrated playback, pink noise at a signal level of 0 dB on the scale will correspond to 83 dB(C) in the listening position.

*) X can be 20, 14, or 12 (see Figure 17.3).

It should be noted that the recording scale goes together with an acoustic listening/monitoring level specified in dB(C).

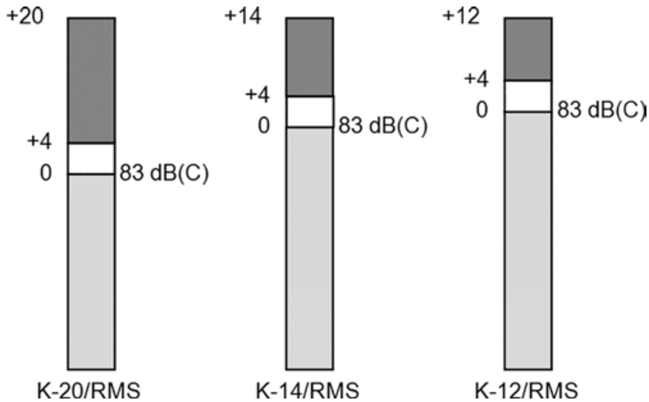

Figure 17.3 The K system. The three scales are specified here with an RMS detector. The scales can also be used in connection with the display of L_{eq} and Zwicker Loudness.

CHAPTER 18

Polarity and Phase Reading

CHAPTER OUTLINE

In the production, recording, and transmission of stereo signals, it is not just the levels of the signals that need monitoring but also the polarity or phase relationship between them. This chapter describes some important instruments that apply to this purpose.

POLARITY

Polarity tells whether the audio signals are “in phase” or “out of phase,” in this case meaning the positive parts of the waveform become negative, and the negative becomes positive. If only looking at – or listening to – one channel (mono) audio being 180° out of phase, it rarely is recognized as a problem. However, in professional audio, no phase inversion should be accepted if not created on purpose.

Looking at the waveform of two channels, the “out-of-phase” situation corresponds to the inversion of only one of the two channels.

In analog audio, this polarity problem is an error far more common than one would imagine. In loudspeaker manufacturing, a certain proportion of units are born with reversed polarity by inadvertently swapping the voice coil leads on the soldering terminals. It is easy to go wrong in the soldering of connectors, patch bays, and so on. Most power amplifiers retain polarity of the signal from input to output; however, others do not. So just by having different amp brands powering the speaker systems may cause problems.

Another area is the microphones. All standards require the pin 2 signal in the XLR-connector to go positive when positive-going pressure is present at the front of the membrane. So far, so good. However, wireless microphones (i.e., miniature microphones connected to a belt pack transmitter) in some cases create a polarity mess. Some major vendors of miniature microphones design them in such a way that the microphones’ unbalanced connection is out of phase. Then the “error” is corrected in the transmitter by inverting the input. The result is that mixing brands (mics and transmitters) may end up creating polarity problems.

All equipment should retain the polarity from input to output. To check this, several clever devices are available: A transmitting device generates a DC impulse (an impulse deflecting in only one direction) either as an electric signal or as the output of a small loudspeaker. The receiver’s microphone or line input detects the direction of the impulse. The readout is a LED display indicating whether the received signal is in phase or out of phase in comparison with the transmitted signal.

Note that in some two-way or three-way loudspeaker systems, the midrange speaker measures as out of phase. This is due to the crossover technique if using second order (12 dB/octave) filters. At the crossover frequency, the phase shift is +90° in one filter and −90° in the other. To make the loudspeaker units work together at this point, the signal to one of the speakers is reversed.

THE PHASE METER

This instrument is used for stereo recording and monitoring. It has a scale ranging from "+1" to "−1" to indicate the current phase difference between the signals in the left and right channels. Instead of presenting the phase angle in degrees, it displays the cosine of the phase angle (the phase difference, see Figure 18.1).

The cosine of an angle of 0° is +1.

The cosine of 90° is 0.

The cosine of 180° is −1.

If the left and right signals are completely in phase (i.e., the phase angle is 0°), then the instrument indicates "1."

If the signals are completely out of phase, (i.e., the phase angle is 180°), then the instrument indicates "−1."

If the phase angle is 90°, the instrument indicates "0" (see Figure 18.2).

If there is only a signal in one of the channels, then the instrument will also display "0" since this position is the neutral position.

When recording stereo signals, the indication should normally lie between "0" and "1."

Normally, the instrument is equipped with a relatively long integration or averaging time, typically 600 ms; this means it has a slow reaction, but it is fast enough to show the phase relationship of the constant and low frequency of varying signals.

This instrument is normally used to assess recordings for gramophone records, where out-of-phase signals at lower frequencies can be synonymous with the stylus just about having to leave the surface of the record if not compensated.

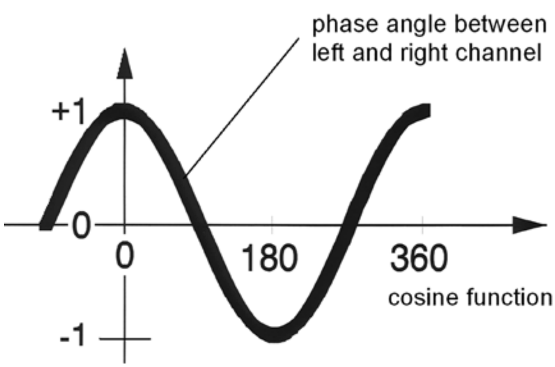

Figure 18.1 The cosine function is shown here. When the phase angle between left and right is 0° the cosine is +1. When the phase angle is 180° the cosine is –1.

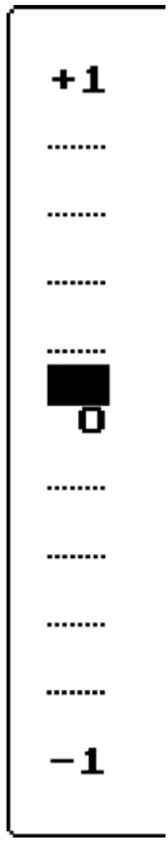

Figure 18.2 The scale of the correlation or phase meter. It has values from +1 to −1.

The instrument is, however, used in all types of downmix for multichannel productions.

The meter may also provide an indication of what happens when signals are played back through matrixed surround systems, where these still exist: When the instrument reads values below 0, the audio disappears into the rear channel.

THE PHASE CURVE

In more complex measurements, it is relevant to measure the phase vs. frequency (e.g., between the actual measured signal and a reference). Just like frequency response is displayed with the frequency on the X-axis and the magnitude on the Y-axis, the phase display exhibits the frequency on the X-axis and the phase on the Y-axis. The unit of

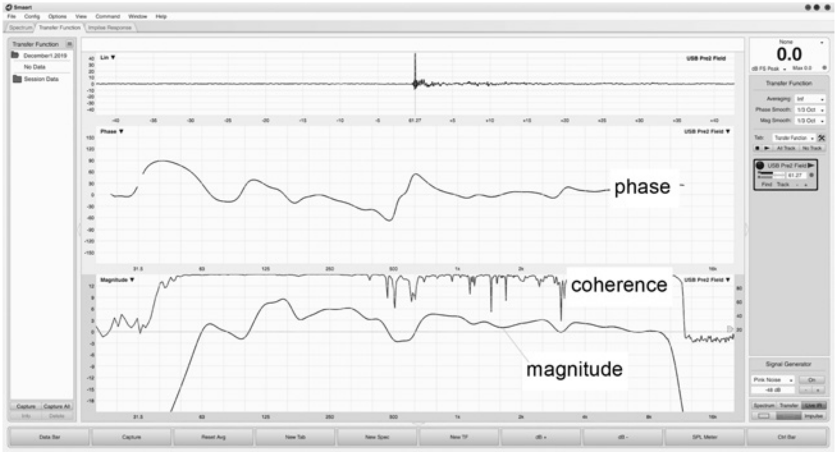

Figure 18.3 The scale on the phase curve normally goes from –180 degrees to +180 degrees. If the phase exceeds these limits, it starts all over, and the reader should add or subtract 360 degrees to the value displayed each time it shifts. Here are two situations where the upper curve shows the phase of the incoming signal emitted from a (two-way) loudspeaker and received by a microphone 1 meter away. The second curve shows the phase curve after synchronization, which is a delay of the reference signal of 3 ms (Screen dump, Smaart Live).

phase is degree. The fundamental scaling on the Y-axis most often is ±180 degrees. If the phase exceeds these values, the curve "starts all over." Thus, phase curves may look noncontinuous, but they are actually continuous – they just are kept within the reading space. It is also common to apply a normalization or a fixed point, where the phase is defined as 0 degrees. For instance, if measuring the transfer function from an amplifier through a loudspeaker and received by a microphone meters away: at the receiving point (the microphone) the sound has been delayed considerably due to the distance through the air. If you want to measure the transfer function, the reference signal must synchronize to the measured signal, which is done by delaying the reference signal.

LISSAJOUS – XY

A Lissajous figure is what you get if you take an oscilloscope and let one audio channel (right) control the X-deflection and let the other channel (left) control the Y-deflection. If the two signals are identical /in phase), the screen displays a sharp line with a slope of +45 degrees. If one of the signals, either left or right, are inverted, the figure is still a sharp line, now sloped −45 degrees. When the phase, is different from 0 degrees in the two channels, the display creates spatial figures (see Figure 18.4). There is one

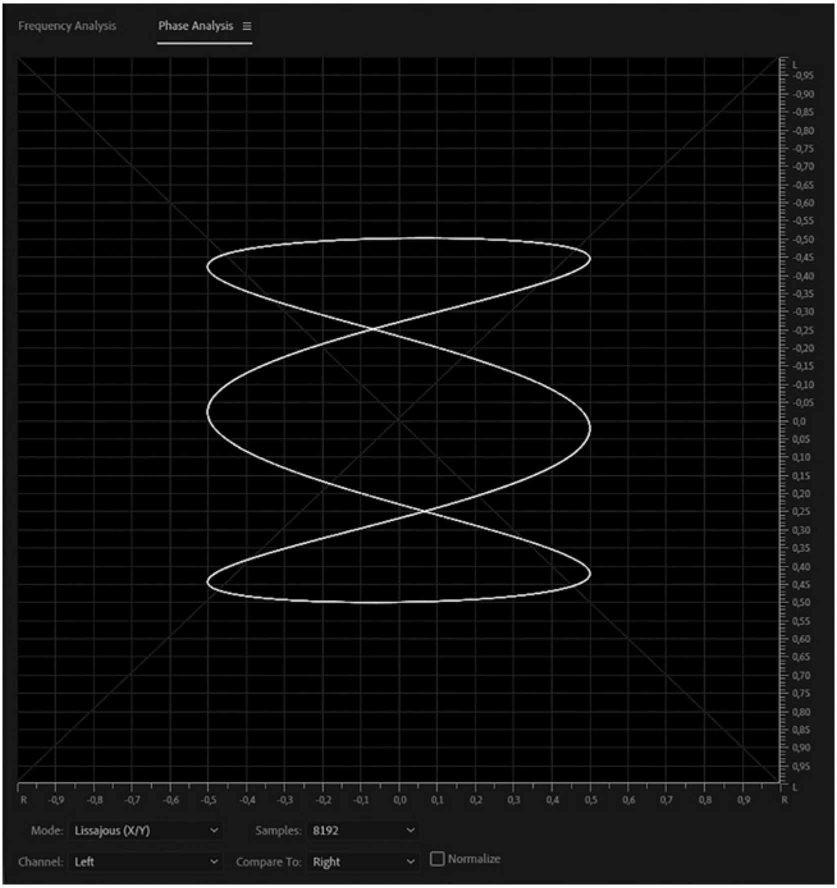

Figure 18.4 Lissajous figure. Input: X = 200 Hz, Y = 600 Hz (Adobe Audition 2020, Phase Analysis).

sine wave for the X-deflection, and the double frequency sine wave for the Y-deflection creates something like a figure of 8. These figures created are the Lissajous, first described by an American mathematician Nathaniel Bowditch looking at the moving pattern of a pendulum. Later this phenomenon was named after a French physicist Jules Antoine Lissajous.

GONIOMETER – AUDIO VECTOR OSCILLOSCOPE – MID/SIDE

The goniometer, or audio vector oscilloscope, or Mid/Side display, is an instrument that can provide a detailed picture of the relationships of a stereo signal – or between two arbitrary signals (see Figures 18.5 and 18.6). Holger Lauridsen, the Chief Engineer at Danish Broadcasting in the 1940s and the early 1950s, developed the idea behind this instrument. It is exactly the same technique as mentioned above (Lissajous), involving the oscilloscope. However, Lauridsen rotated the screen by 45°. Thus, when the same signal is applied on the left and right channels (X and Y), respectively, it leads to a vertical deflection (a vertical line); if the signals are otherwise identical but oppositely phased, then a deflection will occur in the horizontal plane (horizontal line). When the signals are different, the display changes from straight lines to spatial figures (see Figures 18.7–18.10). What is ingenious about the instrument is that it can show many different parameters in the stereo signal simultaneously.

As long as the major portion of the display lies within ±45° around the vertical axis, there is a high degree of mono compatibility (see Figure 18.5). If the curves

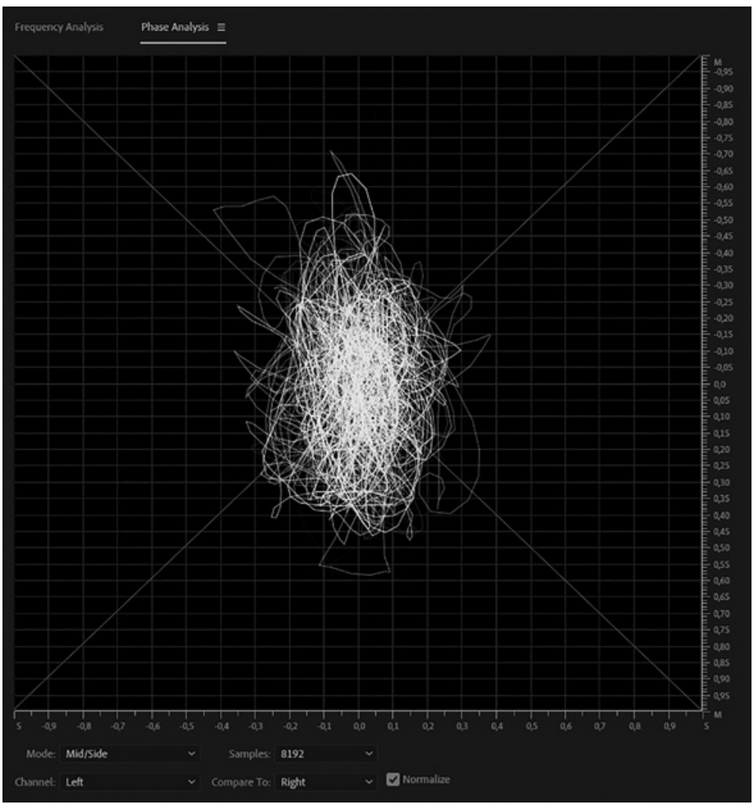

Figure 18.5 The stereo signal is wide, but there is still a high degree of mono compatibility, which is regarded as an optimum (Adobe Audition 2020, Phase Analysis).

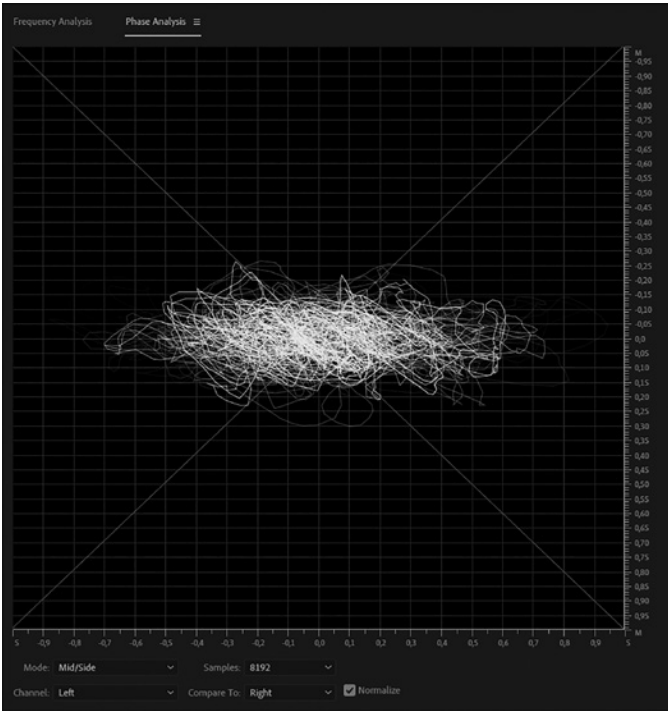

Figure 18.6 The stereo signal is too wide. The mono compatibility is small (i.e., some of the signals will disappear in mono).

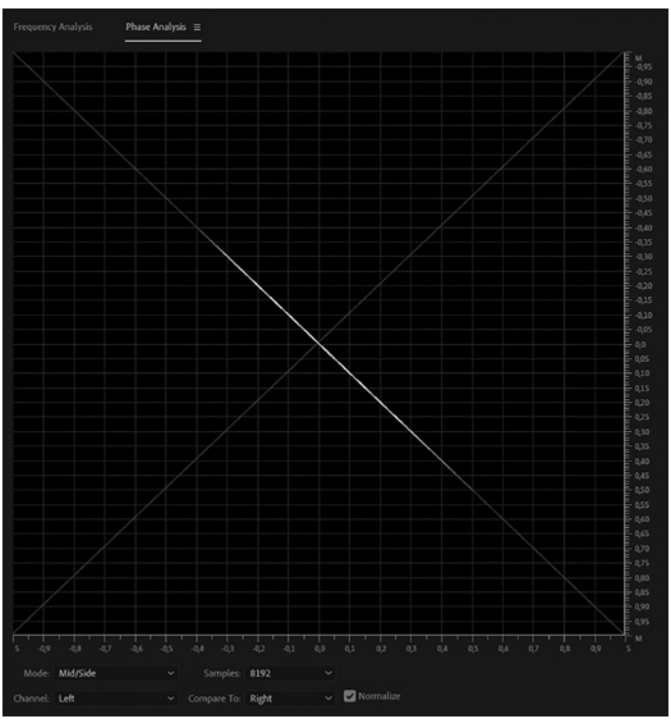

Figure 18.7 Signal only present in the left channel.

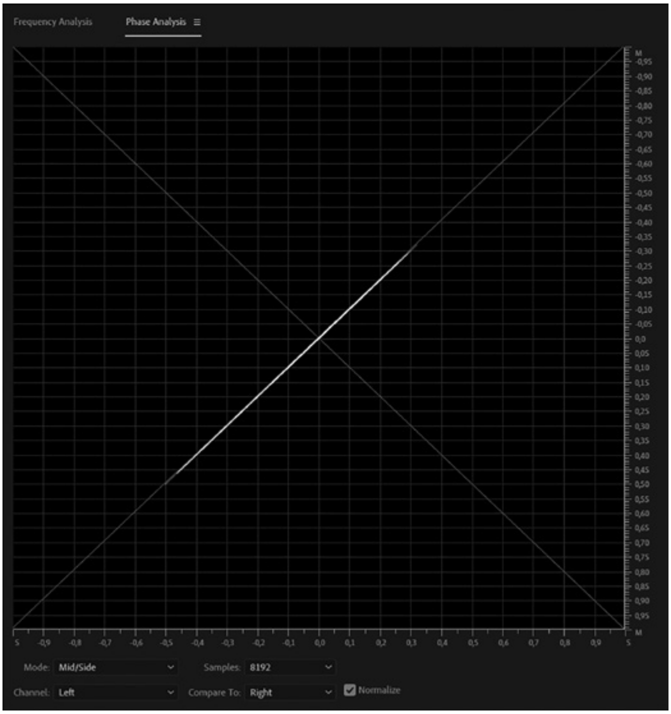

Figure 18.8 Signal only present in the right channel.

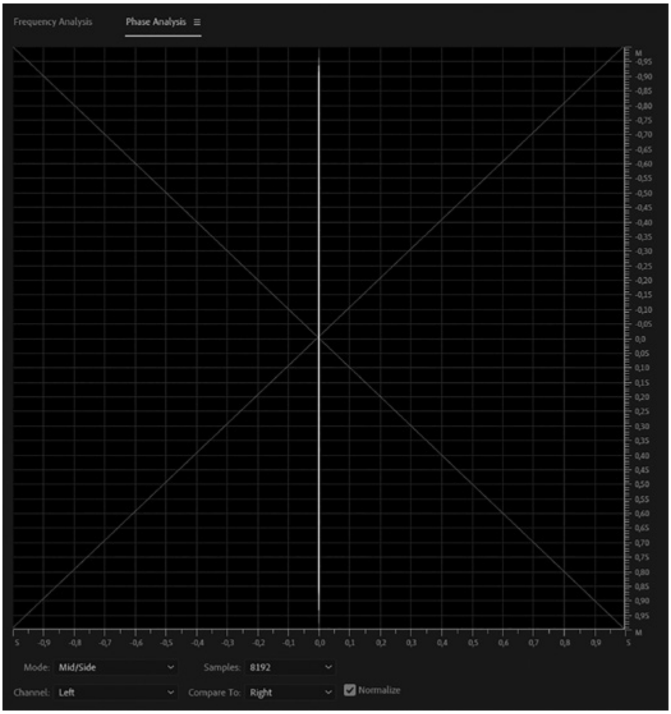

Figure 18.9 Almost a pure mono signal present in the two channels.

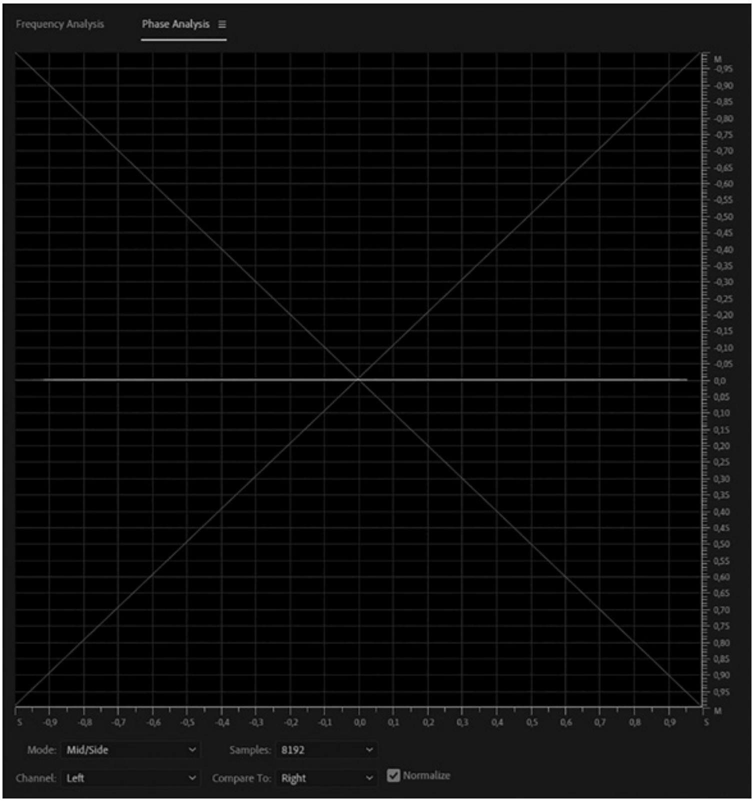

Figure 18.10 Almost a pure mono signal, but in oppositely phased present in the two channels.

begin to bend at the top and bottom, it can indicate nonlinearity. This nonlinearity can arise if overloading or limiting occurs in only one channel. For instance, it may originate from a time delay between the channels that may arise in older digital systems or tape units – or in bad handling of signal processing.

In some DAWs like Adobe Audition, the display is called Mid/Side. In software like iZotope Insight 2, the display is called Lissajous; however, the display is rotated by 45 degrees.

SPECTRAL PHASE DISPLAY

This display is developed as a special analysis tool for DAWs. It shows the phase difference, in degrees, between the left and the right channels recorded. Due to the various

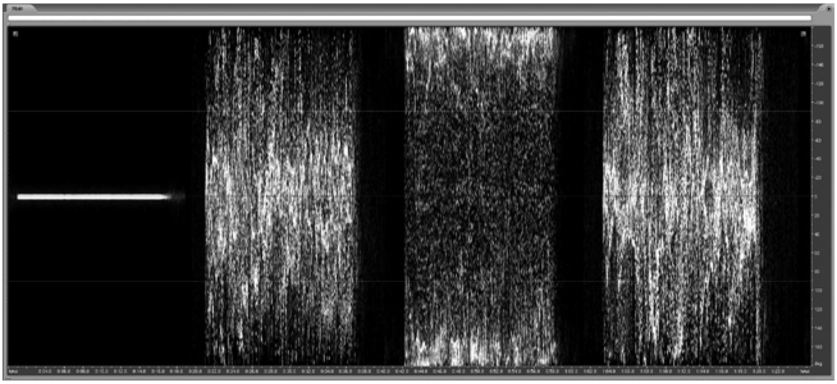

Figure 18.11 This spectral phase display show three sets of microphones recording an organ in a small twelfth century church. The first part is a mono (omnidirectional) microphone, the second part shows the phase distribution of a MS setup placed further away. The third microphone setup is a reversed XY setup. The fourth part shows all three setups recorded simultaneously. This is created in Adobe Audition 3.

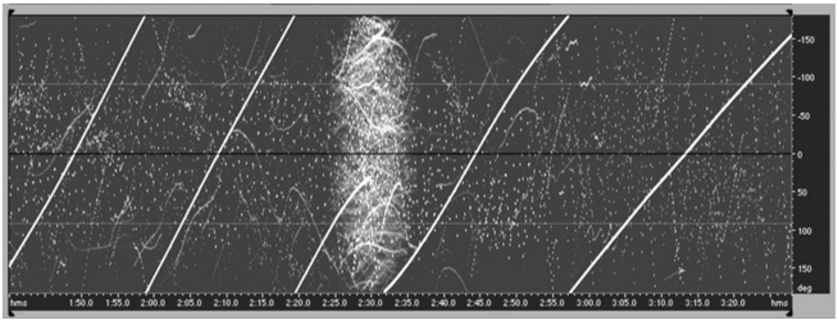

Figure 18.12 In this case, the display is used for phase comparison between an extracted ENF-component (Electric Network Frequency) and a fixed reference. This analysis is performed for authentication of digital recordings in the field of audio forensics; created in Adobe Audition 3.

display colors available, frequency-selective readings are possible. Hence it is possible to distinguish the phase of complex two-channel signals. The horizontal axis is time. The vertical axis is the phase (+/−180°), having 0° in the middle. As mentioned, the colors define different ranges of the frequency spectrum. See Figures 18.11 and 18.12 for practical applications of this display.

CHAPTER 19

Display of Level Distribution

CHAPTER OUTLINE

In several contexts, it is interesting to be able to monitor the signal level over a longer interval of time. The displaying of level can be done either by plotting the level along a time axis or by performing a statistical analysis that shows the level distribution over an interval of time.

LEVEL RECORDER

The level recorder is an instrument in use for nearly all forms of sound analysis. The recorder in its original form would be mechanical with paper and pen. However, in audio, it is more commonly a computerized system, where data is collected and printed out afterward.

The recorder typically must perform some integration of the incoming signal. This “slowing down function” is either obtained by mechanical properties of the level recording system or it is an integration process taken care of by the electrical recording device – either in analog or digital form. An example of NoiseLab, a software-based system for the analysis of environmental noise, is shown in Figure 19.1. Another

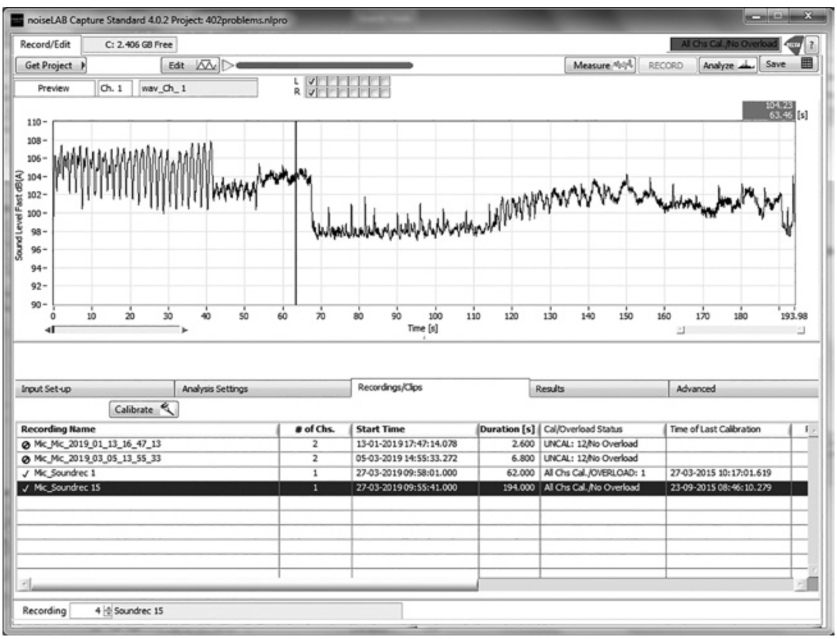

Figure 19.1 A-weighted sound pressure level vs. time. A printout from NoiseLab, a software-based analyzing tool for environmental noise.

Source: (Courtesy dk.madebydelta.com/)DELTA.dk).

Figure 19.2 Example of a level recording in the form of Loudness Level.

Source: (iZotope Insight).

example of software-based analysis of program materials' Loudness is shown in Figure 19.2, iZotope Insight.

HISTOGRAM

The histogram is a graph that shows the statistical distribution within the given time interval considered.

Columns (few or many) show by their width the interval covered and by their length how often this interval is present.

Among other things, the histogram can be used to show level distribution within an audio sequence such as a piece of recorded music. This analysis provides information on the dynamic range of the recording. In signals with a high dynamic range such as classical music, the plot extends across a large portion of the vertical axis, which indicates a large scatter of levels.

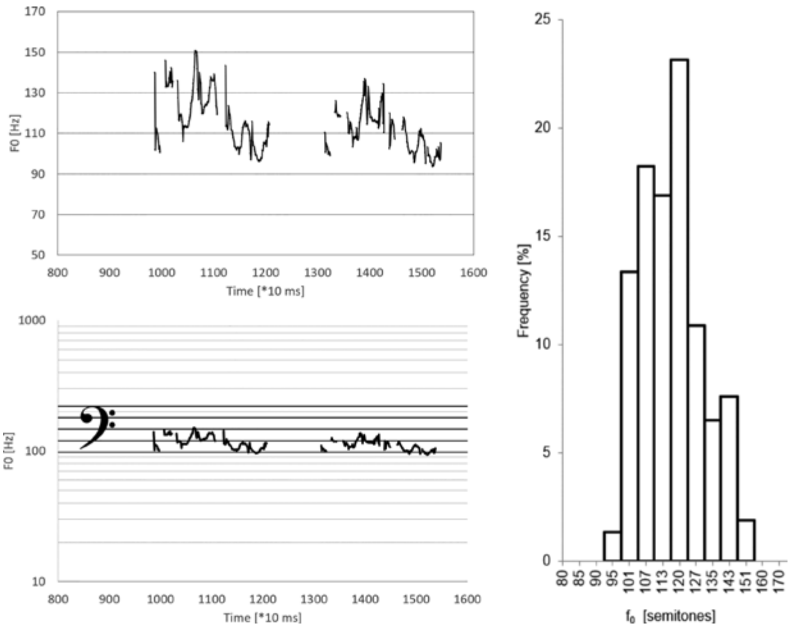

Figure 19.3 Left, upper: Pitch of a person (male) talking for one minute (linear scale). Left, lower: The same data presented on a log scale (music score). Right: The distribution of the data presented in a histogram on a log scale: The units here are semitones.

If much compression is applied, as in pop music, a very narrow figure is traced out. For purposes of comparison, it should be mentioned that a constant tone only results in a single column, at the level concerned.

In histograms for digital recordings of pop music, it is possible to observe levels exceeding 0 dBFS, which in principle is not possible. However, this reading is related to the way the calculation is performed. Presumably, when more than two succeeding samples exceed the maximum, then it might involve overloading, which also can be observed in the D/A-conversion.

Another application for the histogram is for voice analysis. Not only sound levels are examined by applying the histogram. Also, this applies to the examination of pitch or f_0 of a person talking (see Figure 19.3). Each column represents a frequency span (for instance a semitone). The histogram tells whether this was a very monotone (narrow histogram) or lively talker (wide histogram).

OVERLOAD DISPLAY

It is not always possible to constantly keep an eye on the meter scale when checking for possible overloading in a recording. It is, therefore, practical to have a function that

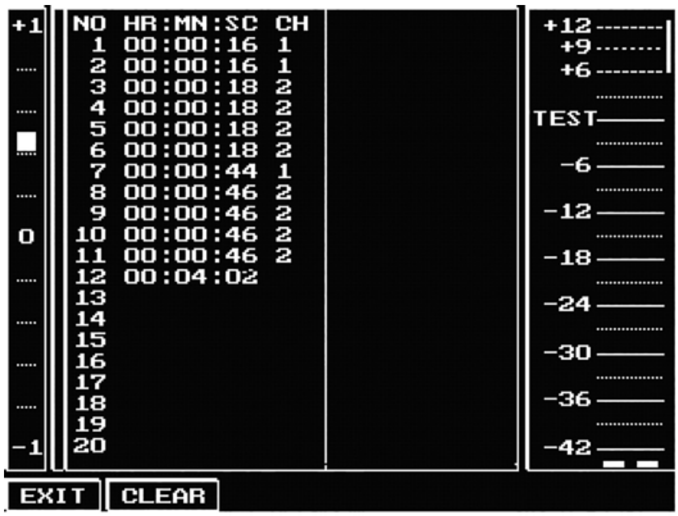

Figure 19.4 Listing of overloads in a stereo recording. Information on time at which the overloading occurs, as well as the channel, is shown (DK-Audio MSD600M).

can log the points in time at which overloading occurs during a recording. By adding time marks, it is possible to go back into a recording and find out whether or not these overloads should be repaired (see Figure 19.4).

As mentioned previously in the Histogram section, overloading does in principle not exist in the digital world, even when people speak of levels beyond 0 dBFS.

Clipping of a signal can be calculated as multiple samples that lie successively at the level of 0 dBFS. By a counting process, it is therefore possible to discern when a clip of the digital signal occurs. In some systems, it is possible to set how many successive 0 dBFS samples must occur before it indicates an overload.

It is possible to log when the overloads occurred by counting samples or applying a timecode (see Figure 19.5).

CUMULATIVE DISTRIBUTION

In connection with noise measurements (i.e., environmental noise or background noise) or measurements such as the calculation of the Loudness Range (LRA), a statistical analysis is applied that provides the cumulative distribution of the sound levels recorded. In the graphical rendering, a graph is shown where one axis is the A-weighted sound pressure level while the other axis represents the

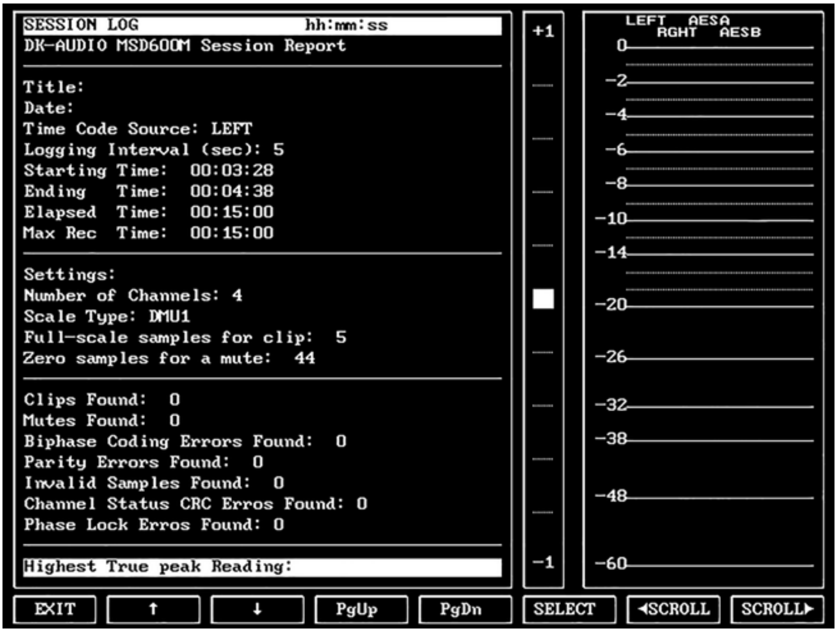

Figure 19.5 An example is shown here of a report covering a session of 15 s. Notice that the definition for overloading – and hence for possible clipping – has been set to 5 successive full-scale samples (DK-Audio MSD600M).

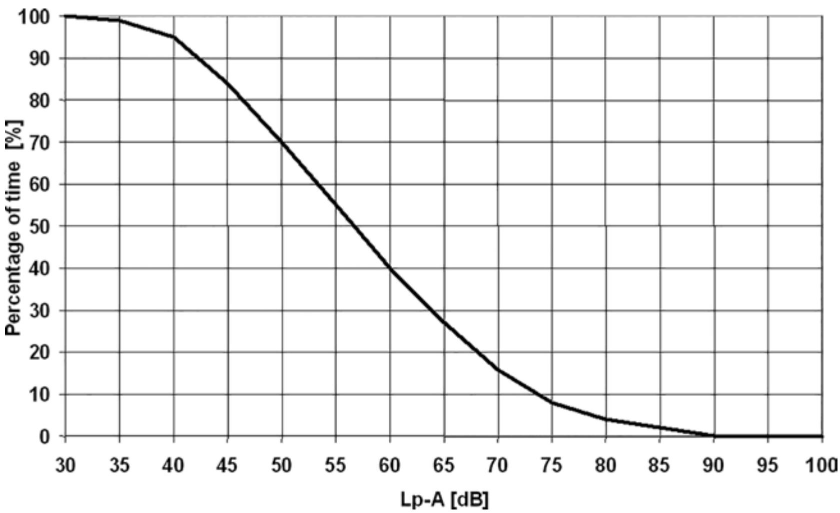

Figure 19.6 Example of the cumulative distribution. The curve shows what part of the time interval under consideration a given level is exceeded. For example, the level exceeds 60 dB(A) for 40% of the time.

percentage share of the time interval under consideration (see Figure 19.6). The curve shows which sound pressure level exceeded for what percentage of the time. The concept of background noise in the external environment typically is defined as the A-weighted sound pressure level that is exceeded during 95% of the time interval observed.

CHAPTER 20

Multichannel/ Immersive Audio

CHAPTER OUTLINE

Audio involving more than two channels is standard in many areas, such as audio productions for TV, cinema, home entertainment, the car, and music formats for Hi-Res CD, DVD, Blu-ray (BD), streaming services, and audio for gaming.

These formats involve sound that may end up in an almost infinite amount of loudspeaker channels. Certain formats are stored and transmitted in just two channels and are "unpacked" when played back. Other formats are produced as discrete channels but are stored and transmitted digitally as a single bitstream. Again, other formats are no longer being produced as channels but rather sound objects that contain metadata about time and direction for the playback. Further, there is no longer a fixed number of loudspeakers. The algorithms are designed to reproduce the audio almost in whatever number of loudspeakers available, even down to single-box speakers such as "soundbars."

On top of that, we have audio for mobile devices providing AR/VR formats produced in many fashions, however, often reproduced in headphones.

In this chapter, we will try to provide an overview of these formats/systems.

DIGITALLY BASED MULTICHANNEL SYSTEMS

This section presents a group of more or less true multichannel systems based on digital technology. What they have in common is that the channels are kept separate during the entire production phase until the point of transmission or media printing. Here, they are packed together – sometimes with the use of bit reduction into a single bitstream. The systems are used both for film sound in cinemas and for film and music production for domestic listening.

THE 5.1 FORMAT

Even though several other formats exist, most multichannel sound production formats are based on the standardized 5.1 loudspeaker arrangement – as a starting point at least. The number "5.1" refers to the fact that there are five channels with a full 20 Hz 20 kHz bandwidth. Also, there is one channel for sound effects at low frequencies. This channel is normally called the LFE channel (Low-Frequency Enhancement or Low-Frequency Effects). It has a limited frequency range. Tomlinson Holman (of Lucas Film at the time, later Apple) thus called it ".1" (even though the frequency range is less than one-tenth of other channels). This name has stuck and is subsequently included in standards and similar official writings. It does not involve an actual subwoofer channel or the like, because all primary channels go all the way down to 20 Hz. However, this LFE channel by playback is amplified 10 dB more than the other channels to provide additional headroom. It may only apply to a "bang" that occurs once, half an hour into the program material.

The LFE or ".1" channel is included due to regard for the better utilization of the dynamic range of the primary channels. In general, it is used mostly in movies and only extremely rarely in music.

This configuration was also called 3/2 because there are three front channels and two rear channels. (This nomenclature was later changed when more layers of

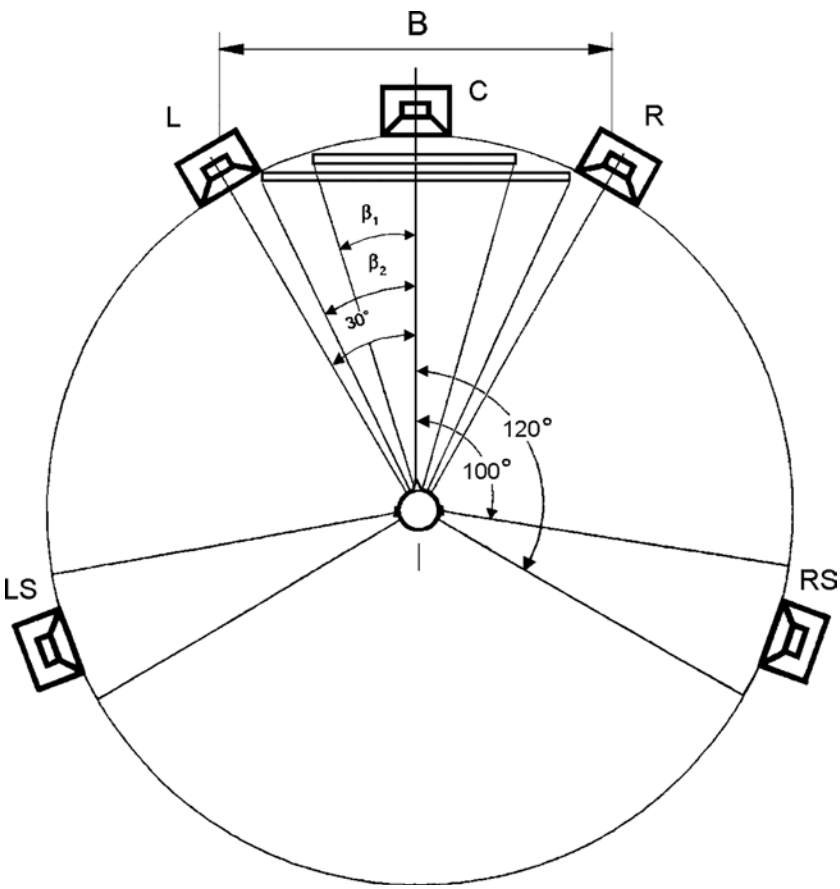

Figure 20.1 5.1 listening setup according to ITU-R BS.775–1.

loudspeakers were added). If the front speakers are in a straight row, it is suggested by the ITU that the center speaker is delayed. The LFE (or .1) channel is not taken into account in this arrangement. The subwoofer can, in principle, be placed anywhere with appropriate regard paid to distance and acoustic conditions of the room (see Figure 20.1).

DOLBY® STEREO SR•D/DOLBY® DIGITAL

Dolby® Stereo SR•D is a sound format on 35 mm film. It contains both analog and digital sound. The analog sound is encoded in Dolby® Stereo with Dolby® SR (Spectral Recording) noise reduction. The digital sound is encoded in Dolby® Digital 5.1. The codec system is Dolby® AC-3 with a bit rate of 320 kbps. The LFE channel encompasses the frequency range 20–120 Hz (see Figure 20.2 A and B).

During playback in a cinema, switching from the digital to the analog tracks is possible if digital errors occur. On DVDs, Dolby® Digital is one of the standardized formats, both on DVD Video, DVD Audio, and Blu-ray.

SURROUND EX®

In connection with the recording of "Star Wars – Episode 1" Dolby and THX®/ Lucasfilm felt a need for an extra rear channel introduced without any large technology-related problems. The Center-Surround channel is encoded into the Left Surround, and Right Surround using the same method as for the encoding of the center channel in Dolby® Surround. The digital encoding and decoding remain unchanged as to Dolby® Digital (see Figure 20.2 D).

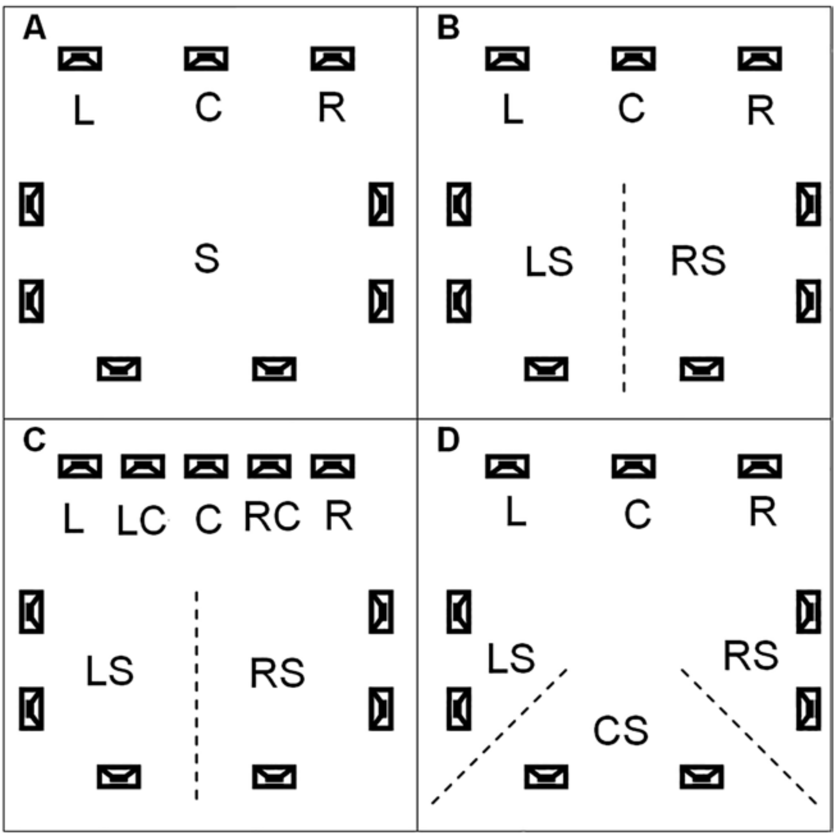

Figure 20.2 A: Dolby® Stereo (analog and matrix-encoded format). Configuration: 3/1 (three front and one rear channel). B: Dolby® SR•D (Dolby® Digital) and DTS® (Digital Theater Systems). Configuration: 3/2. C: SDDS® (Sony Dynamic Digital System). Configuration: 5/2. D: Surround EX™ and DTS®-ES. Configuration: 3/3.

DTS® (DIGITAL THEATER SYSTEMS)

This format was, as the name suggests, developed by Digital Theater Systems (now DTS), an American company. The bit reduction system used, Apt-X, was, however, developed in Ireland. The reduction has a fixed ratio of 4:1 (i.e., only 25% of the original quantity of bits remains). It is fundamentally a 5.1 format (see Figure 20.2 C) used both for film sound and music production. A CD-ROM provides the storage for the content. The consumer version is called Digital Surround.

Arrangements as they occur in a cinema are as follows: For normal feature films, DTS uses two CDs, which makes for a total feature-length of 3 hours and 20 minutes. A timecode is printed alongside the analog soundtracks of the film. This timecode applies to the synchronization of the CD-ROM player and projection machinery. The LFE channel here encompasses the 20 Hz – 120 Hz frequency range. The first film with DTS was Jurassic Park.

DTS®-ES (EXTENDED SURROUND)

DTS found it necessary to follow suit when Dolby developed Surround EX. Hence DTS also became able to offer a format with a Center-Surround channel (see Figure 20.2 D).

SDDS® (SONY DYNAMIC DIGITAL SOUND)

SDDS has eight channels: Left, Right, Center, Left Center, Right Center, Left Surround, and Right Surround, and an LFE channel. It is a 7.1 format. SDDS differentiates itself from the 5.1 format by the fact that it uses two additional speakers placed between left and center and center and right (see Figure 20.2 C). The system can, however, also function in 5.1 or 4/1, or 3/2 formats.

The bit reduction system is ATRAC (Sony's own), which also applied to MiniDisc, a format very popular at the time, but now long gone. The compression is approximately 5:1. The maximum bit rate is 1411 kbps. SDDS is used only for films. The digital information is placed on the film itself.

IMMERSIVE FORMATS

In the search for a higher degree of listeners' envelopment, several formats have entered the scene; first improving the horizontal resolution, next adding height information to this. Followingly the name "Surround Sound" gradually changed to "Immersive Sound." Formats, earlier mentioned as 3D-sound, actually only provided a 2D experience. So, 5.1 and 7.1 emerged into an almost infinite number of formats such as 9.1, 10.1, 11.1, 13.1, 10.2, 7.1.4, and so on.

ITU DEFINITIONS

The ITU has, to some extent, succeeded in the systemization of the many systems. Good sources for information on audio systems beyond 5.1 are the Reports ITU-R BS.2051-0 (2014) "Advanced sound system for programme production" [1], ITU-R BS.2150–8 (2019) "Multichannel sound technology in home and broadcasting applications" [2]. They describe the practical considerations in connection with the production and monitoring of these formats. Also, the assessment of the quality performance of the multichannel systems is included by referring to essential research on the topic.

ITU has provided objective naming for the formats to make sure all are on the same page when discussing the systems, A-H, and describing the configuration of each layer of loudspeakers.

An overview is given in Table 20.1.

Example: Sound System F has 3 upper + 7 middle + 0 bottom speakers. In each layer the speakers are arranged as follows:

Upper layer: 2/0/1: Two front speakers, zero side speakers, and one rear speaker.

Middle layer: 3/2/2: Three front speakers, two side speakers, and two rear speakers.

Bottom layer: 0/0/0.2: Zero front speakers, zero side speakers, zero rear speakers, two LFE speakers.

Some of these formats are designed to match microphone-specific configurations such as HOA (Higher Order Ambisonics) or discrete arrays reproducing natural sound stages. Other formats primarily serve the purpose of creating sound-spheres by manually positioning sound sources utilizing mixing tools like panners, delays, reverbs, and so on.

Table 20.1 Layout for multichannel sound systems as defined by the ITU.

