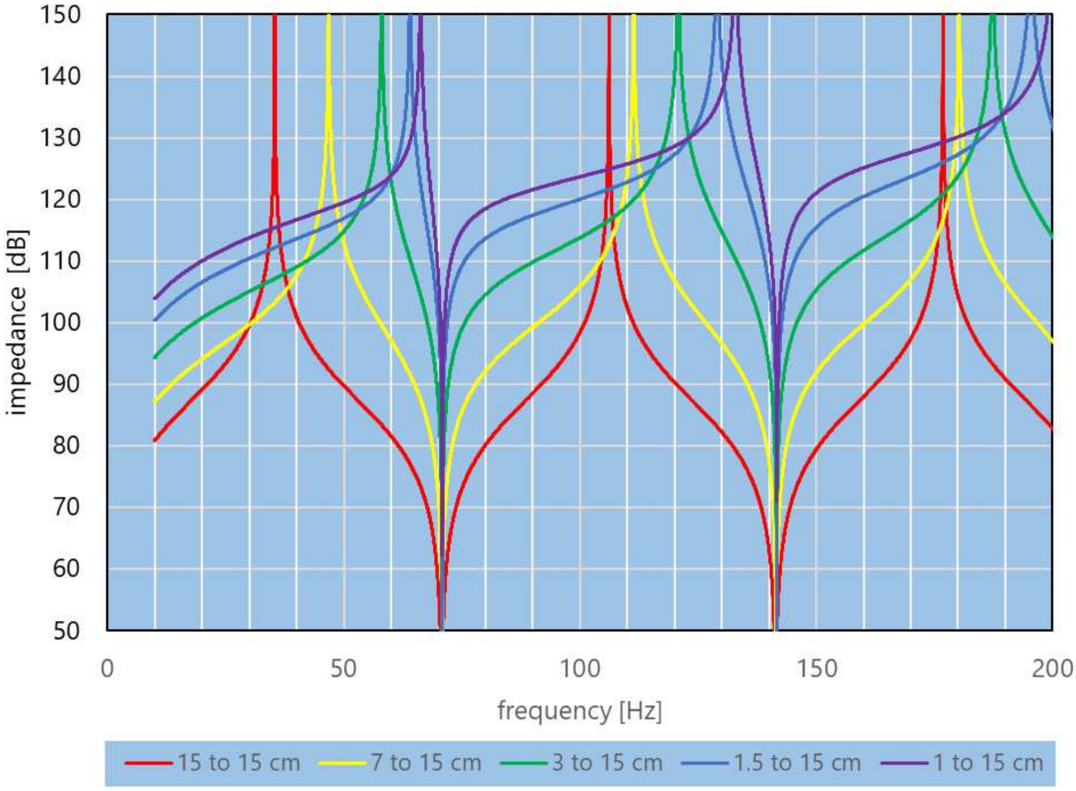
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

227-0478-00L

Lecture Notes on Acoustics II



Kurt Heutschi
Institute for Signal- and Informationprocessing, ISI
Swiss Federal Institute of Technology, ETH
CH-8092 Zurich

2020-08-07

Spice simulation of the input impedance of conical tubes of length 2.4 m.

Contents

1 Electrical-mechanical-acoustical analogies **1**
1.1 General remarks . . . 1
1.2 Mechanical systems . . . 1
1.2.1 Quantities for the description of mechanical systems . . . 1
1.2.2 Mechanical elements . . . 1
1.2.3 Mechanical resonance - spring pendulum . . . 3
1.3 Acoustical systems . . . 5
1.3.1 Quantities for the description of acoustical systems . . . 5
1.3.2 Acoustical elements . . . 5
1.4 Analogies . . . 6
1.4.1 Quantities . . . 6
1.4.2 Impedances . . . 7
1.4.3 Mechanical elements . . . 7
1.4.4 Acoustical elements . . . 8
1.4.5 Measurement of acoustical impedances . . . 9
1.4.6 Examples . . . 9
1.4.7 Long tubes represented by a series of lumped elements . . . 12
1.4.8 Long tubes represented by distributed elements . . . 13
1.5 Coupling of mechanical, acoustical and electrical systems . . . 16
1.5.1 Interfaces . . . 16
1.5.2 Dual conversion . . . 17
1.5.3 Examples . . . 18

2 Microphones **20**
2.1 Electrostatic microphone . . . 20
2.1.1 Principle of operation . . . 20
2.1.2 Complete equivalent network . . . 21
2.1.3 Simplified network . . . 25
2.1.4 Free-field and pressure microphones . . . 26
2.1.5 Power supply of the microphone amplifier . . . 26
2.1.6 Overview of measuring microphones . . . 27
2.2 Dynamic microphone . . . 27
2.2.1 Principle of operation . . . 27
2.2.2 Complete equivalent network . . . 28
2.2.3 Simplified networks for different frequency ranges . . . 29
2.3 Microphone directivity . . . 31
2.3.1 Omnidirectional microphone . . . 31
2.3.2 Figure of eight microphone . . . 32
2.3.3 Cardioid microphone . . . 32
2.3.4 Microphones with switchable directivity . . . 34
2.3.5 Directional microphone . . . 35
2.3.6 Proximity effect . . . 36
2.4 Exotic transducers . . . 36
2.4.1 Microflown . . . 36
2.4.2 Soundfield Microphone . . . 37
2.5 Microphone measurements . . . 37
2.5.1 Frequency response measurements . . . 37