Sound System	Layers distribution	Top layer	Middle layer	Bottom Layer	Example
A	0 + 2 + 0	0/0/0	2/0/0	0/0/0	General Stereo
B	0 + 5 + 0	0/0/0	3/0/2	0/0/0.1	General 5.1
C	2 + 5 + 0	2/0/0	3/0/2	0/0/0.1	7.1
D	4 + 5 + 0	2/0/2	3/0/2	0/0/0.1	9.1
E	4 + 5 + 1	2/0/2	3/0/2	1/0/0.1	11.1
F	3 + 7 + 0	2/0/1	3/2/2	0/0/0.2	10.2 systems
G	4 + 9 + 0	2/0/2	5/2/2	0/0/0.1	AURO 3D, 13.1
H	9 + 10 + 3	3/3/3	5/2/3	3/0/0.2	NHK 22.2

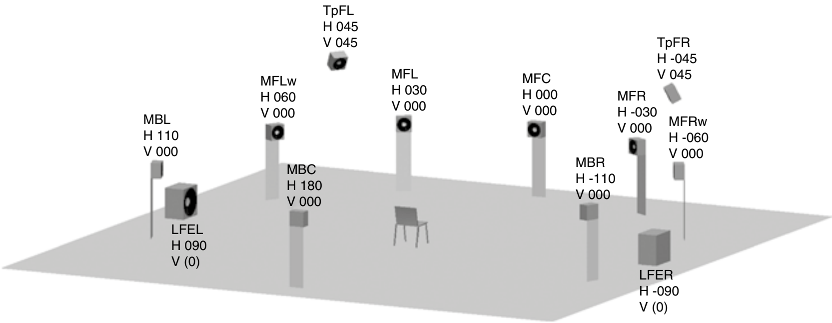

Figure 20.3 TMH 10.2 channels

TMH 10.2

Tomlinson Holman proposed a system which includes 12 audio channels for 12 speakers (2 + 8+0: H 2/0/0 + M 5/0/3 + B 0/0/0.2) (see Figure 20.3).

Seven front channels: Left Wide, Left Height, Left, Center, Right, Right Height, Right Wide.

Three surround channels: Left Surround, Back Surround, Right Surround.

Two LFE channels: LFE Left and LFE Right.

This format was demonstrated and is still applied by researchers for comparison purposes. However, only limited commercial content is produced in this format.

Another 10.2 format is developed in Korea for national television's Ultra High Definition service. The setup takes its starting point in 5.1 but has a is slightly different layout compared to the TMH-version. It has left and right side speakers and L/C/R height speakers.

To distinguish between the systems, the TMH is called 10.2 Type A, and the Korean system is called 10.2 Type B.

AURO TECHNOLOGIES, AURO 3D®

Auro Technologies developed a technology, that – nearly – is lossless packing of digital audio into standardized streams. This packing technique provides the transfer of many audio channels on a limited bit budget – with virtually no loss of quality. Examples of the packing using the Auro-3D Octopus codec: A 24-channel audio multitrack can be encoded into an 8-channel PCM-stream, and 18 tracks of PCM audio can be encoded into a 5.1 PCM-stream.

Auro 3D is available for all platforms, including cinema, home theater, automotive, gaming, and mobile devices.

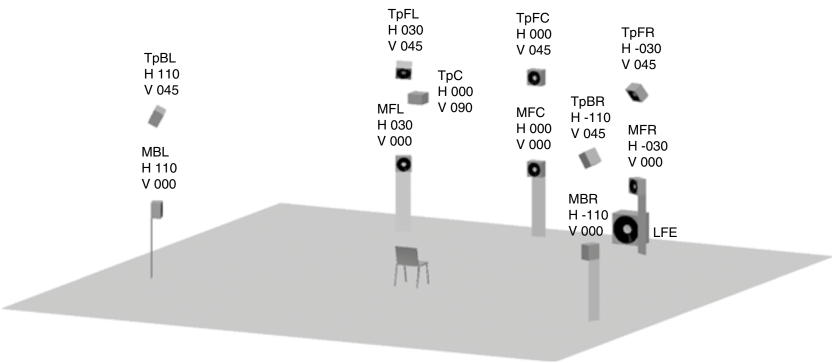

Figure 20.4 AURO 3D 11.1

The basic layout for the loudspeaker configuration begins at the 5.1, but with overhead loudspeakers added above the Left, Right, Left Surround, and Right Surround, this becomes a 9.1 configuration. For the home cinema, an additional center ceiling speaker is an option. However, one of the formats that has become widely used is the 11.1 (see Figure 20.4).

The channels can carry independent and uncorrelated audio signals. The format has proven efficient for natural, microphone-based immersive recordings.

For the home theater, the available decoders found in major vendors' consumer equipment allow the number of reproduced channels to be reduced according to the actual speaker configuration.

There is a lot of content available encoded in the AURO 3D format.

NHK 22.2

The national broadcaster of Japan, NHK, developed a configuration for Super Hi-Vision, a format with much higher resolution than standard HD. The high resolution is also present when it comes to the audio. Not less than 22.2 audio channels are laid down in this format (see Figure 20.5). Compared to other systems, this differs in that it includes bottom speakers, which are placed beneath the front L-C-R speakers.

The horizontal plane has ten speakers, there are nine overhead speakers, and then three bottom (front) speakers. Further, there are two LFE channels.

The digital information is designed to be transported by 12 AES3 streams (or any equivalent "transporter" such as MADI).

This format is in service in Japan. However, many laboratories across the world working on immersive audio have implemented the configuration for research

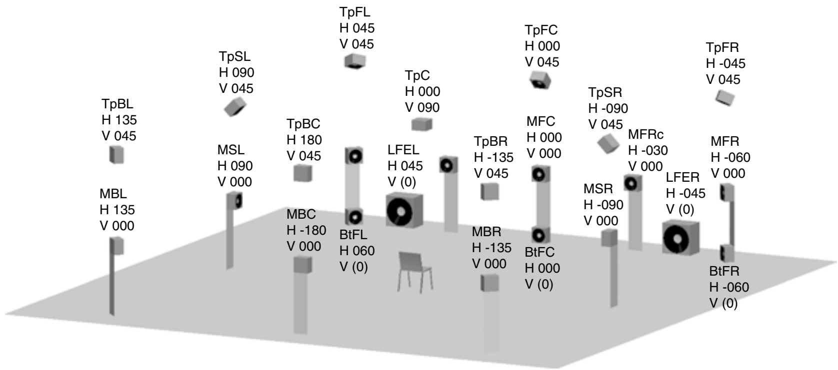

Figure 20.5 NHK 22.2.

purposes. It is specified as Sound system H in ITU-R BS.2051–0: Advanced sound system for programme production (2014).

OBJECT-BASED AUDIO

After increasing the number of (physical) channels over a period of years, immersive audio development took new directions into Object-Based Audio (OBA). So what is an audio object? Essentially, it is any piece of recorded audio – mono, stereo, eight channels, and so on. The main thing here is that metadata goes along with the object. The metadata contains information such as when to execute the object (time code) and where to place it in the sound sphere (providing vectors, see Figure 20.6). All the objects are virtually dropped in a big container, and – if selected – they will play out at the right time, in the right direction, and at the right level.

Examples: A sound effect for a film like a bird's call is recorded in stereo. It is positioned above the audience and has a cue to play at 0:17:10. Or: The second language (from dubbing) of a documentary is defined as an object. If selected by the viewer, this language will follow the program and switch off the native speaker. Or: The dialogue of some TV content is a little vague. However, the dialogue is an object, so it is possible to raise the level of this specifically.

One major advantage is that no object is depending on a given audio channel. The decoder seeks to position the object correctly in the space no matter how many loudspeakers are available. When there are more speakers, it is probably more precise; with fewer loudspeakers, it is less precise but still present in the soundscape.

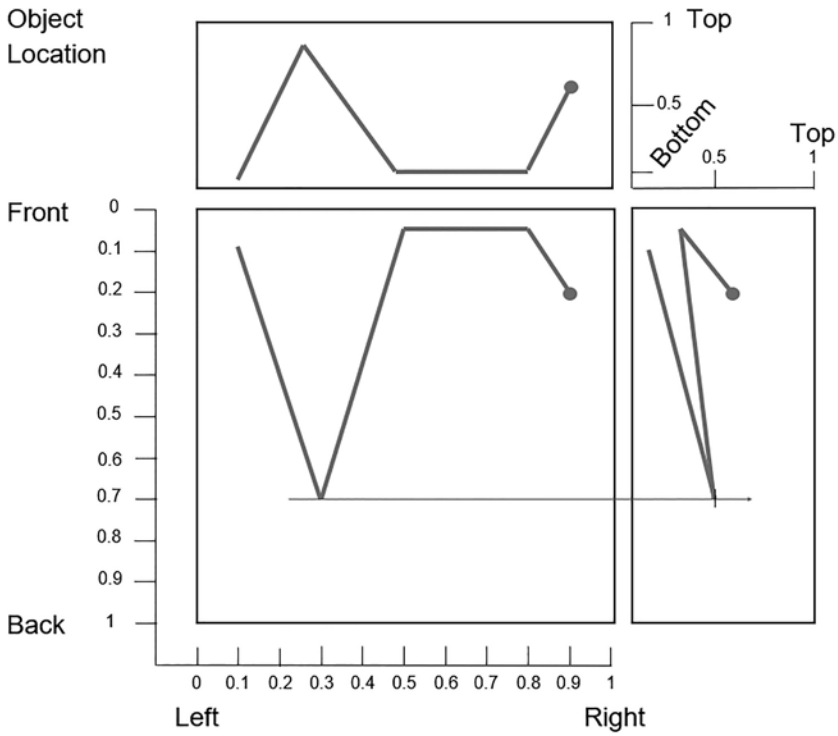

Figure 20.6 Sound object positioned in the room by the help of vectors.

DOLBY® ATMOS

Dolby Atmos is a format for the cinema – and the home – that takes advantage of object-based audio. The basic configuration starts with a 5.1 or 7.1, but the floor layer of speakers expands, and overhead speakers are added. So the nomenclature of an Atmos setup could be 5.1.4, where the "4" indicates the upper layer of speakers, the overhead sound sources.

In the mixing, each audio track can be assigned to an audio channel, the traditional format for distribution, or to an audio object. Dolby Atmos by default has a 10-channel 7.1.2 bed for ambience stems or center dialogue, the traditional channel-based production. This leaves 118 tracks for objects. Dolby Atmos home theaters can be built on traditional 5.1 and 7.1 layouts (see Figure 20.7).

Dolby provides an interesting tool for the correct design of Atmos cinemas/listening rooms. It is named Dolby Audio Room Design Tool (DARDT). It can help you through the calculations for the loudspeaker placement with respect to room size, number of speakers, screen, and so on.

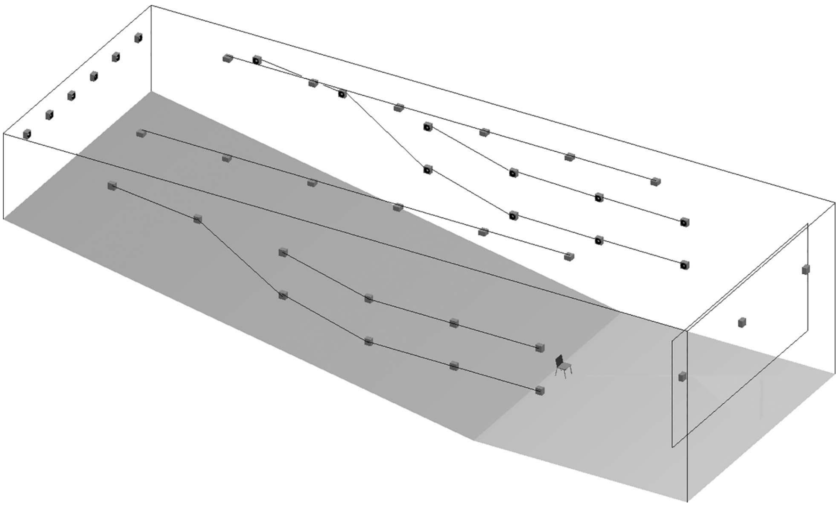

Figure 20.7 Dolby Atmos configuration in the cinema.

SONY® 360

Sony 360 Reality Audio is a new audio format that can run parallel to Dolby® Atmos. It can deliver sound in up to 64 loudspeaker channels. Like Atmos, it offers object-based sound design. It is partly developed together with Fraunhofer Institut für Integrierte Shaltungen. The carrier is MPEG-H 3D, which also is an introducer for music streaming.

Both Dolby Atmos and Sony 360 are valid formats for Amazon Echo Spatial Speakers.

WAVE FIELD SYNTHESIS

Sound Field Synthesis (WFS) is being developed and promoted mainly in the Netherlands by Diemer de Vries and his team at Delft University. The system recreates the sound field in the horizontal plane. One system can easily consist of 64 channels. This system is installed in very few places. However, it has convincing features regarding perceived directionality and distance perception.

OBA FOR TV

Traditional flow TV is fighting to keep up with the possibilities offered in other fields of the entertainment industry. Now many features can be selected on demand. Watching

Formula 1 leaves the possibility of following a specific car, stay in a specific position at the raceway, or getting the comments in a specific language.

Object-based production lies in the extension of the services mentioned earlier like 5.1 for cinema, broadcast, and home. First, exotic multichannel sound mixes are "collapsible" so all have an opportunity to experience the content in the best possible way. Next, the services like channel extraction (dialogue), choice of language, and so on have extra value for the group of people with hearing disorders, a group that may not be interested in overwhelming channel-rides. However, also sound aficionados get a chance to experience exciting sound settings in the TV room.

DIGITAL CINEMA

The goal for Digital Cinema is the establishment of completely file-based formats to be distributed by data networks. Many cinemas have been refurbished to accomplish this goal. However, it has also been an investment to follow the trend. Actually, in many countries, the picture may have been improved by the transition to file-based films. However, the playout sound systems for a long period did not improve at the same speed.

In 2008 the Digital Cinema Initiative LLC, a collaboration between major motion pictures studios, published the "Digital Cinema System Specification, version 1.2," which has been the basis for present cinema systems. Latest version of the DCI specification is 1.3 (2018) plus an addendum on Object-Based Audio.

The basis for the audio is uncompressed PCM, 24 bits/48 kHz or 96 kHz sampling. The file format is WAV frequency (AES3). The stream can transport up to 16 audio channels.

UPMIXING SYSTEMS

Most software-based mastering tools include algorithms to upmix from stereo to surround, both 5.1 and 7.1. This has been a necessity in order to provide content for the many systems found on the market. These algorithms are not all alike and must be assessed in connection with the program material to be converted.

VIRTUAL SURROUND

Virtual surround is a set of techniques where only two speakers or a set of headphones are used to recreate a sound field from many speakers. The effect is often arbitrary.

The surest way to monitor signals is to use the goniometer since the signals typically have high phase opposition content. However, this is very much what is achieved by Dolby Atmos and systems like it.

ANALOG, MATRIX-ENCODED FORMATS

Before the Digital Age, multichannel formats took advantage of matrix encoding/decoding technology. Here, four or five audio channels were mixed down to two. After transmission or distribution on media, the signals were separated again at a later stage, although this was not always as successful as one could wish.

The principle was also called 4:2:4 or 5:2:5, which indicates the number of channels produced, number of stored/transmitted channels, and finally, the number of channels of the reproduction format. The most widespread systems are Dolby® Stereo (or Dolby® Surround), which is a 4:2:4 system. Others include Circle Surround, 5:2:5, and Lexicon Surround, 4:2:5 or 5:2:5.

If you come across any of the old matrix-based systems, try to preserve the phase relation between channels, because this is essential for making it play right.

LOUDSPEAKER POSITIONING – DOLBY® SURROUND

The basic loudspeaker setup for Dolby® Surround was three front speakers – Left, Center, and Right. Normally, the center speaker was positioned beneath the monitor screen. If projection applied, the speaker was positioned behind a partly transparent projector screen.

The surround channel was normally reproduced by two speakers placed diagonally behind the listener. To avoid a perception of sound from rear speakers coming from a point between the two speakers, the signal for one of the speakers might either be phase-shifted or phase-inverted. When mixing in a small room, it was, however, not recommended to use phase-inverting of one of the surround speakers, as it was very difficult to judge the level of the reproduced surround signal. However, a 90° phase shift worked. In THX-specified systems for home use, it was possible to use dipole speakers to attain diffuse sound.

ACOUSTIC CALIBRATION

The decoder's pink noise generator applies to acoustic calibration. The electrical signal levels are at –6 dB re full modulation. Each channel is measured on its own. All speakers in the surround chain are regarded as one channel. In the listening position, a sound pressure level of 82 dB(C) should be measured from each channel.

In a cinema (Dolby Stereo), the measurement is made over a larger area; typically four to five characteristic measurement points are selected. The sound from each channel is then measured at these points. The result is an average of the measurements. Instead of a test generator, a special test film is used with pink noise recorded at −6 dB.

Be aware that new procedures for cinema calibration are on their way.

DISPLAY OF MULTICHANNEL SIGNALS

DISPLAY OF 5.1

With true 5.1 or the equivalent, the channels will be kept separate prior to the final coding into one of the standardized formats. Six bar graph meters can of course show what the individual channels contain. However, it is most practical and more manageable to use an instrument with a Jelly Fish™-like display.

JELLY FISH™ AND OTHER GRAPHICAL DISPLAYS

Jelly Fish™, originally from DK Technologies, and other slightly different designs provide goniometer-like displays on a screen.

The figure in itself does not show the phase between the channels but instead shows the amplitude in each of the channels concerned. The purpose of this display is to create an overview (see Figure 20.8).

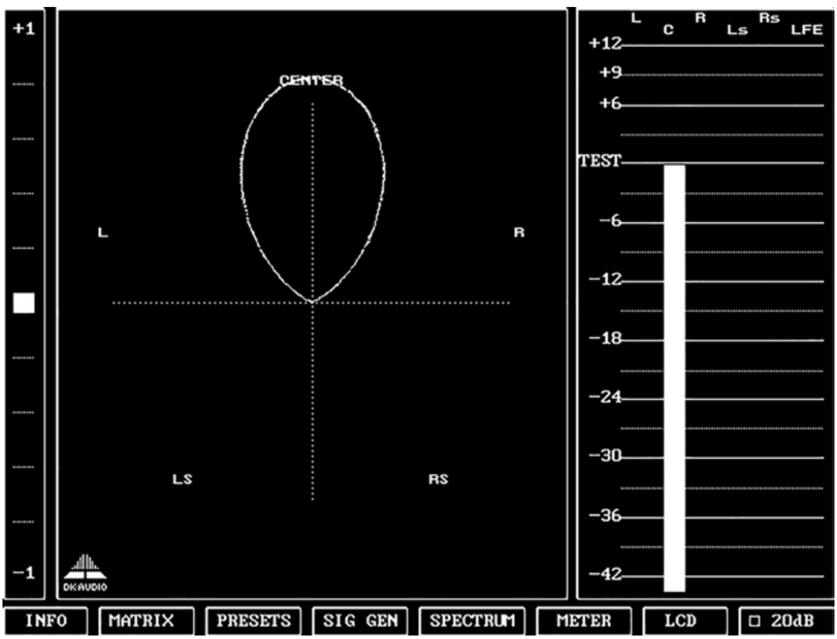

Figure 20.8 Jelly Fish™: signal solely in the center channel.

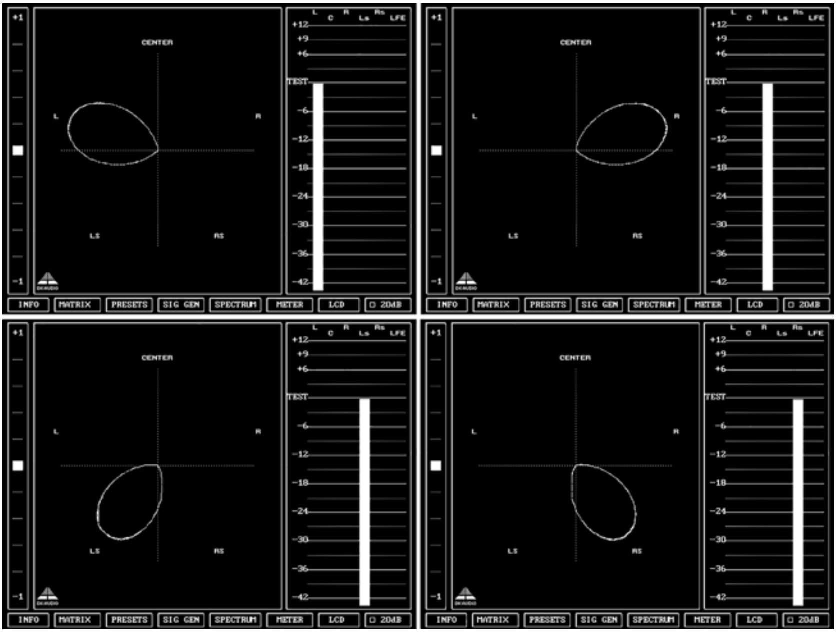

Figure 20.9 Upper: Jelly Fish™: signal solely in the left and right front channels. Lower: Jelly Fish™: signal solely in the left and right surround channels.

On the instrument's screen, a circle is fundamentally established and the magnitudes of the levels in the channels concerned (left, center, etc.) are multiplied into this figure (see Figure 20.9). The figure thus becomes the "most full-bodied" in the direction/channel that has the strongest signal. For the sake of clarity, the instrument will normally have a certain inertia in order for the user to be able to follow that part of the signal that has a certain weight in terms of time.

PHASE DIFFERENCES BETWEEN ADJACENT CHANNELS

The Jelly Fish™ itself does not show anything about the phase. However, it provides a change of color in the transition area between two channels if the phase angle is greater than 90°.

ACOUSTIC CALIBRATION OF MULTICHANNEL SYSTEMS

Calibration of the acoustic sound levels has been a requirement for many years when working with sound for film, although it has not been particularly common in other branches of the sound industry. However, with the widespread use of multichannel formats for all forms of music and film presentation in the home, it has turned out to be beneficial to also calibrate the acoustic levels for these formats.

It is important to differentiate between production for the cinema (including the X-curve) and production for 5.1 channel reproduction in the home based on ITU 775.

The ITU-R BS.2159 describes production monitoring for home entertainment without x-curve.

CALIBRATION OF CINEMA SYSTEMS

In a cinema, the listeners sit far from the speakers. Presumably, the majority sit in the diffuse sound field. In any event, attempts are made to establish a diffuse sound field from the surround speakers. Thus, when the sound pressure is measured inside the cinema or in a mixing theater, it must be averaged over many different measurement locations. The typical basis for the majority of standards is at least four locations. If there are different areas for the audience, for example, the main floor and balcony, measurement should be made in at least four locations in each.

Before performing this measurement, the system's frequency response must be in order. Normally, the ISO 2969 X-curve standard applies as a measure for the characteristics of the system. However, during recent years, it is heavily debated whether the X-curve is the right frequency response for the cinema. But until further notice, this is the standard.

OPTICAL SOUND

Cinema systems for the reproduction of Dolby® Stereo (analog optical sound) normally have a built-in generator with pink noise. This signal is sent out at a level corresponding to half modulation of the optical track (i.e., 6 dB below full modulation).

The generator is used in particular in the mixing theater, where the sound has, of course, not hit the recording media yet. With this, the B-chain can also be checked; that is, that portion of the sound system that encompasses everything from the playback system for the specific cinema up to and including the acoustic space. (The A chain encompasses that portion of the system that lies before the playback system for the specific cinema.)

A test film (Dolby cat. No. 69) with prerecorded pink noise at 6 dB below full modulation and with Dolby noise reduction is run in the cinema's projector to assist in making adjustments to the B-chain. For each of the four channels (L, C, R, and S) adjustments are made for a sound pressure level of 85 dB(C) (integration time: slow) in the inside of the cinema as calculated by a simple average value of the measurement results at the selected measurement locations. This procedure regards a chain of surround speakers as one channel (i.e., all the speakers in this chain must be operating at the same time).

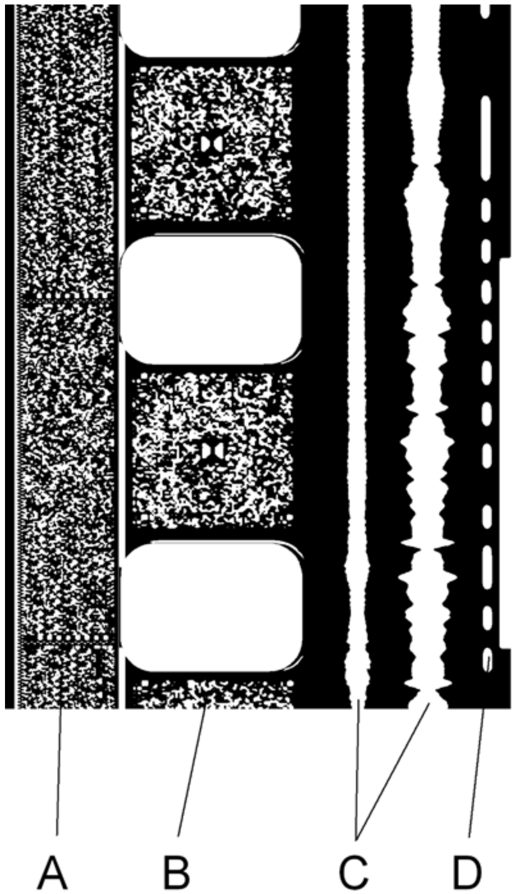

Figure 20.10 A: SDDS (corresponding tracks are found on the opposite edge of the film). B: Dolby Digital. C: Analog soundtracks. D: DTS soundtracks on 35 mm film.

DIGITAL SOUND

Digital sound on film has created a larger dynamic range, of which a large part is used for greater headroom in comparison with optical sound.

The digitally recorded SMPTE standardized test signal (pink noise) lies at −18 dBFS. During playback of each of the front channels, this signal must be reproduced at a sound pressure level of 85 dB(C). The two surround channels are each adjusted to 82 dB(C). This causes the level created by the entire surround chain to equal 85 dB(C) thus.

A 10 dB amplification is inserted in the playback chain for the LFE signal. When the limited bandwidth (20 Hz–120 Hz) pink noise from the LFE channel is played

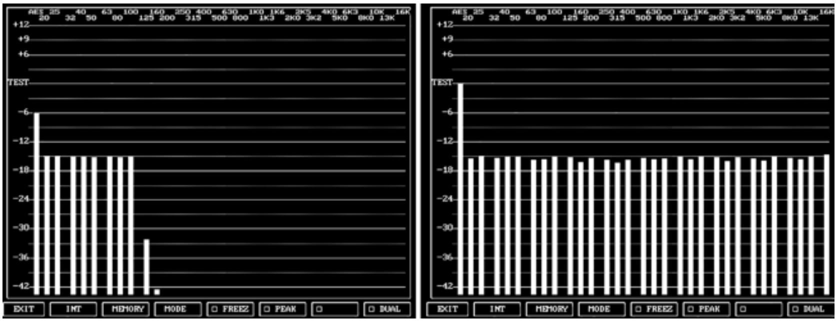

Figure 20.11 The spectra are shown here for pink noise recorded on LFE (on the left) and a main channel (on the right). The LFE only goes to 120 Hz, whereas the main channel has full bandwidth. All columns in the active areas of the channels are of equal height (i.e., the channels have the same level per 1/3-octave). The column furthest to the left in each spectrum shows the total level. The level is 6 dB lower in the LFE channel than in the main channel. This is because there is a smaller frequency range represented here.

back, it is possible with a 1/3-octave spectrum analyzer to see that the individual ranges in the LFE channel are reproduced 10 dB higher than the individual ranges in each main channel. Measured as a C-weighted sound pressure level, the LFE channel will show a level that is approximately 4 dB(C) higher than the level in each main channel.

An example of the pink noise distribution in 1/3-octave bands across the frequency span of a main channel and a LFE channel is shown in Figure 20.11.

Calibration of 5.1 in an ITU-R BS.775 Arrangement

In a 5.1 system based on the ITU arrangement, all main channels have in principle the same conditions: There is one speaker per channel, and each is placed the same distance from the listener.

Internationally, there is, however, agreement neither on the level nor on the bandwidth for the noise signal that applies to acoustic calibration. Pink noise is good since it includes all frequencies; however, it is impractical due to its “unsettled” character, which makes it difficult to measure at low frequencies.

Surround Sound Forum

Surround Sound Forum (SSF) is a German interest group established by Verband Deutscher Tonmeister (VDT, Association of German Tonmeisters), the Institut für Rundfunk Technik (IRT, Institute of Broadcast Engineering), and Schule für Rundfunktechnik (SRT, School of Broadcast Engineering). The SSF has prepared early stage guidelines that are still widely accepted in Europe. Three test signals which are recorded at −18 dBFS (RMS) are specified (see Table 20.2).

Table 20.2 Measurement signals for the main channels in surround sound configuration

	Measurement signals			Listening level	
Signal only in one channel)	PPM level with τ<0,1 ms [dB]	PPM level with τ<10 ms [dB]	RMS level [dB]	Sound pressure level SLOW [dB]	Sound pressure level SLOW [dB(A)]
1 kHz sine	–18	–18	–18		
Pink noise 20 Hz–20 kHz	–9	–13	–18	82	78
Pink noise 200 Hz–20 kHz	–11	–15	–20	80	78

SMPTE

The corresponding standard from SMPTE (RP155 [3]) uses a standard of −20 dBFS for the reference level. Here, the C-weighted sound pressure level ends up at 83 dB.

Bass Management

Bass management consists primarily of filtering out the bass from the main channels and reproducing it in a subwoofer (together with the LFE signal). Frequency response and level must display the same data as if only full range systems were being used in the main channels.

Dolby Stereo, Dolby SR/D, Dolby Digital, Dolby Surround, Dolby AC3, Surround EX, Dolby Atmos are trademarks of Dolby Laboratories. DTS, DTS-ES are trademarks of DTS. Auro-3D is a trademark of Auro Technologies. DSSD, 360 Reality Audio are trademarks of SONY Corporation.

REFERENCES

[1] ITU-R BS.2051-0 Advanced sound system for programme production (2014).

[2] Report ITU-R BS.2159-8 Multichannel sound technology in home and broadcasting applications (2019).

[3] SMPTE RP 155:2014 – SMPTE recommended practice – for motion pictures and television – reference level for digital audio systems.

CHAPTER 21

Standards and Practices

CHAPTER OUTLINE

When program material is exchanged, it is of course practical that both the sender and the receiver know how the levels and channels are placed. It is easiest if there is a standard to which the largest possible number of people agree to follow. Sometimes the standards are set internationally. Other times they are just based on common practices.

A listing of some of the applicable standards and practices follows. Many old formats are included to support those in audio who are working with our audio heritage.

ANALOG TAPE, AUDIO

Even though the majority of recordings today are stored in digital formats, the analog tape recorder is not completely dead. If nothing else, there is some very good archive material stored in the vaults. It would thus be nice to know how to play it back.

1/4 IN., 2 TRACK

Stereo, Left/Right

Track 1	Left
Track 2	Right

Stereo, M/S

Track 1	M(id)
Track 2	S(ide)

1/4 IN. OR CC, 4 TRACK

Slide-Show

Track 1	Left
Track 2	Right
Track 3	Not used
Track 4	Time code/cue

PERFORATED MAGNETIC FILM, 4 TRACK (DIN 15.554/ISO 162)

Dolby® Stereo

Track 1	Left
Track 2	Center
Track 3	Right
Track 4	Surround

PERFORATED MAGNETIC FILM, 6 TRACK (TODD AO)

Track 1	Left
Track 2	Center
Track 3	Right
Track 4	Surround
Track 5	Outer left
Track 6	Outer right

MULTITRACK TAPE

No actual norms exist for the placement of material on specific tracks on multitrack machines, since we are not dealing with master tapes. However, there are a couple of practical rules that should be adhered to for technical reasons.

With sync-recording: Never record on a track adjacent to the track that is being listened to.

With the recording of stereo tracks: Stereo tracks are always placed next to each other, for example tracks 1, 2 or 17, 18. In connection with audio-audio synchronization: Never place two stereo tracks on separate machines.

Time code: Always placed on the outside track, as a rule the track with the highest number.

TEST TONES

The test tones, when properly used, are the key to painless editing and subsequent correct playback. On the finished program, the tones represent what can be expected during the rest of the program.

On analog audio tape, including perforated magnetic film, the tones always refer to a level of magnetization, which is typically specified in nanoweber per meter of track width, abbreviated as nWb/m. If not otherwise specified, then 1 kHz sinusoidal tones are used. This level is also called the nominal level. In addition, there will be headroom, which is determined by the level at which the distortion on the tape exceeds a certain magnitude, such as 3%.

35 MM MAGNETIC FILM (PERF)

320 nWb/m (Europe)

185 nWb/m (USA)

2 IN., 24 TRACK

(76 cm/s or 30 in/s)

510 nWb/m (Europe)

355 nWb/m (USA)

1/4 IN., 2 TRACK WITH DOLBY SR

320 nWb/m (Europe)

200 nWb/m (USA)

1/4 IN., 2 TRACK

(38 cm/s or 15 in/s)

510 nWb/m

320 nWb/m (Europe)

200 nWb/m (USA)

1/4 IN. FULL TRACK, NAGRA

(19 cm/s or 7½ in/s)

320 nWb/m (Europe)

200 nWb/m (USA)

COMPACT CASSETTE

(@ 315 Hz)

250 nWb/m (Europe)

200 nWb/m (USA)

ANALOG TAPE, VIDEO

Professionally, these formats are predominantly used for archiving IF the programs have still not been digitized.

VIDEO, 2 TRACK

Mono program, master: all formats except U-matic

Track 1	Final mix

U-MATIC

Track 2	Final mix

MONO, FOR VOICE-OVER: ALL FORMATS EXCEPT U-MATIC

Track 1	Final mix
Track 2	(IT: International Track)

U-MATIC

Track 2	Final mix
Track 1	IT

COMMENTS

"IT" or "I sound" can contain M (music), E (effects), or D (dialogue, in-vision), but never a feed line (dialogue where the speaker is not in the picture). Hence, a track can also be marked for example as M+E, M+D, or M+E+D.

STEREO, LEFT/RIGHT

Track 1	Left
Track 2	Right

STEREO, M/S

Track 1	M(id)
Track 2	S(ide)

TEST TONES

On analog video tapes, the test tones will refer to a magnetization level on the linear, longitudinal track, whereas a frequency fluctuation ought to be referred to either in kHz or in percentage for the AFM tracks.

3/4 IN. U-MATIC

100 nWb/m

BETACAM SP, LINEAR

100 nWb/m

BETACAM SP, AFM

–9 dB re 100% mod.

SVHS, LINEAR

100 Wb/m

SVHS, HIFI (AFM)

–9dB re 100% mod.

BROADCAST, PROGRAMS

Test tone	1 kHz
Level	6 dB below full modulation. On a PPM instrument with the Nordic scale the modulation should be 0 dBu.
Duration	90 seconds
Start, time code	00:00:00:00 (or 09:58:00:00)
Start, program	00:02:00:00 (or 10:00:00:00)

STEREO

Left/right: Test tone first alone in left channel for approximately 5 seconds, then both tracks for 90 seconds.

NOTE

The EBU test tone interrupts the left channel on a regular, repeating basis. The BBC employ a bespoke stereo identity tone (GLITS) that interrupts the left channel once, followed by two interruptions on the right channel – the three interruptions being equally spaced and repeated on a regular basis.

> **M/S**: 1 kHz, 6 dB below full modulation in both channels. The channels are adjusted before matrix encoding for left and right

NOTE I

If the test tone is recorded as left and right, and thereafter matrix-encoded to M $((L+R)/\sqrt{2})$ and S $((L-R)/\sqrt{2})$ on the tape, then there will only be a signal on the M track, which will be 3 dB above the test level. The S track should then be completely without a signal.

NOTE II

For analog formats, the test tone ought to always be recorded without noise reduction if it influences the recorded level. It should be noted on the tape report whether the test tone was recorded with or without noise reduction (for example, "−NR").

NOISE REDUCTION

The different noise reduction systems for analog tape formats each have their own special adjustment procedures, which of course must be adhered to in order to attain the optimum dynamic range. In the following sections, we will discuss some of the test signals used in existing formats.

DOLBY® A

Tone:	850 Hz FM modulated with ±10% per 0.75 seconds
Duration:	30 seconds
Level:	Normally modulated to a magnetization level of 185 nWb/m regardless of the format if nothing else is specified. However, this is highly dependent on the type of magnetic tape used. On standard audio tape, the level is set to 320 nWb/m. For the 1 inch video formats, the level was 100 nWb/m.

DOLBY® B

Tone:	400 Hz sinusoidal, FM modulated with ±10% per 0.75 seconds
Duration:	30 seconds
Level:	Full modulation

DOLBY® C

Tone:	400 Hz sinusoidal, FM modulated with ±10% per 0.75 seconds
Duration:	30 seconds
Level:	Full modulation

DOLBY® SR

Tone:	Pink noise, interrupted for 20 ms every 2 seconds
Duration:	30 seconds
Level:	15 dB below full modulation

TELECOM C4

Tone:	550/650 Hz, alternating every half second
Duration:	30 seconds
Level:	Full modulation, or 4 dB below full modulation

DIGITAL TAPES, AUDIO

The determining factor that decides the level on digital tapes is based on the use of the program material. Typically, there will be a difference between a broadcasting master at a broadcast level (allowing typically 9 dB headroom) and a master for a CD or DVD (with zero headroom).

EBU*

*for those regions that have not implemented the recommendation EBU R128 on loudness.

When exchanging programs between radio broadcasters, it is important that the tapes can be played directly. Hence under the auspices of European cooperation within the EBU, the following standard was adopted until the R128 was accepted:

Test level: –18 dBFS

Max level: –9 dBFS

Hence, there is, so to speak, 1.5 bits (9 dB) left unused. This habit was primarily followed for reasons of safety, to avoid any risk of overloading, in particular when monitoring signal levels applying analog metering systems.

In the Loudness recommendation R128 (see later) −1 dBTP is allowed (if measured using oversampling).

CD MASTER

Old vices from the vinyl era remain true to form: The level is set as high as possible. In pop/rock productions, a liberal view is taken of digital overloading. Devices are actually manufactured that deceive the mastering system into believing that no overloading is occurring by moving all encoding levels on full scale one step down.

Apart from that, all samples at 0 dBFS are normally logged. Or to be more precise: If a number of samples in a row are at full scale, then the time code is noted before they are delivered for mastering.

POST-PRODUCTION

Tapes to and from post-production ought to adhere to the allocation of channels shown here. It is current in both Europe and the USA.

8-track digital media (such as DA88)

Track	1	2	3	4	5	6	7	8
Channel	L	R	C	LFE	LS	RS	20 bit data	20 bit data

DVD/Blu-ray/DTV master

Track	1	2	3	4	5	6	7	8
Channel	L	R	C	LFE	LS	RS	LT	RT
Test signal:	1 kHz sinusoidal							
Level:	Europe: –18 dBFS							
	USA: –20 dBFS							

DIGITAL TAPES, VIDEO

Tracks and equipment follow the rules of digital audio in their respective areas (Europe/USA).

BLITS

BLITS (Black & Lane's Identification Tones for Surround) provides channel identification tones for all channels within a 5.1 Surround Sound signal. This also helps

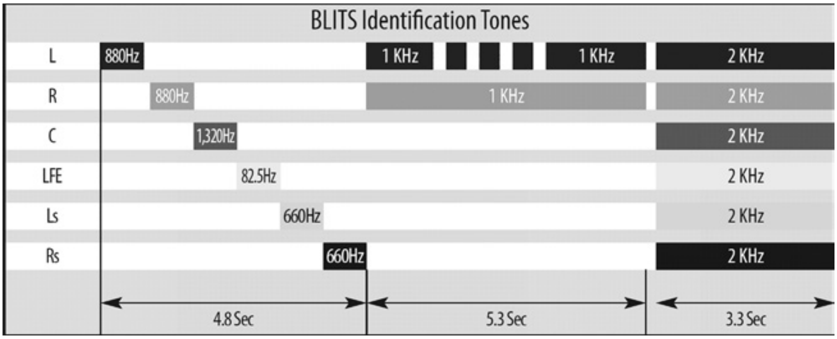

Figure 21.1 The BLITS – Identification Tones for Surround.

to provide information in a stereo downmix – channel presence/absence. The BLITS tones can be used at the start of a program to help identify channels, in OB trucks for channel identification back to the studio, and MCR as well as lineup for storage devices.

The frequencies used are based on the international music standard and are L = 880 Hz, R = 880 Hz, C = 1,320 Hz, LFE = 82.5 Hz, Ls = 660 Hz, and Rs = 660 Hz. The arrangement of the tones is designed to provide sequential and easy-to-read displays on bar graph meters. Basically, this sequence can be generated by the metering system if the software is available (see Figure 21.1).

HARD DISK SYSTEMS/FILE-BASED SYSTEMS

There is no agreement among the producers of hard drive systems on how analog and digital levels should go together. Test levels are placed at −18, −19, and −20 dBFS.

In a working situation, −6 dBFS (corresponding to approximately 50% of full amplitude) should normally not be exceeded. However, metering systems offering reading of true peak (using oversampling) extend this to −1 dBTP for linear PCM. If, however, the audio is bit-reduced, the headroom must be enlarged, as some systems may produce excessive peak levels when converted.

Hard disk-based audio workstations often have a useful function called "normalization." The signal is adjusted to a specific level, defined either by dBFS or by a percentage of full modulation. It should be noted that two different references exist:

1. In most DAWs 0 dBFS has the RMS-value of a full scale sine-wave as the reference [1]. Thus, a full scale square-wave has an RMS value of +3 dBFS.
2. In some DAWs the reference is (or can be set to be) a square-wave. Then a sine-wave with a full scale true peak reaches an RMS-level of −3 dBFS.

This is rather confusing. However, one should go with version 1, as this is the preferred and common standard.

In file-based systems, the Meta Data is very important. This leaves a possibility of batch normalization of the files without actually changing the files themselves. What is changed is a number in the metadata. Thus, the playout systems can correct to standardized signal level or loudness level.

MASTER FOR VINYL

There are some (good old) rules that must be adhered to in order to make a master for subsequently cutting records (particularly if the master is produced from a digital medium).

1. Avoid excessive bass, especially below 80 Hz.
2. Avoid excessive treble.
3. Avoid excessive stereo separation.
4. Avoid too large a dynamic range.

These rules of thumb are due to the limitations of the cutting technology. One should know, for example, that the recording speed is different at the outer and inner peripheries. This is of significance to both the dynamic and the frequency response.

ANALOG MASTERS

30 s.	Leader tape, white + red
10 cm [4 in.]	Leader tape, transparent
10 cm [4 in.]	Leader tape, white + red
5 s.	1 kHz, left channel. 0 VU: 320 nWb/m
30 s.	1 kHz, both channels. 0 VU: 320 nWb/m
30 s.	10 kHz, both channels, −10 dB re 0 VU
30 s.	100 Hz, both channels, 0 VU: 320 nWb/m
(30 s.	NR ref-signal (if NR is used)
10 s.	Leader tape, white + red
	First recording
4 s.	Leader tape, blue
	Second recording
4 s.	Leader tape, blue
	etc.
	Last recording

30 s.	Leader tape, white + red
5 s.	Leader tape, red
10 cm [4 in.]	Leader tape, transparent
30 s.	Leader tape, red

DIGITAL MASTERS

Digital masters follow the same principles as the analog masters (without leader tape of course). However, one needs to be really attentive to the general rules for good mastering for vinyl. The limitations of the vinyl record are unknown in digital technology.

FILM MASTERS

Masters for film are prepared on the basis of the mix to reach standardized playback levels.

FILM, OPTICAL

The sound camera has a clipping level that is determined by the physical limits of the optical track. The test level is set at 50% modulation, 6 dB below full modulation. This level also corresponds to the dialogue level.

FILM, DIGITAL

Digital films will follow the SMPTE standard, that is, dialogue (test level) will be at −20 dBFS. This corresponds to a sound pressure level of 85 dB(C) for each of the front channels. In the US the TASA (Trailer Audio Standards Association) provides specifications for the measurement of cinematic trailers to match the level of the features films they precede.

FROM FILM TO DVD OR BLU-RAY

The acoustic sound level from test and dialogue levels of digital film sound channels lies at −18 dBFS. This is adjusted in the cinema to a sound pressure level of 82 dB(C), for each channel by itself, so that the total level from the two surround channels becomes 85 dB(C). In a home cinema, the adjustment process is such that all channels are adjusted to the same listening level. This means that the surround here will be too strong in relation to the balance in the cinema. Hence, it is normal that each of the surround channels is attenuated by 3 dB before they are transferred to DVD, Blu-ray, or digital TV.

The standard listening level for −18 dBFS (pink noise) is normally set to a sound pressure level of 78 dB(C) for each of the five main channels. When all five channels are playing at the same time, this then corresponds to a sound pressure level of 85 dB(C).

SATELLITE

For an uplink to a satellite, there can be different standards depending on the system in use. What is most important for analog connections is to check whether pre-emphasis is included in the reading of the audio level meter. If not, there is a risk of bringing the connection down, even if there is a program limiter on the output.

BROADCAST, TRANSMISSION

The ITU has established a test signal for transmission lines [3]. This test signal contains three levels, which are defined here and related to EBU R68 [4].

Measurement level (ML) is the level at which measurements are performed on the lines. The level corresponds to −30 dBFS.

Alignment level (AL) corresponds to the test level or nominal level. The level corresponds to −18 dBFS.

Permitted maximum level (PML) is the maximum that can be sent on the line. The level corresponds to −9 dBFS.

BROADCAST, PROGRAM

Recommendations for program levels are defined by the individual broadcaster. In pure analog circuits, most broadcasters will normally imply level setting in accordance with the EBU R68 as mentioned earlier. This may apply to digital transmission as well. However, in general, the loudness-based recommendations have gradually taken over. Broadcasters have been rewriting their technical specifications for program delivery. Several sets of recommendation are in play here. Basically all have the ITU BS.1770-1 (without gating), and BS.1770-2, -3 and -4 (gating included) [5] as the basis and by which the loudness level is stated by one single number; nevertheless, the program contains mono, stereo, or surround sound. The national or regional standards may exhibit minor differences as to the target for integrated loudness level (−23/−24 LUFS/LKFS) and whether or not the gating is applied in this process. Also, some standards distinguish between short-form content (<30–120 sec) and long-form content.

ATSC DOCUMENT A85:2013

Advanced Television Systems Committee Inc. has worked out a set of guidelines intended for the Dolby AC3-based audio [6]. Originally, the level settings were determined by a more complex measure of the Dialogue Level, providing data for the dialnorm (dialogue normalization). These data are sent as metadata along with the program to provide the optimum setting at the receiver. However, after the introduction of the ITU 1770 on program loudness, the k-weighting algorithm has been a valid method to provide data of Dialogue Level for the dialnorm setting. Here are the recommendations:

Integrated loudness	–24 LKFS ±2
Gating	yes
Max true peak level	–2 dBTP

The implementation of the ATSC-standard has encountered some challenges due to various options, especially when it comes to IP-based distribution. So the Recommended Practice is not followed by all content providers.

EBU RECOMMENDATION R 128

The EBU recommendation R 128 [7] [8, 9] is also based on the ITU BS.1770. The gating function was first introduced by the EBU and later included in the ITU BS.1770 recommendation (version 2, 3 and 4).

Integrated loudness	–23 LUFS ±0.5 LU
Gating	yes
Max true peak level	–1 dBTP

The Loudness Level shall be measured with an instrument according to ITU 1770 and EBU Tech Doc 3341. [2, 3, 4, 1, 5, 6, 7, 8].

The Loudness Range shall be measured with a meter compliant with EBU Tech Doc 3342 [10]. (The Loudness Range is program dependent.)

Do note that the assessed loudness is not only dependent on the program loudness. It also depends on how channels are routed in the receiver. If the program, for instance, has one single mono channel – and is normalized according to the loudness of this – and then is played back by the receiver in two channels, the perceived level is too loud [11].

Another technical problem is measuring the integrated loudness of short-form content. EBU has worked out a supplement to R128, that addresses this issue: R128 supplement 1: Loudness parameters for short-form content [12]:

Integrated loudness	–23 LUFS ±0.5 LU
Gating	yes
Max True Peak Level	–1 dBTP
Max short-term loudness	–18.0 LUFS (+5 on the relative scale).
or	

Integrated loudness	–23 LUFS ±0.5 LU
Gating	yes
Max True Peak Level	–1 dBTP
Max momentary loudness	–15.0 LUFS (+8 on the relative scale).

OPERATIONAL PRACTICE, AUSTRALIA, OP-59

Australian Television have, in general, left the OP-48 (VU and peak measures) and moved to a loudness based specification for television. It is very close to the ITU:

Integrated loudness	–24 LKFS ±2
Gating	Yes
Max true peak level	–2 dBTP

THE ASSOCIATION OF RADIO INDUSTRIES AND BUSINESSES (ARIB), JAPAN

By referring to the ARIB TR-B32 Japanese television in general follow the ITU:

Integrated loudness	–24 LKFS ±2
Gating	yes
Max true peak level	–1 dBTP

STREAMING/INTERNET

The era of streaming audio has been pure anarchy. The practice has been that sound files for playback over the Internet must be compressed and normalized to full scale before conversion to the file ingested to the server. This procedure was followed, acknowledging that the dynamic range of mobile devices and computers perhaps were lacking a decent dynamic range – and: everybody wanted to be LOUD.

After the introduction of loudness standards and recommendations, new rules for delivery on the Internet were introduced. However, there are different approaches, and the story is not yet finished. The end-goal seems to be that all content delivered on the Internet should be the same as delivered in other broadcast services. If any compression/reduction of dynamic range is needed, it must be processed by the receiving device. If that happens, it is easier to operate on many platforms, as the technical quality of the content essentially is the same. The reason for the mentioned end-goal is the expectation that the near future will bring rules for all mobile devices regarding the sound pressure level in headphones to avoid damaging users' hearing.

In 2015 the AES worked out recommendations for Loudness of Audio Streaming and Network File Playback [13]. These are the recommendations:

Target Loudness, max:	≤–16 LUFS
Target Loudness, min:	≥–20 LUFS
Short-form Loudness (<60 s.):	≤+5 LU re Target Loudness
Max peak level:	≤–1 dBTP

The "window" of the AES recommendations are greatly followed by most serious content providers until further defined.

Many streaming service providers like Spotify, Tidal, YouTube, and Apple now tend to apply attention-only normalization with no need for limiters. Further, the normalization is based on album-normalization rather than the normalization of the individual tracks. The advantage is, of course, that all tracks on an album retain the inter-track relative levels as intended.

Still there are differences between the services regarding the loudness level and even the way to calculate it. For instance, Spotify uses ReplayGain while Tidal uses EBU-loudness measures. A special tool in the Internet named "Loudness Penalty" and developed by MeterPlugs (www.loudnesspenalty.com) enables you to run your audio files and analyze how Internet services like YouTube, Spotify, TIDAL, iTunes, Amazon Music, Pandora, and Deezer will make corrections to the audio if uploaded. As long as this tool is maintained and updated to reflect the actual practice of these services, it is a good idea to check how your work of art is handled when streamed to the world.

OTT/OVD

OTT: Over-The-Top, the means to deliver video content via streaming, Video On Demand, IPTV, and download mechanisms.

OVD: Online-Video-Distributors defined as any entity that offers video content by means of the Internet or other Internet Protocol (IP)-based transmission path provided by a person or entity other than OVD.

These areas, video with audio, also need general rules. In contrast to pure Internet media production, most of the content is produced as traditional broadcast programs or films. The problem is whether the receiving devices can handle metadata and then set levels according to this, or whether all programs should have identical conditions for playback.

Guidelines have been worked out by the AES in a Technical Document [14]. The rules in Tables 21.1, 21.2, and 21.3 apply for most broadcast services, short-form content, or long-form content.

Table 21.1 Recommendations for short-form content, which is an advertisement, promotional items, and interstitials.

Broadcast Region	Integrated Loudness	Maximum Short-Term Loudness	Maximum True Peak	Anchor Based Measurement	Full Program Mix Measurement
North America	−24 ± 2 LKFS	N/A	−2 dB TP	Not permitted	Recommended
Europe	−23 ± 0.5 LUFS	−18 LUFS and +5 LU relative to IL	−1 dB TP	Not permitted	Recommended
Japan	−24 ± 1 LKFS	N/A	−1 dB TP	Not permitted	Recommended
Australia	−24 ± 1 LKFS	N/A	−2 dB TP	Not permitted	Recommended

Table 21.2 Recommendations for long-form content, which is typical TV programs such as news, sports shows, drama, and movies.

Broadcast Region	Integrated Loudness	Maximum Short-Term Loudness	Maximum True Peak	Anchor Based Measurement	Full Program Mix Measurement
North America	−24 ± 2 LKFS	N/A	−2 dB TP	Recommended	Conditionally Permitted
Europe	−23 ± 0.5 LUFS	N/A	−1 dB TP	Conditionally Permitted	Recommended
Japan	−24 ± 1 LKFS	N/A	−1 dB TP	Not permitted	Recommended
Australia	−24 ± 1 LKFS	N/A	−2 dB TP	Permitted	Permitted

Table 21.3 Recommendations for maximum loudness: [23].

	Integrated Loudness	Maximum True Peak
Short-form content	−16 ± 1 LKFS/LUFS	−1 dB TP
Long-form content	−16 ± 1 LKFS/LUFS	−1 dB TP

Dolby A, Dolby B, Dolby C, Dolby SR, Dolby Stereo, Dolby Digital are trademarks of Dolby Laboratories.

REFERENCES

[1] AES17-2015: 'AES standard method for digital audio engineering – Measurement of digital audio equipment.'

[2] SMPTE RP 155:2014 'Reference Levels for Analog and Digital Audio Systems.'

[3] ITU-R BS.645. 'Test signals and metering to be used on international sound programme connections.'

[4] EBU Technical Recommendation R68-2000. 'Alignment level in digital production equipment and in digital audio recorders.'

[5] ITU-R BS.1770-1, 2, 3: 'Algorithms to measure audio programme loudness and true-peak audio level.'

[6] ATSC Recommended Practice: 'Techniques for Establishing and Maintaning Audio Loudness for Digital Television.' Document A/85:2013.

[7] EBU Recommendation R 128 'Loudness normalisation and permitted maximum level of audio levels.'

[8] EBU Tech Doc 3341 'Loudness Metering.'

[9] EBU Tech Doc 3344 'Practical Guidelines for Distribution.'

[10] EBU Tech Doc 3342 'Loudness Range.'

[11] EBU Tech Doc 3343 'Practical Guidelines for Production and Implementation.'

[12] EBU Recommendation R 128, supplement 1: 'Loudness parameters for short-form content (adverts, promos, etc.).'

[13] AES Technical Document AESTD1004.1.15-10: 'Recommendation for Loudness of Audio Streaming and Network File Playback.'

[14] AES Technical Document AESTD1006.1.17-10: 'Loudness Guidelines for OTT and OVD Content.'

CHAPTER 22

Summation of Audio Signals

CHAPTER OUTLINE

The summation of audio signals can be acoustic or electrical: acoustically, two sound sources in a room recorded at one position; or electrically, microphones connected to a mixer, combined and routed to one output (see Figure 22.1).

The total level of combined signals depends on the nature of the two signals. Some characteristic examples are explained here. (Also see “Addition of dB” in Chapter 6: The dB Concept).

ELECTRICAL SUMMATION, CORRELATED SIGNALS

Correlated signals more or less resemble each other, or at least they have much in common.

Imagine a test tone being recorded on a two-track recorder; it is the same 1 kHz sinus wave fed into both tracks. The signals on the two tracks are highly correlated. They are identical: the same phase and the same level. Combining these signals, the level is twice the level of the individual channels. Specified in dB, a doubling of the level is the same as +6 dB (see Chapter 6: The dB Concept).

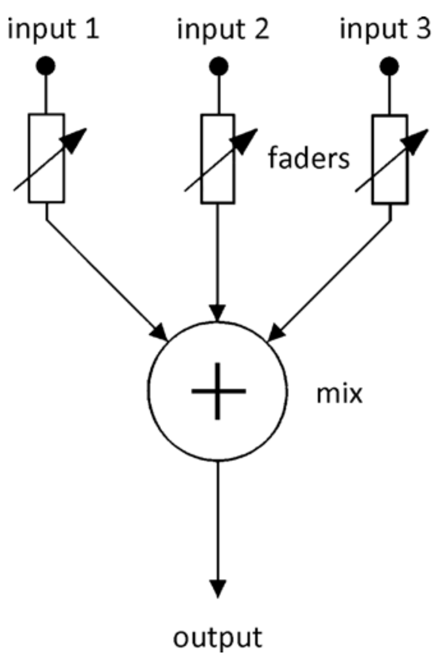

Figure 22.1 A mixer is a summation device.

If, however, we have two identical signals and one of them is phase-inverted (the phase is turned 180°) before the summation, then the result is zero – in other words, nothing. Expressed in dB, the result is –∞ dB.

ELECTRICAL SUMMATION, UNCORRELATED SIGNALS

If recording on a tape recorder, then there occurs some tape noise in the form of hiss. This noise certainly has the same magnitude on both tracks. However, the noise is random and therefore uncorrelated. Thus, by summing the two tracks, the hiss noise level only increases by +3 dB.

Engineers enjoying the ancient art of tape recording are familiar with an old trick: Even when recording mono, feed that same signal on two tracks. The signals add to +6 dB, but the noise only adds to +3 dB. In this way, the signal-to-noise ratio is improved by +3 dB.

SUMMATION FOR MONO

It can be practical to use a dedicated circuitry, a mono converter, for the summation of the two channels of a stereo program.

The simple mono converter works in this way:

1) For stereo signals, with random phase distribution, the summed mono level is calculated: mono = L + R – 3 dB.
2) Same signal in both channels (or highly correlated signals): mono = L + R – 6 dB.
3) Signal present in only one channel: mono = L; or mono = R.

However, this simple summing does not manage program material with an extensive amount of out-of-phase content. That material may originate from matrix-encoded surround programs (4:2:4 or 5:2:5). The same applies to stereo recordings building on MS-techniques if S>>M and to so-called produced stereo with artificially enhanced stereo width obtained by the application of effect processors, which is the source of the out-of-phase material. In these situations, 90° summation is applicable.

90° SUMMATION

By 90° summation, the content of one channel, compared to the other channel, is shifted by 90° at all frequencies. This phase shift also can be explained as a time delay.

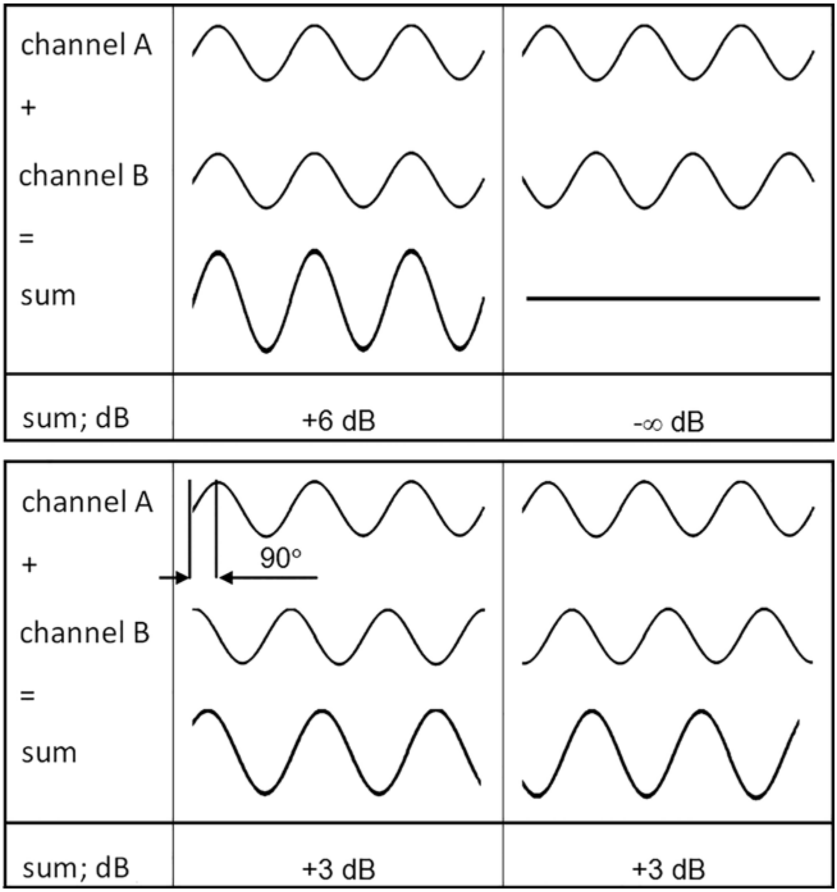

Figure 22.2 Direct summation of signals in phase and 180° out of phase. The upper part of the illustration shows the direct summation. The lower part of the illustration shows the summation by the use of a 90° phase shift, where signals both in-phase and out of phase add to the same level.

However, instead of a time delay that is constant at all frequencies, this involves a delay that is frequency dependent (see Figure 22.2).

TIME DELAY IN A 90° PHASE-SHIFTING PATH

At 20 Hz, one of the signals is in principle delayed by 1/20 × (90/360) s = 12.5 ms.

At 20 kHz, one of the signals is in principle delayed by 1/20.000 × (90/360) s = 12.5 µs.

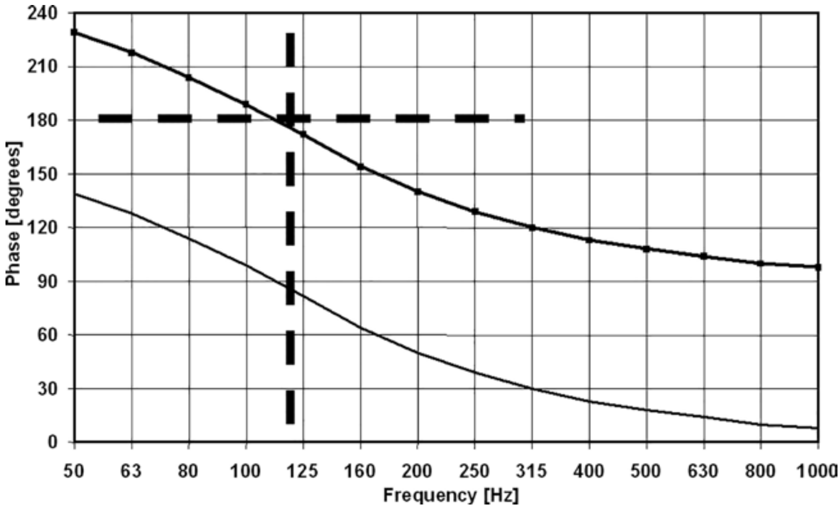

Figure 22.3 An example of one of the very few situations where a 90° summation has unfortunate consequences: Two channels (stereo) are summed. Accidently a low-cut is activated in one channel. This low-cut is hardly audible when played back in stereo. However, the channels are oppositely phased around 115 Hz, which in mono sounds like (and is) a lack of bass.

The phase relationship between two signals a and b is relevant for the calculation of the level of the summed signals (see Figure 22.3):

$$a = b \times \cos\phi$$

where

ϕ = the phase angle (the phase between the two signals)

COMB FILTERING

The filtering function that arises when a signal adds to itself after being delayed is called a comb filtering (see Figure 22.4). The resulting frequency response looks like a comb, hence the name (see Figure 22.5). The comb filter function is rarely intentional, but it is heard all the time in sound productions, where it can arise both acoustically and electrically.

Acoustically, it typically occurs when the sound on its way from source to receiver takes in part a direct path and in part an indirect path via a single reflective surface. The reflection must be attenuated at least 10 dB and preferably 15 dB for it not to

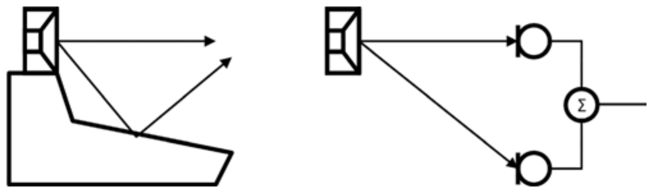

Figure 22.4 Two typical situations in which comb filtering arise, either acoustically or electrically.

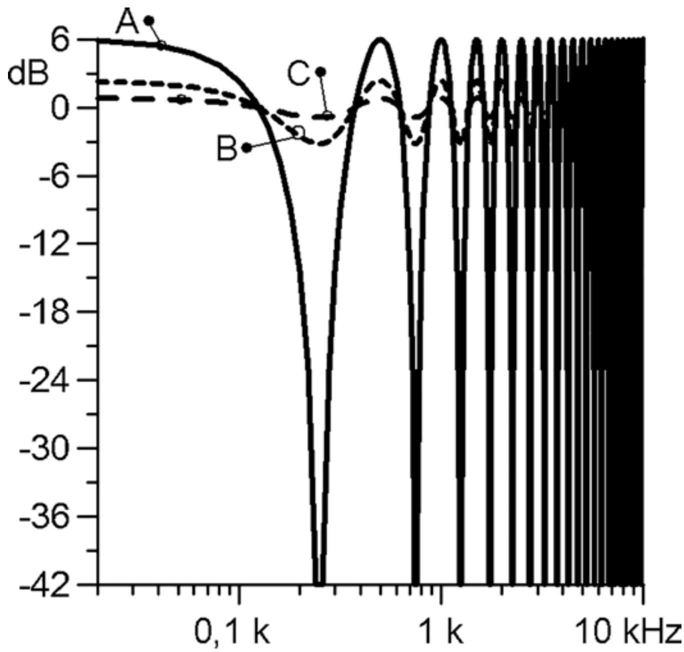

Figure 22.5 An example of a comb filter created by the combining of two signals with the same amplitude but with a time delay between them of just 1 ms. A dip occurs due to cancellation at 500Hz, 1.5 kHz, 2.5 kHz, and so on. The two signals add to double their value (+6 dB) at low frequencies and with a full wavelength's delay at 1 kHz, 2 kHz, 3 kHz, and so on.

influence the sound field at the receiver position. Electrically, the phenomenon arises when two microphones with some distance between them capture the same acoustical signal.

In general: All digital signal processing takes time. In practice, the comb filter effect can arise if you loop a signal via, for example, a compressor and followingly reinsert the processed signal with the original.

DIP FREQUENCIES

Cancellation occurs at all oppositely phased frequencies, when the time delay is comprised of periods of a duration of ½, 1½, 2½, and so on. At 1 kHz the duration of one period is 1 ms; half of that period is 0.5 ms. If a time delay is precisely 0.5 ms, then the cancellation occurs, not just at 1 kHz, but also at 2 kHz, 3 kHz, 4 kHz, and so on.

SUMMATION FOR LOUDNESS LEVEL

When calculating the loudness level of multichannel formats, the audio signals of each channel are summed. This summation is phase insensitive, as it is based on the summation of the power of each channel.

ACOUSTIC SUMMATION OF SIGNALS

The total sound level by the acoustic summation of two sound sources, such as two monitor speakers, depends on both the signal and the acoustics. The sound sources can be correlated or uncorrelated, as mentioned previously (see Figure 22.6).

The listening position (or measurement position) can be either in the direct sound field or the diffuse sound field. In the direct sound field, there is only one sound direction from each source. This direct field exists either in the open, in a reflection-free

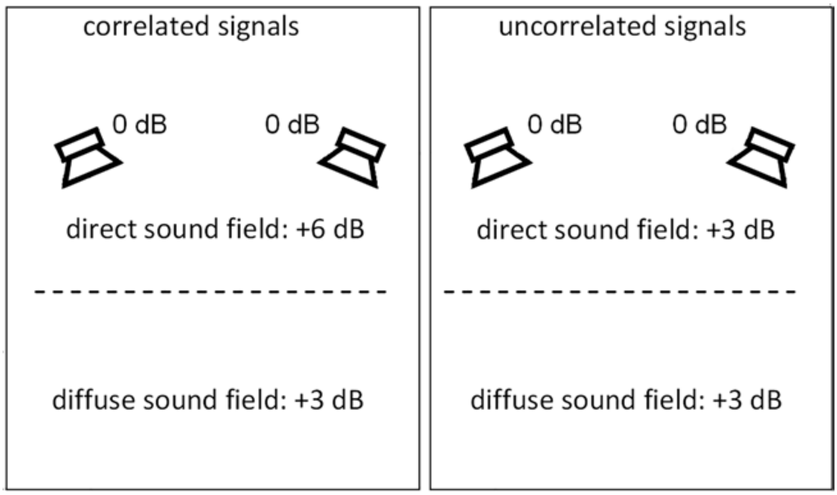

Figure 22.6 Summation of sound from two sources with the same level. The summation in the direct field is determined by whether the sources are correlated or not. The summation in the diffuse field is independent of the correlation of the sources.

room, or close to the speakers. The diffuse sound field dominates in a closed space when you are so far away from the speakers that the portion of the direct sound is less than the sum of all the reflections. The distance from the speakers where the direct sound field and the diffuse sound field are equally large is called the critical distance. In a control room, it can typically be 1–3 meters; however, it varies with frequency. At lower frequencies, the critical distance is shorter compared to that at higher frequencies. The near field in front of the speakers is regarded as a direct field.

In a Left – Center – Right loudspeaker configuration in a small environment, like an OB-van or the like, it is common to skip the center speaker to save the space for a visual monitor. So instead the center signal is attenuated by 3 dB and routed equally to left and right speaker to form a phantom center. However, the problem is in the nearfield, that the split center signal adds and is 3 dB louder (−3 dB + (−3 dB) = +3 dB) compared to the level of one center loudspeaker. In the diffuse field, the level is okay.

CHAPTER 23

Digital Interface

CHAPTER OUTLINE

When transferring digital information from one device to another, a form of a protocol is used to ensure that both the transmitter and the receiver are in agreement on which bit means what. Also, the two devices' physical connection is specified. Together, this defines an interface.

PROTOCOL

Each sample, which is described by several bits arranged in a sequence, can have several extra bits (metadata) appended to it which provide information associated with the actual sample. It could be information about which channel the signal belongs to, whether pre-emphasis of the analog signal was applied, data about the sampling frequency, and much other useful information. Samples and extra information are combined into frames. The protocol defines in part the structure of the individual frames and the combining of frames into blocks of data.

PHYSICAL CONNECTION

The digital information is transferred in the form of a voltage signal, a current signal, or as light impulses on a fiber optic cable. The individual standards define what method or methods that apply.

Even though the connection transfers audio, the frequency content in a connection like that immediately reaches several MHz. Hence audio transferred as digital signals should always be treated as RF.

AES3

The most widely used interfaces for audio were originally initiated and standardized by the Audio Engineering Society and the European Broadcast Union, and hence bore the names of both of these organizations. However, the AES has taken over the maintenance of this standard, so the name has just become "AES3," a standard in four

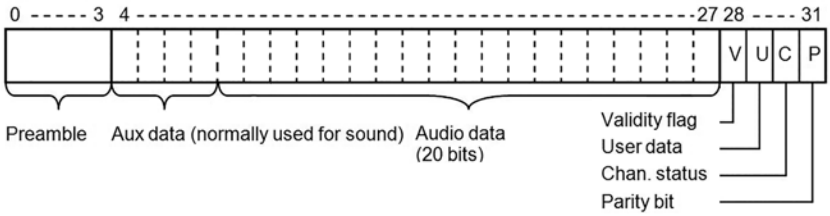

Figure 23.1 AES subframe, "packaging" of each sample. Each subframe begins with a preamble to help with the synchronization. Each subframe ends with four bits containing various useful information.

parts. It is a standard that is constantly under revision as new options in digital audio enter the market.

The format is serial (i.e., even though it contains two channels, it can be run on one wire). During the transmission, a sample from the first channel is transferred, followed by a sample from the second channel, then again a sample from the first channel, and so on.

The physical connection is a balanced 110 ohm cable with XLR connectors. The level is 2–7 Vpp (peak to peak). However, the signal can be read down to 200 mV pp. Other than balanced lines may work as well.

The core of the data portion of the AES3–2 consists of 20 audio bits and four extra bits (aux). These four extra bits have gradually become used for sound, so regardless of how many bits applied by the time of sampling, 24 bits per sample are transferred. For example, if 16 significant bits are involved (as in the CD format), then bits 17–24 are set to the value "0." Each sample (from either the right or the left channel) is inserted into a 32-bit subframe (see Figure 23.1).

PREAMBLE, V, U, C, AND P

The eight "surplus" bits in each subframe applies to synchronization and extra information.

The first four bits comprise a preamble, which indicates which channel the current frame represents or whether it is the first subframe in a block.

The last four bits contain extra information that can say something about the signal that is transmitted. This information is divided up into four individual bits, which are designated V, U, C, and P (see Figure 23.2).

V stands for Validity. This bit indicates whether the associated sample is in order. P stands for Parity. This parity bit is set providing an even number of 1´s always is

Figure 23.2 AES: For every 8 subframes, one byte is generated. One block contains 192 frames (2*192 subframes).

attained in each subframe. If a check on the Parity bit fails, the Validity bit is reset to indicate possible corruption in the subframe data.

C (Channel status) tells about the signal itself, and U (User bit) can, as the name implies, be defined by the user.

C and U are collected on a running basis into 8-bit bytes.

FRAMES AND BLOCKS

Two subframes together comprise an entire frame of 64 bits (2*32) in total. In audio, this will normally be one subframe with data from the left channel and one subframe with data from the right channel. If the sampling frequency is, for example, 48 kHz, then it means that 64 bits are transferred 48,000 times per second. This results in a bitstream of a good 3 Mbps (Megabits per second).

In connection with the expansion of the standard also to be used for higher rate sampling (88.2 or 96 kHz), two subframes can be used to contain samples from a single channel.

A block is created from 192 pairs of samples or frames. At a sampling frequency of 48 kHz, this results in 250 blocks per second.

As mentioned, one byte is created from eight bits. In each block, 24 (192/8) channel status bytes – and the same number of user-defined bytes – can be formed by left channel user data and right channel user data, respectively. Subframe 0–7 forms the first byte, 8–15 forms byte number two, and so on.

Each byte in a block has its own significance. With 250 blocks and thus that many bytes per second, a great deal of extra information can be transferred along with the audio.

CHANNEL STATUS

The information that is contained in channel status can be used to indicate what the actual bitstream consists of, where the individual sample belongs, and further, can provide a full timecode. The data contained may be protected with its own error correction code (CRC).

As mentioned earlier, 24 bytes is established within one block. Of these, we will only look at the first five. The AES3–2 defines this information. However, even the AES3 is very comprehensive; much of this information is similar to what you can find in other interface protocols. It may seem a little nerdy to go through these protocol tables. However, it provides an excellent overview and understanding of the capability of a standard like this.

STATUS BYTE 0

Here you find the most significant difference between the professional and consumer versions. Depending on the value of the first bit, the channel status will have a different meaning. AES deals in principle only with the professional version (see Table 23.1). The consumer information lies in the protocol for IEC 60958, also known as S/PDIF (see a later section).

Table 23.1 Status byte 0.

	Bit	State	Meaning
Use of channel status block	*0*	*0*	*Consumer use of channel status block (see note 1).*
		1	*Professional use of channel status block.*
Linear PCM identification	*1*	*0*	*Audio sample word represents linear PCM samples.*
		1	*Audio sample word used for purposes other than linear PCM samples.*
Audio signal emphasis	*4 3 2*	*000*	*Emphasis not indicated. Receiver defaults to no emphasis with manual override enabled.*
		001	*No emphasis. Receiver manual override is disabled.*
		011	*50 μs + 15 μs emphasis, see ITU-R BS.450. Receiver manual override is disabled.*
		111	*ITU-T J.17 emphasis (with 6,5-dB insertion loss at 800 Hz). Receiver manual override is disabled.*
		All other states of bits 2 to 4 are reserved and are not to be used until further defined.	
Unlocked indication	*5*	*0*	*Default. Lock condition not indicated.*
		1	*Source sampling frequency unlocked.*
Sampling frequency	*7 6*	*00*	*Sampling frequency not indicated. Receiver default to interface frame rate and manual override or auto set is enabled.*
		10	*48-kHz sampling frequency. Manual override or auto set is disabled.*
		01	*44.1-kHz sampling frequency. Manual override or auto set is disabled.*
		11	*32-kHz sampling frequency. Manual override or auto set is disabled.*

STATUS BYTE 1

This byte contains information on the relationship between the audio signals in the two channels (see Table 23.2). After higher rate sampling became an option, an indication was also added for the relationship between subframes.

STATUS BYTE 2

Here information is found on auxiliary bits, word length, and alignment level (see Table 23.3).

Table 23.2 Status byte 1.

	Bit	State	Meaning
Channel mode	*3 2 1 0*	*0000*	*Mode not indicated. Receiver default to two-channel mode. Manual override is enabled.*
		1000	*Two-channel mode. Manual override is disabled.*
		0100	*Single-channel mode, (monophonic). Manual override is disabled.*
		1100	*Primary-secondary mode (Subframe 1 is primary). Manual override is disabled.*
		0010	*Stereophonic mode, channel 1 is the left channel. Manual override is disabled.*
		1010	*Reserved for user-defined applications.*
		0110	*Reserved for user-defined applications.*
		1110	*Single-channel double sampling frequency mode. Subframes 1 and 2 carry successive samples of the same signal. The sampling frequency of the signal is double the frame rate, and is double the sampling frequency indicated in byte 0, but not double the rate indicated in byte 4, if that is used. Manual override is disabled. Vector to byte 3 for channel identification.*
		0001	*Single-channel double sampling frequency mode – stereo mode left. Subframes 1 and 2 carry successive samples of the same signal. The sampling frequency of the signal is double the frame rate, and is double the sampling frequency indicated in byte 0, but not double the rate indicated in byte 4, if that is used. Manual override is disabled.*
		1001	*Single-channel double sampling frequency mode – stereo mode right. Subframes 1 and 2 carry successive samples of the same signal. The sampling frequency of the signal is double the frame rate, and is double the sampling frequency indicated in byte 0, but not double the rate indicated in byte 4, if that is used. Manual override is disabled.*
		1111	*Multichannel mode. Vector to byte 3 for channel identification.*
		All other states of bits 0 to 3 are reserved and are not to be used until further defined.	
User bits management	*7 6 5 4*	*0000*	*Default, no user information is indicated.*

	Bit	State	Meaning
		1000	*192-bit block structure with user-defined content. Block start aligned with Channel Status block start.*
		0100	*Reserved for the AES18 standard.*
		1100	*User defined.*
		0010	*User data conforms to the general user data format defined in IEC 60 958–3.*
		1010	*192-bit block structure as specified in AES52. Block start aligned with Channel Status start.*
		0110	*Reserved for IEC 62 537.*
		All other states of bits 4 to 7 are reserved and are not to be used until further defined.	

Table 23.3 Status byte 2.

	Bit	State	Meaning	
Use of AUX bits	*2 1 0*	*000*	*Maximum audio sample word length is 20 bits (default). Use of AUX bit is not defined.*	
		100	*Maximum audio sample word length is 24 bits (default). AUX bits are used for main audio samples.*	
		010	*Maximum audio sample word length is 20 bits. AUX bits in this channel are used to carry a single coordination signal.*	
		|110	*Reserved for user-defined applications.*	
		All other states of bits 0 to 3 are reserved and are not to be used until further defined.		
Encoded audio sample word length of transmitted signal			*Audio sample word length if maximum length is 24 bits as indicated by bits 0 and 2 above.*	*Audio sample word length if maximum length is 2 bits as indicated by bits 0 and 2 above.*
	5 4 3	*000*	*Word length not indicated (default).*	*Word length not indicated (default).*
		001	*23 bits.*	*19 bits.*
		010	*22 bits.*	*18 bits.*
		011	*21 bits.*	*17 bits.*
		100	*20 bits.*	*16 bits.*
		101	*24 bits.*	*20 bits.*
		All other states of bits 0 to 3 are reserved and are not to be used until further defined.		
Indication of alignment level	*7 6*	*00*	*Alignment level not indicated.*	
		10	*Alignment to SMPTE RP155, alignment level is 20 dB below max code.*	
		01	*Alignment to EBU R68, alignment level is 18.06 dB below max code.*	
		11	*Reserved for future use.*	

STATUS BYTE 3

In multichannel mode, this byte can indicate which channel is involved (see Tables 23.4, 23.5 and 23.6).

Table 23.4 Status byte 3

	Bit	State	Multichannel mode
State	*7*	*0*	*Undefined multichannel mode (default).*
		1	*Defined multichannel modes.*

Either

Table 23.5 Status byte 3 bit 7 = 0

	Bit	State	Channel number, when byte 3 bit 7 is 0
Values		The channel number is the numeric value of the byte, plus one, with bit 0 as the least significant bit.	

Or:

Table 23.6 Status byte 3 bit 7 = 1

	Bit	State	Multichannel mode, when byte 3 bit 7 is 1
Use of AUX bits	*6 5 4*	*000*	*Multichannel mode 0, The channel number is defined by bits 3 to 0 of this byte.*
		001	*Multichannel mode 1, The channel number is defined by bits 3 to 0 of this byte.*
		010	*Multichannel mode 2, The channel number is defined by bits 3 to 0 of this byte.*
		011	*Multichannel mode 3, The channel number is defined by bits 3 to 0 of this byte.*
		All other states of bits 0 to 3 are reserved and are not to be used until further defined.	
Channel number, when byte 3 bit 7 is 1	*3 to 0*	*The channel number is the numeric value of these four bits, plus one, with bit 0 as the least significant bit.*	

STATUS BYTE 4

The last part of this byte was put into use for the indication of sampling frequencies that were introduced after the original standard ones (see Table 23.7).

Table 23.7 Status byte 4

	Bit	State	Meaning
Digital audio reference signal	*1 0*	*00*	*Not a reference signal (default).*
		10	*Grade 1 reference signal (ref AES11).*
		01	*Grade 2 reference signal (ref AES11).*
		11	*Reserved and not to be used until further defined.*
Information hidden in PCM signal	*2*	*0*	*Not indicated (default).*
		1	*Audio sample word contains additional information in the least significant bits (ref AES55).*
Sampling frequency	*6 5 4 3*	*0000*	*Not indicated (normal).*
		0001	*24 kHz.*
		0010	*96 kHz.*
		0011	*192 kHz.*
		0100	*Reserved.*
		0101	*Reserved.*
		0110	*Reserved.*
		0111	*Reserved.*
		1000	*Reserved for vectoring.*
		1001	*22.05 kHz.*
		1010	*88.2 kHz.*
		1011	*176.4 kHz.*
		1100	*352.8 kHz.*
		1101	*Reserved.*
		1110	*Reserved.*
		1111	*User defined.*
Sampling frequency scaling flag	*7*	*0*	*No scaling (default).*
		1	*Sampling frequency is 1/1.001 times the value in bits 3–6 or by byte 0 bits 6–7.*

AES3 AS A GENERAL FORMAT

Regarded as an interface, AES3 also applies to general transport, such as in bit-reduced multichannel sound. The interface applies to Dolby E, for example, which is a standard for bit-reduced sound in six channels with accompanying metadata. Similar possibilities exist for MPEG multichannel formats. Further, the Digital Cinema applies eight pairs of AES3 channels for the delivery of cinema audio.

Figure 23.3 Significant information gathered from channel status and made available on the screen of the instrument. Among other things, you can see that this concerns professional/two-channel audio/48 kHz/24 bit/without emphasis/no time code.

BI-PHASE MODULATION

Before the digital bitstream in the AES3 signal is transmitted, it is transformed for bi-phase modulation, which is the same coding that applies to SMPTE time codes.

With bi-phase modulation, a shift is performed (from 0 to 1 or vice versa) for each new bit in the signal. Also, a shift occurs each time the bit value is "1" (see Figure 23.4). Bi-phase modulation has the advantage that the signal is free of DC, it is independent of connecting polarity, and it is essentially self-clocking.

IEC 60958–3

The IEC 60958–3 (+ amendments), formerly known as the Sony Philips Digital Interface, or just S/PDIF, was defined as the consumer format. The structure of the data stream is the same as in AES3.

There are certain differences in that the format also contains the transmission of four-channel usage. Another significant point where it differs is copy protection (copy-prohibited).

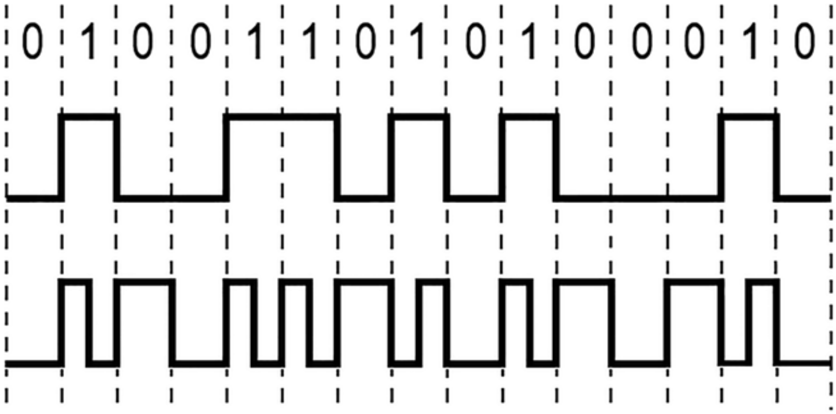

Figure 23.4 Bi-phase modulation. Top: Digital information. Bottom: Same signal, but bi-phase modulated.

In the latest version information on Loudness and loudness alignment of the contained data is also implemented. Further, MPEG-surround now can be buried in the data stream.

The physical connection deviates from AES3 in that a 75 ohm unbalanced coax cable with RCA-phono jacks is used. The electrical level ought to be above 1 volt pp. However, values down to 400 mV pp can be read. In practice, this means that there are abundant opportunities for IEC 60958 and AES3 to be able to understand each other. IEC 60 958 can also run on fiber optic cables.

CHANNEL STATUS

The structure of the bit format is, as mentioned earlier, comparable with AES3. The difference is indicated in channel status byte 0. If the value is 0, then the subsequent bits and bytes in channel status will have meanings that differ from AES3.

OTHER INTERFACE STANDARDS

AES3 is the basis for many applications. There are, of course, other important interfaces besides AES3 and standards in that class. The formats can differ from each other by their primary purpose, the number of channels, the quantity of metadata, and so on.

AES42: A standard for digital microphones. It is very much like the AES3. However, metadata bytes are defined for signaling of the microphone make, type, and settings, and so on.

AES10: AES Recommended Practice for Digital Audio Engineering - Serial Multichannel Audio Digital Interface (MADI), a serial multichannel interface supporting 56 audio channels.

SDI (serial digital video interface), here four AES3 pairs are embedded for audio purposes.

REFERENCES

AES3-1-2009 (r2019): AES standard for digital audio – digital input-output interfacing – serial transmission format for two-channel linearly-represented digital audio data – part 1: Audio Content.

AES3-2-2009 (r2019): AES standard for digital audio – digital input-output interfacing – serial transmission format for two-channel linearly-represented digital audio data – part 2: Metadata and Subcode.

AES3-3-2009 (r2019): AES standard for digital audio – digital input-output interfacing – serial transmission format for two-channel linearly-represented digital audio data – part 3: Transport.

AES3-4-2009 (r2019): AES standard for digital audio – AES standard for digital audio – digital input-output interfacing – serial transmission format for two-channel linearly-represented digital audio data – part 4: Physical and electrical.

AES10-2008 (r2019): AES Recommended Practice for Digital Audio Engineering – serial multichannel audio digital interface (MADI).

AES11-2009 (r2019): AES recommended practice for digital audio engineering – synchronization of digital audio equipment in studio operations (revision of AES11-2003).

AES18-1996 (s2008): AES Recommended practice for digital audio engineering - Format for the user data channel of the AES digital audio interface (revision of AES18-1992).

AES42-2010 (r2015): AES standard for acoustics – digital interface for microphones.

AES50-2005: AES standard for digital audio engineering – high-resolution multi-channel audio interconnection.

AES52-2006 (r2017): AES standard for digital audio engineering – insertion of unique identifiers into the AES3 transport stream.

AES55-2012 (r2017): AES standard for digital audio engineering – carriage of MPEG-Surround in an AES3 bitstream.

Digital Cinema System Specification Version 1.3 – approved 27 June 2018. Digital Cinema Initiatives, LLC, Member Representatives Committee.

IEC 62 537: Interface for loudspeakers with digital input signals based on IEC 60958 (2006).

IEC 60 958-3 Ed. 3.0 b:2006: Digital audio interface – part 3: Consumer applications (+ amendments 1 & 2).

ITU-R Recommendation BS.450-3: Transmission standards for FM sound broadcasting at VHF.

ITU-T Recommendation J.17: Pre-emphasis used on programme circuits (1993).

SMPTE RP 155:2004 – SMPTE recommended practice – for motion pictures and television – reference level for digital audio systems.

CHAPTER 24

Audio-over-IP

CHAPTER OUTLINE

The majority of professional audio networks have adopted IP technology. Essentially, these are the same technologies used in computer networks based on the Ethernet introduced in 1973. Since the 1990s, different protocols for network-based audio distribution have been introduced to the market. The majority of these systems are proprietary. Essentially, it is only the developing companies (and those who buy licenses) that can implement them. These include Allen & Heath's ACE, Aviom's A-Net, Behringers Ultranet, QSC's Q-LAN, Peak Audio's CobraNet, Roland Pro's REAC (Digital Snake), and Audinate's DANTE.

To bring some order to the increasing amount of formats, AES (Audio Engineering Society) has prepared a standard (AES67–2015: AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability). This standard is not in itself a network. However, AES67 is instead considered as a set of guidelines that describe and put together various existing elements and structures for the use of sound exchange through Ethernet-based networks. In this way, several existing networks have gained so many common denominators that one can reasonably establish "translators" and thus exchange data across the types.

Networking is interesting in connection with audio metering, as still more meters connect to larger audio/video systems via Ethernet cables.

AOIP – FOR PROFESSIONAL AUDIO

AoIP is the abbreviation of Audio over Internet Protocol. IP is a technology originally intended for data transmission in general. The principle here is that each device, such as every connected computer in the network, is assigned a unique IP address.

It also requires that communication from a computer to other devices use a common method or model. ISO has defined one model applicable: Open Systems Interconnection model (OSI). The standard establishes communication in, for example, computer systems, regardless of their underlying internal structures, technologies, and makes.

The basis of the OSI model is a seven-layer structure. Each layer has a function, and the hierarchy simultaneously specifies the order of execution of each function (see Figure 24.1).

Layer 1 - physical layer - is the outermost layer comprising the physical connection, (i.e., cable and connectors).

Layer 2 - data link layer - this layer ensures seamless transmission of data packets from one point to another via the physical layer, including resubmission of data that didn't seem to get through in the first place.

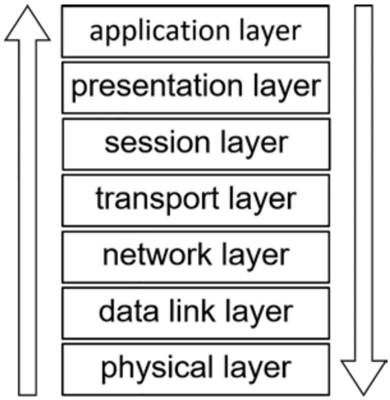

Figure 24.1 Structure of the OSI model.

Layer 3 - network layer - this is where we find the IP addresses. In this layer, the logical addresses translate into physical addresses. Also, it controls subnet (i.e., decides which way data pass in the system). Further, this layer checks the order of the necessary processes to get the data ahead of time.

Layer 4 - transport layer - often referred to as end-to-end layer - ensures transport from point to point by adding the correct address to each data packet transmitted in the system. This layer also ensures any error correction.

Layer 5 - session layer - keeps track of the setup and communication between multiple devices on the network.

Layer 6 - presentation layer - describes the data structure that must be used partly for data to be read by the user and partly used to format data transported in the system. Also, this is the place for any encryption applied.

Layer 7 - application layer - is the software that the user uses to control the system setup.

The network is typically built by connecting all devices via a switch (i.e., a device capable of directing the signals to the correct addresses). In the existing protocol IP_v4, the possible number of addresses is restricted. Therefore, the development of an IP_v6 was a necessity (see Table 24.1). The number of possible addresses is 2^{128} or approximately $3.4 * 10^{38}$. It probably covers near-future needs. The challenge for the various networks is the change over to using the extended addresses.

In analog audio, as is well known, a set of copper wires must be used for each signal. The connection from a stage box to a mixer, therefore, requires many conductor pairs and the overall cable may end up being quite thick. If a version with

Table 24.1 Structure of IP addresses, IP_v4 and IP_v6.

IP version	IP_v4	IP_v6
Introduced	1981	1999
Address size	32-bit number	128-bit number
Address format	Decimal point notation	Hexadecimal notation
	192.149.252.76	3FFE:F200:0234-AB00
Prefix notation	192.149.0.0/24	3FFE:F200:0234::/48
Number of addresses	2^{32}	2^{128}

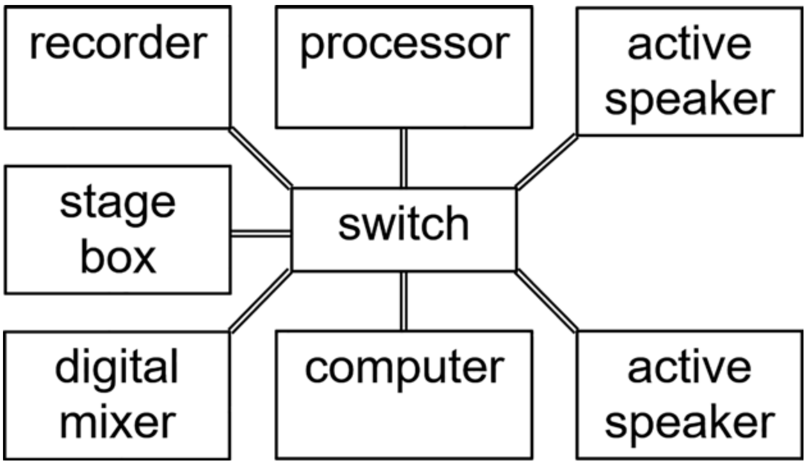

Figure 24.2 Diagram of typical AoIP-installation.

digital network technology applies, you can still handle all the signals with a simple network cable – even with hundreds of audio channels – with a copper connection. This simple network cable (CAT 5, CAT 5e, CAT 6, CAT 6a, CAT 7; see the section on CAT cables later in this chapter) has eight wires and is equipped with the RJ45 connector. Alternatively, fiber optic technology is applied, which is primarily for long-distance transmission. See Figure 24.2 for a typical IP-based network configuration.

Audio data is received and transmitted through the OSI model. Data is broken down into small "packets" containing IP information with addresses, timing, and so on. The packets, like all other data packets, are sent to the network and subsequently put together in the correct order at the recipient address. If the structure is optimal, it is a process completed in a few milliseconds – or faster.

A COMMON DENOMINATOR FOR AOIP: AES67

AES67 has, as mentioned, been established based on existing technologies, as described both in AES standards and in publications from IEEE, IETF, and other parties. In the current versions, communication is based on IP_v4, although future use of IP_v6 is considered. AES67 excludes low capacity networks, which means that ordinary Internet is excluded. A sufficient capacity never can be guaranteed here. Also, data-compressed (bit-reduced) audio is excluded. Precision Time Protocol (PTP) applies for distributing clock information; a standard retrieved from the IEEE. Up to eight audio channels can be streamed (i.e., conveyed in a one-bit stream, usually at 48 kHz, 16 or 24 bits). However, there is no limit to the number of streams to be synchronized down to bit accuracy. In the bitstream, a series of data packets, which essentially are defined with a content of 48 bits, are transported. The choice is a compromise to achieve limited time delay in systems that have limited speed, such as pure software-based virtual PC cards. In principle, the AES67 did not include a procedure for automatically recognizing devices on the network. However, this may come at a later time. Therefore, so far, there may be different procedures for setting up the various networks that apply to link to other networks via AES67.

NETWORKS

Some networks, such as Ravenna, are based on open platforms for AoIP, in contrast to others based on proprietary standards (i.e., standards that typically require licensing if anyone wants to implement them in their products). Dante is an example of this.

RAVENNA

The requirement for the physical networks is the need for sufficient bandwidth in cables, switches, and so on. As with all data networks, Ravenna prefers to limit the load to 50% of its absolute capacity. A multicast protocol is applied: IGMP (Internet Group Management Protocol). A protocol like that ensures that the individual data packets are routed directly to the addresses for which they are intended. Almost all networks need to avoid unnecessary traffic in the system, with data packets floating around without the ability to find the right address. It is therefore also important that the individual – physical as well as virtual – units are set up with correct addresses. As an IP-based network that may apply to services other than audio transmission, the system is required to have a QoS (Quality of Service) protocol. Essentially, you attach information that tells which audio data packets should be queued first; this is to ensure timing in the system so all data packets containing audio reach their target on time.

In all networks, it is incredibly important that the transmission of audio data does not involve more time delay than is necessary. Therefore, latency in the transmission of audio data is an important parameter that is sought to be kept low. In Ravenna, the delay is typically about 1 ms. Ravenna has the most prevalence in broadcast.

DANTE

The Australian company Audinate is behind the development of Dante. Dante was not born to be compatible with the AES67 foundation, but this has become an option in versions after 2015. Dante has gained extremely broad acceptance, although it is a proprietary system. The widespread use is partly because common hardware solutions such as routers and switches may apply. Also, the continuous development of many tools, typically pure software, makes the connection possible with other network formats. In practice, this means that Dante becomes attractive at the price while still becoming more compatible with other systems. It is recommended to use Gbps networks (also called Gigabit networks) and not just the more limited 100 Mbps. In principle, there is the possibility of an almost infinite number of audio channels. The sound can be resolved up to 32 bits and a sampling frequency up to 192 kHz. There are networks that perform above Tera bits per second capacity, which means it is difficult to see any restrictions. However, the final resolution may depend on the selected products and, moreover, whether you choose to run AES67-compatible (max. sampling frequency at 48 kHz and a fixed latency of 2 ms.). Dante supports both unicast and multicast (i.e., respectively, from one point to another or signals between many points). The setup is generally simple. If applying a computer, a small piece of software must be installed, a Dante controller (freeware). After that, many manufacturers' devices that support Dante can be put together in the network. Dante uses settings where the network itself finds the relevant addresses in the system. Dante also offers a virtual sound card, also a piece of software that provides direct contact to most DAWs. Dante Via is another option to connect between different systems. Dante has the option of redundancy (i.e., setting up a parallel system which – for safety's sake – may be appropriate in large systems). The system can handle glitch-free transitions between the redundant systems. However, this feature does not exist if run in AES67 mode. Dante is the most widely used network for AoIP.

AVB

AVB, Audio Video Bridging, is a set of standards, initiated by the IEEE. In this context, it is now the transport not only of sound but of images. The standards must ensure synchronized data streaming through IEEE 802 networks with minimal time delay. The further development, certification, and implementation of products is partly laid down by the organization AVnu Alliance. However, AVB is open, so that anyone can use the protocols without a license.

As a sound distribution system, AVB includes the principle of infinitely many audio channels, with formats up to 32 bit/192 kHz. Delay in the system is a minimum of 0.25 ms.

Please note that not all AVB devices are compatible. Therefore, it is a good idea before purchasing to check whether the product is certified by the AVnu Alliance.

CONNECTORS AND CABLES

Here, the relevant connections for AoIP are reviewed.

PLUG: RJ45

If the connection is established using copper cable, only one specific plug type is used, namely, RJ45 (see Figure 24.3).

CAT CABLES

The cable type used for Ethernet is referred to as CAT, which is an abbreviation for "Category." The subsequent number Cat5, Cat6, and so on indicates the quality – not only the bandwidth but also other characteristics such as cross talk, impedance, and so on. The cables are standardized according to ISO/IEC 11801. Usually, the type name is printed on the outside of the cable. A higher number means higher quality.

A Cat cable has eight wires that are twisted in pairs but unshielded (UTP, Unshielded Twisted Pair). A common screen may surround the four wiring pairs.

The Cat1–Cat4 categories have no application in audio.

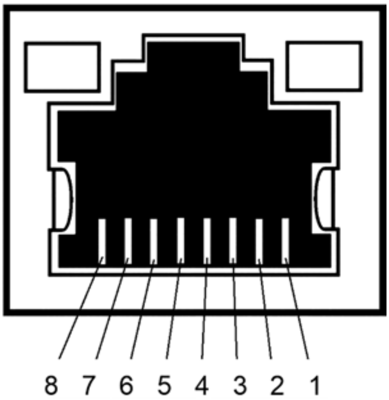

Figure 24.3 The connections in an RJ45 socket (seen from the outside).

CAT5

This cable has a bandwidth of 100 MHz and a data capacity up to 10/100 Mbps, which is sufficient for regular data transmission and some early small AoIP networks (e.g., Digital Snake) but should not be used for newer AoIP installations.

CAT5E

The letter "e" indicates that this is an enhanced version of Cat5. The cable has a bandwidth of up to 350 MHz and can handle ten times more data than the standard Cat5. Maximum length 50 m.

CAT6

The cable conductors have a slightly larger diameter, compared to the Cat5e. The bandwidth is 550–1000 MHz (depending on the source) and can handle up to 10 Gbps. However, the length is a limitation. Maximum cable length is also here approximately 50 m.

CAT6A (ISO/IEC 11801 CATEGORY 6A/CLASS E)

An improved version of Cat6 (augmented Cat6), but with the same bandwidth and capacity. However, thicker cable sheath means that the cable can be pulled up to 100 m.

CAT7 (ISO/IEC 11801 CATEGORY 7 / CLASS F)

The bandwidth is ≥700 MHz.

CAT7A

Often referred to as "augmented Cat7," an improved version.

POE – POWER OVER ETHERNET

PoE is a way of conveying a supply voltage via a wired version of a local area network (Ethernet LAN) to a device that needs power. Typically, switches work as a supply unit. These are connected to devices via the Cat cables, which receive both data and supply voltage in the same cable. Thus these devices do not need a power supply from other power supplies. For example, it is the control boxes, cameras, and the like in the network that benefit from getting the supply this way.

The voltage is nominally 48 volts but in practice must range from 44 to 57 volts. Here, too, there are several versions performing differently. The connections can allocate either the RJ45's two middle pairs (Mode A), the two outermost pairs (Mode B), or all four (4 pairs) (see Figure 24.4 and Table 24.2).

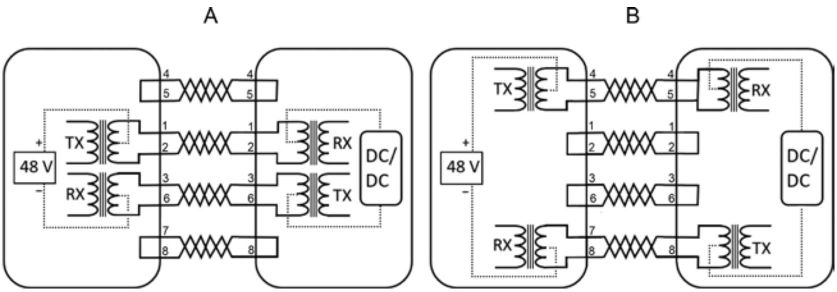

Figure 24.4 Coupling of RJ45 for distribution of PoE, IEEE802–3, alternative A (via data cables) or alternative B (via auxiliary cables).

Table 24.2 Power over Ethernet.

Designation	Voltage (nominal) [V]	Current [mA]	Power PSE [W]	Power PD [W]	Mode	Standard
PoE, type 1	48	≤350	15.4	12.95	A, B	IEEE 802.3 af
PoE+, type 2	48	≤600	30.0	25.50	A, B	IEEE 802.3 at
4PPoE, type 3	48	≤600 per par	60.0	51.00	A, B, 4-par	IEEE 802.3 bt
4PPoE, type 4	48	≤960 per par	100.0	71.00	4-par	IEEE 802.3 bt

OPTICAL CONNECTION, FIBER

In networks where high-speed connections (Gigabit Ethernet) or connections over longer distances (> 100 m) are required, it is advisable to use an optical fiber connection. It is laser light that is transmitted in the optical fiber. The wavelength of the laser is crucial for the permittable length of the cables. The IEEE standard allows for up to 70 km of cable with a wavelength of 1550 nm. Reflections in the cable can weaken the light. Often a different coding of the signals is used, partly to increase capacity and partly to avoid misreading of many data with the same value (0's or 1's). There are different versions of 1000BaseSX.

REFERENCES

AES Whitepaper: Best Practices in Network Audio (2009).

AES47-2006 (s2017): AES standard for digital audio – digital input-output interfacing – transmission of digital audio over asynchronous transfer mode (ATM) networks.

AES47-Am1-2008: Amendment 1 to AES47 – AES standard for digital audio – digital input-output interfacing – transmission of digital audio over asynchronous transfer mode (ATM) networks.

AES51-2006: AES standard for digital audio – digital input-output interfacing – transmission of ATM cells over ethernet physical layer.

AES52-2006 (r2017): AES standard for digital audio engineering – insertion of unique identifiers into the AES3 transport stream.

AES67-2018: AES standard for audio applications of networks – high-performance streaming audio-over-IP interoperability.

DANTE: https://www.audinate.com/node/128.

Digital Cinema System Specification Version 1.3 – approved 27 June 2018. Digital Cinema Initiatives, LLC, Member Representatives Committee.

IEEE 802-1-2011 Standard for local and metropolitan area networks – audio video bridging (AVB) systems.

IEEE 802.3-2018 – IEEE standard for ethernet.

ISO/IEC 7498-1, Second edition 1994-ll-l5: Information technology – open systems interconnection – basic reference model.

ISO 7498-2 Second edition 1994-ll-l5: Information technology – open systems interconnection – security architecture.

ISO/IEC 7498-3 Second edition 1994-ll-l5: Information technology – open systems interconnection – naming and addressing.

ISO/IEC 7498-4 Second edition 1994-ll-l5: Information technology – open systems interconnection – management framework.

RAVENNA: https://www.ravenna-network.com/using-ravenna/overview.

CHAPTER 25

Where to Connect a Meter

CHAPTER OUTLINE

A level meter can be applied at various points in the signal chain. This facility is very useful when optimizing the signal path.

ANALOG CONNECTION

Before the instrument is connected, it is worthwhile to look at a couple of fundamental concepts concerning measurement technique. One must be aware of whether the measurement instrument by accident influences the measurement itself. The keywords here are voltage matching and impedance matching.

VOLTAGE MATCHING

A basic electronic circuit showing the output of a device connected to another device (a level measuring instrument in this case) is lined out in Figure 25.1.

The rule is that the largest voltage lay across the largest resistance or impedance. When measuring the magnitude of the output signal of the "device," then, it is important that the measurement system does not load the output. Therefore, the input impedance must be high compared to the source impedance. A good rule of thumb says that the input impedance of the measurement system must be at least ten times

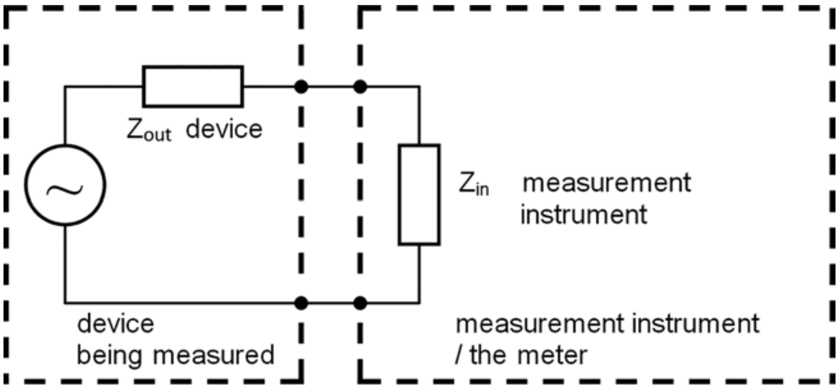

Figure 25.1 The interconnection of two devices is a division of voltage between the source impedance (Z_{out}) and the input impedance (Z_{in}).

that of the output impedance of the device that is measured. This rule applies as long as we are dealing with LF (low frequency). Here, LF means the audio spectrum below 20 kHz; further, the cable lengths must be less than a few hundred meters/yards.

A typical example: The output impedance of a mixer's line output or auxiliary send can be 100–200 Ω. The impedance of an analog input of a measurement instrument is specified to be >20 kΩ. The input impedance is at least 100 times higher than the output impedance; thus, the measurement instrument displays the correct signal voltage on the output concerned.

IMPEDANCE MATCHING

Impedance matching is needed when dealing with AF in cables with lengths of kilometers/miles. Further, the impedance matching is needed when the cables carry RF (radio frequency). In the category of audio RF-signals are antenna signals for wireless microphones, digital audio signals, video signals, and so on. Impedance matching means that the source impedance and the load impedance must be equal. If there is an impedance mismatch, reflections and signal degradation occur in the cable.

The consequence of impedance matching is that the voltage becomes equally divided between the two impedances in terms of voltage. Or, you could say that the equally sized output impedance and input impedance are in parallel, which halves the overall impedance and therefore loads down the voltage measured across the total impedance. In other words, there is half of the voltage (−6 dB) across each (see Figure 25.2). To measure the correct level with a measurement instrument with high input impedance, the output that requires impedance matching must be terminated

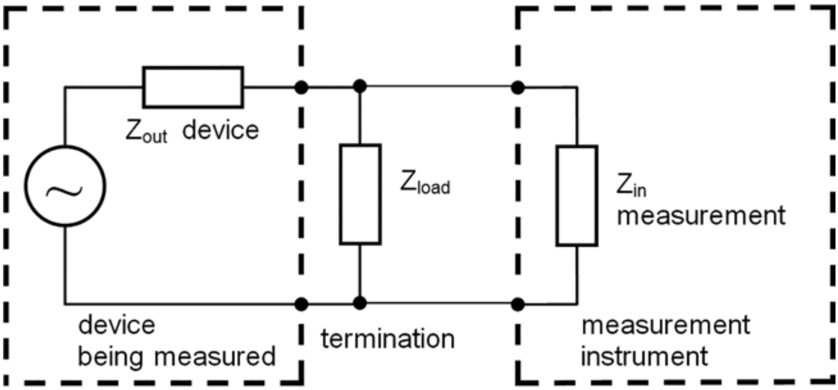

Figure 25.2 Impedance matching: The level is measured correctly when the output is terminated (loaded) with the nominal impedance. Without loading, double the voltage is measured when the measurement instrument exhibits high input impedance.

correctly (i.e., loaded with the nominal impedance). If this is not the case, the measurement instrument shows a value that is up to 6 dB too high. If, for example, an output is specified +4 dBm, it refers to the fact that the signal is 1.23 V, although only when the output has a termination of 600 Ω.

THE MIXING DESK

The need for level measurement arises first and foremost when controlling levels within the sound mixing desk.

Inputs The signals at the inputs are either at microphone level (sensitivity around -60 dBu) or line level (sensitivity around 0 to +6 dBu). Here, an overloading indicator is of primary importance. Drive each channel optimally but without overloading the input preamp.

Master outputs (main out) Here, it is always natural to measure the level depending on the requirements of the connected devices. The input of recording media should not be overloaded; the frequency deviation of an FM transmitter should be kept within certain limits; clipping should be prevented in processors for PA, and so on.

Group outputs (group out), aux and bus, outputs Larger audio mixers will normally either provide meters on these outputs or include the option to allocate one or more meters for this purpose.

Insert points and aux-return Here, it is less common to facilitate metering. It can, therefore, be beneficial to have the instrument(s) connected to a patch panel so that it is possible at any time to monitor whether the return channels run at an appropriate level.

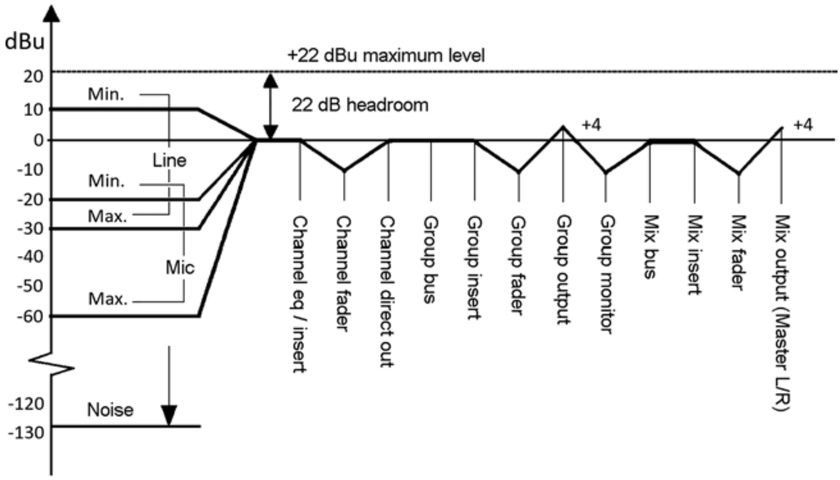

Figure 25.3 An example of a level diagram for an analog sound mixer.

There can be a difference between the nominal levels at different points of the mixer's signal path. Because the level at the insert points may be lower than the group and master outputs, it is a good idea to be familiar with the mixer's level diagram (see Figure 25.3).

OUTPUTS WITH EMPHASIS

Often the level meter is used to measure a signal that subsequently is emphasized (i.e., equalized in certain parts of the spectrum). One example involves the output from a radio station subsequently connected to an FM transmitter. Here, pre-emphasis is performed by raising the treble. On reception, the treble is lowered correspondingly. A form of noise reduction is attained by doing this (see the discussion on emphasis curves in Chapter 9: Frequency Weighting and Filters). A second example involves an audio signal to be transmitted by satellite, where the ITU J.17 standard is applied. Emphasis may even be utilized in digital media. AES/EBU includes the possibility of signaling using two different standards for emphasis, and S/PDIF includes the possibility for one.

Most emphasis curves were developed when it was not particularly common to have extreme modulation in the treble range. Today, the problem can be that the emphasis can still cause the transmission line to be overloaded, precisely because of the raising of the treble. Quite literally, one can risk burning the FM transmitter out or provoke the satellite transmission to drop out.

To protect against accidents, it is thus not just an advantage but instead a necessity to be able to connect an instrument after the emphasis. In the same regard, this also goes for the program limiter (i.e., the limiter normally connected to the output as protection against overloading).

BALANCED/UNBALANCED

To run an electrical signal from one device to another, two wires must be used. Otherwise, a circuit is not established. In a balanced connection, these two conductors are physically identical. The shielding around the conductors can comprise a third wire even though the shielding is not included in the signal path. In an unbalanced connection, there is one conductor involved that is "hot," and the shielding works like the other conductor in the circuit. The shielding and hence the chassis becomes a part of the signal path.

When connecting the meter to the output, it can occur in some instances that the reading is not what you expected. On the following pages, different situations are described.

UNBALANCED OUT/UNBALANCED IN

In an unbalanced connection, the correct level is, in principle, always being measured. There can, however, be unfortunate situations in which there might be problems with low-frequency noise (hum) if both systems are connected to ground via, for example, the power plug. A low-frequency noise loop can arise in the earth/frame/shielding circuit (see Figure 25.4).

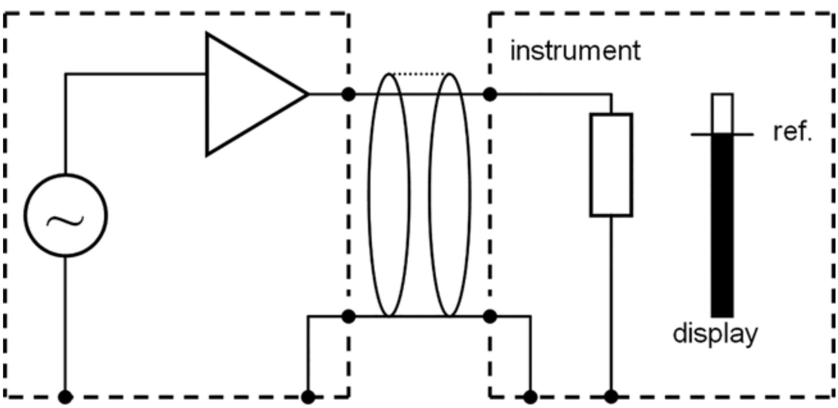

Figure 25.4 Unbalanced output; unbalanced input.

UNBALANCED OUT/BALANCED IN

The connection can be problematic if the pins of the connector are not configured correctly. The aim is, of course, to establish a circuit. Next, the concern is to preserve the balancing on the wires where possible (see Figure 25.5).

If both leads of the balanced connection are not connected, then no circuit is established, and the instrument does not display anything (see Figure 25.6).

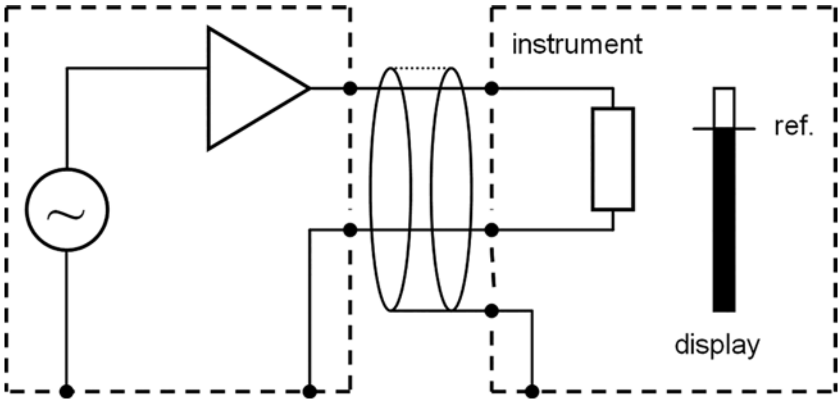

Figure 25.5 Unbalanced output; balanced input. A circuit is established. The signal is present, and the meter reads the level correctly. Note that the shielding is only connected to the receiving device.

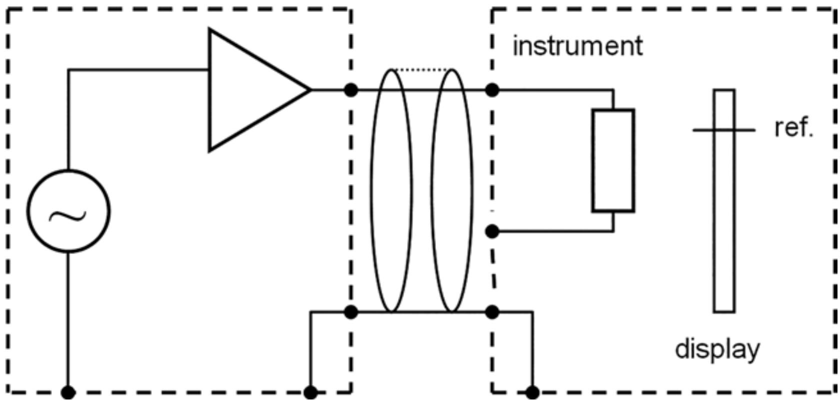

Figure 25.6 Unbalanced output; balanced input. A circuit has not been established. Hence nothing is displayed.

BALANCED OUT/UNBALANCED IN

With a balanced output connected to an unbalanced input, there are several options for errors that relate to the design of the balanced output. For electronic balancing, there are different possible circuits. The balancing may also be attained by using a transformer.

The main problem is the electronic balancing, as shown in Figure 25.7. However, if a signal must go through to an unbalanced input, then the connection must be implemented, as shown. The two frames must not be connected.

In Figure 25.8, the connection goes wrong because the signal is grounded before it reaches the inverter.

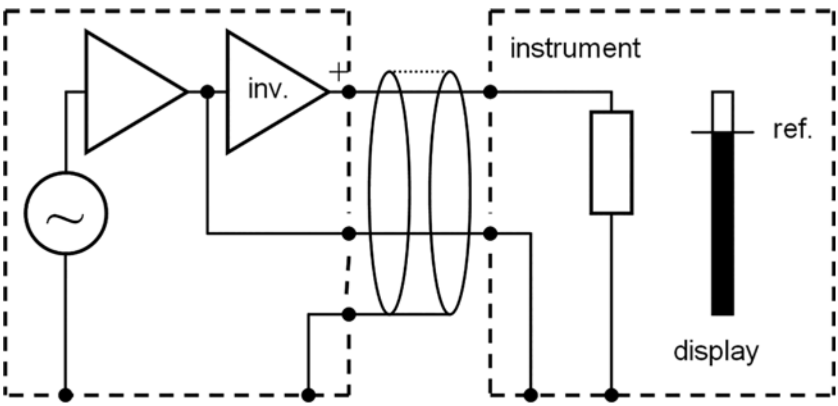

Figure 25.7 "Balanced" output; unbalanced input. The output is not correctly balanced, but the signal is present as long as the two chassis are not connected.

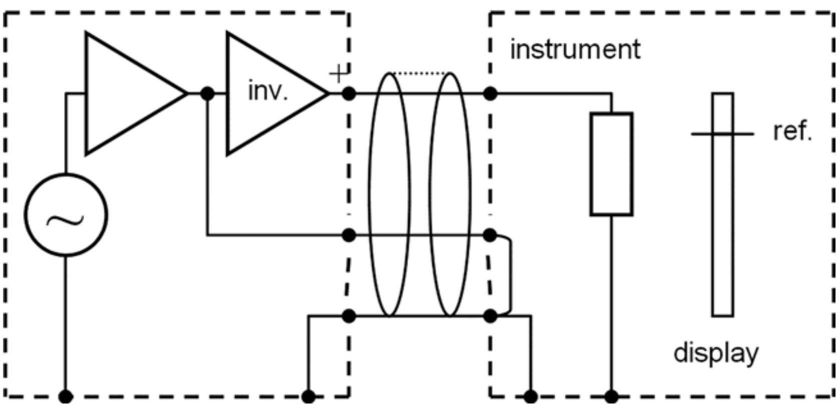

Figure 25.8 "Balanced" output; unbalanced input. No reading! Because of the poor design of the output circuit, the signal is grounded due to connection to the chassis.

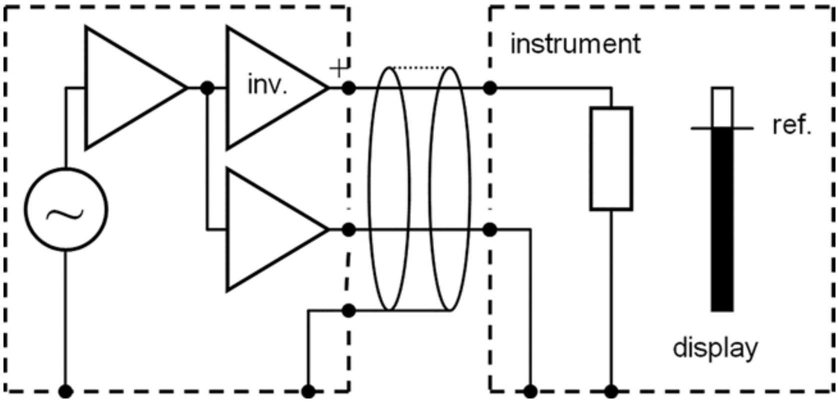

Figure 25.9 "Balanced" output; unbalanced input. The circuit is established, and the display is OK. However, the balancing is referenced to the chassis.

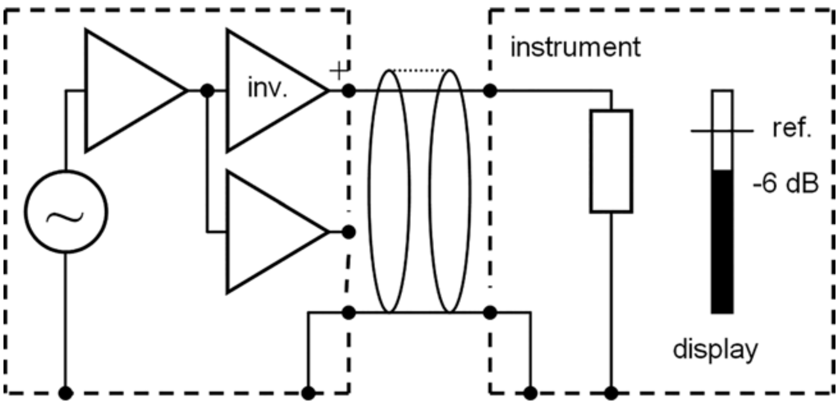

Figure 25.10 "Balanced" output; unbalanced input. Because the output is still referenced to the chassis, a signal arrives at the output, but only half of it (–6 dB)!

Figure 25.9 shows a more common type of electronic balancing.

The following connection (Figure 25.10) is dangerous because there is a risk of only measuring half the signal, which happens if one lead is not connected.

When an output is balanced by a transformer (Figure 25.11), the circuit is somewhat more manageable.

BALANCED OUT/BALANCED IN

When both the input and output are balanced, there is little that can go wrong. The circuit is created independently of whether the shielding is run through or not. It is

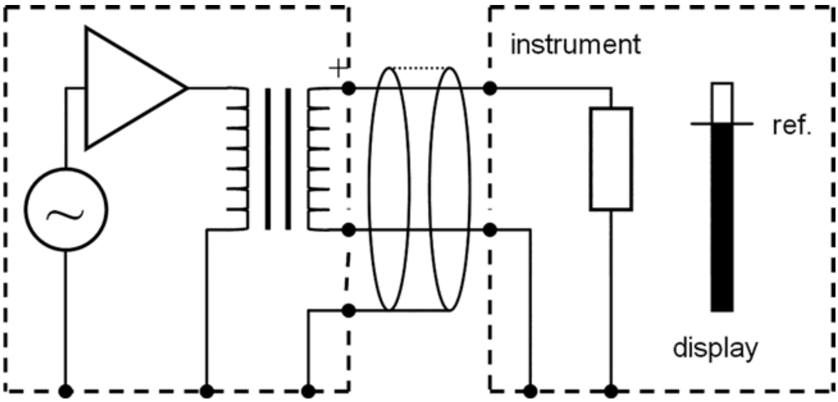

Figure 25.11 Balanced output; unbalanced input.

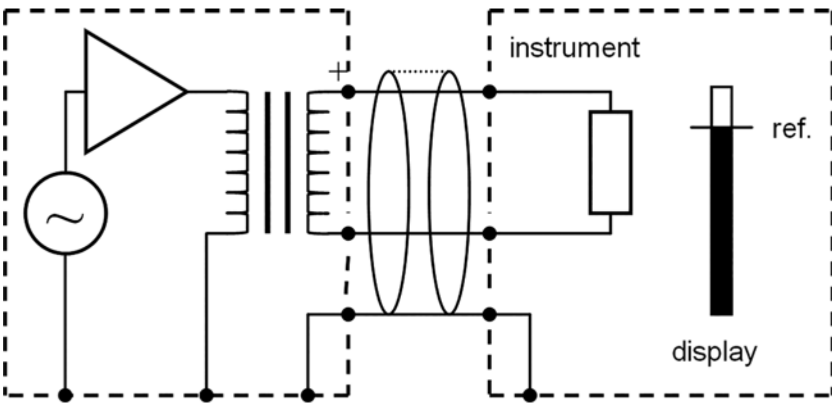

Figure 25.12 Balanced output; balanced input. A circuit is created regardless of whether the shielding is run through or not. The signal is present, and the reading is OK.

then possible to consider ground and shielding strategies without thinking about the signal itself (see Figure 25.12).

CONNECTIONS VIA A JACK FIELD (PATCH BAY)

If you have an installation with a jack field (patch bay), it is always practical to have access to a level meter via that route. The main meter can be fed through the jack field via so-called "half normalization." By this type of connection, the signal commonly routed to the meter is broken when a jack (carrying the external signal) is inserted. Instead, the external signal is routed to the metering device from the output you wish to check.

LIVE SOUND

Regardless of whether you are producing live sound on a PA system or recording "live to hard drive" it is necessary to monitor the sound. This necessity applies in particular to critical sources such as wireless microphones or sources that are extremely variable in level.

It is not just the levels that are kept under control but also the phase relationship between two outputs in a stereo setup, particularly if the signal is played back in mono. A goniometer/phase meter (see Chapter 18: Polarity and Phase Reading) is almost a must-have in this situation.

THE LEVEL METER AS A MEASUREMENT DEVICE

When you have a well-calibrated meter, it is possible to use it as a measuring instrument. As a rule, an instrument that displays RMS can show the level of all signal voltages in the audio spectrum. This application works particularly if the meter has an appropriately high input impedance (>10 kΩ).

One must know the significance of the scale (i.e., whether it is calibrated in dBu, such as with the Nordic scale on a PPM instrument). A level of 0 dBu corresponds to 0.775 volts. Many instruments have the option for amplifying the input signal, typically by 20 dB, which corresponds to amplification of precisely ten times. With the instrument's logarithmic scale, it hence becomes possible to measure signals at less than 1/1000 of the instrument's full scale. If the instrument ultimately is used for larger signals, then an attenuator (of, for example, 20 or 40 dB) can be acquired. By attenuation, it also becomes possible to perform measurements in installations such as 50, 70.7, or 100-volt loudspeaker systems, systems with several loudspeakers on the same string, and where the max SPL is reached when the amplifier delivers the mentioned voltage.

DIGITAL CONNECTION

If you have a digital meter, then there is the possibility of monitoring several functions simultaneously. A digital meter is here understood as some of the level meters mentioned in earlier chapters. A more specific analysis of how a transmission line behaves takes special analyzers. However, analyzing the setting (mono/stereo, sampling rate, bit depth, etc.) and the content of a digital signal is quite useful. It is the meter that the eye first falls on if the digital signals go haywire. A digital instrument is as a rule equipped with an AES3 (AES/EBU) or IEC 60958-3 (S/PDIF) input or is equipped for network connection.

It is important to notice that the meter always should be synchronized with the audio measured. For instance, connecting a digital meter via an asynchronous sample rate converter makes it impossible to quantify true peak levels.

THE METER IS THE SLAVE DEVICE

When connecting it, the meter is designated as a slave. In other words, the meter is clocked from the equipment to which it is connected. The AES/EBU interface is self-clocking. Thus, if the meter is connected, then the instrument can follow along automatically.

It sounds simple, but it can be a problem if you have to use your mixer's only digital output for an external instrument, meaning that only analog signals are available for any other purpose. That, of course, is not optimal. Therefore, it can be quite practical to have a router or a switch, giving the option to make a connection as required. In many smaller studios, however, this is not the type of thing that money is used on first, so other solutions must be found.

An inexpensive variation could be to connect the meter to an S/PDIF coax output. It ought to be possible to get a digital meter to read the data transmitted over the interface, even though the voltage is a little lower than AES/EBU. Certain information – like time code – is lost; however, you still have the information on the level, on the sampling frequency, on the number of bits per sample, on whether the units are locked to each other, and of course the main content, the audio data.

SAMPLE RATE CONVERTER

In everyday terms, the many sampling frequencies applied in practice can certainly be a bit cumbersome. Because of this, a sample rate converter is frequently available. This device allows all arbitrary digital signals to be converted to a common standard.

Where the digital level meter gets its signal from must be known. Some converters are not transparent, and thus not all data come through, even if only changing the sampling frequency. Connecting via the sample rate converter, even when synchronized, changes the (max) sample peak level.

CLOCK DISTRIBUTION

The ideal for a setup with multiple digital devices connected is to have a distributed central clock. There is no guarantee that the level meter has an input for an external

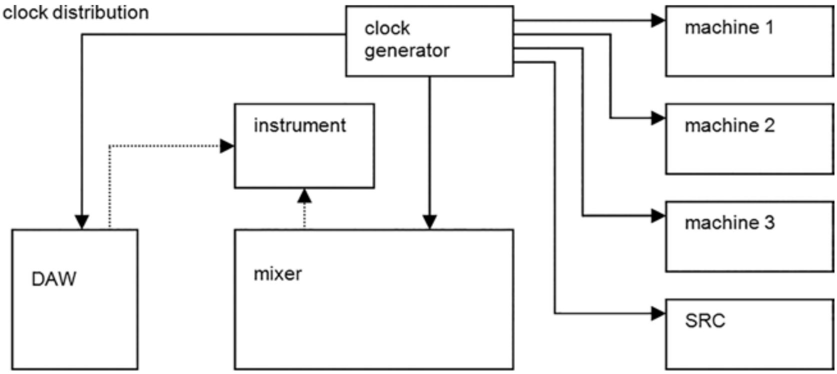

Figure 25.13 Setup of digital equipment with clock distribution. There is a clock generator shared by as many devices as possible, such as a mixer, digital audio workstation (DAW), and sample rate converter (SRC).

clock. However, if the rest of the system runs on the distributed clock, then it can be significantly easier to connect the meter as required (see Figure 25.13).

NETWORKING

Analog audio needs a separate physical circuit for each channel. With digital audio, many channels can be carried in one circuit. The digital interface AES10 (MADI) can carry up to 64 channels. In digital audio networking, it is possible not only to transport audio data but also to address the individual units attached to the network and to circulate lots of control data. Several networks exist for audio – both proprietary and open standards. Among them are dedicated networks like Ravenna, Dante, and AVB.

When connecting to a network, the meter – or the device including a meter – can be attached at any point in the chain. Normally, an IP address is allocated to the meter (see Chapter 24: Audio-over-IP).

SOFTWARE METERS/PLUG-INS

DAW BUILT-IN LEVEL METERS

The standard level meters in digital audio workstations are often peak reading or rather sample-based reading. They may, however, read RMS and further have some extra functionality added, such as peak hold, clip/overload indication, and valleys (the minimum value during a predefined time interval). In some workstations, it is possible to insert the meters pre or post faders.

You must be aware of how (whether) the meter is calibrated and how levels are defined. In some DAWs there are two options for reference, either a full-scale sine wave or a full-scale square wave.

First and most important: The standard is to apply the sine wave as a reference: 0 dBFS means 0 dB RMS of a full scale sine wave! That's a definition. This may not be logical, but this *is* the definition. So, a square wave recorded at an identical peak level as the 0 dBFS sine wave will show +3.01 dBFS!

Let's assume you record a sine wave at −10 dBFS (RMS), and the reference for your scale during recording is a sine wave. The reading of your meter is of course −10 dBFS. Now, if you then change the reference to a full-scale square wave, the RMS level of your recorded sine wave is no longer −10 dBFS but −13 dBFS (RMS).

Is it a problem? Yes. However, the conclusion is that you have to stay with a sine wave and never change reference during production (or never at all). You may end up delivering productions at a wrong level. Also, loudness calculations may be wrong, either internally calculated by native algorithms or by external plug-ins. So: check your DAW.

DAW PLUG-IN METERS

Many software-based meter plugs-ins are available for the DAWs. These meters are usually inserted at the output of the mix. Also, several loudness meters that fulfill the ITU/EBU standards are available.

PROGRAMMABLE PROCESSORS

Freely configurable DSP engines like BSS SoundWeb, Yamaha DME-series, and others may have simple level meters to be applied during programming. These engines provide several inputs and outputs. What is in between is the choice of the user. When setting up a device like that, it becomes handy to be able to follow the signal through the processor. This is typically possible by the aid of simple (peak reading) level meters.

LOUDSPEAKER/HEADPHONE MONITORING

Monitoring audio is not only a question of using visual displays. Listening is just as important – at least! So it is very practical to establish a relationship between reading and the listening level. Further, it is important that the monitoring SPL of speakers or headphones is well defined and calibrated. The calibration secures the right balance during mixing and keeps levels from damaging hearing.

Regardless of the type of signal distributed, analog or digital, a monitor controller with a built-in meter calibrated to a reference listening level is an excellent device

Figure 25.14 An example of a combined D/A converter, meter, and monitor control unit for loudspeakers and headphones. Digital inputs are S/PDIF, AES3, TOS, and ADAT (including the ability to confirm whether those inputs are synchronous or not). (TC Electronic, BMC-2).

in all installations. Experience shows that less attention has to be paid to the visual meter when the engineer or editor is confident with the listening level.

Further, using excessive level monitoring may lead to listener fatigue and eventually to hearing loss. These are reasons for keeping SPLs in control. An example of an monitor control device is shown in Figure 25.14).

REFERENCE

AES17–2015: AES standard method for digital audio engineering - Measurement of digital audio equipment.

C H A P T E R 26

FFT, Fast Fourier Transformation

CHAPTER OUTLINE

FFT is an acronym for "Fast Fourier Transformation," named after the mathematician Jean-Baptiste Joseph Fourier. The purpose of a Fourier Transformation is to transform waveform (impulse response/waveform) into frequency; a methodology implemented in many kinds of audio analysis.

BACKGROUND

Jean-Baptiste Joseph Fourier (1768–1830) was a French mathematician who became one of the founders of mathematical physics. He developed the so-called Fourier series, which apply for the mathematical treatment of periodic functions. John W. Tukey and James W. Cooley from IBM developed the FFT algorithm. It was originally designed to perform computations on stored data. As the operating speed of signal processors gradually increased, it also became possible to perform FFT analysis in real time.

PERIODIC SIGNALS

The basis for the Fourier transformation is the fact that an infinite series of harmonics can describe every (infinite) periodic signal (see Chapter 4: Signal Types).

By the transformation, the signal resolves into a set of constituents based on the fundamental frequency. Each of these constituents contains information on both amplitude and phase. The periodic signal has a discrete frequency spectrum, where the distance between each of the lines in the spectrum corresponds to the fundamental frequency (see Figures 26.1 and 26.2).

However, not all signals are periodic. Nonperiodic signals do not repeat themselves. One can, however, establish the convention that the periodic signal constitutes a single period within the space of time under consideration. As a basis, the space of time under consideration is infinitely large. This assumption, in turn, leads to the

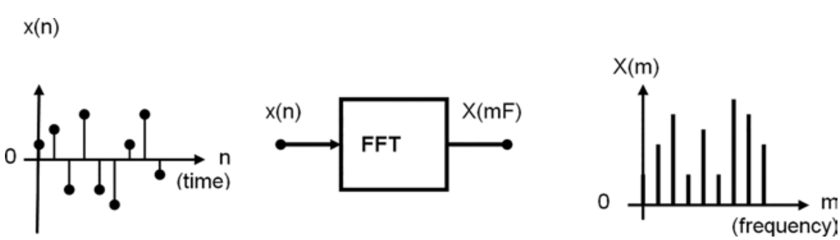

Figure 26.1 Principles of spectrum analysis performed using FFT.

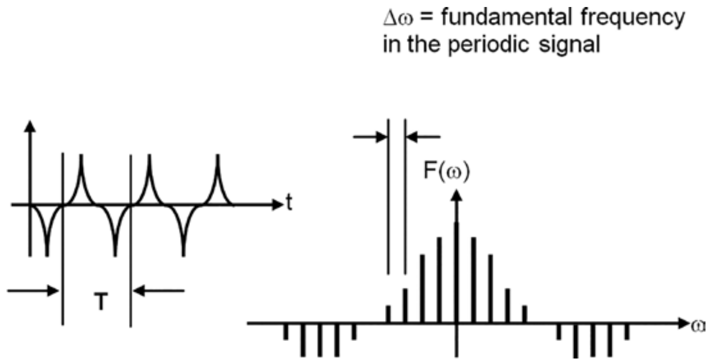

Figure 26.2 Left: Periodic signal. Right: Discrete spectrum of the periodic signal. The frequency distance (Δω.) is then 1/T, where T is the period for the fundamental frequency.

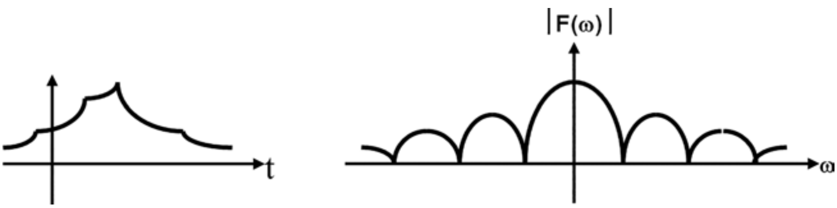

Figure 26.3 Left: Nonperiodic signal. Right: Continuous frequency spectrum for a nonperiodic signal.

distance between the frequency lines in the corresponding spectrum being infinitely small. It is expressed like this:

$$\boldsymbol{\Delta\omega} = 1/\boldsymbol{T}$$

where

$\boldsymbol{\Delta\omega}$ = the fundamental frequency

$\boldsymbol{T}$ = period

Hence, if $\boldsymbol{T} \to \infty \Rightarrow \boldsymbol{\Delta\omega} \to 0$

It thus no longer involves a discrete spectrum but a continuous spectrum instead (see Figure 26.3).

FFT ANALYSIS

FFT is in principle a sampled version of the continuous spectrum. What is most significant about FFT is that the calculation can be performed rapidly when, in each block

Table 26.1 Relationship between the choice of sampling frequency and the number of samples (fft size) for a given resolution in the spectrum.

Sampling frequency	Number of samples (FFT size)	Frequency resolution	Duration
44,100 Hz	2048	21.5 Hz	46 ms
44,100 Hz	1024	43.0 Hz	23 ms
22,050 Hz	2048	10.7 Hz	93 ms
22,050 Hz	1024	21.5 Hz	46 ms
11,025 Hz	2048	5.4 Hz	186 ms
11,025 Hz	1024	10.7 Hz	93 ms

of samples (also called a frame), 2^n samples are selected, where n is an integer. Thus, the number of samples can be 2, 4, 8, 16, . . . 1024, 2048, and so on.

The sampling rate is, of course, crucial to the frequency resolution of the signal (see Table 26.1). Thus, in practice, a sampling frequency is chosen, providing the fundamental sampling rule is adhered to (the rule that the sampling rate must be at least two times the highest frequency that one desires to reproduce).

It is the combination of the sampling frequency (sampling rate) and the number of samples (FFT size) in a frame that determines the frequency resolution of the spectrum. The following expression describes the relationship:

$$\mathbf{f_{res} = f_s/N}$$

where

$\mathbf{f_{res}}$ = frequency resolution in Hz

$\mathbf{f_s}$ = sampling frequency in Hz

$\mathbf{N}$ = the number of samples (FFT size)

In practice, this means that details in the spectrum between the relevant frequencies are not shown. This phenomenon is called "picket fencing" (i.e., equivalent to viewing the world through a picket fence where you can only see between the slats). In the spectrum, all frequencies can be represented by the lines displayed, but if a more detailed display is desired, then a higher resolution must be applied in order to obtain fewer cycles between values. Changing the resolution of the display is done by reducing the sampling frequency (but also the upper boundary frequency) or by increasing the number of samples per frame (FFT size).

In FFT analysis, there is always the same distance in cycles between the individual lines in the spectrum. Since it is usually preferred to view a frequency scale with a logarithmic axis, this means that the lines of the spectrum are shown increasingly closer to each other at higher frequencies (see Figures 26.4 and 26.5).

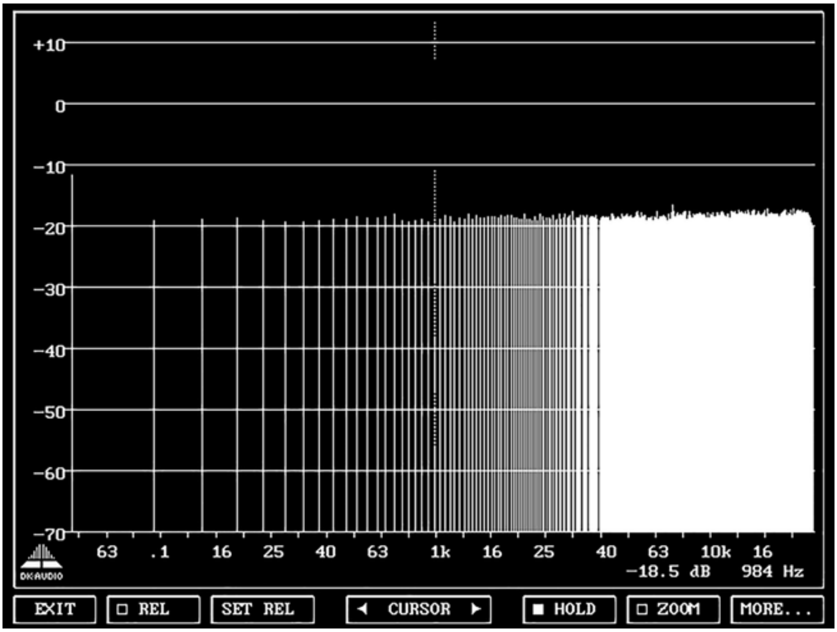

Figure 26.4 White noise is shown here in the form of an FFT spectrum. All lines are of the same height because white noise has constant energy per Hz. The lines are closer together at higher frequencies because the scale is logarithmic.

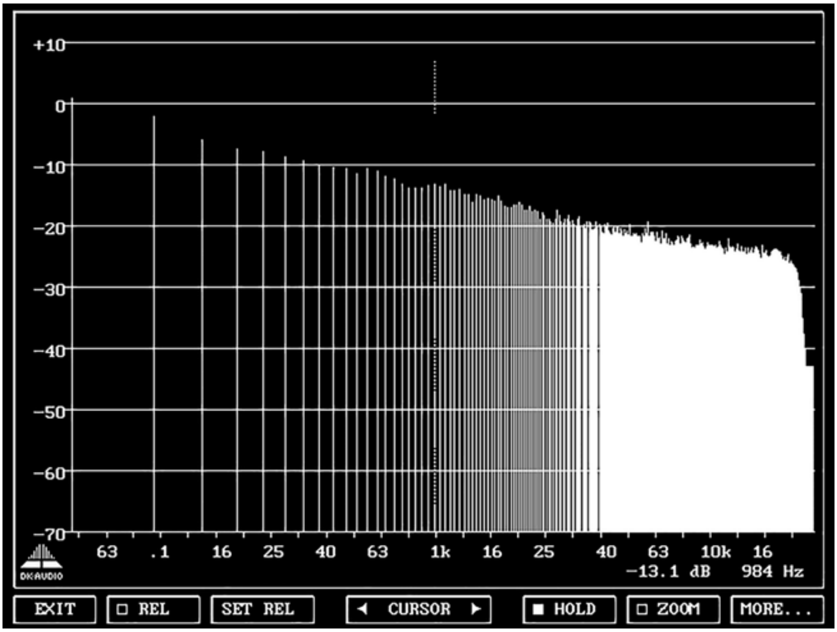

Figure 26.5 Pink noise is shown here in the form of an FFT spectrum. The height of the lines decreases toward higher frequencies because the energy in the signal is decreasing toward higher frequencies. However, the signal has constant energy per octave or decade.

DATA WINDOW FUNCTIONS

When performing an analysis of an extract of a signal of longer duration, errors arise in the analysis if a "transition softening" is not performed between the boundaries. One could call this a kind of "fade in" and "fade out" and compare it to normal editing: If you cut out a chunk of sound and play it back, you may hear a "click" at the beginning and the end (if the cuts were outside zero-crossings). The clicks are errors created by the hard cuts. By adding fades, these errors disappear.

This operation is called a data window function. Each sample in the selected portion of the signal (each frame) is multiplied by the corresponding value in the data window function.

The result of a "raw clip" is shown in Figure 26.6. All samples in the data window function have the value 1. This window function is also called the rectangular window. It can be seen that the frequency function on the right includes a lot of side lobes. These lobes come into view as frequency components in the analysis performed, even though they are not present in the signal that is being measured. These frequency components arise when considering a section or extract of a continuous signal at full amplitude across the entire length of the selected section.

Other types of calculated windows can suppress these phenomena to an appropriate extent. Among them are the triangular (Bartlett), Hanning, Hamming, Blackman, Blackman/Harris, Kaiser, Parzen, and Welch. Each of these types has its advantages.

If there are doubts about which data window function to choose, then a good basic one is either the Hamming or Hanning. Figure 26.7 shows a Hamming window.

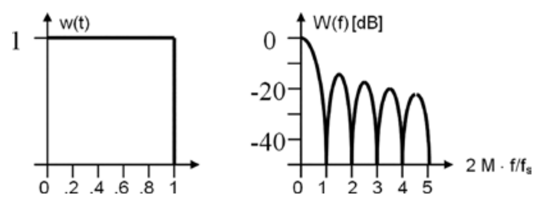

Figure 26.6 Left: Rectangular window. Right: The resultant amplitude spectrum with strong side lobes, the first at –14 dB.

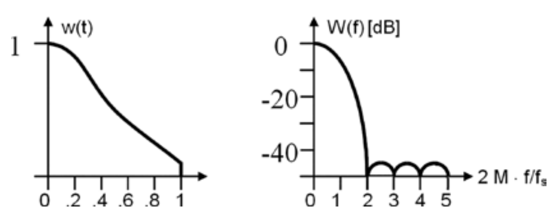

Figure 26.7 Left: The Hamming window. Right: The resulting amplitude spectrum with attenuated side lobes, the first at –43 dB.

Please note that the individual side lobes have a far lower level here than was the case with the rectangular window.

OVERLAP

In certain analyses, in addition to the frame concerned, a little of the prior frame is also included. This operation is done for the sake of continuity. It is called overlap. A percentage specifies the degree of overlap. For example, 50% overlap means that those samples that were located in the last half of the prior frame are included in the current computations. The overlap is in particular applied in analyses where a contiguous course of events is to be analyzed such as spectrograms for the analysis of speech. The overlap is also used in the zoom function to provide greater resolution at low frequencies (Figure 26.8).

OTHER ANALYSES BASED ON FFT

Most audio analyses are or can be, based on FFT. Even though FFT provides linear frequency resolution, FFT applies to logarithmic frequency divisions with a relative bandwidth, such as octave analyses. Here, the measured linear values are converted, ensuring at the same time that there are a sufficient number of points available for each frequency band.

INVERSE FFT

It is also possible to invert an FFT, then called IFFT. A starting point is taken here in the frequency content of the signal, and the result is the signal's time function. IFFT is used to implement digital filters in which the process first involves an FFT (transforming from the time domain to the frequency domain). Once in the frequency domain, the signal processing operations are performed modifying the spectrum. Subsequently, an IFFT is performed by which the signal returns to being a function of time.

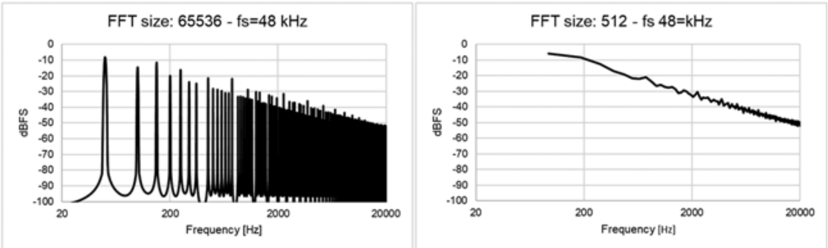

Figure 26.8 Low frequency resolution depending on the FFT-size. Left 65536 samples, right 512 samples (Adobe Audition 2020).

DIGITAL FILTERS

With digital filters, it is possible to implement functions not realizable with analog devices. Among other things, these filters can be made phase linear in contrast to analog filters. It is also possible with digital technology to go much further, such as in the form of adaptive filters. Based on statistical methods, the coefficients of the filters can be determined on a running basis using the FFT analysis of the incoming spectrum.

CHAPTER 27

Spectrum Analyzer

CHAPTER OUTLINE

In audio engineering, it has always been important to be able to describe the frequency content of a signal. A spectrum analyzer is an important and widely used tool for this purpose.

BASIC ARCHITECTURE

The traditional (analog) method was to use one filter bank containing several filters with relative bandwidth. Commonly applied filter bandwidths were (and are) 1/1 octave, 1/2 octave, 1/3 octave (see Chapter 9: Frequency Weighting and Filters), or perhaps with narrower bandwidths depending on the purpose. The signal to be analyzed passes through one filter at a time. Once the detection in one filter is completed, its value read or printed, the process would move on to the next signal. This was the process in the old days, and this is the procedure if you have a DAW and you are able to create filters that meet the requirements from the measurement standards (like FFT filters or Scientific Filters). Then you'll have to perform one filtering at a time, read the level of the filtered file, undo the filtering of the signal, and move on to the next frequency.

A more practical variant was the so-called parallel analyzer, where each filter had a detector. With this, the values were read simultaneously throughout the complete frequency range. Figure 27.1 shows a basic block diagram of the parallel analyzer, in which the same signal runs through many filters simultaneously.

Today the older analog technologies by and large are substituted by digital realization. The measurement results are essentially the same. However, the filtering of the signal is based on FFT technology and then recalculated to obtain the relevant (relative) filter bandwidth.

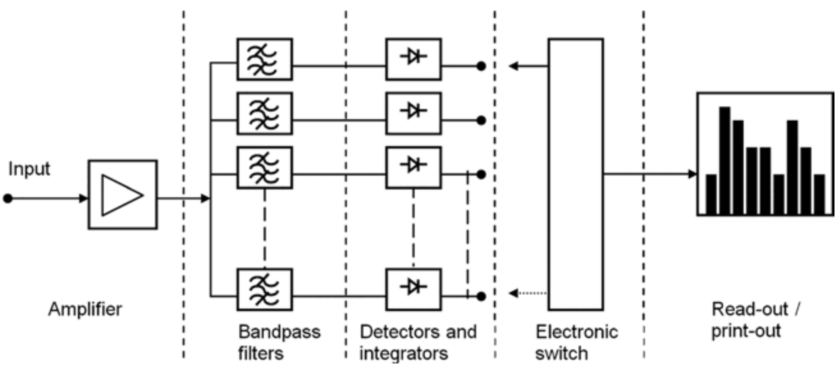

Figure 27.1 Principles of the parallel analyzer. The filters have a relative bandwidth, such as 1/3 octave.

BANDWIDTH

Filters with relative bandwidth especially find their way into practical sound engineering when the measurement relates to reproduced sound. The hearing more or less perceives sound in logarithmically spaced frequency bands of relative bandwidth. In many control rooms and PA systems, it is normal to use analyzers based on 1/3 octave. It is a classical way to measure or monitor loudspeakers' frequency responses.

In room acoustics, the specification of materials' absorption coefficients applies 1/1 octave band. Hence, it is relevant to measure the reverberation time of a room in octaves. Alternatively, an average of the values in three 1/3 octaves can be calculated to obtain the octave-band value (see Figure 27.2).

When analyzing background noise in rooms and presenting the results in the form of noise rating (NR) or noise criterion (NC) curves, 1/1 octave-band analysis always applies.

ANALOG OR DIGITAL

Due to the advantages of digital technology, largely no analog filters apply to measurements (although they still do for creative sound design). As mentioned above, most filters used for measurements build on FFT analysis. The result of the constant bandwidth spectrum is recalculated to achieve the relative bandwidth.

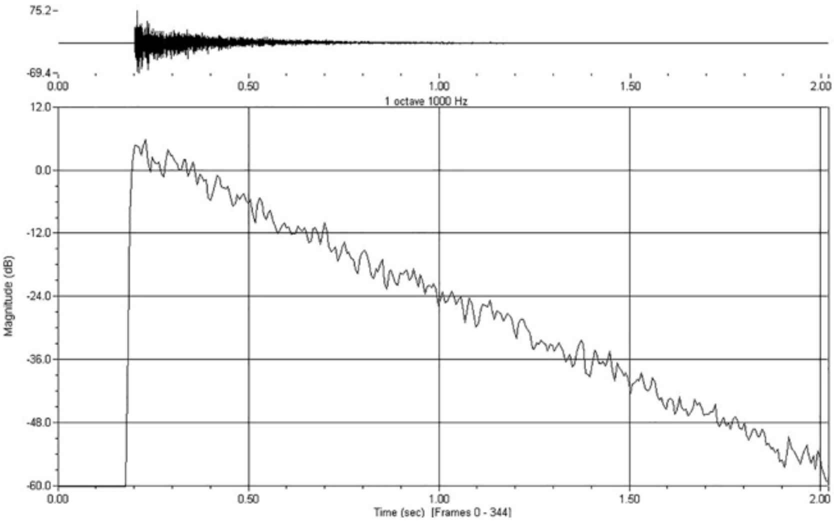

Figure 27.2 The measurement of a room's reverberation time can be performed by recording an impulse like the popping of a balloon in the room. Subsequently, the decay is analyzed in octave or 1/3 octave bands. Top graph: The recorded waveform. Bottom graph: The signal transformed to level (dB-scale), and the slope expresses the decay per time unit, the reverberation time.

One can, however, due to the technology applied, find octave band or 1/3 octave-band analyzers that have different filter slopes (skirts) to work with: One corresponds to the ISO standard for measurement filters. The other represents state of the art, with much sharper digital filters if compared to conventional analog filters.

It is quite common that the more advanced sound level meters have built-in spectrum analyzers. Also, many level meters for audio production may include a 1/3 octave-band analyzer (see Figures 27.3, 27.4, and 27.5).

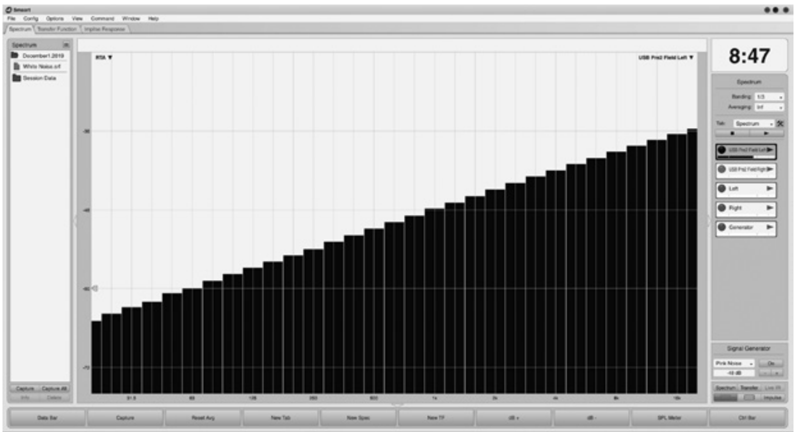

Figure 27.3 The spectrum of white noise analyzed in 1/3 octave bands is shown here. White noise has constant energy per Hz. However, the filters have a constant relative bandwidth. In this way, the frequency span of each filter gets wider with frequency, and each column shows a higher level (a higher frequency band "contains" more frequencies). Analyzer: Smaart Live, version 8.

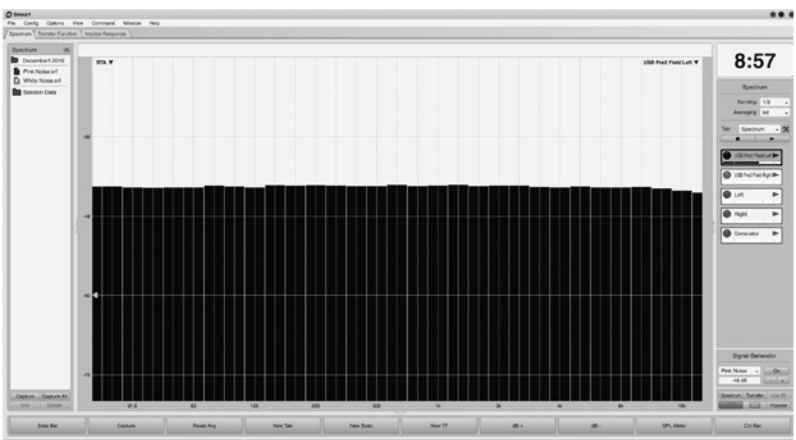

Figure 27.4 The spectrum of pink noise analyzed in 1/3 octave bands is shown here. Pink noise has constant energy per octave; hence, the flat horizontal curve formed by the columns. Analyzer: Smaart Live, version 8).

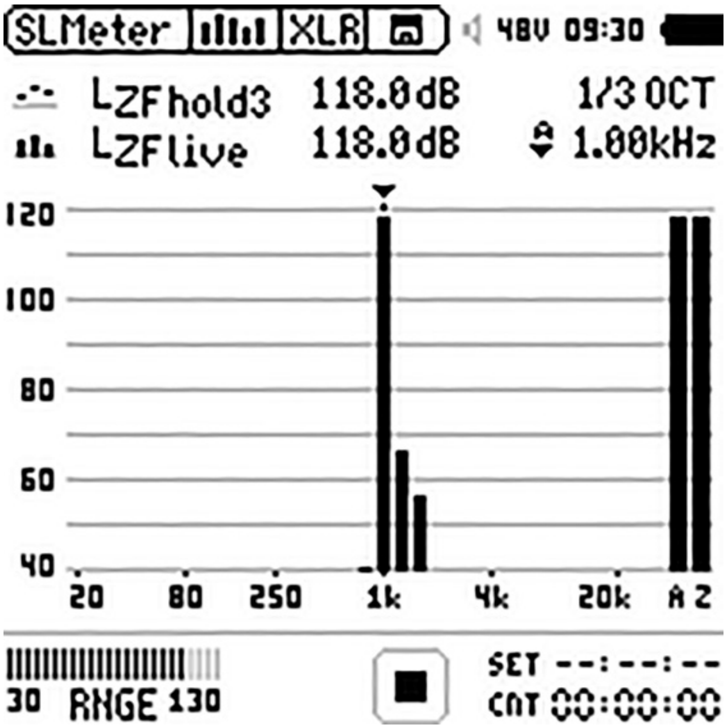

Figure 27.5 Analysis of a 1000 Hz sinusoid using a 1/3 octave analyzer. This bandwidth is obtained using a conversion from an FFT spectrum. Regardless of whether the analysis originates from computed FFT filters or analog filters, there is some form of "skirts" on the sides when analyzing pure tones. (Screenshot from NTI XL2 Analyzer)

PREFERRED FREQUENCIES

The related center frequency names the octave or 1/3 octave bands. The standard ISO 266:1975 specifies the frequencies in the relevant range. These values apply to analyzers and graphic equalizers as well. 1000 Hz is the starting point, and the center frequency values are calculated from that (and rounded) (Table 27.1).

TRANSFERRING DATA

For documentation purposes, it is practical (read: time-saving) to transfer the results from a measurement on digital form. Data are transferred as screen dumps (see Figure 27.6). It is fast and easy. However, it is not possible to work with the data (averaging, statistical calculations, etc.) Thus, it is very common to transfer data

Table 27.1 Preferred frequencies for measurements as per ISO 266:1975.

Preferred frequency [Hz]	1/1 oct	1/3 1/5 oct	Preferred frequency [Hz]	1/1 oct	1/3 1/5 oct	Preferred frequency [Hz]	1/1 oct	1/3 1/5 oct.
16	x	x	200		x	2500		x
20		x	250	x	x	3150		x
25		x	315		x	4000	x	x
31.5	x	x	400		x	5000		x
40		x	500	x	x	6300		x
50		x	630		x	8000	x	x
63	x	x	800		x	10000		x
80		x	1000	x	x	12500		x
100		x	1250		x	16000	x	x
125	x	x	1600		x	20000		x
160		x	2000	x	x			

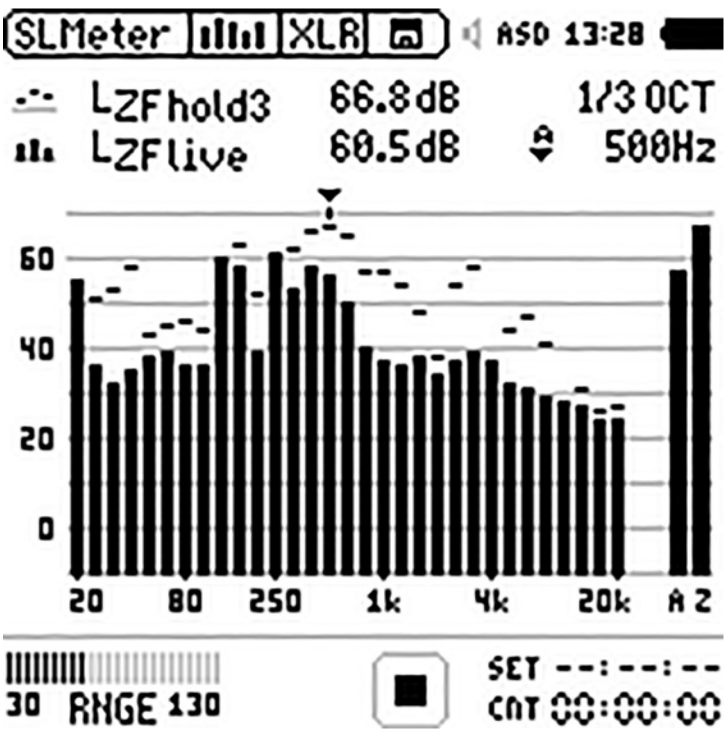

Figure 27.6 Screen dump from an audio analyzer, NTI Audio XL2, handheld audio analyzer.

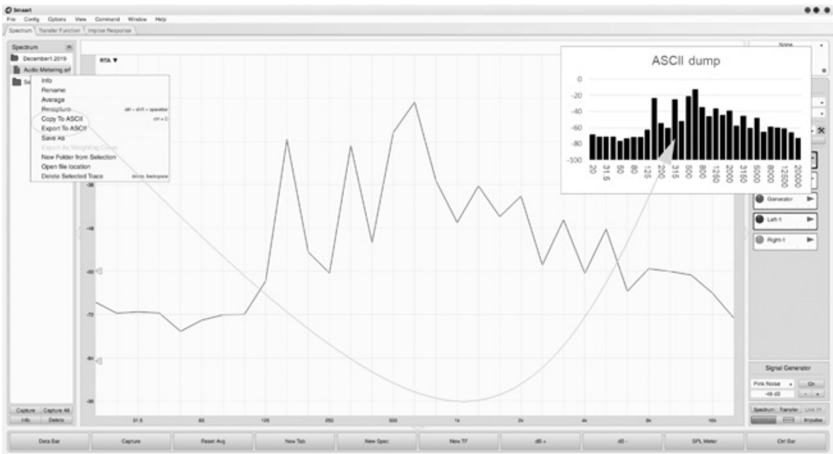

Figure 27.7 Stored spectrum, transferred to spread sheet. Smaart Live v8, ASCII dump.

from measuring device to a spreadsheet for further treatment and presentation (see Figure 27.7).

REFERENCE

ISO 266: 1975 Acoustics – Preferred frequencies for measurements.

CHAPTER 28

Other Measurement Systems

CHAPTER OUTLINE

Spectrum analyzers are very handy. However, more sophisticated measurements are often needed. For that purpose, there are several advanced, commercially available, and easy-to-use software packages to consider. Essentially, all apps run on computers with a decent front end, a sound card.

IMPULSE RESPONSE (IR)

An impulse is a sound of short duration but with the widest possible frequency content. Thus the impulse *response* is how a device, sound system, or room reacts to that impulse.

The impulse response of an acoustical space – or a transducer – provides characterization data of the objects measured. The impulse response contains information that can be "unwrapped" by different transformations of the impulse recorded. For instance, adding a Fast Fourier Transformation (FFT) to the impulse provides a frequency response.

Ideally, the impulse generated is infinitely short. However, in practice, the response time of the object measured may determine the kind of impulse required.

Analyzing the impulse response of a room provides information on the reverberation time of that room. The reverberation time is counted in seconds. The impulse itself is not that critical, except it must contain a sufficient amount of low-frequency energy. A popping balloon often is suitable for the measurement of reverberation in rooms not too big. If you want to study individual reflections in the room, the impulse must be rather short and well defined.

If the object is a microphone, the response time is measured in fractions of a millisecond; then a spark generator may seem like a better choice for that measurement. Sparks may be short and loud, but the energy is limited, and their repetitive precision often depends on humidity and other factors in the surrounding environment.

In room acoustics and sound system control, more "handy" and steady methods are preferred. Then loudspeakers are applied for the room excitation – or the loudspeakers are already a part of the system investigated. The application of loudspeakers may call for other signals than just a "big bang," as there is a need for a sufficient amount of energy in a broad frequency band. This "other" signal is then closely associated with the methodology applied. One advantage of using signals of longer duration is the possibility of emitting more sound energy into a space or device.

For the measurement of the impulse response, different signals and related methods may apply, such as MLS (Maximum Length Sequence), IRS (Inverse Repeated Sequence), Time-Stretched Pulses, and various forms of SineSweeps [1]. These

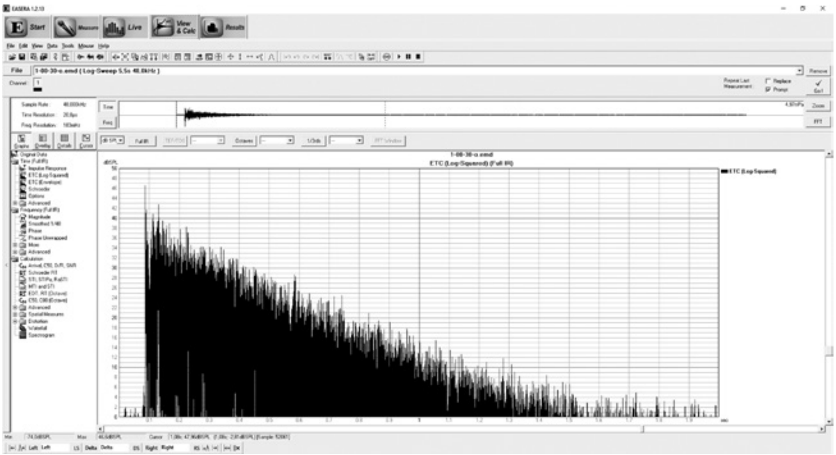

Figure 28.1 Impulse response as calculated from a repeated frequency sweep emitted into the room from a loudspeaker. Software: EASERA by Ahnert Feistel Media Group [2].

technologies take advantage of signals of longer duration, typically in the range of several seconds. It is a little like stretching a rubber band. When stretched, it is very long. In the associated analysis, it is like the rubber band getting released. It is recalculated to become a very short impulse.

A program like EASERA (see Figure 28.1) allows you to create a wav file containing an impulse, originating from a frequency sweep or one of the other signals implemented in that system.

MLSSA/MLS

MLSSA is pronounced "Melissa" and is an acronym for Maximum-Length Sequence System Analyzer. It is a system for acoustical analysis of rooms and electroacoustical components and systems. In contrast to other methods, MLSSA may apply to measurements over a large frequency range and a long time span at the same time.

MLSSA, as well as many other systems, is based on the MLS (Maximum Length Sequence) method. The excitation signal is a periodic sequence of pseudorandom binary digits. A measurement system is typically based on a sound card in a computer. Measurements can be performed very fast, and various post-processing options for the measurement data can provide information concerning impulse response, the reverberation time in a room, frequency response, and speech intelligibility for electroacoustic systems. MLS-based measurement systems can also emulate more traditional systems, such as the 1/1 octave or 1/3 octave analyzer.

One example of a popular MLS software is WinMLS developed by Morset Sound Development.

TRANSFER FUNCTION

Some measurement systems are based on two-channel input. This can be for synchronization purposes, or it can be for the analyses of differences between two points in a sound chain. The measurement of the transfer function (TF) is a measure that requires two inputs.

The TF-measurement is originally related to the analyses of electrical circuitry from input to output. However, with the rising interest in highly sophisticated tools for PA and SR systems, the measurement of transfer function has become quite common as a tool for controlling reproduced sound in the listening area. The idea is to measure the signal on its way out of a mixing console and in one or more listening positions after passing processing, amplifiers, loudspeakers, and the room. The analysis displayed is the difference – or the transfer function in the frequency domain.

Originally Meyer Sound introduced the SIM system (Source Independent Measurement), a "heavy" processing tool that includes ten FFT analyzers working in parallel, one for each octave band. This system provided new possibilities for the PA/SR business. Later, many of the ideas from SIM were implemented in Smaart Live, a software that today is published by Rational Acoustics, which also leads continuous development of the software, which runs on a single laptop with an associated sound card [3].

It is possible to use (almost) any kind of signal for the measurement. It just needs to have a frequency content that covers the relevant range. Hence, a flat magnitude curve indicates that there is no difference between the two measuring points, no matter the spectrum of the signal transferred. The advantage is that, for instance, the live music reinforced by a PA system is sufficient as a measurement signal.

The output must be aligned to the input; comparing signals requires the signals to be synchronized in time. Two impulse responses derived from the two inputs provides a precise indicator for this.

In a concert situation, the receiving microphone(s) hear not only the signal from the loudspeakers but any background noise, such as a cheering audience, traffic noise, and reverberation. Thus a coherence function is added. Essentially, coherence on a scale – from 0% to 100% – indicates how much transmitted sound is recognized on the receiving side by the microphone. The coherence must be high enough to trust the measured signal.

When time-aligned, it is also possible to extract the phase difference between the signals.

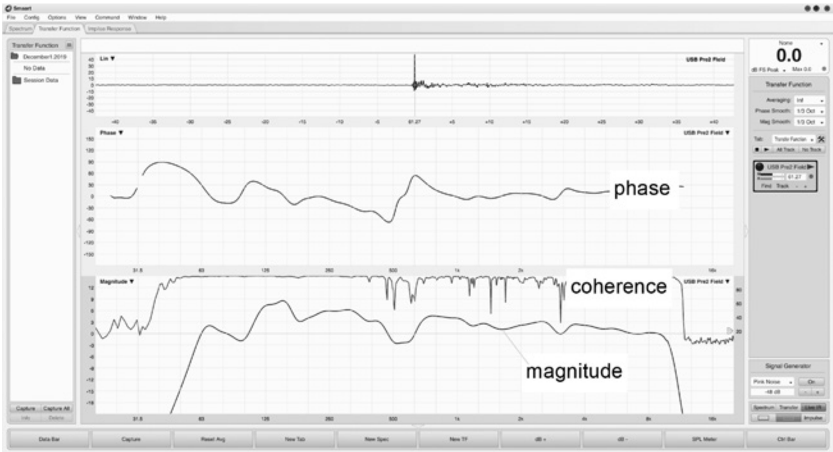

Figure 28.2 Transfer function window, Smaart Live. The upper part is the impulse provided for the automatic alignment of the signals. The next part is the phase vs. frequency. The lower window is the transfer function and on top of that the coherence function, which indicates little coherence below 100 Hz and above 12 kHz. (Courtesy of Rational Acoustics)

An example is of a measurement is shown in Figure 28.2.

SPECTROGRAPH

The spectrograph is an instrument – or software package – that can simultaneously display three parameters or dimensions of an audio signal. It shows time on one axis (the x-axis) and frequency on another axis (the y-axis), and then the degree of density indicates the level. It's time domain and frequency domain at the same time. Almost all spectrographs are based on FFT analysis and run on computers even though the concept has an impressive analog past.

The spectrograph is used in particular for voice analysis (both humans and animals), utilizing filters with both narrow and relatively large bandwidth. However, the spectrographic display (b/w or in colors) also forms a basic editing window for "spectral editing," where specific frequency components can be altered without affecting other parts of the frequency range.

Figures 28.3 and 28.4 show analyses of the author's utterance: "Audio Metering," applying different filter bandwidths.

There are several software apps available on the market, even as freeware: Praat by Paul Boersma and David Weenink [4] and Wavesurfer by Jonas Beskow and Kåre Sjölander [5] being the most popular.

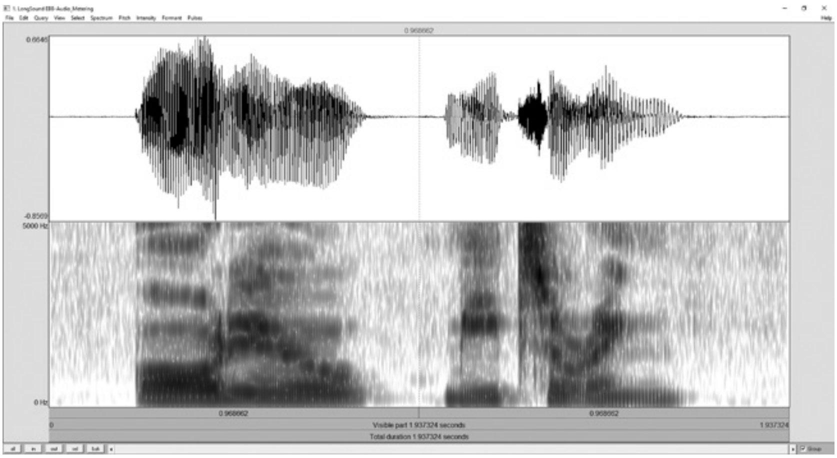

Figure 28.3 Spectrogram of the author's utterance: "Audio Metering." The horizontal axis is time, and the vertical axis is frequency (linear scale), whereas the degree of density indicates the level. A wide filter of 100 Hz was used, so that the individual harmonic overtones in the voice are not seen, but rather the formants, which are comprised of multiple harmonic overtones within a certain frequency range (Praat).

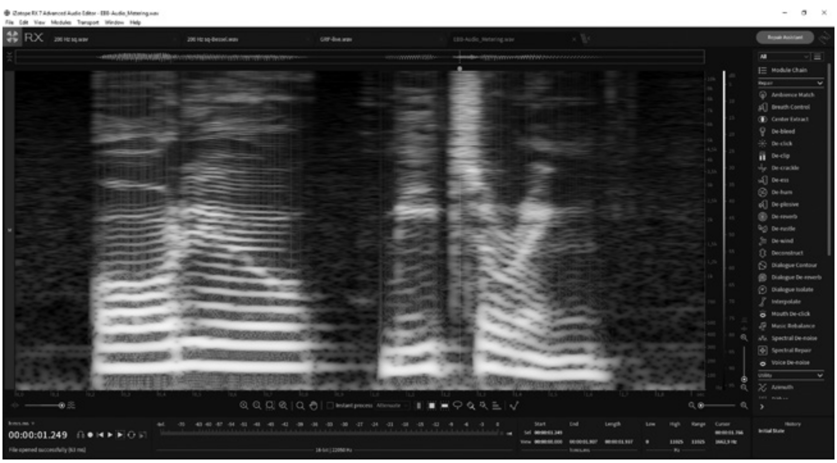

Figure 28.4 An analysis of the same sequence as in the preceding figure is shown here, but a narrow filter has been used. It now shows a plot of the overtones instead. The analysis was performed with iZotope RX 7 Audio Editor.

TRANSIENT ANALYSIS

A very large part of the current measurements and analyses of sound involve finding the RMS values of the signals, either in the entire spectrum or some selected frequency components. The signals are regarded as constant within the period concerned.

Nevertheless, events can occur in the acoustic sound that is audible but difficult to measure using the traditional methods. Examples of this are things like a single click or "glitch" in the sound due to erroneous sampling, drop-out, or clipping. There can also be mechanical rattle from objects in the listening room that are set in motion by the oscillations of the loudspeakers, or by the loudspeakers themselves. Due to the moving coils scraping against the magnet, loudspeakers produce some audible, but not easily measurable, noise. Transient analysis is the tool that makes a large number of these phenomena "visible." Instead of measuring the RMS value of the signal, the signal's instantaneous change in energy is measured.

This technology applies even to audio forensics when looking for possible edit points in the process of the authentication of digital audio recordings (see Figure 28.5).

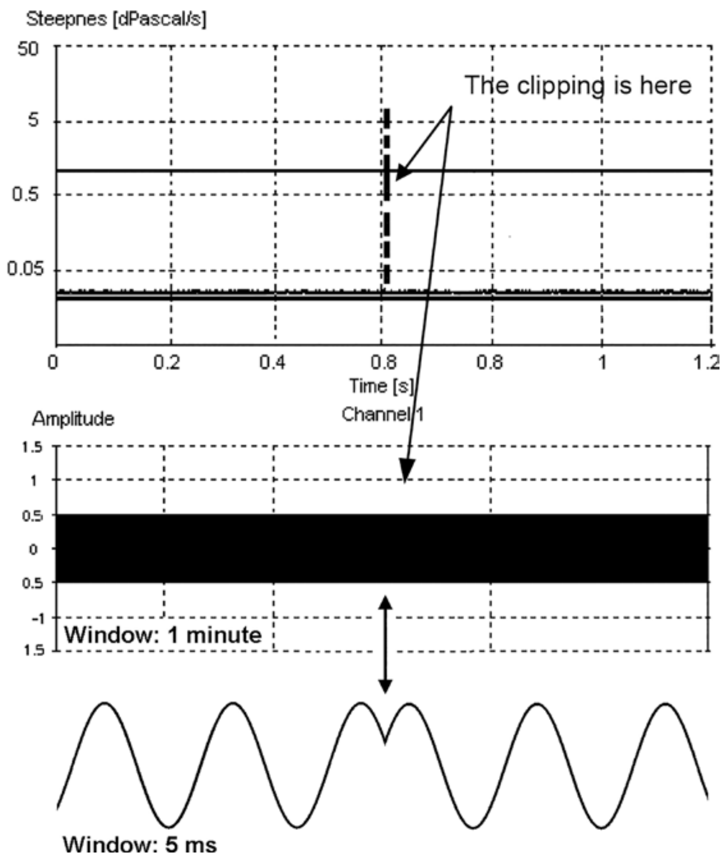

Figure 28.5 Example of transient analysis of a sinusoidal tone where an edit has occurred. The analysis clearly shows where this audible click occurs, even though it cannot be seen on the time signal unless specifically zooming in. The analysis was performed with Harmoni™ Lab., a technique now adapted by NTI Audio.

SPEECH TRANSMISSION INDEX (STI)

At the beginning of the 1970s, it was proposed that MFT, Modulation Transfer Function, be used as a method to describe the intelligibility of speech in a room. MFT expresses an apparent signal/noise ratio since not just background noise but also echoes and the reverberation of the room is regarded as noise in the signal transmitted. The method shows how "unscathed" a speech signal is, from emission to listening position. MFT technology used for speech transmission is named STI, short for Speech Transmission Index. The method is used both for the computation and the measurement of speech intelligibility.

This method uses the seven octave bands from 125 Hz to 8 kHz, which in total cover a frequency range corresponding to the spectrum of human speech. Each octave band is modulated by 14 different low frequencies, which in 1/3 octave steps go from 0.63 Hz to 12.5 Hz. The low-frequency modulation corresponds approximately to the modulation in speech. For each combination of the carrier frequency (octave-band noise) and modulation frequency, the MFT is computed or measured. This procedure gives in all 98 sets of data, which are then calculated as a single number to be the STI value for the combination of the sound source and listening positions in the room investigated (see Figure 28.6).

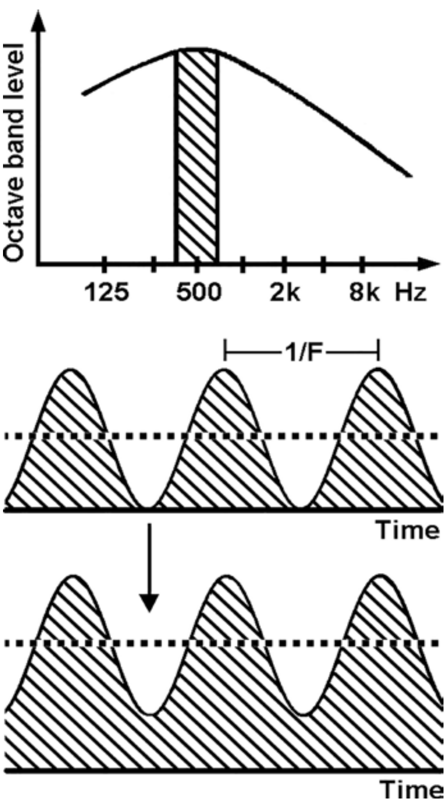

Figure 28.6 Principles of STI: Octave-band noise (here 500 Hz) is modulated by several frequencies. When measuring STI, seven octave bands apply. The modulation index is determined for each modulation frequency, and all results combine into one single number.

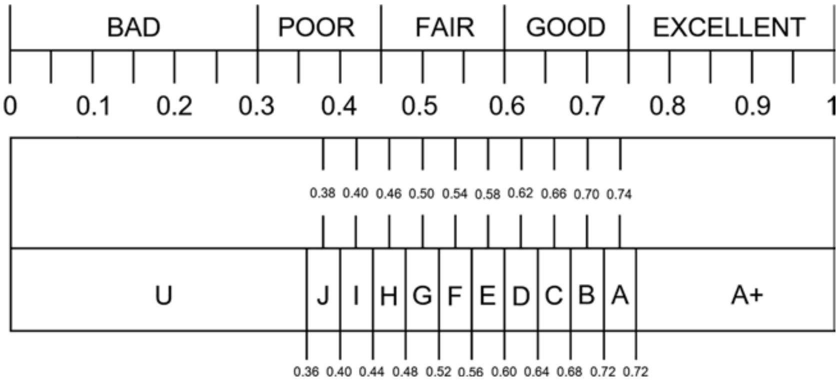

Figure 28.7 This scale indicates the relation between the STI value and the originally implemented subjectively assessed speech intelligibility. The bands A+, A, B, and so on are the scaling implemented in the latest version of the IEC standard.

SUBJECTIVE SCALE OR BANDS

In the IEC standard 60 268–16 Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index [6], the STI values were originally related to a subjective scale, with designations running from "Bad" to "Excellent." An STI value of 0.6 or better is normally what is aimed for. This corresponds at least to "Good." Even being easy to understand, the scale leads to problematic interpretations of the intervals. Each step, for instance, from 0.59 to 0.6 going from "Fair" to "Good" is almost inaudible, but would in many installation projects be the difference between failure and success. Thus, in the latest revision of the standard, it was decided to convert the scale into bands. The difference between "E" and "D" is less "emotional" than "Fair" to "Good" (see Figure 28.7)."

RASTI

Since a complete STI measurement is quite comprehensive and processor-consuming, a reduced version called RASTI was developed and applied when practical measurements were first introduced. RASTI stands for RApid Speech Transmission Index. (Some called it Room Acoustic Speech Transmission Index.) Only the most significant combinations of carrier and modulation frequencies were included in it, amounting to nine combinations. However, the principles were otherwise the same as for STI.

STI-PA

STI-PA is short for Speech Transmission Index for Public Address Systems. Like RASTI, the STI-PA is a derivative of STI. However, this reduced methodology provides results that are sufficiently comprehensive to be compared to complete STI measurements. It was developed to cope with the nonlinear processing environment common to advanced sound systems and to reduce the measurement time required to a practical level.

STI-PA supports fast and accurate tests with portable instruments that can evaluate speech intelligibility within 15 seconds per position. The STI-PA signal can be generated by a file-based signal generator. Implementing an artificial voice source, an end-to-end system check – including the microphone – can be performed (see Figure 28.8).

The procedure includes different routines to get the correct result: If the STI-PA value measured is less than 0.63, two more measurements must be made, and all three results are averaged. If the results differ more than 0.03, three further measurements must be made, and all six readings are averaged for the final results. Results that in the same position vary more than 0.05 should not be accepted. This kind of error may occur if the background noise is very impulsive or otherwise is varying.

Figure 28.8 STI-PA readout from a handheld device (NTI AL1 Acoustilyzer). The display shows that the measurement is finished and that the value is 0.89, which corresponds to "Excellent" or A+.

Measurements can be made at a time with less background noise, and the results can be corrected using dedicated software.

EMULATED STI MEASUREMENTS

There are other types of measurement instruments, based on TDS and MLS, that are used to measure STI. Handheld microphones are essentially not allowed. Additionally, time-varying background noise must be prohibited. Still, when comparing the results of the measurements from these systems, they are not completely in agreement with the "real" measurements and should not be used for documentation.

STANDARDS

The following are some of the standards that require STI data:

ISO 7240 Fire detection and alarm systems.

NFPA 72 National Fire Alarm Code 2002

BS 5839–8 Fire detection and alarm systems for buildings. Code of practice for the design, installation, and servicing of voice alarm systems.

DIN 60849 System regulation with application regulation DIN VDE 0833–4.

LOGGING SYSTEMS

Noise regulations in the entertainment industry have been introduced in many countries. Even "awesome" music can be considered noise. Especially in venues for rhythmic music, it is essential to control the sound level during a concert. The front of house (FOH) engineer should have proper information so she or he can keep the concert rolling and avoid being stopped by authorities due to excess of local limits.

In the measurement of environmental noise, the limit is normally defined by the A-weighted equivalent sound pressure level measured during a given interval of time ranging from minutes to hours. However, to control the level, the time interval usually is in the range of 5–15 minutes.

Due to legislation, very often the levels measured have to be reported. Thus a non-corruptible logging system has to be installed.

Systems and programs for this special purpose have been developed. The best of these perform to the same standards as integrating sound level meters and include a calibrated microphone. Additionally, clear displays are provided showing the actual A-weighted sound pressure level and the C-weighted peak level. The A-weighted L_{eq} is calculated from the start. In some cases, a special readout can tell if you are safe or if you are close to the limits and should "hold your horses" (see Figure 28.9).

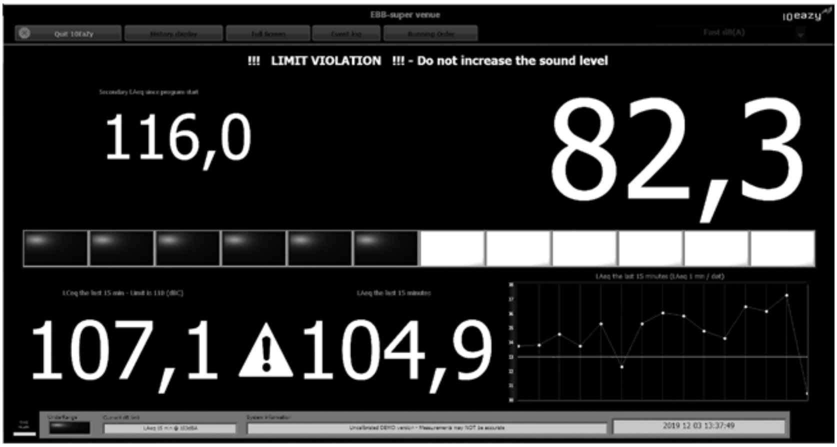

Figure 28.9 Screenshot of readout from the measurement system "10 eazy."

When the job is done, an encrypted data file with logging data can be extracted from the system. Even if the connection to the microphone was broken during the measurement, this would be noted in the file.

European legislation is based on Directive 2003/10/EC of the European Parliament and of the Council [7] (see Figure 28.10).

REFERENCES

[1] Wavesurfer: https://www.speech.kth.se/wavesurfer/man.html.

[2] EASERA tutorial, by AFMG Technologies GmbH, Berlin (2012).

[3] Stan, Guy-Bart; Embrechts, Jean-Jacques; Archambeau, Dominique: Comparison of different impulse response measurement techniques (2002).

[4] Praat: http://www.fon.hum.uva.nl/praat/.

[5] IEC 60 268-16 IEC 60268-16 Ed. 4.0 b:2011: Sound system equipment – part 16: Objective rating of speech intelligibility by speech transmission index.

[6] Directive 2003/10/EC of the European parliament and of the council of 6 February 2003 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise).

[7] Smaart v8, User guide, rational acoustics (2018).

CHAPTER 29

Measurement Signals

CHAPTER OUTLINE

The following is a short overview of some measurement signals that are useful to the sound engineer. Signals normally available are described.

SINUSOIDAL TONES

The sinusoidal tone – or sine wave – is a practical test signal because it can be kept constant and because it contains one (and only one) frequency at a time. It applies most of all to checking and calibration of audio meters and other audio equipment.

In testing analog devices, the preferred frequencies have values that follow the center frequencies of standard octave or fractal-of-octave filters. One frequency that (almost) always applies is 1 kHz. One reason for applying this specific frequency is that most weighting filters are neutral at 1 kHz. Also, the frequency is right in the middle of the efficient frequency range of most audio devices.

However, when it comes to testing of digital equipment, normally slightly different frequency values may apply, for instance, 997 Hz and not 1 kHz. In general, it is recommended to apply prime numbers for the selected frequencies, especially for the measurement of distortion. Doing this avoids conflicts with the sampling frequency (see Table 29.1).

It should be mentioned, though, that the ATSC A/85 recommends a 440 Hz sine wave tone for calibration, as this frequency is in the flat middle portion of the frequency range of the BS.1770 frequency response curve.

Acoustic calibrators mounted directly on measurement microphones also take advantage of the constant level and well-defined frequency. Here, in most cases, the frequency/level is 1 kHz/94 dB re 20 μPa for the (small) loudspeaker type of calibrators and 250 Hz/114 dB for pistonphone types of calibrators.

Specific problems in rooms, such as uneven sound distribution due to standing waves, can easily be spotted using sinusoidal tones; sweep-tones are especially efficient for that. For a start, put a low-frequency loudspeaker in the corner of the room, where there is a maximum for all the standing waves. Then listen – or measure – in various positions across the room. The SPL variations from one position to another indicate the (un-)evenness of the sound distribution.

A sweep can vary in duration. You can define the start frequency and the stop frequency. Also, the sweep can vary in time. If the frequency varies linearly per time

Table 29.1 List of frequencies for testing of digital equipment, especially measurement of distortion and linearity, as proposed by Robert Finger in [1].

19 Hz	41 Hz	101 Hz	317 Hz	499 Hz	997 Hz
3,163 Hz	6,301 Hz	10,007 Hz	16,001 Hz	19,001 Hz	19,997 Hz

interval, it is called a white sweep, and if the frequency doubles (or halves) per time interval, it is called a pink sweep.

Most DAWs can generate sweeps. If you don't have measurement equipment, it is a good idea to generate short sweeps, for instance, each within the band of one octave. Or even better: Create a slow sweep from 200 Hz to 20 Hz, and add small 1 kHz beeps each time the sweep passes the center frequency of a standard one-third octave. Then it is possible to "count" your way through the frequency span when investigating standing waves.

The sine wave is, however, not practical for general level calibration of loudspeaker systems because the room influences the signal due to reflections, standing waves, and so on. One single frequency is too unstable for acoustical measurements because the level typically varies depending on the position in the room.

BURSTS

Tone bursts at well-defined intervals are the foundation for checking level-reading instruments. Either burst generators or downloadable measurement signals are applicable. Also, DAWs are perfect tools for creating tone bursts.

What is most important is that the frequency must be relatively high (5–10 kHz) if short bursts shall contain a sufficient number of complete periods. Another thing to remember is always to cut the sine wave at zero-crossings.

NOISE

Broadband noise signals, primarily in the form of white noise and pink noise, which have constant energy per Hz and constant energy per octave, respectively, are indispensable in practical sound work. You can listen to the signals, and you can use it (primarily pink noise) together with spectrum analysis for simple alignment of loudspeaker setups, and so on.

Noise is also relevant for reverberation measurements: Feed pink noise bursts into a loudspeaker placed in the corner of the room. Then make recordings with an omnidirectional microphone at a position away from the loudspeaker. Record the sequence. Repeat a couple of times, to reduce uncertainties, as the noise has a random character. Move both loudspeaker and microphone to other positions and record new bursts. The recorded decays are analyzed.

Pink noise generators may perform differently. Hence some standards define the crest factor of the noise. Typically the required crest factor is four, which is the same as 12 dB.

Due to the different versions of pink noise available, a special pink noise has been described in a standard issued by the SMPTE (ST-2095–1). This standard defines a digital pink noise signal for calibrating the sound pressure level and the electroacoustic response of a cinema B-chain system.

Also the use of band-limited pink noise has become a standard for alignment of monitoring loudspeakers [2]. A two-octave-band pink noise centered at 1 kHz applies to main speaker alignment, and a single 40 Hz octave band pink noise applies to LFE speakers.

The M-Noise is a broadband noise signal that has an increased crest factor at higher frequencies for testing loudspeaker systems (see Chapter 4: Signal Types).

CLICK GENERATORS

A click sound is a short impulse with a broadband spectrum. The click can be generated electrically and reproduced by a loudspeaker.

It could also be a click that is generated acoustically by a clapper board or perhaps one of those "clickers" found as (dangerous) toys for children. The latter is incredibly effective in examining coincidence in stereo microphone placements, and so on. Be careful with your ears. The impulses are very short but also very loud (>120 dB peak re 20 µPa @ 10 cm typical). The short duration of the clicks fools the ears' level perception.

A click is also practical for examining echo phenomena in larger rooms. Further, the click is efficient when checking delay settings in distributed loudspeaker systems.

Clicks generated by spark generators are efficient. Dedicated spark generating devices often apply in connection with acoustical scale models, due to the contents of high frequencies in the acoustical sound of the spark (typical >100 kHz). Also, the impulse response of microphones can be examined by the application of spark sounds.

Phase checking systems often utilize unidirectional (positive or negative) clicks from an electric generator (short time DC offset).

POPS/ BLASTS

A somewhat more powerful impulse with a fair amount of low-frequency content is a balloon pop or perhaps a pistol shot (a blank) to examine large rooms and as a sound source for reverberation measurements. The recording is regarded as an impulse response of the room.

OTHER SIGNALS

Most measurement signals, however, are related to the dedicated measurement systems, as mentioned in Chapter 28: Other Measurement Systems. The application of Wavelets, Maximum Length Sequence (MLS), Time Delay Spectrometry (TDS), and so on need dedicated measuring systems to become applicable.

REFERENCES

[1] Finger, R.A.: Review of frequencies and levels for digital audio performance measurements. *Journal of the Audio Engineering Society* Vol. 34, 1–2 (1986).

[2] ATSC recommended practice: Techniques for establishing and maintaining audio loudness for digital television (A/85:2013).

CHAPTER 30

Sound Level Meters

CHAPTER OUTLINE

Sound levels can be measured with a sound level meter, abbreviated SLM. A sound level meter contains a microphone, an amplifier circuit, a filter circuit, and a detector. The readout is either a moving coil type instrument or, more likely, a digital display. Also, an SLM may contain some form of computational functions. Then it is possible to perform relevant calculations/statistics of recorded sound levels. Thus, there is a differentiation between sound level meters, integrating sound level meters, and dedicated analyzers.

MICROPHONE

The microphone applied in the SLM is a pressure type. A pressure microphone is a microphone with an omnidirectional pickup pattern. However, large diameter microphones (such as ≥1 inch) exhibit increased on-axis directivity at higher frequencies. This phenomenon occurs when the wavelength of the sound becomes comparable to the dimensions of the microphone. To retain omnidirectionality, smaller diameter capsules may be applied or, in some cases, the microphone grid is replaced by a so-called nosecone. Depending on the kind of measurement performed, the microphone selected should exhibit a flat free-field response (such as for measuring the frequency response of a loudspeaker) or a flat diffuse-field response (such as for the measurement of sound power in a reverberant room).

The pressure microphone has in principle a closed chamber behind the diaphragm; thus it by nature works down to barometric pressure. However, a small vent is there to limit the lower frequency response to 20 Hz. If the SLM is intended for infrasound measurements (sound below 20 Hz), a special microphone that goes at least down to 1 Hz is needed. Also, extremely low levels or extremely high levels need special microphones – and an SLM to accommodate the specific requirements.

More advanced SLMs can identify the microphone and set the correct gain according to the microphone's sensitivity and frequency response.

For measurements outdoors, a foam windshield is applied, preventing the wind from generating noise in the microphone. However, the shields have different properties. If the foam is not sufficiently open, high frequencies are attenuated; if the foam is too open or too small, it does not block the wind.

To be able to check the accuracy of the sound level meter, an acoustic calibrator is used. It is a small sound generator, with a small built-in loudspeaker. The calibrator provides well-defined frequency and level, typically 1 kHz, 94 dB SPL. The calibrator opening fits over the microphone capsule. Note that a calibrator in general only fits the microphones for which it is designed. This fact is due to the influence of the volume of the cavity between microphone diaphragm and loudspeaker membrane, which affects the sound pressure generated.

AVERAGING

The detector rectifies the signal, averages it, and calculates an RMS value. There are two forms of averaging. One is linear averaging, and the other is exponential averaging. With linear averaging, the RMS value is found for a fixed time interval called the integration time. It is this form of averaging that applies to level-reading instruments in audio production. For example, the standardized integration time is 5 ms for a PPM instrument, according to IEC standard 60268–10.

In connection with the measurement of acoustic sound, exponential averaging is used. The incoming signal is weighted so that the meter "remembers the past," but events that lie further back in time have less weight than events that just occurred. The averaging time, or the time constant, is a measure of how fast the exponential function "decays." More precisely, the averaging time specifies the time it takes before the exponential function is reduced to 69% of the starting value. Internationally, the time constants applied include 125 ms (called "F" for "Fast") and 1 second (called "S" for "Slow").

Integrating SLMs can perform measurements over long periods (minutes/hours). The result expresses the energy which a constant sound would have had during the measuring period, also called the equivalent level or L_{eq}.

FREQUENCY WEIGHTING

Since the sensitivity of the ear is not the same at all frequencies, filters are inserted in the SLM to emulate this. The most frequently used filter performs an A-weighting, providing attenuation of the lowest and highest frequencies while gaining frequencies in the 1–4 kHz range. Other filters that also apply include B-, C-, D-, and Z-weighting filters.

In recent years, the problems of infrasound (i.e., frequencies below 20 Hz) have been investigated. To accommodate measurements of this specific frequency range, the G-weighting filter has been applied to the SLM. This filter covers a frequency range from 1 Hz to 20 Hz (see Chapter 9: Frequency Weighting and Filters).

READING

The readout of results is always expressed in dB. As a reference level for the measurement of sound pressure, the value 20 μPa ($2 \cdot 10^{-5}$ Pa) is the common ground. This level is close to the threshold of hearing.

When, for example, reporting an A-weighted SPL measurement, it is written like xx dB(A) re 20μPa or L_{pA} = xx dB re 20μPa. The extension "re 20μPa" is not always stated if it is noted, because it is implied that SPL always has the same reference.

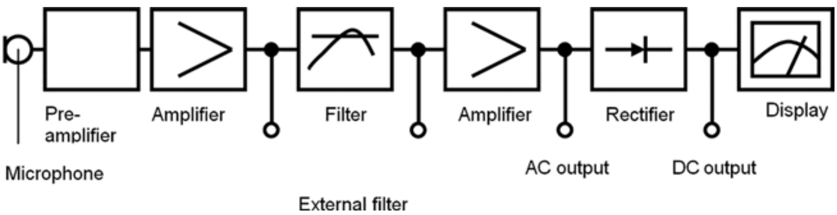

Figure 30.1 Basic diagram of the principles of a sound level meter.

OUTPUT

Often, an AC output is available. Thus, the SLM can function as a microphone preamplifier. This functionality makes it possible to record the signal for later analysis. Some SLMs may also have a DC output for level displays or printout (see Figure 30.1).

CLASSES OF SOUND LEVEL METERS

The following standards specify the requirements for precision of sound level meters and for integrating sound level meters:

IEC 61 672 Sound level meters

IEC 60 804 Integrating-averaging sound level meters

ANSI S1.4 Sound Level Meters

ANSI S1.43 Integrating averaging sound level meters

The norms specify nominal values and tolerances for several characteristics:

Type 0 is intended as a laboratory reference standard.

Type 1 is intended for precise measurements in the laboratory and the field.

Type 2 is envisioned for general measurements in the field.

Type 3 can only be used for a rough orientation concerning noise levels.

Type 0 has the narrowest tolerances. Most countries require that measurements in the external environment be performed with measurement equipment that adheres to the type 1 tolerances.

ANALYSES

In connection with the measurement of noise, different analyses are performed, including some of a more statistical nature. The most significant ones are listed here.

L_p SPL or Sound Pressure Level. The RMS value of the sound pressure concerned expressed in dB re 20 μPa.

L_{eq} The equivalent sound pressure level (i.e., the average value [on an energy basis] of the sound pressure level registered over a period).

L_E (previously called SEL). An expression of the total energy in the period concerned.

L_n With a basis in the cumulative distribution, L_n specifies the sound level that has been exceeded in *n%* of the time under consideration. *n* ranges from 1 to 99. L_n is often applied for the description of the background noise during periods with varying noise; for example, L_{95} = 36 dB means that this value was exceeded during 95% of the measurement period.

L_{DEN} This is a long-term average measure of environmental noise. L_{DEN} is an equivalent SPL covering a period of 24 hours. “DEN” is an abbreviation for Day, Evening, Night. The three periods are defined as Day: 06–18 (6 AM – 6 PM), Evening: 18–22 (6 PM – 10 PM), Night: 22–06 (10 PM – 6 AM). As the intention is to keep evenings and nights quiet, a penalty is “awarded” before summing the noise of the three periods, 5 dB is added to the noise level measured during evening time, and 10 dB is added to the noise level measured during the night time. The L_{DEN} is described in Directive 2002/49/EC (2002), issued by the European Community.

AUTOMATION

It is often very time-consuming to make measurements. For instance, when performing building acoustic analyses like sound insulation, several individual measurements must be done, and a lot of results have to be put into spreadsheets and “massaged” to reach a final result.

Advanced SLMs have gradually become full audio analyzers that can easily be controlled directly, by an external computer, or by a smartphone.

In case of measurements for sound insulation between two rooms, background noise is measured and analyzed; reverberation time is measured and analyzed; a noise source is started and stopped; sound levels in different positions are measured, averaged and analyzed. After this, a major calculation takes place, and the result is presented as a curve.

NTI Audio XL2 is an example of an instrument that can perform all this and present the result in a finished form.

Another example of a comprehensive analyzer disguised as an SLM is the Brüel & Kjær 2270, which is a special SLM providing two-channel input, which makes it an analyzer for sound intensity, building acoustics, and much more.

SMARTPHONE + APP AS SLM

Lots of apps are available to convert a smartphone into a sound level meter or even an audio analyzer. In principle, this can work. However, if the processing of the app is good enough, the weak points are the smartphone's lack of calibration, converters that have limited dynamic range, and microphones that are not up to standard.

It is possible to apply one of the available external converters such as the DPA MMA-A Digital Audio Interface for IOS devices. If attaching a high-quality miniature microphone that can be calibrated, the smartphone can be converted into a Type 2 instrument and trustworthy measurements are within reach.

REFERENCES

(Standards are listed in the text).

Celestina, Metod; Hrovat, Jan; Kardous, Chucri A.: Smartphone-based sound level measurement apps: Evaluation of compliance with international sound level meter standards. *Applied Acoustics* Vol. 139, 119–128 (October 2018).

CHAPTER 31

NR, NC, PNC, RNC, and RC Curves

CHAPTER OUTLINE

Noise Rating (NR), Noise Criterion (NC), Preferred Noise Criterion (PNC), Room Noise Criterion (RNC), and Room Criteria (RC) curves by and large serve the same purpose, the measurement of acoustic noise. However, their origins are different: NR is used in Europe, whereas the others predominantly are used in North American countries and related to slightly different applications.

NOISE RATING (NR)

Noise Rating curves are a set of curves based on the sensitivity of the ear. They are developed to arrive at a single number (rating) for noise, based on sound levels measured in octave bands.

C.W. Kosten and G.J. van Os developed the NR as requirement curves for noise sources both indoors and outdoors. Now the curves are predominantly used in connection with requirements for, and documentation of, background noise indoors. Here, they particularly apply with noise from ventilation systems and other similar noise sources in building installations.

NR curves were standardized in ISO/R 1996:1971 Acoustics – Assessment of noise with respect to community response (now withdrawn, but still the reference for existing standards). The NR curves encompass the standard octave bands from 31.5 Hz to 8 kHz. Normally increments of 5 from NR-0 to NR-100 apply to the standard set of curves. However, the standard has defined values with the increments of 1 dB. The value at 1 kHz gives the curve its name, like “NR-10” (see Figure 31.1).

This is how to use the curves: Measure the noise as sound pressure level applying octave-band filters. Plot the measurement result of each octave band on the NR curve sheet. Then find the lowest of the NR curves not exceeded by the measured values (see the example at the end of this chapter). This curve is then the result of the measurement (in the example shown in Figure 31.2, it is called NR-25).

Sound level meters/analyzers that perform the complete measurement and prepare a printout of the results in one go are of course available.

Please note that NR values cannot be compared directly with dB(A) measurements. In the measurement and assessment of noise from ventilation systems, where the NR curves often apply, the A-weighted sound pressure level typically lies 4–9 dB above the NR value based on the measurement of the sound pressure level in octave bands. In this regard, note that the NR values are always specified or read in increments of 5 dB. However, the 1 dB increment curves apply only for comparison of different noise situations in the same room.

For some applications, the noise is measured in 1/3 octave bands.

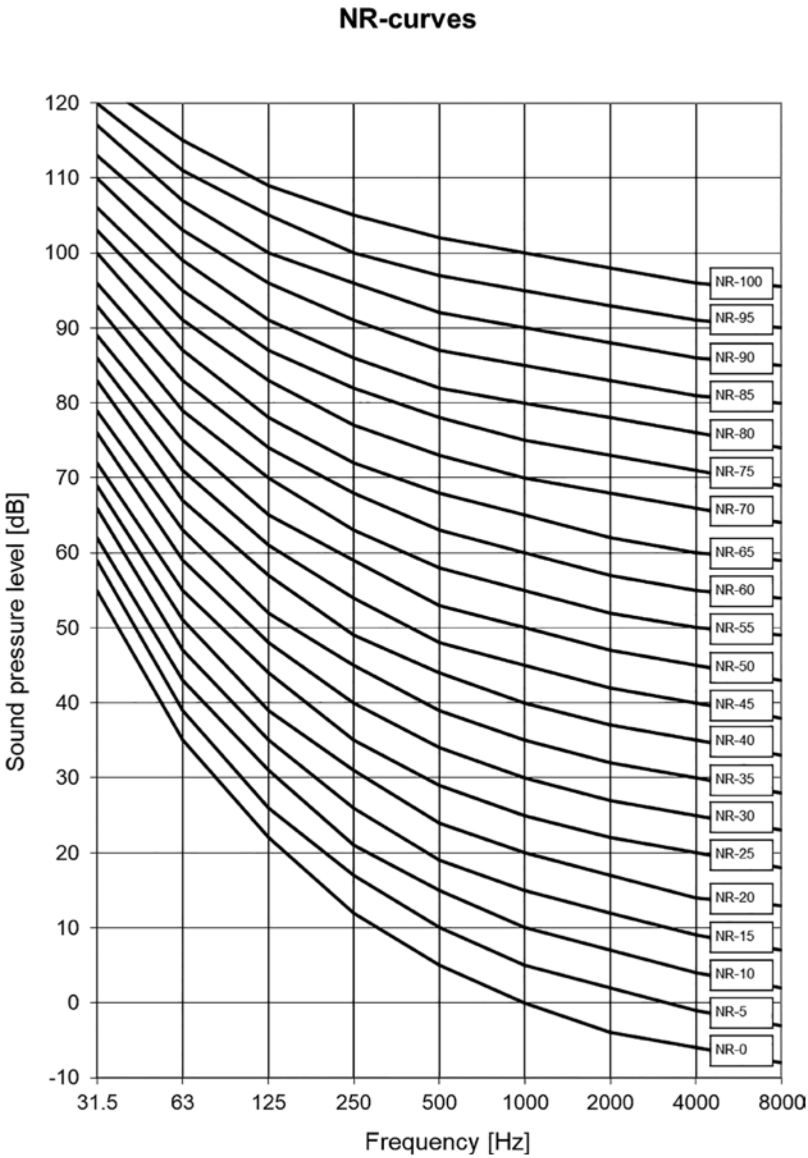

Figure 31.1 Noise Rating Curves.

NR REQUIREMENTS AND RECOMMENDATIONS

A list of predominantly European requirements and recommendations for Noise Rating is shown in Table 31.1.

Table 31.1 NR, requirements, and recommendations. If the noise contains audible tones, then the requirement is tightened by 5.

NR Curve	Application	Reference
NR-10	Listening rooms, recommended	IEC 60.268–13
	Control rooms for classical music, recommended	EBU tech. 3276
NR-15	Listening rooms, required	IEC 60.268–13
	Control rooms for classical music, required	EBU tech. 3276
NR-20	First-run cinemas, recommended	FSI, class 1
	Radio studios, concert halls	
NR-25	Cinemas, required	FSI, class 1
	Theaters, churches	
NR-30	Cinemas, required	FSI, class 2
	Private dwellings, hospitals	
NR-35	Libraries, museums, courtrooms	
NR-40	Halls, corridors, restaurants	
NR-45	Department stores, canteens	

NR-calculation, Example

Noise is measured in octave bands, and the resulting NR value must be found. The following levels have been measured:

31.5 Hz:	*55 dB*
63 Hz:	*49 dB*
125 Hz:	*28 dB*
250 Hz:	*33 dB*
500 Hz:	*11 dB*
1 kHz:	*20 dB*
2 kHz:	*14 dB*
4 kHz:	*7 dB*
8 kHz:	*12 dB*

The levels measured are plotted on the data sheet with the curves (see Figure 30.2*). The lowest NR curve not exceeded is NR-25. The octave band on the curve that comes the closest to the measured level is the 250 Hz octave. The measured noise thus corresponds to NR-25.*

A measurement of the A-weighted sound pressure level for the same noise source would give a result of 29 dB(A).

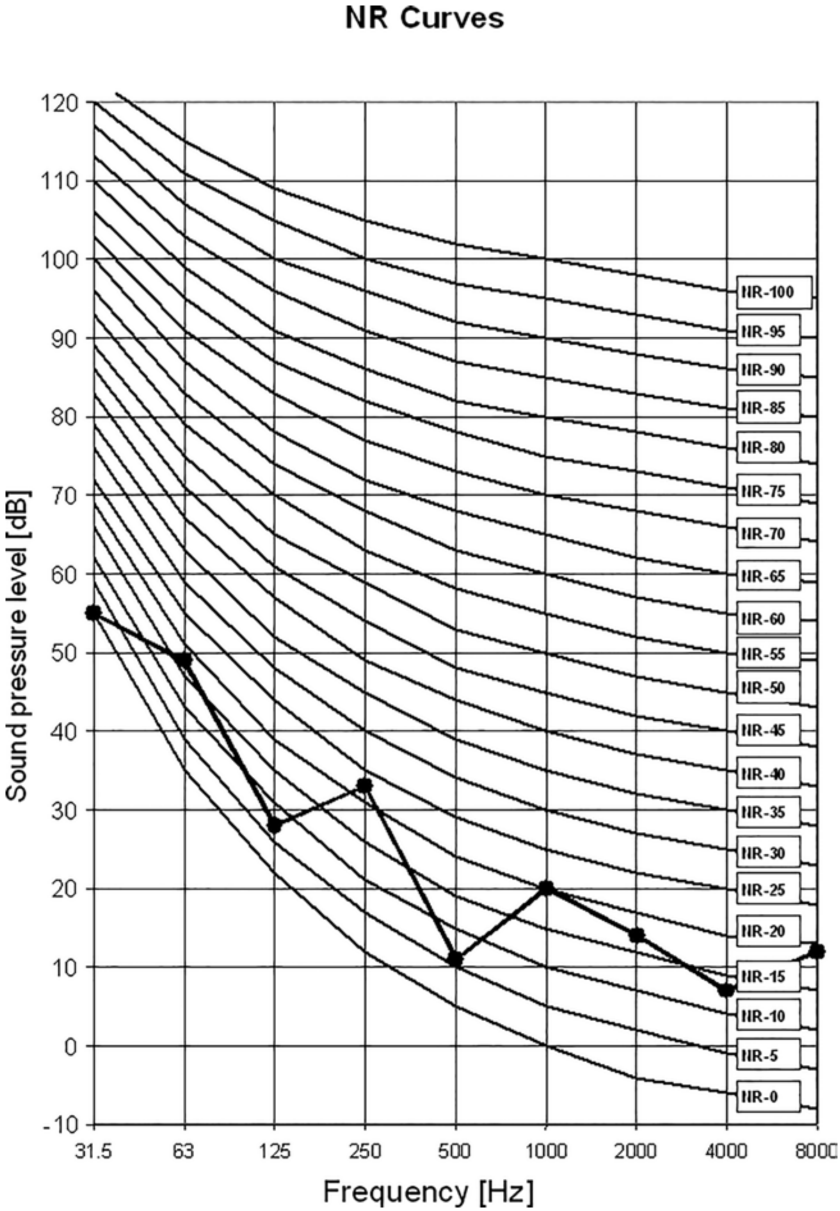

Figure 31.2 The values given in the example are plotted on this data sheet (NR-25).

NOISE CRITERION (NC)

Leo Beranek of BBN defined the Noise Criterion (NC) curves in 1957 and later standardized by ANSI (ANSI S12.2 2008 and 1995). These curves are constructed in the same manner as NR curves. However, the NC curves should only be used for specification values or as measurement curves for indoor noise sources.

NC curves are also based on the sensitivity of the ear; only the values are a little different than the corresponding NR curves. In general, NC curves are a little flatter in comparison to NR curves (i.e., they have slightly lower values at low frequencies and slightly higher values at higher frequencies).

The curves were originally defined in octave bands running from 63 Hz to 8 kHz. The curves are defined in increments of 5, from NC-15 to NC-70. Measurements are made with the standardized "SLOW" integration time (see Figure 31.3).

The reading takes place in the same manner as for NR curves: The octave levels are plotted, and the lowest curve that is not exceeded is the result of the measurement (extrapolated curves are used).

NC REQUIREMENTS AND RECOMMENDATIONS

A list of predominantly American requirements and recommendations for Noise Rating is shown in Table 31.2.

PREFERRED NOISE CRITERION PNC

The Preferred Noise Criterion curves (PNC) are regarded as an extension to the NC curves. The frequency range of the PNC includes the 31.5 Hz octave band as well as the 16 kHz frequency band. The reading of the curves is the same as with NC (see Figure 31.4).

ROOM NOISE CRITERION (RNC)

In the ANSI/ASA standard, additional criterion curves are mentioned. The Room Noise Criterion curves (RNC) are designed to describe octave-band sound pressure levels that, when modulated by fluctuations at low frequencies, may cause perceptible vibrations or rattles in lightweight constructions. These curves are extended downward to include the 16 Hz octave band. This measurement often is mentioned as the "rumble criterion."

Please notice that it requires special equipment to measure frequencies as low as that. It is always problematic to measure such low frequencies. As the

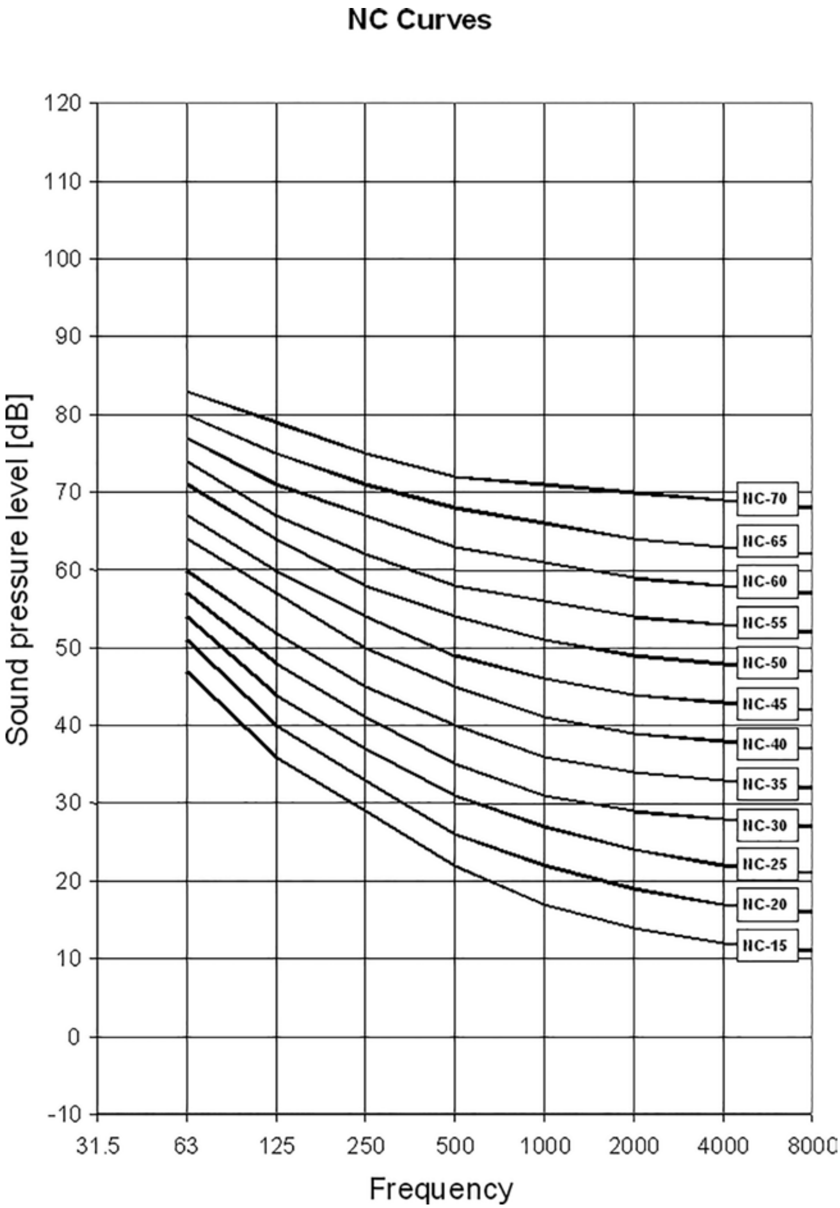

Figure 31.3 Noise Criterion curves.

Table 31.2 NC, requirements and recommendations.

NC Curve	Application	Reference
NC-15	Radio broadcast, recording studios, rooms for acoustical music performance, recital halls	
NC-20	TV broadcast studios (NC15-NC-25)	ANSI/ASA
	Large auditoria, large churches (20–25)	
NC-25	Cinematography: Dubbing and reviewing rooms, as well as first-run movie theaters	ISO 9568; 1993 (E) Dolby, THX (recommended)
	Small auditoria (25–30)	ANSI/ASA
NC-30	Cinemas, requirement	Dolby, THX
	Small churches (30–35)	ANSI/ASA
NC-35	Libraries, museums, courtrooms	
NC-40	Halls, corridors, restaurants	
NC-45	Department stores, canteens	

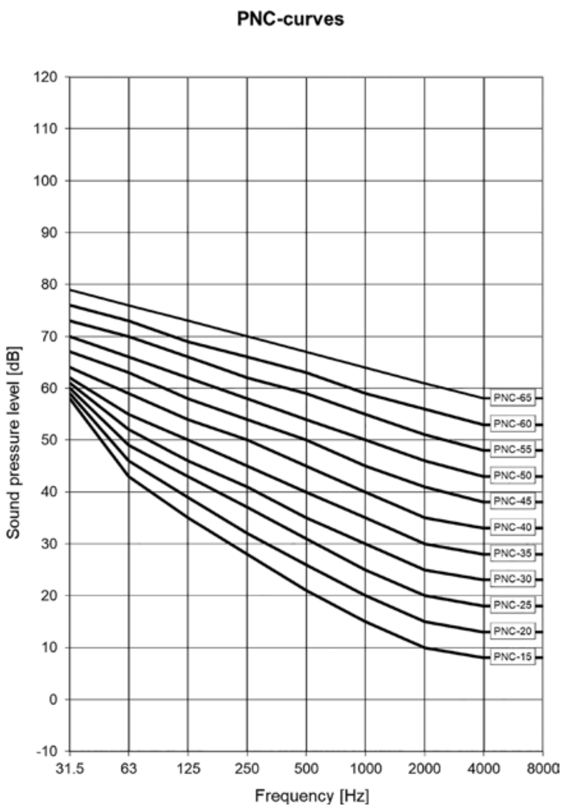

Figure 31.4 Preferred Noise Criterion curves.

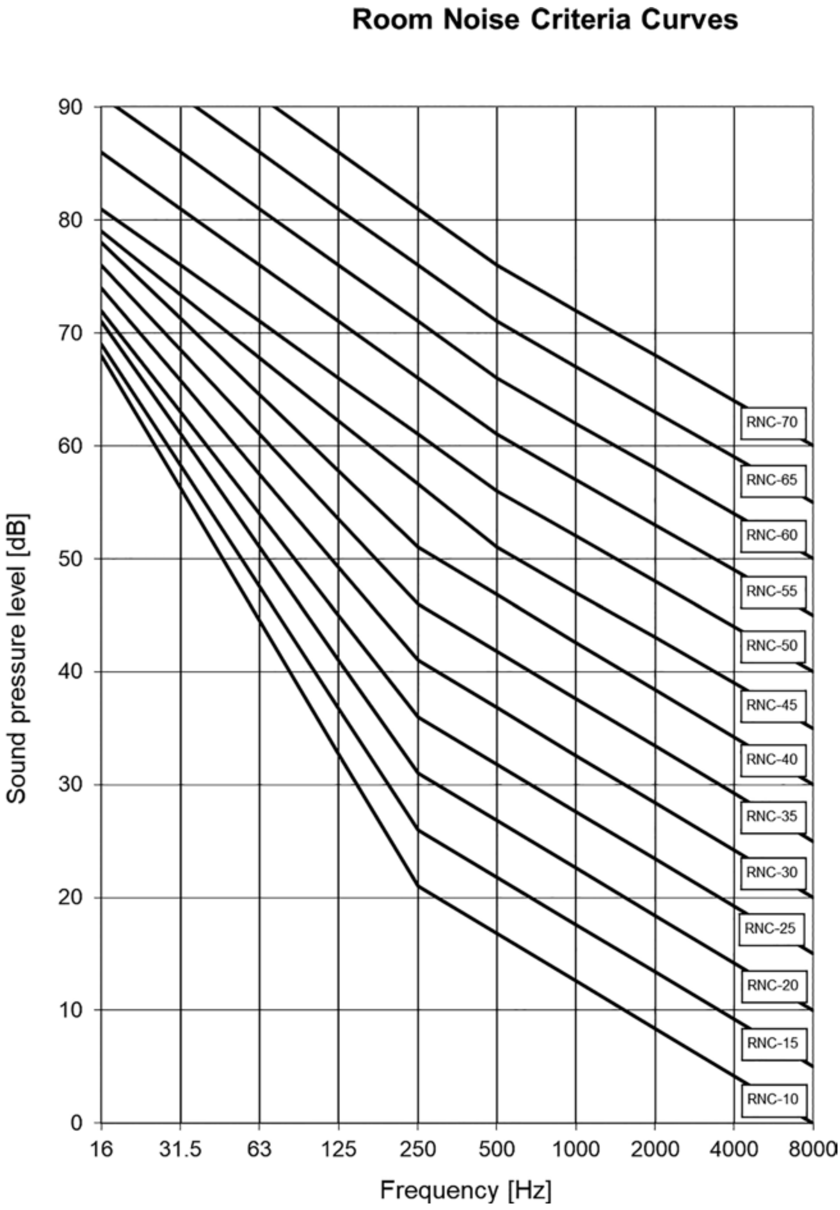

Figure 31.5 Room Noise Criterion curves.

sound may fluctuate, several successive measurements are taken to ensure a valid measurement.

The curves are shown in Figure 31.5.

ROOM CRITERIA (RC)

The Room Criteria curves (RC) are intended for the measurement background noise in buildings over a frequency range of the 16 Hz octave band to the 4 kHz octave band. Also, it is intended for mid-frequency sound pressure levels in the range of 25–50 dB re 20 μPa. This set of curves is, in particular, designed for the measurement of noise from ventilation, air-conditioning, and heating (HVAC). The standard is ANSI S12–2 - 1995 (see Figure 31.6).

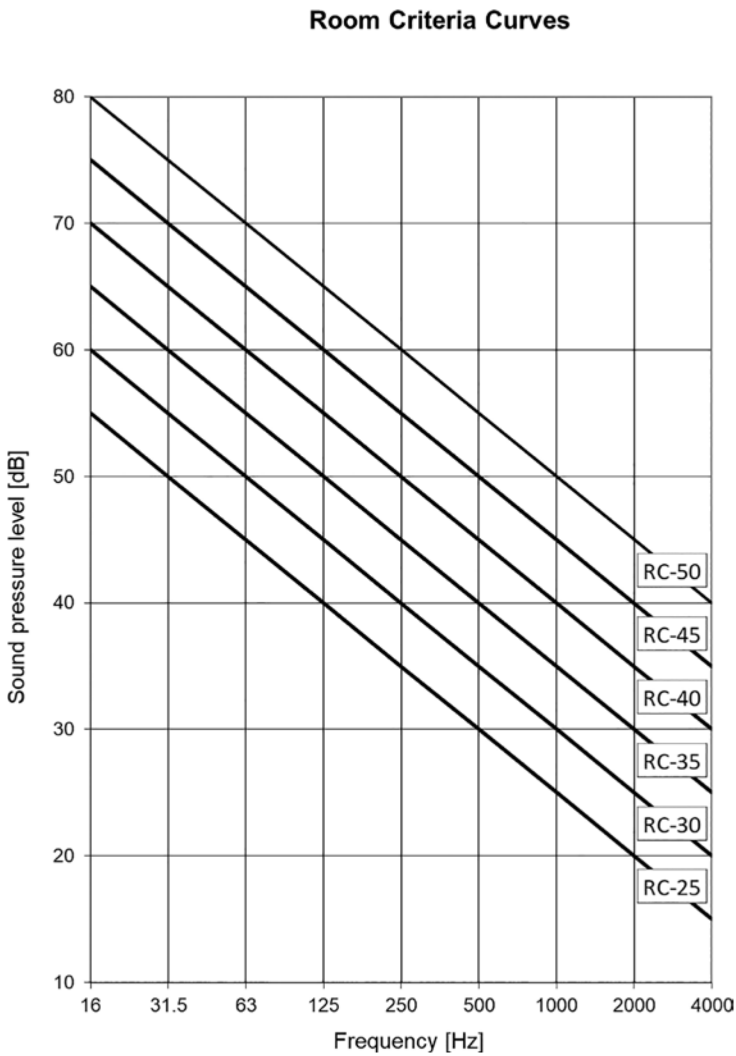

Figure 31.6 Room Criteria curves

REFERENCES

Beranek, L.L.: *Noise control*. McGraw-Hill, New York (1954).

Kosten, C.W.; van Os, G.J.: *Community reaction criteria for external noises (. . .)* Proceedings the Control of Noise. National Physical Laboratory, Symposium No 12, Paper 5:373–387 (1962).

ISO/R 1996:1971 Acoustics – assessment of noise with respect to community response (now withdrawn, but still the reference for existing standards).

Anbefalinger til biograflyd (Recommendations for cinema sound). Filmbranchens Samarbejds- og Ideudvalg (1992).

Noise Curves. NTI Audio.com.

CHAPTER 32

Room Acoustic Measures

CHAPTER OUTLINE

Room acoustic measures provide information about how sound behaves inside a room, how the sound is distributed, and about the influence of size, shape, surface materials, and so on. Many of the measures are related to the way we perceive and assess the sound as listeners, as performers in the room, or as recording- or SR-engineers. In this chapter, some basic room acoustic parameters and the related measures are briefly described.

In room acoustics, in general, there is a distinction between small rooms like listening rooms, control rooms, recording booths, and so on, and large rooms like concert halls, theaters, rock venues, and so on. This is why different parameters, to some degree, apply to different room types.

GENERAL RULES FOR GOOD ACOUSTICS

When designing rooms for audio, there is a general set of rules that can be looked at to find the starting point for the achievement of "good" acoustics:

- appropriate reverberation time
- appropriate sound distribution
- low background noise
- no echoes (flutter echoes)
- appropriate control of early reflections

Although this set of rules apply to acoustic design, in general, the solutions may turn out to be different depending on the application of the room. For instance, good sound distribution in an auditorium is vital; so is it in a classroom. However, it is contrary to the requirement for open plan office environments with multiple workstation installations as found in many broadcast facilities these days.

SCHROEDER FREQUENCY

Before moving to a closer look at the good rules, we have to consider the distribution of the sound waves within a room. When the physical size of the wavelength at lower frequencies may approximate the dimensions of the room, the sound waves do not move freely. What we have are the so-called room modes or eigenfrequencies. As with any other acoustical resonance phenomenon, there is an emphasis of these frequencies that literally "fit" into the room. The result is an uneven distribution of the sound in this frequency range, the modal region. At higher frequencies, in the statistical region, the sound waves move freely and geometrically.

A practical approach to a definition of the transition between the modal region and the statistical region is the Schroeder frequency equation. The Schroeder frequency is defined in relation to the number of modes within a given frequency band. It is dependent on the volume of the room and reverberation time. However, this frequency is also dependent on the room shape. That is, the sphere and the cube exhibit a higher Schroeder frequency compared to other shapes. The Schroeder frequency ($f_{schroeder}$) is defined as follows:

$$f_{schroeder} = 2000 * \sqrt{\frac{T}{V}} \left[\mathrm{Hz} \right] \text{ (unit: meter)}$$

or

$$f_{schroeder} = 11885 * \sqrt{\frac{T}{V}} \left[\mathrm{Hz} \right] \text{ (unit: foot)}$$

where

T is the reverberation time [s]

V is the volume [in m^3 or f^3]

It is not possible to describe this frequency exactly, as the room shape may affect the frequency distribution. However, Figure 32.1 is a diagram that shows the relationship between Schroeder frequency, volume, and reverberation time.

REVERBERATION TIME

The reverberation time is the single most important parameter regarding the room acoustics in general. Reverberation time is defined as the time it takes the sound field to attenuate by 60 dB after the sound source has stopped. Hence we also use the expression RT60 as a name for reverberation time.

If a sound impulse is emitted from a sound source in a closed space, reflections bouncing from the surfaces can be observed. The more reflective (or less absorbing)

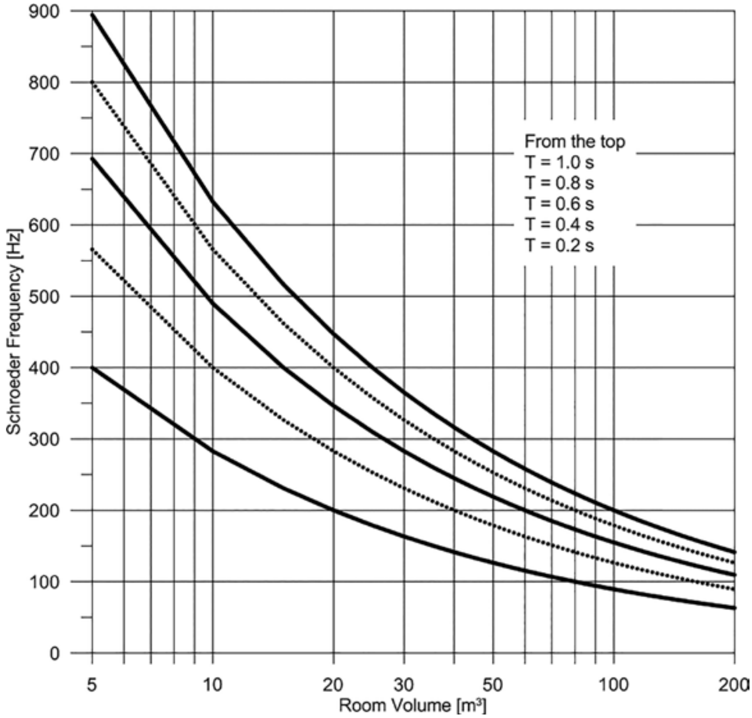

Figure 32.1 Relationship between room volume and Schroeder frequency for T (reverberation time) = 0.2, 0.4, 0.6, 0.8 and 1 s.

the surfaces are the longer time it takes for the sound to die. The bigger the room, the longer the distance the sound must travel to meet the either reflecting or absorbing surface (see Figure 32.2).

Figure 32.2 shows the impulse response of a room. In this case, it is the blast of a balloon. It can be seen on the waveform that after the blast a lot of gradually diminishing impulses pass the recording microphone. When this response is converted to level (reading in dB), we can observe a linear decay (with minor bumps). It is the slope – measured in seconds per 60 dB – of this decay that expresses the reverberation time. However, we do not then necessarily need a 60 dB decay to find the reverberation time.

The RT60 (the reverberation time) is commonly defined by the part of the decay that has been observed or evaluated:

T30: Slope in seconds per 60 dB evaluated over the −5 to −35 dB range of the decay curve.

T20: Slope in seconds per 60 dB evaluated over the −5 to −25 dB range of the decay curve.

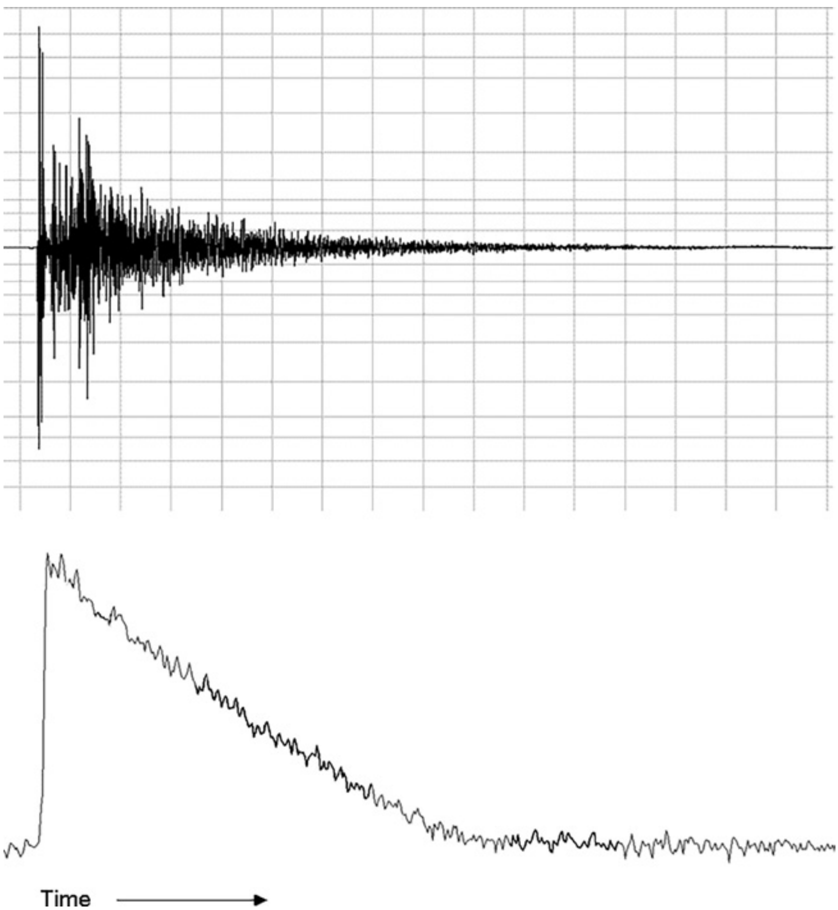

Figure 32.2 A: Impulse response in a room (waveform of a balloon blasted). All individual reflections can be seen. B: The signal above has been converted to level (now the reading is in dB).

EDT, Early Decay Time: Slope in seconds per 60 dB evaluated over 0 to −10 dB range of the decay curve.

The T30 is used in general for analysis of the reverberation time. However, the EDT is more related to the perceived reverberation time, especially if the slope sometimes changes at lower levels which may occur in complex rooms. In that case, there are different values for EDT and T30 (see Figure 32.3).

The reverberation time is, in principle, only relevant above the Schroeder frequency. However, we do try to measure reverberation below this frequency. The

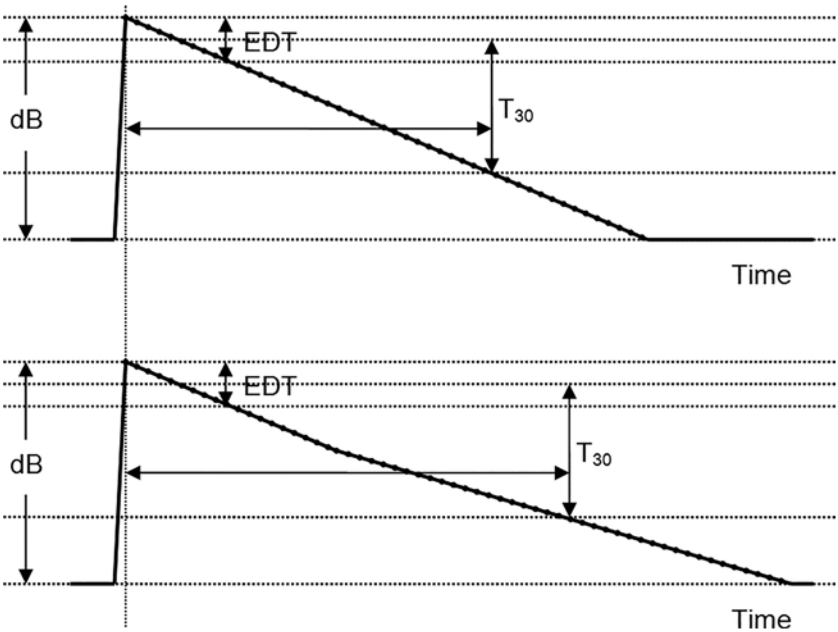

Figure 32.3 On the upper plot, the decay is a straight line. EDT and T30 are identical. On the lower plot, two different slopes can be observed. EDT and T30 will have different values in this case.

large variations that might be observed below the Schroeder frequency are, to some degree, compensated for by averaging measurements in many different positions of the room.

In general, the crucial point is how the reverberation time varies with frequency (see Figure 32.4). Depending on the application, the reverberation time as a starting point should have the same value at all frequencies. In many cases, it can become a major problem if the reverberation time at lower frequencies exceeds the midrange. Normally it is allowed that frequencies below 125 Hz may reach values of 1.5 times the midrange in studios, control rooms, and medium-sized room for music. However, in rock venues, it is preferred that the low-frequency reverberation should be kept at the same level as – or even lower than – the midrange.

When characterizing the reverberation time of a room with one single number, this is nominally the reverberation time around 500 Hz or in the midrange (i.e. 250–1000 Hz). Table 32.1 shows generally preferred reverberation times @500 Hz depending on the application of the rooms.

When designing rooms for audio, especially the smaller ones, the target reverberation time is linked to the room size. In one commonly used recommendation for

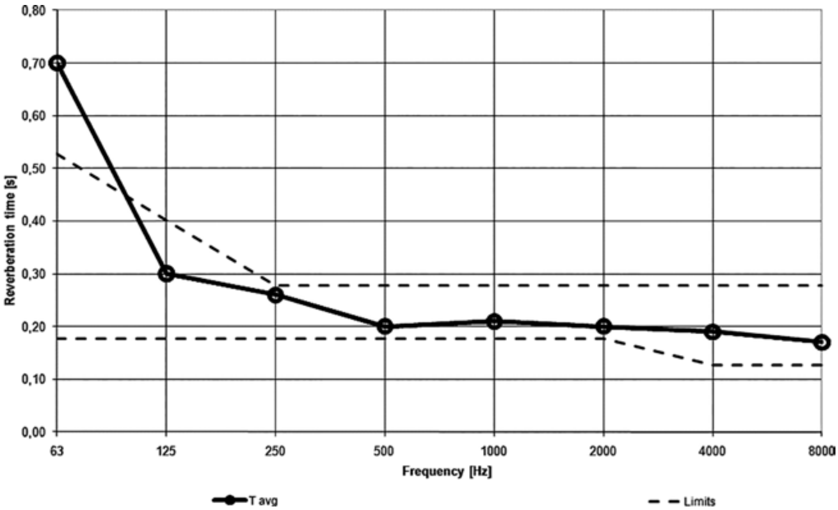

Figure 32.4 An analysis of the reverberation time as a function of frequency. Octave-band filtering is applied. In this example, the reverberation time at lower frequencies (63 Hz octave band) exceeds the limits provided by a recommendation for small rooms for audio editing.

Table 32.1 Application dependent preferred reverberation times @500 Hz.

Application	Reverberation time	Comment
Vocal booth	*0.1–0.2 sec.*	*The most difficult room to design; sounds always as a small box (boominess).*
Control room	*0.2–0.4 sec.*	*If the room is also used for the recording of acoustical instruments the reverberation should be slightly longer.*
Recording studios	*0.4–0.6 sec.*	*Rhythmic music.*
Living room	*0.4–0.5 sec.*	
Lecture room	*0.6–0.9 sec.*	*Provides level to the speech but keeps it intelligible.*
Cinema (larger)	*0.7–1.0 sec.*	*Must provide a fair reproduction of the audio.*
Rock'n' roll (smaller venues)	*0,6–1,6 sec.*	*Linear relation. Room sizes from 1.000 m^3 to 10.000 m^3 or 35,300 f^3 to 353,000 f^3.*
Theater	*1.1–1.4 sec.*	*Provides level to the speech but keeps it intelligible.*
Opera	*approx. 1.6 sec.*	*The reverberation must sustain the singing but still retain some degree of intelligibility.*
Concert hall for classical music	*1.8–2.2 sec.*	*May vary with the size of hall and music genre.*

listening rooms/control rooms (also found in the EBU tech. 3276 and the EBU tech. 3276 supplement 1), the nominal reverberation time T_m is defined as:

$0.2 < T_m < 0.4$ [s], and

$T_m = 0.25(V/V_0)^{1/3}$ [s]

where

V is the volume in cubic meters (or cubic foot)

V_0 is a reference volume of 100 m^3 (or 3531 $f^{3)}$

Further to this, limits to the variation of the reverberation time vs. frequency is given. In the Figure 32.5 these limits are shown.

There are several standards and recommendations for listening rooms and control rooms. The "IEC 60268–13: Sound system equipment - Part 13: Listening tests on loudspeakers" is one of them. Another one is given by the THX PM3 proprietary standard.

For larger auditoria, the relation between the reverberation time at lower and higher frequencies is sometimes expressed as the Bass Ratio, BR, calculated like this:

$$BR = \left. \frac{T_{125\ Hz} + T_{250\ Hz}}{T_{500\ Hz} + T_{1000\ Hz}} \right]$$

where

$T_{125\text{ Hz}} \ldots T_{1000\text{ Hz}}$ is the reverberation time in the octave bands 125 Hz . . . 1000 Hz, respectively.

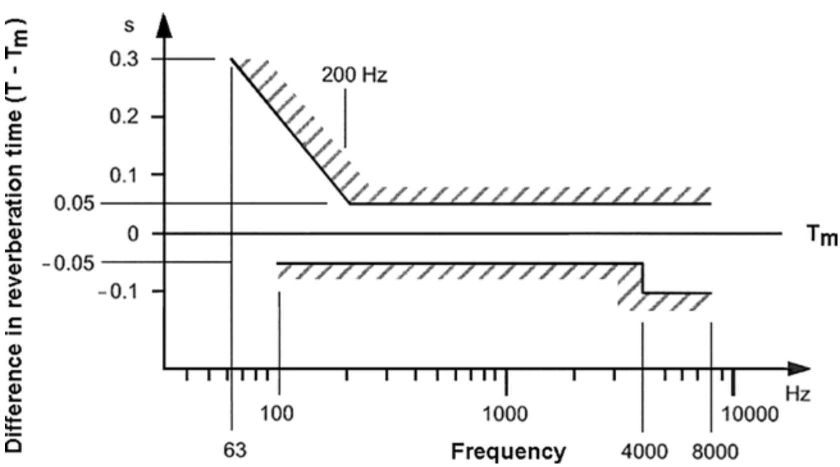

Figure 32.5 General recommendation for reverberation time for control rooms with an indication of the tolerance limits. (Reference EBU Tech 3276).

For classical music, the BR should be in the range of 1.0–1.3. For speech and rhythmic music, this value should be in the range of 0.9–1.0.

CALCULATING THE REVERBERATION TIME

When calculating the reverberation time, we utilize a simple equation, also known as the Sabine equation:

$$T = \frac{0.161 \cdot V}{A}\left[s\right] \text{ (unit: meter)}$$

or

$$T = \frac{0.049 \cdot V}{A}\left[s\right] \text{ (unit: foot)}$$

where

T is the reverberation time in seconds

V is the volume in cubic meter or cubic foot

A is the absorption:

$(\alpha_1 \cdot S_1) + (\alpha_2 \cdot S_2) + (\alpha_3 \cdot S_3) + \ldots + (\alpha_n \cdot S_n)$

where:

α is the sound absorption coefficient of a given part of the surface/material ($0 \leq \alpha \leq 1$)

S is the area of that given surface/material.

A sound absorption coefficient of 1 (α=1) is the same as an open window by which the sound leaves and never returns. An absorption coefficient of 0 (α=0) is comparable to a hard reflective surface like concrete; sound hitting the surface is reflected by 100%.

The unit for absorption is Sabine; so one square meter of a material that has an absorption coefficient of one equals one square meter Sabine.

Or: 1 m^2 of material with α=1.00 $\Rightarrow$ 1.00 m^2 Sab.

Hence 3 m^2 of material with α=0.50 $\Rightarrow$ 1.50 m^2 Sab.

In foot:

1 f^2 of material with α=1.00 $\Rightarrow$ 1.00 f^2 Sab.
Hence 3 f^2 of material with α=0.50 $\Rightarrow$ 1.50 f^2 Sab.

This absorption "A" is normally defined in frequency bands (i.e., octave bands or 1/3 octave bands).

The use of the simple Sabine equation is an approximation, and it assumes that the absorption material is well distributed among the faces of the room to get the result right. Still, if we try to calculate the reverberation time of a room which is treated with a material of 100% absorption, the reverberation time is *not* 0.0 second. If the volume is in the range of 45–50 m^3 (1600–1750 f^3), the reverberation time calculated is around 0.1 s.

The Sabine equation has been modified with the purpose to reduce uncertainties. Two of the most utilized are here.

Eyring equation:

$$T = \frac{0.161 \cdot V}{-S \cdot \ln\left(1-\bar{\alpha}\right)}\left[s\right] \text{ (unit: meter)}$$

$$T = \frac{0.049 \cdot V}{-S \cdot \ln\left(1-\bar{\alpha}\right)}\left[s\right] \text{ (unit: foot)}$$

Fitzroy equation:

$$T = 0.161 \cdot \frac{V}{S^2}\left[\frac{-x}{\ln\left(1-\alpha_x\right)} + \frac{-y}{\ln\left(1-\alpha_y\right)} + \frac{-z}{\ln\left(1-\alpha_z\right)}\right] \text{[s] (unit: meter)}$$

$$T = 0.049 \cdot \frac{V}{S^2}\left[\frac{-x}{\ln\left(1-\alpha_x\right)} + \frac{-y}{\ln\left(1-\alpha_y\right)} + \frac{-z}{\ln\left(1-\alpha_z\right)}\right] \text{[s] (unit: foot)}$$

where

x, y, and z are three sets of parallel faces in a (box-shaped) room.

The absorption coefficients provided for the room designer are normally obtained by placing the materials in a special reverberant chamber. The reverberation time is measured with a known area of the material placed in the room. The material is removed, and the reverberation time is measured again. Knowing the volume and surface area of the room, the absorption is calculated from the reverberation times measured. That is the standard methodology. However, the characteristics of a given absorbing material should also be measured in a space providing the same conditions as the room to be calculated. This is at least true regarding the frequency range below the Schroeder frequency.

The conclusion is that it is okay to make a simple calculation for the estimation of the reverberation time. However, in the final room you may expect some deviations from the predicted values, as the materials may behave differently in the lab and in the real room. (This is where experience comes in).

ABSORBERS

First and foremost, we use sound-absorbing materials when we want to control the acoustics (i.e., change the reverberation time of a room). However, we must acknowledge that different materials may exhibit properties that are frequency dependent. This is why we have to use different types of absorbers.

In general, there are three groups of absorbers at hand for the acoustical treatment of the room: porous absorbers, resonance absorbers, and membrane absorbers.

POROUS ABSORBERS

These absorbers include materials like foam, fabric, mineral wool, polyester fiber, and so on. They absorb frequencies in the high range, and they are very efficient, having absorption coefficients typically above 0.7. If the porous absorbers, however, are placed at a distance (25–30 cm or approximately 1 f) from a hard surface (wall or ceiling), they will provide efficient absorption down to approximately 100 Hz (see Figure 32.6). The porous absorbers are efficient when the surface of the absorber is at least one-quarter of a wavelength from the hard wall or ceiling behind it. So putting a carpet on the floor of a room will not damp low frequencies.

In some cases, the specification sheets of fancy-looking products like foam with the pyramid-shaped surface may exhibit a α-value above 1.00. This is not an error. However, this high absorption exists only above a certain frequency. The explanation

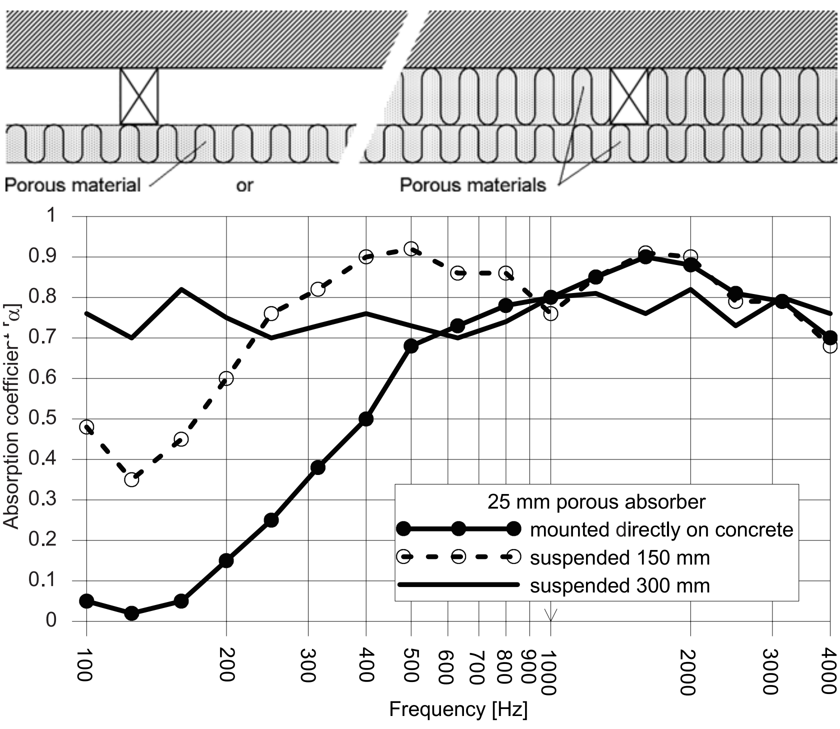

Figure 32.6 Porous absorber and generic absorption curve.

is that high frequencies can "see" a larger surface (the pyramid faces) in comparison to the area that the material covers. The lower frequencies can only "see" the average depth of the absorber (hence this absorber might be more efficient at lower frequencies if turned inside out).

RESONANCE ABSORBERS

Resonance absorbers are based on the principles of Helmholtz resonator and include perforated plates, perforated bricks, slitted walls, and so on. They are efficient in the midrange (200 Hz–5 kHz). The advantage of this absorber is the possibility of tuning the resonance. However, normally, it is preferred to design it to have a more broadband absorption around the resonance frequency by placing damping porous materials to the cavity (see Figure 32.7).

MEMBRANE ABSORBERS

Membrane absorbers work at low frequencies. They can be a part of the building construction: light walls, windows, floating floors, and so on. They are not very efficient.

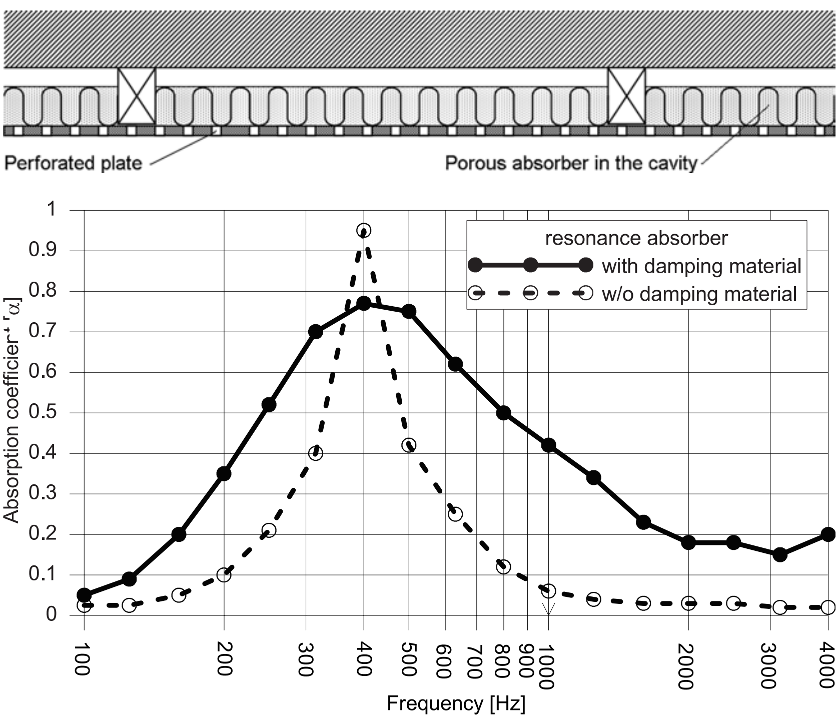

Figure 32.7 Resonance absorber and generic absorption curve.

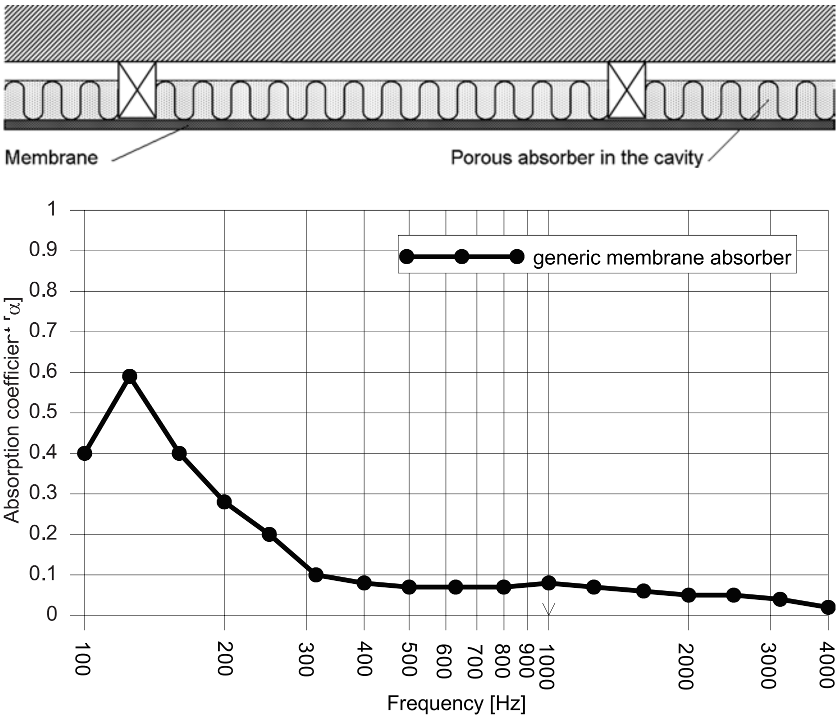

Figure 32.8 Membrane absorber and a generic absorption curve.

At the resonance, the absorption coefficient is in the range of 0.2–0.3 (measured in a diffuse sound field). However, if the membranes are a part of the construction, you already may have large areas available (see Figure 32.8).

In praxis, when placing the membrane absorber in a room, the apparent absorption coefficient of an efficient membrane absorber may range from below 0.5 to 2.5, depending on the placement in the room! The membrane absorber is most efficient when placed in the corners where the maxima (pressure) of the room modes are found.

SOUND DISTRIBUTION

Requirements regarding an appropriate sound distribution are normally related to larger rooms, like auditoriums, concert halls, and so on. However, in smaller rooms, this requirement can be related to frequencies in the modal region, as mentioned earlier. In most rooms for audio, it is vital to obtain a good low-frequency response not just in one single point but over a wider listening area or at least in a couple of listening positions.

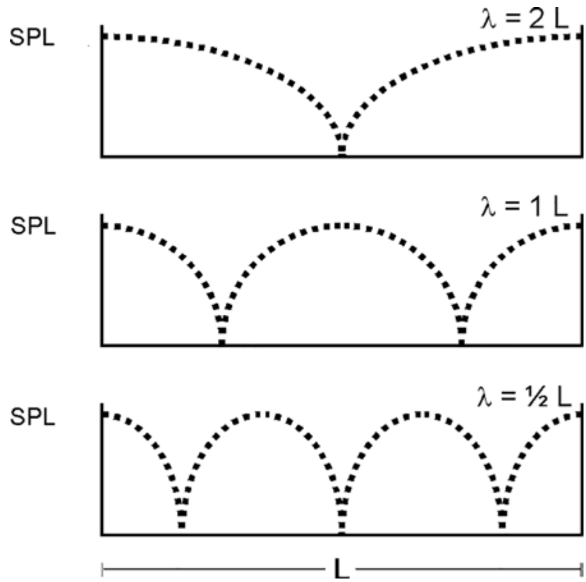

Figure 32.9 First-, second-, and third-order room modes in one dimension.

Figure 32.9 is a graphic representation of the first three axial modes – also called standing waves – through a single room dimension for an instant in time. Sound pressure maxima always exist at the room boundaries (i.e., the left and right side of the figure). The second-order mode has a maximum at the center as well, while the first- and third-order modes pass through a minimum at this point. The point where the sound pressure drops to its minimum value is commonly referred to as a "null." If there is no mode damping at all, the sound pressure at the nulls drops to zero. However, in most real rooms, the response dip at the nulls are typically in the −20dB range.

What we find in praxis is very often rooms that are not exactly box-shaped and rooms that do not have identical absorption on opposite walls. This may be a result of an attempt to prevent (reduce) the influence from the standing waves. For instance, this is done by building nonparallel walls. Just a 5-degree offset of a wall is changing the buildup of standing waves dramatically. This is fine; however, we get more or less unpredictable results when there is no longer symmetry along the median plane. Hence we may end up with different responses for each of the speakers in a stereo or a 5.1 channel setup, making it very difficult to calibrate for proper monitoring.

It is worth noticing that there always are standing waves (modes) present in a room. It is just a question of how strong the modes are.

The sound distribution in larger rooms like auditoria for speech and music is related to the amount and equality of the sound received and perceived across the audience area or across the stage. Some of the measures are described in the "Large Room Parameters" section.

BACKGROUND NOISE

In general, most control rooms and studios have to be kept quiet. Noisy equipment must be built into racks having airtight doors, and a cooling system placed somewhere else. Hopefully!

However, from time to time, cables seem to grow out of the rack box so the door cannot be closed properly.

In larger studios, concert venues, and churches, it is the HVAC or outside traffic noise that set the limits for the activity (see Chapter 31: NR, NC, PNC, RNC, and RC Curves).

ECHOES AND FLUTTER ECHOES

Single reflections that arrive 30–50 ms (or later) after the initial sound is perceived as an echo. This is a typical experience when playing music from a stage and receiving the echo from a reflective rear wall of the room. However, a real (slap-) echo will never occur in smaller rooms due to the limited distances between the walls.

Flutter echoes, however, exist in all room sizes but mostly in smaller rooms. It can occur between parallel surfaces like two walls, between control room windows, between floor and ceiling, between control room window and backside of loudspeakers, between computer monitors, between table and lamp screens, and so on. The repeated reflections create a kind of tone itself due to the fast repetition.

Flutter echoes should never be present in any room for audio. The phenomenon is easy to remove: Offset of parallel surfaces, supply sound absorption if necessary, or install diffusing elements that remove the flutter but retain the sound energy in the room.

EARLY REFLECTIONS

The definition of early reflections and their influence on the perceived sound depends on the size of the room and thus the level and time of arrival. In the small room, early reflections, in general, should be avoided. In the large room (for music) the early reflections provide valuable information on room size and directional definition.

EARLY REFLECTIONS IN THE SMALL ROOM

Early reflections in the small room may affect the perceived frequency response in the listening position, typically in the midrange. The problem is comb filtering and should always be avoided (see Chapter 22: Summation of Audio Signals).

It is not always the best solution to remove reflections by adding absorption. In most rooms, too much porous absorption is introduced. The result is a room with too short reverberation time at higher frequencies. Hence it is a better idea to involve diffusion to retain the high-frequency energy in the room. There are several solutions for this, the Schroeder diffuser being one. These devices may introduce diffusion in one or two dimensions. Most diffusers also exhibit some absorption in the active frequency range.

EARLY REFLECTIONS IN THE LARGE ROOM

In the large room, especially the concert hall, early reflections are important both to the listeners in the room as well as to the musicians and even the recording engineer when placing and aligning microphones.

At a given distance from a sound source, the first sound received is the direct sound. This is followed by individual early reflections; later again, the higher-order reflections arrive from all directions and are characterized as the diffuse sound field. The time between the direct sound and the early reflections is called the initial time delay gap or ITDG (see Figure 32.10).

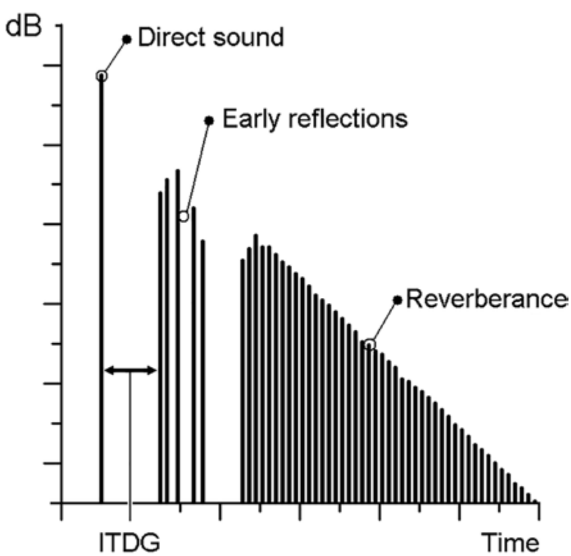

Figure 32.10 The reflectogram shows the direct sound, the Initial Time Delay Gap, the early reflections, and the reverberance tail.

LARGE ROOM PARAMETERS

In auditoria design and evaluation, there has always been an effort to find quantities that objectively describe parameters that express what is subjectively perceived. Some of these measures are briefly mentioned here. They in general describe the relation between a sound source – it can either be a generic sound source, a human voice, a musical instrument, a loudspeaker, and so on – and the sound field received in the point of observation. Several of these parameters can be calculated n simulation software like EASE® and ODEON®, which are used by many sound engineers.

D/R-RATIO

The D/R-ratio expresses the ratio of the level of direct sound from a sound source received in a given position of the room to that of the level of reflected/diffuse sound received in the same position. The directivity of the sound source has to be considered.

REVERBERATION RADIUS AND CRITICAL DISTANCE

The reverberation radius r_H is positions in the room where the amount of direct sound equals the amount of reflected sound. (D/R-ratio = 1). For a spherical sound source, the reverberation radius r_H can be expressed by this:

$$r_H = \sqrt{\frac{A}{16\pi}} \approx \sqrt{\frac{A}{50}} \approx 0.141\sqrt{A} \approx 0.057\sqrt{\frac{V}{T}}$$

where

A is the absorption in m^2 (or f^2)

V is the volume in m^3 (or f^3)

T is the reverberation time in seconds.

The name reverberation "radius" (in German Hall Radius, hence r_H) is only valid if the positions around a given sound source are forming a fixed radius. However, this is not the case if the sound source is directive (as is for instance, with PA/SR speakers). Hence we use the expression critical distance, r_R or D_c:

$$D_c \approx 0.141\sqrt{A \cdot Q_s} \approx 0.057\sqrt{\frac{V \cdot Q_s}{T}}$$

where:

Q_s is the directivity factor of the sound source. (An omnidirectional source has a Q_s of 1).

It can be seen that the critical distance from a (monitor-) loudspeaker will change with frequency as most speakers are more or less omnidirectional at low frequencies but directive at higher frequencies.

Clarity, C_{80}, C_{50}, C_7

Clarity is a way to evaluate the amount of first arriving sound to later arriving sound. Originally this was related to predicting clarity of different modes of music and expressed as the ratio in dB of the energy arriving before and after 80 ms relative to the time of the first arrival: C_{80}.

Later other measures using the same methodology have been introduced:

C_{50}: The ratio in dB of the energy arriving before and after 50 ms relative to the first sound. This calculation is used to predict the articulation of human speech.

C_7: The ratio in dB of the energy arriving before and after 7 ms relative to the first sound. This calculation is used to predict the strength of direct sound sources for localization in auditoria.

Strength, G

The sound strength (or relative sound level) *G* is measured using a calibrated omnidirectional sound source and is the ratio in dB of the sound energy of the measured impulse response to that of the response measured at a distance of 10 m from the same sound source in a free field.

The strength measurement is a practical way to evaluate the sound distribution in an auditorium. For further analysis, the received sound can be evaluated in relation to the time of arrival. That is, G_{80} expresses the strength of the sound within the first 80 ms relative to the arrival of the first sound. The sound strength of the late-arriving sound, G_L, consists of sound energy arriving at the receiver more than 80 ms after the direct sound.

Further analyses of the early and late arrival of lateral sounds, G_{EL}, and G_{LL}, respectively, can be obtained when changing the standard omnidirectional measurement microphone with a figure-of-eight microphone. Performing measurements perpendicular to the main axis can provide information on the envelopment of the sound.

BIBLIOGRAPHY

Adelmann-Larsen, Niels Werner, et al.: Suitable reverberation times for halls for rock and pop music. *Journal of the Acoustical Society of America* Vol. 127, No. 1 (January 2010).

Bech, S.: *Perception of reproduced sound: Audibility of individual reflections in a complete sound field, II*. AES 99th Convention (1995). Preprint 4093.

Beranek, Leo L.: Sabine and Eyring equations. *Journal of the Acoustical Society of America* Vol. 120, No. 3 (2006).

Cox, Trevor J.; D'Antonio, Peter: *Acoustic absorbers and diffusers*. Spon Press, London (2004).

EBU Tech. 3276 and EBU Tech. 3276 supplement-1: Listening conditions for the assessment of sound programme material: multi channel sound.

Fuchs, H.V.: Zur absorption tiefer Freqyenzen in Tonstudios. *Sonderdruck aus Runfunktechnische Mitteilungen*, No. 36, 1–10 (1992), Heft 1.

Geddes, Earl R.: *Small room acoustics in the statistical region*. Proceedings of the AES 15th International Conference: Audio, Acoustics & Small Spaces (1998).

Hunecke, Jörg; Fuchs, H.V.; Zhou, Xiarou; Zhangg, Ting: Einsatz von Membran-Absorbern in der Raumakustik. Bericht. 17. Tonmeistertagung (1992).

IEC 60268–13 Sound system equipment – Part 13: Listening tests on loudspeakers.

ISO 16283–1; Acoustics – Field measurement of sound insulation in buildings and of building elements – Part 1: Airborne sound insulation.

ISO 3382–1:2009; Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces.

ISO 3382–1:2009; Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms.

ISO 3382–3 2012; Acoustics – Measurement of room acoustic parameters – Part 3: Open plan offices.

Manual for EASE 4.3 (Enhanced Acoustic Simulator for Engineers) by Ahnert Acoustic Design/AFMG Technologies GmbH.

Manual for EASSERA xxx

Randeberg, R.T.: Adjustable slittet panel absorber. *Acta Acoustica united with Acoustica* Vol. 88, 507–512 (2002).

Rasmussen, Birgit; Rindel, Jens Holger; Henriksen, Helge: Design and measurement of short reverberation times at low frequencies in talk studios. *Journal of the Audio Engineering Society* Vol. 39, No. 1–2 (January–February 1991).

Sabine, Wallace Clement: *Collected papers on acoustics*. Dover Publications, Inc., New York (1964).

Schroeder, M.R.: Frequency-correlation functions of frequency responses in rooms. *Journal of the Acoustical Society of America* Vol. 34, No. 12 (1962).

Schroeder, Manfred R.: Diffuse sound reflection by maximum-length sequences. *Journal of the Acoustical Society of America* Vol. 57 (January 1975).

Voetmann, Jan: *50 years of sound control room design*. AES Convention 122 (2007). Preprint 7140.

Voetmann, Jan; Klinkby, John: *Review of the low-frequency absorber and its application to small-room acoustics*. AES Convention 94 (1995). Preprint 3578.

Walker, R.: *A controlled-reflection listening room for multi-channel sound*. AES Convention 104 (1998). Preprint 4645.

Walker, R.: *Optimum dimension ratios for small rooms*. AES Convention 96 (1994). Preprint 4191.

Walker, R.: *Room modes and low frequencues responses in small enclosures*. AES Convention 100 (1996). Preprint 4194.

CHAPTER 33

Listening Tests

CHAPTER OUTLINE

The most valuable tool for the assessment of the reproduced sound is the hearing. The final judge of any sound production or any piece of audio gear is your ears. There is no piece of measurement equipment to substitute all aspect of human perceptual habits. Thus it is obvious to employ listening tests when appropriate. However, this is not exactly an easy task. If you want to obtain valid results, many factors must be considered – and controlled. If there are too many uncertainties or even a few, the test has no value – other than considering it as a pilot for the next. It takes serious planning to conduct reliable listening tests. This chapter goes through some aspects and tries to provide an introduction to simple tests related to reproduced sound. If you want to go deeper into the topic and are planning to design listening tests, please look up the references.

INITIAL CONSIDERATIONS

In the field of audio, there are various needs for testing. Electric/electroacoustical measurements provide nice repeatable results if performed correctly. However, the results of these measurements do not necessarily closely relate to the perceived sound. If we want to make perceptual measurements, we must introduce a perceptual model, a model that provides objective results based on how our ears and brain would conclude on the stimulus presented.

We have some "semi-perceptual" models such as for the measurement of speech intelligibility expressed by the Speech Transmission Index (STI) (see Chapter 28: Other Measurement Systems). The acoustical signal designed for these measurements, combined with the associated analysis, to some degree include or emulate conditions for the perception of the spoken word. Does it fully encompass all involved parameters? No. However, we are happy to use this methodology. Why? Because this is fast. Normally, it is extremely resource-demanding to make measurements involving humans. We'll save that option for more advanced experiments.

So measurements based on human perception are practical. However, we are not quite there yet even though lots of research, including the ongoing development of artificial intelligence, brings us closer every day. Until then, (and presumably beyond) we use "the real thing," listening tests.

Initially, we have different options to consider, depending on who we ask, and about what. The **hedonic** test is all about like/dislike and is essentially related to decisions based emotion being in the affective domain. The subjects most likely are untrained, so-called naïve listeners, but they do have an opinion. **Discrimination** is a rather simple test to tell whether A is different from B. This may involve experienced or trained listeners. **Quantification** is a question of how much. Now it definitely takes trained listeners to deliver trustworthy results. **Description**, the subject must describe how it sounds. Again, the subjects must be well trained and in this context described as expert listeners.

If deciding on making listening tests, in many cases, various standards provide recipes on what and how. However, sometimes you have to design the test by yourself, as no standards exist. One area missing a standard is listening tests for PA/SR systems [1]. These tests often become "shootouts" with too little control of too many conditions.

BIAS

Let us start with an issue that too often has too little attention. If considering listening tests to be objective, avoid any bias. Easy to say, less easy to practice, however. What is bias? Bias is any propensity of making decisions due to preferences not related to the test question. So what creates bias? Almost everything. It is the experimental design that determines how well you keep bias out of the test. Here are some examples:

When comparing the sound of two loudspeakers A and B, some subjects may like A better because:

- A is black, B is green, and in your opinion black is nicer than green. Precaution: Make the loudspeakers invisible. Perform the test as a blind test. Or even double-blind test where neither the assessors nor the test administrator knows what is presented.
- A is more expensive than B. A must be better. Precaution: You should not provide any price information to the assessors before the listening test.
- A is a renowned brand, B is a "no-name" brand. Precaution: You should not provide any information about brands before the listening test, neither

intentionally nor unintentionally. Even keep empty packaging out of sight to the assessors participating in a listening test.
- A sounds louder than B. Precautions: Devices under test – including loudspeakers – must be aligned to the same loudness. If the question was to define the sensitivity, this is more easily achievable by acoustical measurements.
- When activated, there is a small "click" in the sound reproduced by B. Precaution: Make sure that the sound samples, timing, level setting, and assistive devices are identical for each stimulus presented.
- A must be better because my fellow assessor just moaned when he heard B. Precaution: Instruct assessors to be silent during the listening session – or separate them.

And the list of possible bias issues continues. For more on bias, see [2, 3, 4, 5]

TEST QUESTION

There is a reason for any test. It presumably starts with a hypothesis such as "These two sound systems sound equally good." This hypothesis, which can be proven true or false, is challenged by a research question that also may turn into a test question. The question can be quite simple, like "Which system sounds best?" However, what does the term "best" mean in this case? The question may have such a general character that is not possible to answer directly, or the assessors create answers based on their personal understanding: A sounds best - implied understanding: it has more bass. And another answer: B sounds best - implied understanding: it has less distortion. Thus the main question must be broken down to simpler subsets of test questions to counteract divergence. Or the terms and attributes applied must be so well-defined that all assessors have a common understanding of their meaning.

The questions also, to some degree, is related to the assessors' listening skills. Naïve listeners essentially only can answer one question, "Which do you like best?" However, often, we are interested in how the sound is perceived, not opinions. If assessors do not understand the question or the task, the outcome of the test is not reliable. It is a little like measuring SPL with a thermometer. The thermometer can be a fine instrument; however, it cannot provide any relevant measurement results.

ATTRIBUTES

Attributes are defined as a perceived characteristic of a hearing event according to a given verbal or written definition. Daily much sound is described by words like "cool," "awesome," or "sucks." However, when it comes to qualified and qualitative descriptions, a more specific and scientifically agreed vocabulary is needed. To reach a proper result from a listening test, each of the attributes should only describe one dimension or one parameter. In practice, far more than 100 words are acknowledged to describe sound. These descriptors must be selected and organized to cover the relevant aspects of the reproduced sound.

THE SOUND WHEEL

To find and define descriptors for the evaluation of reproduced sound, the Sound Wheel is a good starting point. Wheels like that are generally known in many areas – from wine aroma to concert hall acoustics to psychology. The Sound Wheel has been developed by SenseLab of DELTA (now FORCE) for the purpose of listening tests (see Figure 33.1).

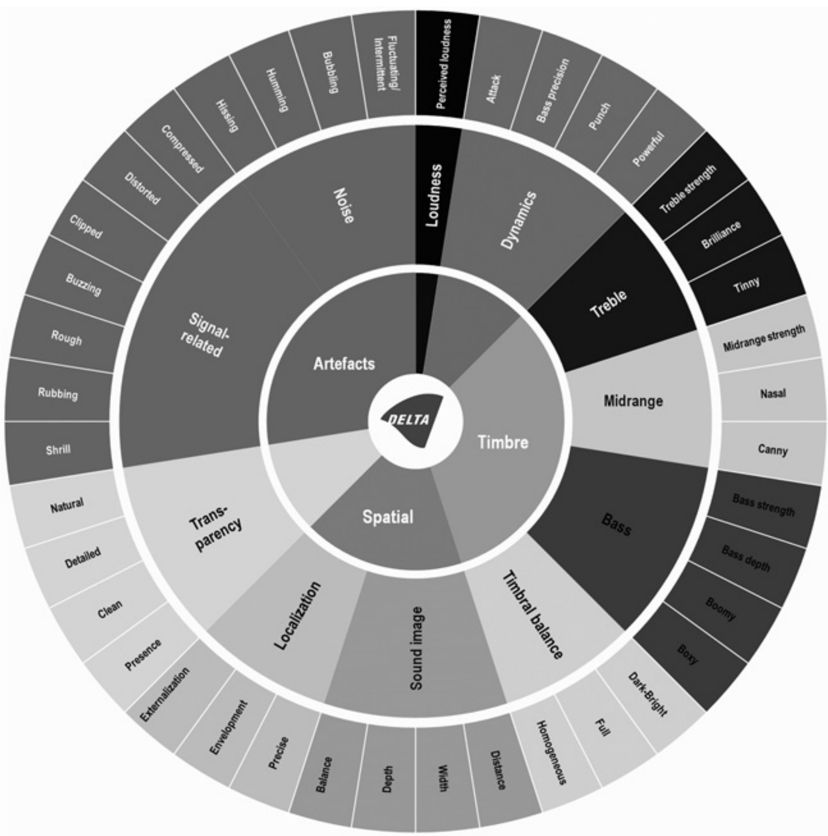

Figure 33.1 The Sound Wheel for reproduced sound. Inner ring: Main Groups, middle ring: Categories, outer ring: Attributes.

Source: (Courtesy of SenseLab/Force). (Further info: www.madebydelta.com/senselab).

The Sound Wheel provides a rather detailed set of descriptors that may apply to the reproduced sound. In general, the assessors must be trained to have the same understanding of the descriptors. Especially when differences become very subtle, the training is highly necessary. The structure of the Sound Wheel guides to understand the meaning of the individual words by grouping in categories. For more convenient reading, please see Table 33.1, where the content of the wheel is listed:

Table 33.1 Content of the Sound Wheel. A detailed description of the individual attributes is found in Pedersen and Zacharov [6].

Main group	Categories	Attributes
Loudness	Loudness	Perceived loudness
Dynamics	Dynamics	Attack
		Bass precision
		Punch
		Powerful
Timbre	Treble	Treble strength
		Brilliance
		Tinny
	Midrange	Midrange strength
		Nasal
		Canny
	Bass	Bass strength
		Bass depth
		Boomy
		Boxy
	Timbral balance	Dark-Bright
		Full
		Homogenous
Spatial	Sound image	Distance
		Width
		Depth
		Balance
	Localization	Precise
		Envelopment
		Externalization
Transparency	Transparency	Precence
		Clean
		Detailed
		Natural

Main group	Categories	Attributes
Artifacts	Signal-related	Shrill
		Rubbing
		Rough
		Buzzing
		Clipping
		Distorted
		Compressed
	Noise	Hissing
		Humming
		Bubbling
		Fluctuating/intermittent

The attributes selected must be relevant to the test; for instance, “Externalization” is primarily related to the assessment of headphones and/or spatialized sound. In the group of signal-related artifacts, distortion and clipping are two of the attributes described. Clipped audio is distorted; however, distorted audio is not necessarily clipped.

The Sound Wheel does not, of course, cover all descriptors applied to listening tests. For instance, in the field of spatial audio, a broader selection of terms often is necessary [7, 8].

METHODOLOGIES

When designing a test, it is essential to find a methodology that provides relevant answers. The assessors’ listening skills and the stimuli to be presented are design factors. The expected magnitude of the perceptual differences are also parameters that may determine how to design the test. If the differences are too obvious, it is a problem. A test wasn’t needed then. Or the opposite way, if the expected differences are so small, the answers are by chance rather than by perception, it is also a problem.

RANKING

One method used in perceptual testing is to perform a ranking of the test objects. If presenting five sound objects – A, B, C, D, E, the task is to rank these objects based on the attribute(s) defined for the test. A method like that works nicely if each of the assessors individually can control to which of the objects he or she is listening. If several subjects simultaneously take part in the test, it is problematic not being able to control what and when. Thus this method does not directly apply to groups of listeners.

AB

The AB-method is the most applied for listening tests: Here is A and here is B, which one represents this attribute the best? Essentially, the assessor must determine whether it is A or B. However, what if A and B are equally good? Or the assessor can't hear a difference? Should that leave the possibility of a third question: They are equally good? Or should the assessor be forced to choose one of the two? Both solutions apply to AB-tests.

Experiments [9] have shown that the order of presentation of the sound objects also matters, even if the assessor has a free choice of order. It is known that the so-called "anchor effect" has an influence [10]. A sound sample presented becomes an anchor for the next sample. The brain tries to create an audible "white balance" of the sound presented which implies that you will perceive how the second sample differs from the first rather than being able to assess a specific descriptor. In any AB presentation, the AB setting should be repeated. Further, the opposite order, BA, should be included to reduce the influence of bias.

In evaluating more than two sound objects, each object must be compared against all others. Having A, B, and C leads to these comparisons: AB, AC, BC, BA, CA, and CB (in a randomized order).

It should be noted that in many tests many more repeats are carried out to measure the consistency of the subject's response.

Note: To experience the anchor effect, try this: Record 20 seconds of pink noise in your DAW. In a middle section of this recording (from time 8 sec to time 12 sec) apply a one-octave band cut (−24 dB) at 1 kHz. Now play from the beginning of the file and ask yourself whether the second section of the pink noise (after the cut) sounds the same as the first part of the pink noise (before the cut). Repeat with another octave band such as 4 kHz. You will experience that your acoustic memory, in this case, is less than four seconds!

ABX

One way to overcome some of the AB problems is by applying the ABX (a variant is known as the 3 Interval Forced Choice).

Two different sound objects or samples (A and B) are presented together with a third (X), which is either A or B randomly selected. It is an adaptive method with one question for every task: Is X identical to A or B? The advantage is that assessors do not necessarily need to have the same understanding of descriptors. They only need to hear a difference. However, to reach a decent result, all assessors should have the same training.

Another way to ask a question when presenting three samples, of which two are identical, is: Which one is different?

Essentially the data acquired also may provide information about the individual listeners' ability to assess the material presented. If the result of one assessor seems to be by chance, while all others clearly can hear the differences, then there is the possibility to exclude the results of that assessor.

SCALES

In some tests, it is preferred to apply scales. Depending on the test and test objects, typically more detailed information can get acquired from a test like that.

The hedonic scale is for tests where the answer is based on like/dislike. It is a 9-point scale that ranges from "Like Extremely" to "Dislike Extremely," see Figure 33.2a.

There are more ways to employ a scale. One way is to establish two opposing values or two anchors to create a semantic differential scale. The opposing anchors could be warm/cold, light/heavy, soft/hard, and so on.

The IEC recommendation for an evaluation of the fidelity of consumer loudspeakers suggests a scale from 0 to 10. The outer values are the anchors of the test, and subjects should use the values from 1 to 9 for their assessment (see Figure 33.2b).

Grade	Score
Like extremely	9
Like very much	8
Like moderately	7
Like slightly	6
Neither like nor dislike	5
Dislike slightly	4
Dislike moderately	3
Dislike very much	2
Dislike extremely	1

Figure 33.2a The 9-point Hedonic Scale.

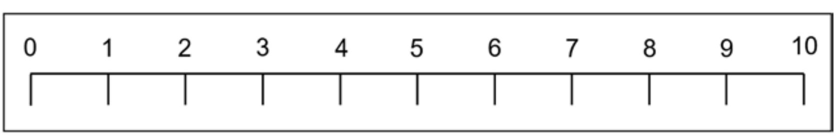

Figure 33.2b The IEC scale with end points as anchors (extreme values).

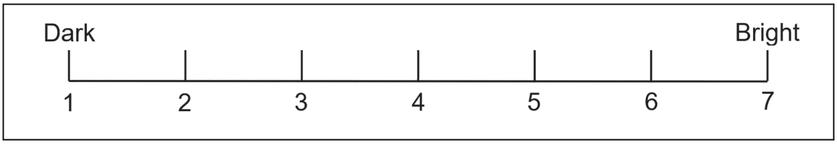

Figure 33.2c Semantic scale with free choice between two of an attribute.

Grading value	Estimated	Perceived impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Figure 33.3 ITU-R five-point continuous impairment scale.

Returning to the Sound Wheel, in the category of timbral balance, we find the attribute “dark - bright.” By applying these terms as the opposing poles, any value between the two can be selected (see Figure 33.2c).

One of the most applied scales is the ITU-R 5 point continuous impairment scale ranging from “very annoying” to “imperceptible” (see Figure 33.3). This scale mostly applies to the detection of artifacts of the audio signals such as found in bit-compressed audio.

It does not necessarily require a standardized scale to make tests. There are methods for organizing assessors’ feedback in a way based on her or his personal impression and perception, so-called free-choice profiling [2]. The data may become more accurate; however, there is a special task to calibrate the data in these cases [2].

STANDARDS AND RECOMMENDATIONS

There are international standards and recommendations for listening tests. Which standard you may find appropriate depends on what you need to test. The ITU-T (International Telecommunications Union - Telecommunication Standardization Sector) and ITU-R (- Radiocommunication Sector) has several recommendations and guidelines essentially developed for listening tests of equipment and algorithms applied in the telecommunication business. ITU-T P.800 series are methods for subjective determination of transmission quality [11]. ITU-R BS.1116–1 is a method for testing of intermediate quality [12], and the ITU-R BS.1534–3 Multi Stimulus test with Hidden Reference and Anchor, (MUSHRA) is designed to test for small impairment, such as in codecs [13].

As with the ITU standards, others such as EBU Tech. 3276 / 3278 [14, 15], the AES 20 [16], and IEC 60.268–13 [14] also provide information on the conditions for the listening tests, such as size and acoustical conditions of the listening rooms, quality of equipment, and so on.

SELECTING SUBJECTS

In many experiments, it is the availability that determines which subjects participate in the test. If you are at a private lab, if you do not have a trained panel of assessors, you may appoint your colleagues for the job. If you are at a college or university, it is usually possible to get students to attend, perhaps even students who are interested in audio. It has the advantage that it is a homogeneous group, in which (hopefully) everyone has good hearing. The IEC 60.268–13 standard for home speaker listening tests recommends including subjects of both sexes and all ages. Normally gender does not affect the listening test itself. It may do so, however, if you ask for opinions. The number of participating subjects may vary depending on the nature of the test. However, essentially, the larger the listening panel, the greater the weight the result gets. The individual standards and recommendations usually provide an appropriate number.

HEARING TEST

It is important to know about the subjects' hearing. All people are subject to age-related hearing loss. Some people also may suffer from noise-related hearing loss. Obviously, if you want to perform listening tests on hearing aids, it is relevant to include reference groups of people with hearing loss. If you are dealing with basic psychoacoustic research or it is a matter of being able to determine small differences in the stimuli presented, then the subjects should have normal hearing.

In a listening test [1] involving 80 subjects in which all had an interest in audio but otherwise were regarded as untrained listeners, all subjects were asked to indicate whether they had hearing problems. It turned out that approximately 25% of the participants had an acknowledged problem, but this did not, however, affect the result!

TRAINING

If a trained listening panel is required, it is normally important that everyone undergo the same training. All subjects should have the same understanding of the concept and the attributes applied. Further, the ability to hear and interpret the differences and artifacts that may occur in a test is important. If you do not know what to listen for, you may not hear it [18, 19, 20, 21].

SELECTING STIMULI

The sound samples chosen as stimuli for listening tests must have quality so high that it is possible to determine differences when reproducing the signals reproduced by the equipment under test. Otherwise, it is regarded as a bias for the test. Typically, the stimuli exhibit the kind of complexity that occurs in music or speech. In their paper [22], the authors identify four domains for characterization of stimuli:

The **physical domain**, which is basically what can be measured with a sound level meter, such as level, frequency range, and spectral distribution. For example, if you need to hear the difference between reproduction in speakers or headphones, it is an important part of the study to know the content of bass in the original recording compared to the actual reproduction.

The **sensory domain**, where it is important that the selected material contains information that takes into account the attributes that are being investigated. To determine how much bass is reproduced by a loudspeaker, the recording must contain bass. If testing the dynamic reproduction of a headphone, appropriate dynamics must be present in the stimuli.

The **affective domain**, which by and large, is an overall assessment of whether you like or dislike the quality of what you are presented with.

The **statistical domain**, where the material's relevance to the media must be taken into account. This could mean that there should not be too many string quartet recordings among the stimuli selected for a Hip-Hop speaker evaluation.

The EBU has developed a set of audio stimuli to be applied in listening tests: Sound Quality Assessment Material (SQAM), recordings for subjective tests[23]. This set of recordings primarily consists of single musical instrument recordings and voice recordings. These recordings are relatively short but are long enough to present the characteristics of the given sources.

LISTENING CONDITIONS

As previously mentioned, in the standards, there is good information on how acoustical listening conditions should be for listening tests, especially concerning loudspeaker evaluation. This is of course important, as loudspeaker-reproduced sound is not only dependent on the speakers but also on the space in which the speakers are installed. Furthermore, one must also consider how the subjects are placed in the room, in relation to providing an optimal listening position. Thus, it easily becomes critical when assessing multichannel formats, where there may only be one single sweet spot [24].

REVERBERATION

The reverberation time of the listening room is of great importance. The optimal reverberation time is often determined by the volume of the room. Broadcast standards such as the EBU and ITU employ the same calculation for the optimum reverberation time:

$$T = 0.25\ (V/V_0)^{(1/3)}\ [\mathrm{s}]$$

where

V = Volume of the listening room [m^3]

V_0 = A reference volume of 100 m^3.

Thus a space of 50 m^3 has a target reverberation time of 0.2 s.

For evaluation of loudspeakers for domestic use, the IEC recommends an average reverberation time of 0.3–0.6 s in the frequency range of 200 Hz to 4 kHz. Outside this frequency range, the tolerance field is slightly wider.

While the reverberation time more or less describes the diffuse sound field of the room, it is important to reduce strong single reflections at the listeners' position. Different rules exist, however, a 15 dB suppression of reflections within the first 15 ms after the direct sound (or 20 dB within the first 20 ms) are practical goals [25].

BACKGROUND NOISE

Keeping the background noise low is, of course, extremely important in obtaining a decent acoustical signal-to-noise ratio in the test. The noise sources are in general either installed ventilation or heating systems, or sources outside the room such as road traffic or activities in adjacent rooms.

A typical requirement to the background noise provided by the IEC is <NR15, which is in the range of 20–25 dB(A). This requirement is for the evaluation of consumer products. In broadcast, the EBU recommends the noise level to be less than NR10, and required under no circumstances to exceed NR15.

Even for listening tests using headphones, the background noise level should be kept low. The isolation of the earmuffs is not always particularly efficient.

CALIBRATION

The perceived frequency balance depends on the level presented. For instance, we do not hear bass that well at low levels (see Chapter 7: The Ear, Hearing and Level Perception). Thus an awareness of the level of the reproduced sound is important. First of all, stimuli to be compared must (in most cases) exhibit the same loudness. Second, listener fatigue due to excessive SPLs is also an issue to consider [26].

Calibration of a setup secures decent listening levels and provides important information for the evaluation of the test. The SPL of the stimuli can be measured as an SPL in the listening position, applying a sound level meter, or it can be measured applying an artificial head (and torso) which has calibratable microphones.

INSTRUCTIONS

Besides establishing the best possible physical and acoustical conditions, it is extremely important to provide precise instructions to the subjects. It is normal to present written information on the purpose and the character of the test. The instruction also must provide all the information about the subject's task, how to manage the feedback system in this test, the time available, and so on. It is also normal to read the text aloud to the subjects. Further, all subjects in a test must have exactly the same information.

DATA ACQUISITION

When acquiring answers during a listening test, this naturally can be carried out using pencil and paper. However, it takes only a few answers before the post-processing work of transferring data to spreadsheets becomes an excessive workload. Therefore, the starting point is often computer-based data collection using a graphical user interface (GUI) specially designed for this purpose. In this way, data can be collected directly in the database. If you have some experience with spreadsheets, you can design your own system for simpler testing. However, there are a number of companies that offer software for the purpose, such as webMUSHRA (see Figure 33.4).

If there are many assessors in the same session, alternative methods may come into play. In [1], 80 assessors used their mobile phones as a terminal for the webservice from SurveyMonkey®. A method like that requires a high-capacity WiFi.

FINDING THE RESULTS

When all the data are collected, and a database is created, then comes the part that provides all the answers, hopefully.

There are many ways to find the results. Sometimes the only thing needed is to count points, so many preferred A and so many B. Perhaps recalculate to a percentage. Other times more analyses are needed to find out how consistent data are. For instance, are there any outliers to be excluded from the set of data? How do these data translate into a perceptual scale, and so on?

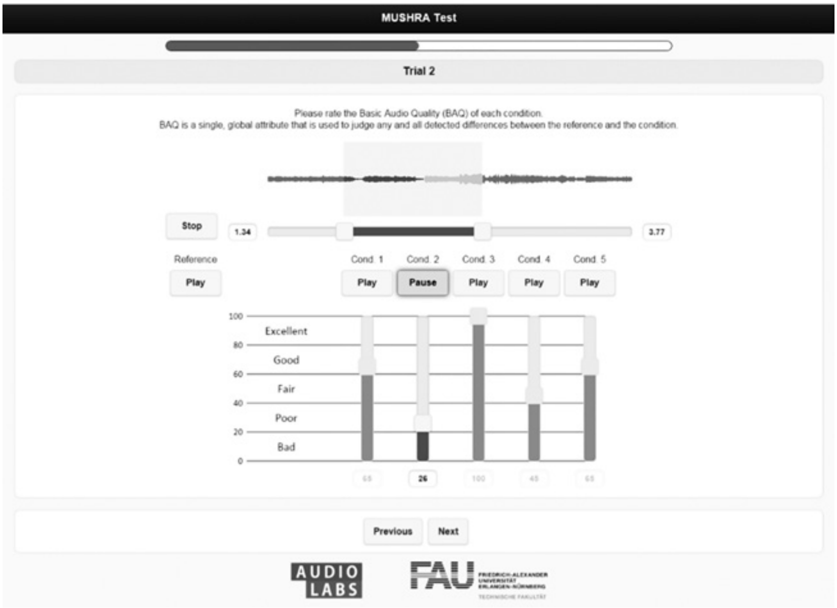

Figure 33.4 An example of software for listening test (MUSHRA).

We will not go into detail on statistics but will just make a short explanation of a few simple terms and expressions often seen in connection with the analysis of data acquired from listening tests. Many of these tests can be carried out by standard spreadsheet programs with a statistics module.

ANOVA, Analysis Of Variance - tests whether the means of two or more populations are equal by comparing the variance between the group-means versus the variance within groups. This test is generally regarded as a basic test for most datasets. Let us assume in a test we have three loudspeakers and five subjects. We have the hypothesis that the loudspeakers sound the same. We then can look into (calculate) the variance of the mean results for each speaker, and we can look at the variance between subjects. We can find the ratio of "between groups" divided by "within groups." The higher the ratio, the more likely it is that groups have different means and by that rejecting our hypothesis.

Average - see Mean

BAQ, Basic Audio Quality - as defined by ITU-TR, this single, global attribute is used to judge any and all detected differences between the reference and the object.

Confidence interval - in statistics, one of the most common ways of indicating the uncertainty of a measured parameter. This uncertainty can be described by a confidence interval, which is determined so that the interval with a certain probability (often selected 95%) contains the true value of the parameter.

Mean - is the same as "Average," which is a summation of the numbers in a list, divided by the number of numbers of that list. *Example: List of numbers: 1, 2, 3, 5, 7. The average is 3.6.*

Median - in a list of numbers, this is the middle number. Example: List of numbers: 1, 2, 3, 5, 7. The median is 3.

MOS, Mean Opinion Score - is the arithmetic mean value of all individual values on a predefined subjective scale that a subject assigns to his opinion when assessing the performance of a system's quality. A typical predefined, absolute category scale is found in Table 33.2.

Further definition and application is found in the ITU-T Recommendation P.800.

OLE, Overall Listening Experience - is an affective attribute which incorporates all aspects important to the individual assessors, including their preference for music genre and audio quality.

Outlier - an outlier in a set of data is a value that is at an abnormal distance from other values in the set. Most often the outliers are defined with reference to the values of the lower and upper quartiles (the 25th and 75th percentiles) also named Q1 and Q3. The Inner Quartile range (IQ) is defined as the difference between Q3 and Q1.

Additionally, "fences" are defined from the Q1, the Q3, and the IQ values:

Lower inner fence: Q1–1.5*IQ

Upper inner fence: Q3 + 1.5*IQ

Lower outer fence: Q1–3*IQ

Upper outer fence: Q3 + 3*IQ

Points outside an inner fence are considered mild outliers. Points outside an outer fence are considered to be extreme outliers. In each listening test, considerations of whether removal of outliers from the dataset are relevant.

Table 33.2 Absolute category scale.

Rating	Label
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

t-test - a statistical measure to define whether there is a significant difference between the mean value of two sets of related data (and not more than two sets).

FUTURE OF LISTENING TESTS

We must acknowledge that we still do not know everything about human hearing and auditory perception. In recent years, research findings indicate that we should pay more attention to the time in listening. Our ability to focus on sound depends on the stimuli presented to us from the earliest age, of which language and music are important factors. Thus, in listening tests, differences between listeners' should be taken into account to a larger degree. Lund and Mäkivirta suggest in their paper [27] that we should consider "slow listening" when we design listening tests. For instance, a phenomenon like listeners' fatigue does not show immediately but comes with the duration of listening. Also, it takes time to learn how the impairment of audio actually may sound. Typically the sound of lossy formats sound absolutely fine when first introduced. However, over time, years, the more we listen and the more we get acquainted with the sound and the audible artefacts, we find the impairment more annoying.

The listening test is still the most important way to assess audio. However, we must learn to design our tests to satisfy our perception [28].

REFERENCES AND FURTHER READING

[1] Brixen, Eddy B.: *Designing listening tests of SR/PA systems, a case study*. Engineering Brief, AES 147th Convention (2019).

[2] Bech, Søren; Zacharov, Nick: *Perceptual audio evaluation: Theory, method, and application*. John Wiley & Sons, Chichester, England; Hoboken, NJ (2006). ISBN 0-470-86923-2.

[3] Toole, Floyd E.: *Sound reproduction – the acoustics and psychoacoustics of loudspeakers and rooms*. Focal Press (2009). ISBN 978-0-240-52009-4.

[4] Toole, F.E.; Olive, S.: *Hearing is believing vs. believing is hearing: Blind vs. sighted listening tests, and other interesting things*. AES 97th Convention (1994). Preprint 3894.

[5] Zielinski, S.; Rumsey, F.; Bech, S.: On some biases encountered in modern audio quality listening tests - a review. *Journal of the Audio Engineering Society* Vol. 56, No. 6 (2008).

[6] Pedersen, Torben H.; Zacharov, Nick: The development of a sound wheel for reproduced sound. *Audio Engineering Society Convention* Vol. 138, 1–13. Audio Engineering Society, Warsaw, Poland (2015).

[7] Berg, Jan; Rumsey, Francis. *Identification of perceived spatial attributes of recordings by repertory grid technique and other methods*. AES 106th Convention (1999). Preprint 4924.

[8] Rumsey, F.: Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society* Vol. 50 (September 2002).

[9] Brixen, Eddy B.: *Distance perception based on the spectra of the spoken voice*. AES 108th Convention (2000). Preprint 5076.

[10] Letowski, Tomasz R.: Anchor effect in an optimum timbre adjustment. *Journal of the Audio Engineering Society* Vol. 40 (September 1992).

[11] ITU-T P.800 Methods for subjective determination of transmission quality (1996).

[12] itu-r bs.1116 Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems (2015).

[13] ITU-R BS.1534–3: MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA). Method for the subjective assessment of intermediate quality level og audio systems (2014).

[14] EBU Tech. 3276: Supplement – 1: Listening conditions for the assessment of sound programme material: Multi-channel sound. European Broadcast Union, Geneva, CH (2004).

[15] EBU Tech 3278: Listening conditions for the assessment of sound programme material: Monophonic and two-channel stereophonic. European Broadcast Union, Geneva, CH (1998).

[16] AES20-1996: *AES recommended practice for professional audio – Subjective evaluation of loudspeakers*. Audio Engineering Society (1996).

[17] IEC 60268-13: Sound system equipment – part 13: Listening tests on loudspeakers.

[18] Brixen, Eddy B.: *Spectral ear training*. AES 94th Convention (1993). Preprint 3474.

[19] Head acoustics, application note: Conducting listening tests. https://www.head-acoustics.com/downloads/eng/application_notes/ListeningTests_e.pdf.

[20] ISO 8586-2: Sensory analysis – general guidance for the selection, training and monitoring of assessors – part 2: Expert sensory assessors (2008).

[21] Shlien, S.; Soulodre, G.: *Measuring the characteristics of "expert" listeners*. AES 101st Convention (1996). Preprint 4339.

[22] Ekeroot, Jonas; Berg, Jan; Nykänen, Arne: *Selection of audio stimuli for listening tests*. AES 130th Convention (2011). Preprint 8445.

[23] EBU Tech. 3253: Sound quality assessment material recordings for subjective tests (2008).

[24] Rumsey, F.: *Spatial audio*. Focal Press (2001). ISBN 0-240-51623-0.

[25] Bech, S.: *Perception of reproduced sound: Audibility of individual reflections in a complete sound field, II*. AES 99th Convention (1995). Preprint 4093.

[26] Schatz, R.; Egger, S.; Masuch, K.: The impact of test duration on user fatigue and reliability of subjective quality ratings. *Journal of the Audio Engineering Society* Vol. 60, No. 1–2 (2012).

[27] Lund, Thomas; Aki Mäkivirta: *On human perceptual bandwidth and slow listening*. VDT & AES, Tonmeistertagung (2018).

[28] Zacharov, Nick, ed.: *Sensory evaluation of sound*. CRC Press (2018). ISBN 9781498751360.

Glossary

3D sound See *immersive audio*.

3:1 rule A rule for placing microphones in front of several sound sources: The distance from one sound source to the next microphone should be three times to that of the nearest microphone. This rule is to prevent creation of a *comb filter*.

5.1 surround sound An audio format involving five channels of full-bandwidth audio: center, left, right, left surround, and right surround. A special low-frequency effects channel (LFE or ".1") covers a frequency range from 20 Hz to 120 Hz. The basic setup for music production is defined in ITU 775.

AAC Advanced Audio Coding. Special variant of the MPEG standard.

AB (1) Microphone placement for spaced microphones intended for time difference stereo. (2) Manner in which microphones are provided with power (in German: Tonader). (3) A method for comparison of two versions, such as in listening tests.

absorption A property of materials that reduces the amount of sound energy reflected; unit: sabin. 1 m^2 full absorption (such as an open window) equals 1 m^2 sabin.

absorption coefficient (symbol: α) The practical unit between 0 (no absorption) and 1 (full absorption) expressing the absorbing properties of a material. This is basically specified per octave or 1/3 octave. Absorption may exceed 100% (or α>1) when the surface area seems larger than the area it covers (applies typically only at high frequencies).

ABX A method for comparison of two versions A and B, such as in listening tests, where X can be either A or B.

AC Alternating current, as opposed to DC, direct current.

AC-3 Digital Audio Compression Standard.

A-chain In the cinema: consists of the audio tracks on the film and the related equipment to establish the line-level signals for the PA system.

acoustics The interdisciplinary science that deals with the study of sound.

A-D Analog-Digital.

ADC Analog-to-Digital Converter. A circuit that converts an analog signal to a digital signal.

ADPCM Adaptive Differential Pulse-Code Modulation. Encoding form for digital signals.

ADR Automatic Dialogue Replacement. Automatic replacement of location dialogue with studio dialogue.

AES Audio Engineering Society, worldwide interest organization for all branches of audio.

AES/EBU The AES and the EBU jointly created a standard for the transfer of two channels of digital sound. The signal is bi-phase modulated, self-clocking, and runs on a balanced cable (max. 50 m or 150 ft) or via optical fiber. The standard currently is called AES3.

AES67 AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability. Published by the Audio Engineering Society 2013 and revised in 2015. This standard brings together several common techniques that can be used as a common denominator for individual network manufactures to approximate simpler interconnection of different systems.

AF Audio Frequency. Audible frequencies, basically in the range of 20 Hz to 20,000 Hz.

A-filter See *A-weighting*.

AFM Designation for a video recorder's frequency-modulated sound track (Audio-FM).

A-format A microphone arrangement with four coincident cardioid capsules. Through use of a dedicated matrix, this signal can be converted to B-format. See *sound field microphone*.

AGC Automatic Gain Control. Method by which the amplification in a circuit is controlled by the input voltage or another parameter.

AIFF File format for audio.

AM Amplitude Modulation. Transfer of information by the variation of the amplitude of a carrier wave.

ambience Spatial effect, notably mixing, of remotely placed microphones in order to include the atmosphere of the room on the recording.

ambient noise (Acoustic) noise from the surroundings.

Ambisonics A principle that utilizes the sound field microphone involving B-format recording (X, Y, Z, and W).

ampere, A Unit for electrical current.

amplitude For an alternating signal; the maximum deviation from zero (positive or negative).

analog Quantities in two separate physical systems having consistently similar relationships to each other are called analogous. One is then called the analog of the other (for example, sound pressure in front of a microphone and the electrical output of that microphone).

analog-to-digital (A/D) Conversion: The process of "digitizing" an analog signal by sampling its amplitude at regular intervals. This process almost always involves limiting the frequency content of the digitized signal to a maximum of one-half the sampling rate, as this provision enables perfect reconstruction of the original band-limited signal from its samples.

anchor element The perceptual loudness reference point or element around which other elements are balanced in producing the final mix of the content, or that a reasonable viewer would focus on when setting the volume control.

anechoic room without echo; the reverberation time is ideally close to zero seconds.

ANSI American National Standards Institute. A federation of American organizations concerned with the development of standards.

anti-aliasing Low-pass filter for the removal of frequencies that otherwise would create "alias" frequencies that were not present in the signal originally.

AoIP Audio over Internet Protocol. The term for Audio in IP-based networks.

apt-X Bit reduction system from Audio Processing Technologies used by DTS. 4:1 or 3:1 compression.

artifact Unwanted effects that arise due to technical limitations.

ASCII American Standard Code for Information Interchange. Character set encoding used in data transfer.

asynchronous sample rate converter When two (digital) devices cannot be synchronized (for example, a Blu-ray player with a mixer), even a small deviation between the clock frequencies of the two devices will cause occasional glitches due to the accumulation of shortages or excesses. This creates a small "crack" in the sound. The asynchronous sample rate converter can perform interpolation and hence create the missing intermediate values so that glitches are avoided.

ATR Audio Tape Recorder.

ATRAC Adaptive Transform Acoustic Coding. Bit reduction system from Sony originally for use on MiniDisc. In stereo, 300 kbps is used. The system is also used for films (see *SDDS*).

ATSC Advanced Television Systems Committee.

attack The beginning of a sound; the initial transient of a musical note.

attenuation Variable or fixed downward adjustment of a signal level.

attenuator A device used to control (decrease) the level of an electrical signal.

audio Normally any audible information. Recent understanding as defined by the SMPTE: An electrical (or digital) representation of sound. By this definition, the audio itself cannot be perceived by hearing and thus is not audible.

autolocate Facility on a tape unit's transport system.

AUX Auxiliary jack, an extra input or output.

AVB Audio Video Bridging. Designation based on IEEE standard for both audio and video networks.

A-weighting The connection of an A-filter (IEC-A) for the acoustical and electrical measurement of noise. Produces a result that roughly corresponds to the human ear.

azimuth Angle of a magnetic head gap in relation to the direction of travel of a magnetic tape.

balancing When a signal is run independently of the frame/shielding and both terminals have the same impedance to the frame. See *CMRR*.

band-pass filter See *filter, band-pass*.

bandwidth The distance between the 3 dB cutoff frequencies on a response curve. Expressed either in octaves or in Hz.

batch The number of sound or video media produced in the same fabrication process.

BCD Binary Coded Decimal. Decimal digits converted to the base two system.

B-chain In the cinema: The reproduction chain from room level control to the acoustical room. It includes equalizers, processing, and loudspeakers.

bel, B Relative logarithmic unit for the measurement of sound levels. Normally one-tenth of a bel is regarded as the main unit: decibel, dB.

Betacam Tape-based video format that used the Betamax cassette with a tape speed of 10.15 cm/s [4 in./s.] and two analog sound tracks. Betacam SP: two analog and two FM modulated sound tracks. Betacam SP Digital: one analog, two FM, and two digital sound tracks.

B-format (microphones) A recording format virtually based on three orthogonal oriented figure-of-8 microphones (X, Y, Z) and one omnidirectional microphone (W). By combining these signals, a variety of directional characteristics can be obtained. Most B-format microphones are designed with cardioid microphones forming the A-format and from here converted into the B-format. See *sound field microphone*.

bias (1) In tape recording: High-frequency alternating current (typically 100 kHz) that is added to the recording head in a tape recorder together with the sound signal. The use of bias improves the quality of the recording with respect to both the frequency response and the distortion. (2) In testing: Any propensity of making decisions due to preferences not related to the test question.

binary Number system that only contains two digits.

binaural recording Stereo recording using either an artificial head with built-in microphones or a real person with small microphones in, or close to, the ear canals. For correct monitoring, the recording must be reproduced by using headphones to eliminate crosstalk.

bi-phase modulation A form of modulation that, among other things, is used by the AES/EBU interface and time code. In the bitstream, each bit shift is marked with a level shift (from high to low or vice versa). If the bit value is high, this is marked with a level shift in the middle of the bit concerned. Some of the advantages include that the DC component in the signal is minimized, and the signal is self-clocking.

bit Binary digit in digital technology, "0" or "1."

bit companding A technique for digital audio via which greater resolution is obtained for a given lower number of bits.

bit reduction As linear quantization can result in a larger number of bits per second than there is room for in a transmission channel or on a storage medium, the number of bits is reduced, preferably in a manner so that it cannot be heard in the audio signal.

bitstream converter See *Delta-Sigma converter*.

BNC connector Baby N-Connector; coaxial connector.

BS.1770 Formally ITU-R BS.1770. This standard specifies an algorithm that provides a numerical value indicative of the perceived loudness of the content that is measured. Loudness meters and measurement tools that have implemented the BS.1770 algorithm will report loudness in units of "LKFS."

buffer In computer science, a data buffer (or just buffer) is a region of physical memory storage used to temporarily store data while it is being moved from one place to another.

bus A common set of conductors where many signals are gathered, such as in a mixer console.

byte Data word, consisting of a number of bits, normally 8.

c/s Cycles per second = Hz.

calibration The process of measuring to determine the accuracy of the measurement chain.

calibrator, acoustic A device that produces a known sound pressure on a microphone in a sound level measurement system, typically 94 dB SPL, 1 kHz.

camcorder A contraction of the words camera and recorder; the recording medium and the camera are built together as a unit.

cartridge Special tape cassette with 1/4-inch tape earlier used for jingles, ads, and so on, at radio stations. Format: mono, stereo, or 8-track versions.

CC See *Compact Cassette*.

center frequency The arithmetic center of a constant bandwidth filter, or the geometric center (midpoint on a logarithmic scale) of a constant percentage bandwidth filter.

center post The center lead in a contact device.

channel separation The attenuation of one channel appearing in neighboring channels.

clean sound Atmospheric sound.

clipper See *limiter*.

clipping An electrical signal is clipped if the signal level exceeds the capabilities of the signal chain or recording device. It is a distortion of the signal.

CMRR Common Mode Rejection Ratio. Quantifies the suppression of common-mode signals at a differential input, such as a balanced microphone or line input.

coherence In acoustical and electrical measurements: The coherence expresses an estimation of the signal-to-noise ratio and the linearity of a system, such as in transfer function measurements. Normally expressed as a percentage.

comb filter A distortion produced by combining an electrical or acoustic signal with a delayed replica of itself. The result is a series of tops and dips across the frequency response that makes it look like a comb.

Compact Cassette, CC. Registered name for the cassette developed by Philips with a 3.81 mm [0.15 in.] audio tape.

compand Describes a process that involves compression and expansion successively.

compander A contraction of compressor and expander. A device that can perform both functions.

compression Reduction of the dynamic range of recorded audio.

compressor A device or plug-in that provides a reduction of the dynamic range of recorded audio.

content 0 Material or essence used for distribution by an operator.

corner frequency Transition frequency of a filter.

CRC Cyclic redundancy check. Error correction.

crest factor The term used to represent the ratio of the peak (crest) value to the RMS value of a waveform. For example, a sine wave has a crest factor of 1.4 (or 3 dB), since the peak value equals 1.414 times the RMS value. Music has a wide crest factor range of 4–10 (or 12–20 dB). This means that music peaks can be 12–20 dB higher than the RMS value, which is why the headroom is so important in recording and in audio design.

critical band In human hearing, only those frequency components within a narrow band, called the critical band, will mask a given tone. Critical bandwidth varies with frequency but is usually between 1/6- and 1/3-octave. The ears act like a set of parallel filters, each with its own bandwidth.

critical distance, D_c The distance from a sound source in a room at which the direct sound and the diffuse, reflected sound has the same level.

cross fade The audio equivalent of what is known as dissolve in video.

crossover frequency In a loudspeaker with multiple radiators, the crossover frequency is the −3 dB point of the network dividing the signal energy.

crosstalk Undesired energy in one signal (or channel) introduced from an adjacent signal or channel.

cue Audio or visual information that concerns timing or synchronization.

cue wheel Control button for use with slow forwards and backward winding with cueing.

cutoff The cutoff frequency of a filter. The frequency at which a filter begins to attenuate. Often defined by the attenuation being 3 dB at the frequency concerned.

D-A Digital-Analog (digital-to-analog).

DAC Digital-to-analog converter. A circuit that converts a digital signal into an analog signal.

DAT Digital Audio Tape (-recorder). Digital tape format (cassette). Normally understood as R-DAT (Rotary-head DAT) as opposed to S-DAT (Stationary head DAT), which never became a widespread format.

dB See *decibel*.

dBFS Decibels, relative to full-scale sine wave (1) Concerning level meters and measurement of digital equipment (AES17, IEC61606, ITU R BS. 1770–4, ITU-T P.38x): Levels reported in dBFS are always rms full scale level. 0 dBFS, is the RMS level of a dc-free 997-Hz sine wave whose un-dithered positive peak value is positive digital full scale, leaving the code corresponding to negative digital full scale unused. It is invalid to use dBFS for non-RMS levels. Because the definition of full scale is based on a sine wave, the level of signals with a crest factor other than that of a sine wave may exceed 0 dBFS. (2) Confusingly concerning some DAWs and measuring software: Levels reported as 0 dBFS refer to peak levels. Accordingly, the RMS level of a full-scale sine is −3 dBFS and the RMS of a full-scale square is 0 dBFS. In some software the user may define which practice to follow.

dBTP Decibels, true-peak relative to full scale (per ITU-R BS.1770 Annex 2).

dBm Logarithmic ratio with a reference of 1 mW/600 ohm.

dBu Logarithmic ratio with a reference of 0.775 V.

dBV Logarithmic ratio with a reference of 1 V.

dbx A particular brand of audio processing equipment, including noise reduction.

DC Direct current, as opposed to AC, alternating current.

DC offset The change in input voltage required to produce zero output voltage when no signal is applied to a device. Basically, this is an unwanted phenomenon in audio; however, DC offset often occurs in lower-quality sound cards.

D-connector Multiconductor plug with D-shaped collar.

decade Ten times any quantity or frequency range. The range of the human hearing is about 3 decades.

decay rate A measure of the decay of acoustic signals, expressed as a slope in dB/second. Essentially, the rate at which a signal drops off.

decay time See *reverberation time*.

decibel, dB Logarithmic ratio between two values.

decimation Suitable restructuring of data, such as swapping bandwidth (sampling frequency) for bit depth.

de-emphasis See *emphasis*.

de-esser Signal processing to reduce the sharp sound of the letter 's' in vocal recordings.

delay Time delay. (1) Electrical circuit that can delay a signal, in practice from fractions of a millisecond up to multiple seconds; used for sound effects. (2) Unwanted effect of time-consuming digital conversion, processing, and so on; see also *latency*.

Delta-Sigma converter (ΔΣ converter) Serial conversion at a high sampling frequency, where each bit specifies whether the current value of the signal is higher or lower than the prior one. After this process, this bit stream is converted to standard values such as 24-bit format.

DFT See *Discrete Fourier Transform*.

DHCP Dynamic Host Configuration Protocol. A DHCP server in the network provides connected devices with the network information needed to operate.

DI (1) Digital In, digital input. (2) Direct Injection, buffer circuitry. (3) Directivity Index, directional index for acoustic transducers such as loudspeakers and microphones. (4) Dialogue Intelligence™, a speech-gating technology that measures loudness only on the segments of a program that contain dialogue; trademark by Dolby Labs.

dialnorm An AC-3 metadata parameter, numerically equal to the absolute value of the dialogue level carried in the AC-3 bit stream.

dialogue level The loudness, in LKFS units, of the anchor element.

diffraction The distortion of a wavefront caused by the presence of an obstacle in the sound field; the scattering of radiation at an object smaller than one wavelength and the subsequent interference of the scattered wavefronts.

diffuse sound The sound field contains no directional information and spreads randomly in the room.

diffuser A device that provides scattering when sound hits the surface.

digital Concerning a state wherein an electrical signal has been converted to a series of impulses according to a specific code. The signal has come to exist in "tabular form."

digital full scale See *Full Scale (FS)*.

digital interface between digital systems. A number of standards exist, of which AES3 is the most common.

DIN plug Contact connection as per the DIN norm.

DIO Digital In/Out, digital input and output.

Directivity factor The ratio of the mean-square pressure (or intensity) on the axis of a transducer at a certain distance to the mean-square pressure (or intensity) that a spherical source radiating the same power would produce at that point.

Directivity Index (DI) Directivity factor expressed in dB (10 · log [directivity factor]).

Discrete Fourier Transform, DFT A mathematical technique for determining the spectral content of complex waveforms. The DFT essentially compares the signal being analyzed to a series of sine and cosine waves at regularly spaced (harmonic) intervals to determine how much energy is present at each harmonic frequency.

dissolve Mixing (of images). Corresponds to crossfade in audio.

distance double law When doubling the distance from a point source, the sound pressure is halved (reduced by 6 dB).

distortion factor Percentage measure for harmonic distortion; for example, k3 = 3% means that the third harmonic overtone is 3% of the fundamental tone.

distortion When a signal is changed from its original form, such as due to nonlinearities in the transmission chain.

distribution amplifier Can distribute a signal to multiple inputs without the signal source becoming overloaded.

dither Noise that is added to the lowest bit in the digital signal in order to reduce the distortion at low levels of the audio signal. Dither can be noise shaped in order to be less audible.

DO Digital Out, digital output.

Dolby E An audio data-rate reduction technology designed for use in contribution and distribution that also conveys Dolby E metadata.

Doppler effect or Doppler shift The apparent upward shift in frequency of a sound as a noise source approaches the listener or the apparent downward shift when the noise source recedes, such as the sound of a speeding car passes by.

DRC Dynamic Range Control. For instance, implementing digital TV with ATSC by the use of accompanying metadata.

DRC profile A collection of parameters that describes how dynamic range control metadata is calculated.

drop frame Variant of time code, where a frame is periodically skipped in order to preserve synchronization.

drop-out Short-duration loss of the signal on a tape due to faults in the tape's magnetic coating.

DSB Digital Satellite Broadcasting.

DSP Digital Signal Processor. A circuit that can perform manipulation of data.

D-sub A plug type that can be used, among other things, for the transfer of digital signals. The flange around the connector pins are D-shaped, hence the name.

DTS Digital Theater System. Digital sound system originally for movies. The sound information is stored on a CD-ROM, which is controlled by a time code on the film. Uses APT-X bit reduction.

dubbing Mix of one signal with another.

ducking Automatic compression, such as when a speaker's insert dims the music signal.

dynamic range The relationship between the strongest and weakest passages in the program material. Used for both the acoustic and the electrical signal.

EBU European Broadcasting Union. An association of European radio broadcasting stations.

echo Sound impulse arising from a reflection with such a strength and time delay after the direct sound (normally >50 ms) that it is perceived as a repetition of it.

edge track Designation for the longitudinal (sound) tracks of a video tape, since these are located on the edge of the tape.

EDT In acoustics: Early Decay Time. See *reverberation time*.

EFM Eight-to-fourteen modulation. Digital modulation form, used for CDs, among other things.

EFP Electronic field production. The production form typically applied for drama and documentaries produced outside the studio.

emphasis A technique often used in analog transmission systems (wireless microphones, FM broadcast, etc.) to enhance dynamic range by raising the level of higher frequencies. De-emphasis is introduced on the receiving side.

ENG Electronic News Gathering. News production in which electronic media is used, such as video tape, HD, or flash recorders (as opposed to film).

EQ See *equalizer, graphic; equalizer, parametric; equalizing*.

equal loudness contour A contour representing a constant loudness for all audible frequencies. The contour with a sound pressure level of 40 dB at 1000 Hz is arbitrarily defined as the 40-phon contour.

equalizer, graphic Electronic equipment for "equalizing." Built from a number of 1/1 octave or 1/3 octave band-pass filters that can each amplify or attenuate, and essentially used to "flatten out" or alternatively to obtain a desired frequency response.

equalizer, parametric. Electronic equipment for "equalizing." Constructed from a set of filters where the center frequency, bandwidth, and amplification/attenuation can be adjusted independently for each filter.

equalizing (1) The process that consists of modifying the frequency balance in the amplifier chain for the purpose of obtaining a flat frequency response, minimizing noise, or achieving an artistic effect. (2) Equalization of nonlinearity (in frequency response).

expander (1) Electronic equipment in which the output signal's dynamic range is increased in relation to that of the input signal. (2) Designation for a controllable synthesizer without a keyboard.

far field A region in free space at a much greater distance from a sound source than the linear dimensions of the source itself where the sound pressure decreases according to the inverse-square law (the sound pressure level decreases 6 dB with each doubling of distance from the source).

FFT See *Fast Fourier Transform*.

Fast Fourier Transform an efficient method of estimating the frequency spectrum of a signal.

file-based scaling device A device used to apply an overall gain correction to audio content stored as files.

filter A device or algorithm for separating components of a signal on the basis of their frequency. It allows components in one or more frequency bands to pass relatively unattenuated, and it attenuates components in other frequency bands.

filter, band-pass A filter that passes all frequencies between a low-frequency cutoff point and a high-frequency cutoff point.

filter, high-boost A filter that amplifies frequencies above a specific frequency.

filter, high-cut A filter that attenuates frequencies above a specific cutoff frequency.

filter, high-pass A filter that passes all frequencies above a cutoff frequency but attenuates low-frequency components. They are used in instrumentation to eliminate low-frequency noise and to separate alternating components from direct (DC) components in a signal.

filter, low-cut A filter that cuts off low-frequency signals below the cutoff frequency with a certain attenuation (roll off). In microphones, the filters are typically active below 80–300 Hz, and the slope is typically 6 or 12 dB/octave. See also *filter, high-pass*.

filter, low-pass A filter that passes signals below the cutoff frequency and attenuates the signal above that frequency. An anti-aliasing filter in a digital system is an example of a low-pass filter with a very steep roll off.

filter, notch Narrow-band filter with very strong attenuation in a very narrow frequency range. Used to remove individual frequencies such as hum.

filter, octave Filter with a bandwidth of an octave.

filter, shelving A type of filter that gives constant amplification or attenuation from the corner frequency.

filter, third-octave A filter whose upper-to-lower passband limits bear a ratio of $2^{1/3}$, which corresponds to 23% of the center frequency.

flanging Sound effect based upon the direct signal mixed together with itself using varying time delays. Originally made using a tape recorder, where the source reel is slowed down by a finger placed on the flange of the reel.

FM Frequency modulation. Modulation principle in which a carrier wave is varied about its center frequency in proportion to the frequency of the modulating wave and where the oscillation of the carrier wave is proportional to the amplitude of the modulating wave.

FOH Front of house - is the part of a performance venue that is open to the public. The FOH sound engineer controls the sound projected to the audience by the PA/SR system.

fold back The musicians on the scene have a need to be able to hear themselves and the others in a quite specific manner. They are given a fold back or monitor loudspeaker, where the sound is specially mixed for the purpose.

Format In broadcast/streaming: A name, such as “contemporary hit radio” or “news,” that describes the type or style of a given long-form program. A stream often has only one format, but some netcasters transmit different formats at different times.

FQTSS Forwarding and Queuing for Time-Sensitive Streams. Term used in AVB.

frame In video, an image. The smallest unit of a time code.

framesync Short for “frame synchronizer.”

free field An environment in which there are no reflective surfaces within the frequency region of interest, and the sound is isotropic and homogeneous.

frequency response Figure that shows the relationship between amplitude and frequency.

frequency The number of cycles per second. Its reciprocal is the period. Specified in hertz (Hz).

frequency weighting Modification of the frequency spectrum of a signal by means of a filter having a conventional characteristic known as A, B, C, D, RLB, K, CCIR/ITU, and so on.

FS See *dBFS*.

Full Scale (FS) See *dBFS*.

fundamental (1) The basic pitch of a musical note. (2) Fundamental frequency, the lowest frequency of a vibrating system. The spectrum of a periodic signal will consist of a fundamental component at the reciprocal of the period and possibly a series of harmonics of this frequency.

gain Amplification (in a circuit).

gate See *noise gate*.

gear Encompasses all equipment.

genre A name, such as "full-bandwidth speech," "telephone-grade speech," "popular music," "classical music," and so on, characterizing a program element that has a homogenous texture and style.

glitch When a bit is skipped, it can lead to a little "crack" in the sound.

GPI General Purpose Interface.

graphic equalizer See *equalizer, graphic*.

Haas effect Also called the precedence effect or principle of first arrival.

harmonic A discrete sinusoidal (pure-tone) component whose frequency is an integer multiple of the fundamental frequency of the wave. If a component has a frequency twice that of the fundamental, it is called the second harmonic, and so on.

harmonic distortion Changing the harmonic content of a signal by passing it through a nonlinear device. Clipping results in harmonic distortion (uneven harmonics).

harmonic overtones Tones with the frequency of an integer multiple of the fundamental frequency.

HDBaseT is a standard for transmission of ultra-high-definition video and audio, Ethernet, control signals, USB, and up to 100W electrical power over a single long-distance cable (Cat 6 or higher).

HDMI High Definition Multimedia Interface. An interface standard for the exchange of both sound and images, especially over shorter distances (usually max. 15 meters in cable).

headroom Overloading reserve. The amount of signal above nominal level that can be permitted before overloading arises with distortion as a consequence. In digital audio: The ratio of 0 dBFS to the integrated program level of a segment.

hertz, Hz Measurement unit for frequency.

HF High frequency.

high-boost filter See *filter, high-boost*.

high-cut filter See *filter, high-cut*.

high-pass filter See *filter, high-pass*.

house sync signal that is distributed (in-house) so all digital devices can run at the same speed.

HRTF Head Related Transfer Function. The head's influence at sound received at the ears.

HVAC noise Heating, ventilation, and air-conditioning noise. The word is used in connection with requirements concerning noise in control rooms, cinemas, and so on.

Hz See *hertz*.

iCheck Integrity check. Reveals if the signal is spatially compromised (e.g., because of data reduction, such as MP3 or AAC encoded at too low a bit rate).

IEC International Electrotechnical Commission. A standardization commission.

IEEE Institute of Electrical and Electronics Engineers. Professional organization with recognized standardization groups. Stands, among other things, for Essential Ethernet Standards (IEEE 802–3).

IETF Internet Engineering Task Force. An interest group that also develops standards and recommendations.

IGMP Internet Group Management Protocol. (Network): A multicast protocol that ensures that individual data packets are guided directly to the addresses they are meant for.

IM Intermodulation.

immersive audio Immersive audio encompasses all surround-sound formats, channel-based formats reproduced in 5.0/5.1, 7.1, 9.1, and so on, and formats that include height information, either channel-based or object-based. Also binaural audio is regarded as immersive audio. Alternative definition by TASA: A soundtrack format for digital cinema with greater channel count or greater audio data capacity than 7.1 channels.

impedance Electric resistance: A material's opposition to the flow of electric current; measured in ohms.

impedance matching Maximum power is transferred from one circuit to another when the output impedance of the one is matched to the input impedance of the other. Impedance matching is generally only relevant for RF and electrically coupled digital interfaces.

impulse response The response of a system to a unit impulse. The Fourier transform of the impulse response is the frequency response.

induction The property that an electrical current is produced in a conductor when it is exposed to a varying magnetic field.

inductor Coil wound on an iron core.

infrasound Sound at frequencies below the audible range, that is, below about 16–20 Hz.

initial time-delay gap (ITDG) In acoustics: The time gap between the arrival of the direct sound and the first sound reflected from the surfaces of the room.

in-phase Two periodic waves reaching peaks and going through zero at the same instant are said to be "in phase."

insert (1) Insertion of a clip (scene) in an existing recording. (2) A breakpoint in a mixer channel where an external device can be inserted.

Integrated Loudness The electrically measured average loudness between two points in time. Integrated Loudness according to the international standard ITU-R BS.1770–4 uses a gated algorithm. If Integrated Loudness is measured over the entire length of a program, the result is called Program Loudness (PL).

interpolation Computation of intermediate values in relation to fixed values (for example, values between two samples).

inverse square law A description of the acoustic wave behavior in which the mean-square pressure varies inversely with the square of the distance from the source. This behavior occurs in free-field situations, where the sound pressure level decreases 6 dB with each doubling of distance from the source.

IP Internet Protocol. That part of a network-based system that describes the structure.

ips Inches per second.

IPv4 Present standard for Internet Protocol; however, limited by the still growing demand for unique IP addresses, which is beyond what this standard can deliver.

IPv6 A revised standard that provides far more unique IP addresses compared to IPv4.

IR (1) Infrared. IR light is used as a control signal in remote control units. (2) Impulse response (see *impulse response*).

ISO International Organization for Standardization.

ITU International Telecommunication Union.

jack plug One-legged connector with two or three contacts. Normal dimensions are 6/3.5/2.5 mm.

jack-bay A patch panel with jacks that are used to make connections between different devices easy to establish.

jam sync Method for recording new time codes from a source tape.

JEITA Japan Electronics and Information Technology Industries Association.

jitter Expresses that the sample timing deviates from a fixed rate. Jitter is measured in seconds (typically ns or ps). Jitter causes noise in the signal reproduced. The worst kind of jitter is sampling jitter, as this cannot be corrected later. Transmission jitter can to some degree be compensated for.

kbps or kb/s Kilobits per second.

kHz, kiloherz One thousand hertz, see *hertz*.

K-system Dynamic metering system proposed by American engineer Bob Katz.

K-weighting A weighting filter used for the measurement of program loudness (ITU/EBU).

LAN Local Area Network. Typically Ethernet-based applying Cat5/6 cabling and RJ45-connectors.

latency In digital systems, delay due to processing.

layback A post-production technique where audio is rejoined with the associated video after editing, mixing, or "sweetening."

LCD Liquid crystal display.

LED Light-emitting diode.

Leq Equivalent continuous sound pressure level.

LF Low frequency: (1) Oscillations in the audible frequency spectrum (below 20 kHz). (2) Oscillations in the low-frequency part of the audible frequency spectrum.

limiter A signal processing unit that attenuates the output when the input exceeds a given threshold. This provides a fixed maximum output level.

line driver Intermediate amplifier for studio and stage use that brings a low voltage (such as from a guitar pickup) up to line level. At the same time, provides a matching for long cables.

line level The voltage that the signals are amplified up to (such as in the mixer) before they are routed to an output amplifier (or transmitter).

linear phase response The phase is constant with frequency through the device or circuit.

linear quantization Conversion to a digital signal where all bits represent identical steps in level (as opposed to nonlinear quantization and bit compression).

lin scale linear scale. A scale on which there is a constant value between the units, such as 10, 20, 30, 40.

lip-sync The technique of miming to a prerecorded program in connection with a TV recording, and so on.

LKFS Loudness, K-weighted, relative to full scale, measured with equipment that conforms to ITU recommendation 1770.

load If an electric circuit has a well-defined output terminal, the circuit connected to this terminal (or its input impedance) is the load.

locate To find a specific position, such as on a track, digital or analog.

log scale Logarithmic scale. A scale on which there is a constant ratio between the units, such as 10, 100, 1000, 10,000 (a.k.a., 10^1, 10^2, 10^3, 10^4).

long-form content Show or program related material or essence. The typical duration is greater than approximately two to three minutes.

longitudinal Lengthwise.

loudness A perceptual quantity; the magnitude of the physiological effect produced when a sound stimulates the ear.

loudness level (1) In acoustics; measured in phons, it is numerically equal to the median sound pressure level (dB) of a free progressive 1000 Hz wave presented to listeners facing the source, which in a number of trials is judged by the listeners to be equally loud. Loudness level can be calculated according to ISO 532B. (2) In audio production; measured in loudness units (LU) according to ITU 1770.

loudness normalization A working practice in which a specific loudness level defines the reference point (also known as the "target"), and the loudness of individual signals is assessed so that they can then be adjusted to match the target loudness on replay. This approach allows a margin for peaks and thus encourages dynamic variations if a reasonably low target is chosen. (See also *Normalization; peak normalization*).

low-cut filter See *filter, low-cut*.

low-pass filter See *filter, low-pass*.

LRA Loudness range. Describes the variation of loudness levels within a program on a macroscopic scale. Not to be confused with "dynamic range," which is the distance between the noise floor and the highest possible peak of a signal path. It is based on statistics and uses the 3-second short-term loudness levels. Reference: EBU Tech 3342.

LTC Longitudinal Time Code. The time code that is recorded on a tape's longitudinal track.

LU Loudness Unit. A-weighted measure used in program metering. Defined in ITU-1770. A step of 1 LU is the same as the step of 1 dB. If one stream's target loudness is −16 LUFS and another stream is set to −18 LUFS, then they are 2 LU apart. Also a relative loudness value where the target level is 0 LU. If the target level is, for example, −23 LUFS = 0 LU, then an audio signal with a program loudness of −19 LUFS could be written as PL being +4 LU. Reference: EBU Tech 3341 and ITU-R BS.1771–1.

LUFS Loudness Units with reference to digital full scale. The EBU preferred term for the absolute loudness value, which includes a psychoacoustic filter shape known as K-weighting, explained in ITU-R BS.1770–3.

MAC Media Access Control is a unique address for a unit applied in a network.

MADI Multichannel Audio Digital Interface; a.k.a. AES10. Interface standard where up to 64 audio channels can be transferred serially on a cable, coax, or optical. NRZI encoding, 125 Mbps.

magnitude The absolute value of amplitude.

matrix A circuit in which the output signal is an encoded version of the input signal. A MS-matrix will transform MS to LR or vice versa.

Mbps or Mb/s Megabits per second.

measured loudness The magnitude of an audio signal when measured with equipment that implements the algorithm specified by ITU-R BS.1770/EBU R-128. It is an approximation of perceived loudness.

meter Designation for a measurement instrument for the control of signal levels.

MIDI Music Instrument Digital Interface.

mixing level Indication of the absolute sound pressure level calibration of the mixing studio that produced the content.

mode (1) A room resonance. Axial modes in rectangular rooms are associated with pairs of parallel walls. Tangential modes involve four room surfaces and oblique modes all six surfaces. Their effect is greatest at low frequencies and for small rooms. (2) A specific setting for a device (such as record mode, sleep mode, etc.)

modulation An analog process by which the characteristic properties of a wave (the modulating wave) are mixed into another wave (the carrier wave).

modulation noise in the signal chain that varies with the signal strength. Particularly a problem with audio tape recorders.

MOL Maximum output level. For tapes, the highest attainable output.

monitor A control or monitoring device; in this context, a loudspeaker for the assessment of recording quality.

mono compatible A stereo signal, if no significant loss of level and no coloration (comb filtering) occur when summing the channels to one mono signal.

mono Signal in a single channel.

MPEG Motion Picture Experts Group. Standardization organization that sets standards for techniques such as digital audio compression.

MPX Multiplex filter, for removing an FM radio's stereo pilot tone.

MSB Most significant bit.

MS-technique A stereo recording technique using M (mid) and S (side) signals generally from one cardioid and one figure-8 microphone. To obtain L/R format, the signals are processed like this: L = M + S, R = M – S.

MTC MIDI time code.

MUSICAM Masking pattern adapted Universal Subband Integrated Coding And Multiplexing. An (older) bit reduction system.

mV Millivolt, 1/1000 V.

MVPD Multichannel Video Programming Distributor. Includes DBS service operators, local cable system operators, and cable multiple system operators.

N See *newton*.

NAB or NARTB National Association of Radio and Television Broadcasters. American standardization organization.

NBC Non-backward compatible. Compression standards that are not backward compatible (i.e., it is not possible to "unpack" stereo to 5.1 channels having first gone from 5.1 to two-channel stereo).

newton Measurement unit for force.

noise (1) Unwanted sound. (2) Technically, sound without tonal components.

noise floor A measure of the signal created from the sum of all noise sources and unwanted signals within an audio system.

noise gate Special signal processing unit whose output signal is 0 until the input signal exceeds a specific preset value.

noise shaping A special technique providing frequency weighting of dither used in connection with reducing high-resolution digital audio to a lower bit format.

normalization The process of adjusting the loudness of incoming material to conform to a target loudness.

notch filter See *filter, notch*.

NR Noise reduction.

NRZ Non-return to zero.

nWb Nanoweber = 10^{-9} weber. See *weber*.

Nyquist Frequency In digital sampling, the highest frequency it is possible to reproduce with respect of the sampling frequency. Equal to half the sampling frequency. Named after Harry Nyquist.

OBA Object Based Audio, immersive audio format.

octave A range of frequencies whose upper frequency limit is twice that of its lower frequency limit. For example, the 1000 hertz octave band contains noise energy at all frequencies from 707 to 1414 hertz. In acoustic measurements, sound pressure level is often measured in octave bands, and the center frequencies of these bands are defined by ISO and ANSI. Fractional portions of an octave band are also often used, such as a 1/3 octave band.

octave filter See *filter, octave*.

ohm [Ω] Unit for electrical resistance.

one-bit converter See *Delta-Sigma converter*.

operator A broadcast network, broadcast station, DBS service, local cable system, or cable Multiple System Operator (MSO).

oscillator Signal generator for the generation of, for example, pure sinusoidal tones with constant amplitude for use in measurements.

OSI Open Systems Interconnection. A standardized model with seven steps that determine the communication in computer systems independent of underlying internal structures and technologies.

OTT Over-The-Top: The means to deliver video content via Streaming, Video On Demand (VOD), Pay TV, IPTV, and download via IP mechanisms.

output impedance The alternating current resistance a circuit's output represents. May also be called source impedance.

overdub A process whereby a track is recorded while listening to already recorded material on the same or a different tape recorder.

overlap Refers to the amount of data each successive FFT frame shares in common with the one before.

override See *ducking*.

oversampling The use of a sampling frequency that is a number of times higher than is necessary. This makes it easier to make low-pass filters that ensure that no alias frequencies arise in the sound.

overtone A harmonic frequency component of a complex tone at a frequency higher than the fundamental. This can involve both harmonic as well as nonharmonic overtones.

PA Public Address, in sound communities understood as the loudspeaker system used for the amplification of speech and music when addressing larger audiences.

PAC Perceptual Audio Coding. Bit reduction system from Lucent: 96 kbps.

pad Circuit in the input module (such as in a mixer) with the purpose of pre-attenuating the signal.

PAL Phase Alternation Line. European TV system.

pan-pot Potentiometer in the mixer for “moving” the signal between the left and right channel.

parallel port Digital input or output port with multiple lines allowing multiple bits of data to be transferred simultaneously.

parametric equalizer An equalizer in which both the relative gain/attenuation and frequency and bandwidth of filters can be controlled individually.

partial One of a group of frequencies, not necessarily harmonically related to the fundamental, which appear in a complex tone. Bells, xylophone blocks, and many other percussion instruments produce harmonically unrelated partials.

patch Connection of circuits via external connections.

PCM Pulse-code modulation. A form of modulation in which the information is described by the number and duration of impulses.

PD Powered Device. The device that receives supply voltage from PoE.

peak normalization The practice of adjusting the peak level of each program to (usually) full scale or 0 dBFS. This results in widely varying loudness levels. This practice has been cited as the original cause of loudness wars in broadcast and music production.

peak The maximum positive or negative dynamic excursion from zero of any time waveform. Sometimes referred to as “true peak” or “waveform peak.”

peak hold Circuit that for a shorter or longer period of time is in a position to maintain a peak value display on an instrument.

perceptual coding A principle for low bit-rate coding based on the masking abilities of the ears.

periodic signal A signal is periodic if it repeats the same pattern over time. The spectrum of a periodic signal always contains a series of harmonics.

phantom supply/power Method for DC power supply of condenser microphones via the signal cable (ideally 48 volt).

phase The different values that alternating current or alternating voltage runs through during a period.

phase shift A timing difference in a signal, typically expressed in degrees.

phon Logarithmic unit for loudness.

phono plug/jack Plug and jack for one conductor and shielding.

pink noise Electrical noise signal for test purposes; contains constant energy per octave.

PIPU Punch-In Punch-Out.

pitch Subjectively perceived tone frequency.

PL See *program loudness*.

PLL Phase-locked loop. A circuit that ensures a stable frequency in relation to a reference.

PLR Peak-to-loudness ratio. The ratio of maximum true peak level of a program segment to its integrated ITU-R BS.1770–3 loudness.

PMP Portable Media Player.

polarity Referring to the positive or negative direction of a signal. In all kinds of stereo and surround-sound production, it is important that microphones have the same polarity or else the imaging is totally blown.

post production The entire post-processing phase (including editing, sound effects, music, etc.) for video and TV productions.

pot See *potentiometer*.

potentiometer Adjustable resistance used, for example, in controlling electrical levels.

PPM Peak Program Meter. A meter with a time constant of 5 or 10 ms.

precedence effect See *Haas effect*.

pre-emphasis See *emphasis*.

presence The frequency range around 2–5 kHz. Highlighting of this range causes voices and instruments to stand out in the acoustic image.

pressure zone As sound waves strike a solid surface, the particle velocity is zero at the surface and the pressure is high, thus creating a high-pressure layer near the surface.

program loudness A measurement of Integrated Loudness over the entire length of a program. Abbreviated PL.

program A section of a stream or a broadcast having a single format such as news/talk, or popular music with DJ introductions and commentary.

PSE Power Sourced Equipment. The device delivering PoE (Power over Ethernet).

psychoacoustics The study of the interaction of the human auditory system and acoustics.

PTP Precision Time Protocol. Is a part of Audio-over-IP transmission.

punch-in A rerecording of a short sequence in an already recorded program.

punch-out See *punch-in*.

Q factor Measure for the slope of a resonance top.

QoS Quality of Service, a code applied to prioritize packages of audio over other kinds of data transported in a network.

quantization Digitalization of a signal. Conversion of the analog signal to numbers that express the values measured at the time of the samplings.

rack Cabinet or frame in which devices are installed.

RAID Redundant array of inexpensive disks. Method for simultaneous operation of multiple hard disk drives so that they work as a fault-tolerant unit.

RAM Random-access memory. Digital storage where data can be arbitrarily stored and retrieved.

RASTI Rapid Speech Transmission Index. See *STI*.

R-DAT Rotary-Head Digital Audio Tape recorder.

refraction The bending of a sound wave from its original path, either because it is passing from one medium to another with different velocities or caused by changes in the physical properties of the medium, such as a temperature or wind gradient in the air.

ReplayGain a method to calculate loudness published in 2001. Spotify uses this algorithm for their internet audio delivery.

resolution The number of bits per sample of a digitized signal.

reverberation time In acoustics, the time it takes sound to decay by 60 dB after the sound has stopped. Often only a part of the decay is applied for the analysis: T10, T20, and T30 expresses the RT60 based on the part of the decay from −5 to −15 dB, −5 to −25 dB, and −5 to −35 dB, respectively. The EDT, Early Decay Time, encompasses the decay from 0 to −10 dB.

RF Radio frequency. Name for frequencies in the range of 30 kHz to 3000 GHz. This is the range for electromagnetic waves applied to radio communication.

RLB Revised version of B-weighting curve. Applies to the measurement of program loudness.

RMS Root mean square, the effective value of a signal.

roll off The attenuation of a high-pass or low-pass filter is called roll off. The term is mostly used for high-frequency attenuation.

room mode The normal modes of vibration of an enclosed space. See *mode*.

routing switch A switch for the routing of signals.

RT60 Reverberation time based on the 60 dB attenuation of the sound after the sound source has stopped.

RTP Real-time Transport Protocol. Protocol for the transfer of data for immediate use.

s/n See *signal-to-noise ratio*.

S/PDIF Sony/Philips Digital Interface.

sample rate converter (synchronous) Converts digital signals with the same reference from one sampling frequency to another.

sampling frequency The number of samples that are extracted per second. In professional sound equipment: 48 kHz. For CDs and the like: 44.1 kHz.

sampling In digital technology, entry of the signal into discrete values in a table. A single value is called a sample.

sampling rate The number of times per second that the amplitude of a signal is measured within a given period of time in analog-to-digital conversion.

SCART plug A special 21-pin plug/jack on some videotape recorders and TV units that is no longer used.

SCSI Small Computer System Interface. Ultimately, a purely computer-related standard, for the controlling of hard disk drives and so on. It can, however, also carry sound information.

SDDS Sony Dynamic Digital System. Digital sound system for movies with a total of 8 sound channels: L, CL, C, CR, R, SL, SR, and subwoofer. SDDS is bit-reduced with the ATRAC system.

SECAM Systeme En Couleur Avec Memoire. Television system.

sensitivity Expresses the size of a signal that a device requires in order to reach nominal level.

separation Degree of signal-related separation between tracks, channels, and so on.

sequencer A device on which information about sequences can be recorded, but not the sequence itself.

SFX Special effects.

shelving filter See *filter, shelving*.

short-form content Advertising, commercial, promotional, or public service related material or essence. Also termed “interstitial” content. The typical duration is less than approximately two to three minutes.

Short-Term Loudness As defined in EBU Tech 3341, Short-Term Loudness uses the ITU-R BS.1770–3 algorithm but with no gating and a sliding rectangular time window of length 3 s. The update rate for “live meters” must be at least 10 Hz.

shuttle mode As related to tape transport or DAW track handling, the running of the recording back and forth between two points.

signal In audio, commonly understood time-domain information, where voltages or numeric values representing amplitude.

signal-to-noise ratio of the nominal signal level to the undesired noise level. Expressed in dB.

sinusoidal tone The most simple harmonic tone, consisting of only one frequency.

slap back A discrete reflection from a nearby surface.

slating Marking on a recording before a take.

SLM Sound Level Meter.

SMPTE Society of Motion Picture and Television Engineers. International association.

snake For PA systems, the bus cable between scene microphones and the mixer in the concert hall.

SNG Satellite news gathering. Direct news transmission to main stations via satellites.

SOL Saturation output level. The magnetization level at which a magnetic tape is saturated.

sound field microphone A single unit microphone with four cardioid condenser microphone capsules shaped like a tetrahedron. This configuration is also called the A-format. By processing the signals the B-format is formed equivalent to three figure-8 microphones and one omnidirectional microphone, all in a coincident position. A countless number of directional characteristics are obtained by combining the signals in different ways.

sound pressure A dynamic variation in atmospheric air pressure. Sound pressure is measured in pascals (Pa).

sound pressure level (SPL) The sound pressure level of a sound in decibels, equal to 20 times the logarithm to base 10 of the ratio of the RMS sound pressure to the reference sound pressure 20 μPa.

SP Special performance. Relates to improved versions of existing video formats.

spectrograph A data plot, displaying frequency vs. time. A third dimension (level) is provided by color or black-white shading.

spectrum The frequency content of a given signal.

speed of sound: The speed at which sound waves propagate through a transmission medium such as air or water. This quantity has dependent actors such as temperature and density of the material of propagation. Useful rule of thumb values for the speed of sound in air at room temperature are 1130 ft/sec, or 344 m/sec. In Smaart, the speed of sound is used primarily to calculate distance equivalents for time differences.

Speech Transmission Index (STI) Measurement method based on modulation transfer functions (MTF) for the objective determination of speech intelligibility in a room by applying 13 modulation frequencies to each of the seven octave bands 125 Hz–8 kHz. A practical version implemented for handheld measurement is STI-PA (Speech Transmission Index, PA systems). To reduce the measurement time required, STI-PA only uses two modulation frequencies. In a good room, the STI value ought to be greater than 0.6 (on a scale of 0–1) or D on a scale from U to A+.

spherical wave A sound wave in which the surfaces of constant phase are concentric spheres. A small (point) source radiating into an open space produces a free sound field of spherical waves.

SPL see *sound pressure level*.

splitbox A unit that distributes the signal from a microphone to multiple mixers or tape recorders.

square waves Waveform for acoustic or electrical signals that in terms of frequency contain a fundamental tone and all the odd-numbered harmonics.

SR Sound reinforcement.

SRP Stream Reservation Protocol. Term applied in AVB.

stereo A technique intended to created a spatial sound impression by the use of two or more channels.

STI See *Speech Transmission Index*.

STI-PA See *Speech Transmission Index*.

stream A continuous transmission over a network (typically the Internet) that consists of one or more programs presented sequentially. Analogous to a "radio station" in over-the-air broadcasting.

streamer A content provider offering a streaming service to customers.

streaming audio Sound that can be played back immediately in (almost) real time from the Internet.

subharmonic Subharmonics are synchronous components in a spectrum that are multiples of 1/2, 1/3, or 1/4 of the frequency of the primary fundamental.

supersonic See *ultrasound*.

surround sound Designation for certain multichannel systems for film and TV stereo sound.

sweet spot The optimum listening position in front of the stereo or surround speakers.

symmetry attenuation Also known as common-mode range rejection. The attenuation of induced electrical noise signals that is attained in a balanced connection.

synchronizer A unit that by the use of a time code can deliver control signals for the time-related locking of, for example, tape units.

T10, T20, T30 See *reverberation time*,

tape speed The speed at which a magnetic tape passes the magnetic heads in a sound or video tape recorder. Specified in cm/s or inch/sec.

tape width The width of a magnetic tape.

taping Outlet from mixer channel, such as aux-send. The expression is used by radio broadcasters.

target loudness A specified value for the anchor element (in some systems dialogue level), established to facilitate content exchange from a supplier to an operator.

target In program material, the intended Integrated Loudness of a entire stream. In mixed-format streams it can also refer to the intended Integrated Loudness of programs having a given format within the stream.

TASA Trailer Audio Standards Association. Provides standard for motion picture trailer volume.

TB Talk back. A system via which the control room can give oral messages to the studio.

TC See *time code*.

TDIF Tascam Digital Interface. Factory-initiated interface, 8-channel parallel, unbalanced in a D-sub connector, cable max. 15 m [45 ft].

telephone hybrid Unit that adapts a telephone to a mixer console.

THD Total harmonic distortion.

THD+N Total harmonic distortion plus noise. An electroacoustic measure of (unwanted) signals from distortion (occurring, for instance, by the clipping of the signal) and the noise in the given audio channel. Specified in dB below the main signal or in a percentage.

third-octave band A frequency band whose cutoff frequencies have a ratio of $2^{1/3}$, which is approximately 1.26. The cutoff frequencies of 891 Hz and 1112 Hz define the 1000 Hz third-octave band in common use.

third-octave filter See *filter, third-octave*.

tie-line Connection line for signals (such as between editing rooms).

TIM Transient intermodulation distortion; occurs when the feedback circuitry is not acting fast enough when the amplifier is exposed to fast transient signals.

timbre The quality of a sound related to its harmonic structure.

time code Digital code that marks time with information on frames, seconds, minutes, hours, and so on. Examples: SMPTE, EBU, MTC.

time constants In transmission and recording, corner frequencies of de-emphasis are often defined by time constants, such as 3180 µs. The frequency is 50 Hz, since the time constant $\tau = 1/2\pi f$.

time window The amount of time required for and/or represented by a measurement or other process. Often used interchangeably with Time Constant.

tinnitus Ringing in the ear or noise sensed in the head typically caused by excessive exposure to high sound levels.

tolerance The maximum permissible deviation from the specified quantity.

TP Abbreviation for true peak (unweighted). The term is used in digital audio metering and is then related to full scale. See *true peak*.

transcode To convert material from one coded representation to another. In the worst case this is complete decoding of the material, and re-encoding. For example, to convert from one bit-rate to another, or to insert processors in the signal path. Transcoding is frowned upon because it can multiply the artifacts of codecs.

transfer function The output to input relationship of a structure. For instance, in a PA system from the output of a mixing console to the sound in a listening position.

transformer An electronic device consisting of two or more coils used to couple one circuit to another.

transient A relatively high amplitude, suddenly decaying, peak signal level.

true peak level See ITU-R BS.1770–3, Annex 2.

true peak The maximum absolute level of the signal waveform in the continuous time domain, measured per ITU-R BS.1770. Its units are dB TP, meaning decibels relative to nominal 100%, true peak.

truncation Cutting off of superfluous bits without weighing their value.

twisted-pair A way to ensure balancing in cables.

ultrasound Sound at frequencies above the audible range, that is, above about 20 kHz.

unbalanced A connection where the chassis/ground is a part of the circuit between electroacoustic devices. Also referred to as single-ended.

UTP Unshielded Twisted Pair; unshielded twisted pairs, such as those used in Cat5 cables.

varispeed Variable speed, generally in playback systems.

VCA Voltage-controlled amplifier.

VCR Video cassette recorder.

voiceover See *ducking*.

volt, V Unit for electrical voltage.

Vpp Measure of the magnitude of an electrical voltage from peak to peak.

VSC Virtual Sound Check. A method for checking PA/SR setups applying prerecorded sound sources, recorded using the same microphones and instruments as the act is going to use. Typically recorded during an earlier concert.

VTR Video tape recorder.

VU Volume Unit, a level meter (standard volume indicator) for audio recording; time constant: 300 ms.

watt, W The unit of electrical acoustic power.

wav File format for sound.

waveform The waveform is the shape of a time-domain signal as seen on a DAW screen or an oscilloscope screen. It is a visual representation or graph of the instantaneous value of the signal plotted against time.

weber, Wb Measurement unit for magnetic flux.

weighting A measurement is weighted when the signal has passed through a filter (e.g., an A filter).

white noise A random signal that contains all frequencies, with constant energy per Hz.

window function A mathematical function applied to Fourier Transform analysis. Can be understood as a fade in/fade out, to ensure a minimum of spectral artifacts from analyzing a finite and not infinite signal. The Hanning window is an example of a window that works in most situations.

woofer Bass loudspeaker.

working-level That signal level at which a circuit is nominally modulated. Sometimes called the "0-level."

WRMS Weighted RMS value. The abbreviation is also used for the RMS value of the power.

X-curve A target curve for the frequency response of reproduced sound in cinemas, standardized by SMPTE.

XLR Professional cable connector for audio. Pin 1: shield, pin 2: +, pin 3: −.

Zwicker Loudness A technique developed by Dr. E. Zwicker for calculating a real-time estimate for the loudness of sound as perceived by the human ear. This methodology is primarily used in connection with calculation and perception of noise exposure.

Index

Note: **Boldface** page references indicate tables. *Italic* references indicate figures.