2.5.2 Microphone directivity . . . 38
2.5.3 Non-linear distortion - maximum sound pressure level . . . 38
2.5.4 Self-noise . . . 38

3 Loudspeakers **39**
3.1 Transducer principles . . . 39
3.1.1 Electrodynamic loudspeaker . . . 39
3.1.2 Electrostatic loudspeaker . . . 39
3.2 Sound radiation of a membrane . . . 40
3.2.1 Radiated power . . . 40
3.2.2 Radiation directivity . . . 40
3.3 Electrodynamic loudspeaker . . . 42
3.3.1 Principle of operation . . . 42
3.3.2 Equivalent network . . . 42
3.3.3 Nonlinearities . . . 44
3.3.4 Mounting in a cabinet . . . 44
3.3.5 Frequency response of sound pressure . . . 46
3.3.6 Thiele-Small parameter . . . 47
3.3.7 Velocity and displacement of the loudspeaker membrane . . . 49
3.3.8 Bass reflex cabinet . . . 49
3.3.9 Bandpass cabinet . . . 49
3.3.10 Feed-back strategies . . . 50
3.3.11 Digital equalizing . . . 50
3.4 Horn loudspeaker . . . 50
3.5 Electrostatic loudspeaker . . . 52
3.6 Loudspeaker systems . . . 53
3.7 Loudspeaker measurements . . . 54
3.7.1 Cumulative decay spectrum . . . 55

4 Sound storage media **58**
4.1 Vinyl records . . . 58
4.2 Analog tape recorder . . . 59
4.3 Compact Disc . . . 63
4.4 DVD Audio and Super Audio CD . . . 63
4.4.1 DVD Audio . . . 64
4.4.2 Super Audio CD . . . 64
4.5 Harddisk recorder . . . 64
4.6 Compressed audio formats . . . 64
4.6.1 Lossless compression . . . 65
4.6.2 Lossy compression . . . 65

5 Recording technique **68**
5.1 Stereo recordings . . . 68
5.1.1 Intensity stereophony . . . 68
5.1.2 Time of arrival stereophony . . . 69
5.1.3 Mixed stereophony . . . 69
5.2 Positioning of microphones . . . 70
5.3 Surround sound . . . 71

6 Reproduction of audio signals **72**
6.1 Stereo loudspeakers . . . 72
6.2 Headphones . . . 72
6.3 Loudspeaker reproduction with crosstalk cancelation . . . 73
6.4 Room simulators . . . 73
6.5 Surround sound . . . 74
6.6 Wave field synthesis . . . 75
6.7 Ambisonics . . . 76
6.8 Auro 3D . . . 80
6.9 Audibility of transmission errors . . . 80
6.9.1 Distortion of the amplitude response . . . 80

6.9.2 Distortion of the phase response . . . 80
6.9.3 Non-linear distortion . . . 80
6.9.4 Residual noise . . . 81
6.10 Differential signaling . . . 82
6.11 Digital connections, AES/EBU, SPDIF . . . 82

7 Public address systems **84**
7.1 Introduction . . . 84
7.1.1 Considerations of sound pressure levels . . . 85
7.1.2 Maximal amplification . . . 85
7.1.3 Examples of sources with high directivity . . . 88
7.1.4 Feed-back suppression . . . 89
7.2 Speech intelligibility . . . 90
7.2.1 Articulation Loss . . . 91
7.2.2 Speech Transmission Index . . . 92
7.2.3 Relation between Articulation Loss and RASTI . . . 92
7.2.4 Deutlichkeit D50 . . . 92
7.2.5 STI-PA . . . 92
7.3 Localization . . . 93
7.4 Types of PA systems . . . 93
7.4.1 Centralized PA system . . . 93
7.4.2 Distributed PA system . . . 93
7.4.3 100 V systems . . . 93
7.5 Tools for the simulation of PA systems . . . 94
7.6 Electroacoustical adjustment of reverberation . . . 94

8 Audio signal processing **95**
8.1 Effects . . . 95
8.1.1 Equalization . . . 95
8.1.2 Reverbs . . . 95
8.1.3 Compressors / Limiters . . . 95
8.1.4 Noise Gates . . . 96
8.1.5 Flanger . . . 97
8.1.6 Chorus . . . 97
8.1.7 Phaser . . . 97
8.1.8 Leslie . . . 97
8.1.9 Pitch-Shift and time-scaling . . . 97
8.2 Audio watermarking . . . 98
8.2.1 Introduction . . . 98
8.2.2 Requirements for audio watermarking . . . 98
8.2.3 Methods for audio watermarking . . . 98
8.2.4 Tests of audio watermarking . . . 100

9 Appendix **101**
9.1 dB conventions in electroacoustics . . . 101

Chapter 1

Electrical-mechanical-acoustical analogies

1.1 General remarks

A major topic in electro-acoustics is the analysis of vibrating systems, such as microphones or loudspeakers. These systems usually contain mechanical, acoustical and electrical subsystems. A common approach is to split the structures into small units that can be described by concentrated elements. There is no principal difference between the fundamental differential equations that describe the behavior of a mechanical, an acoustical or an electrical system. It is therefore possible to transform a system of one type into an equivalent system of an other type. To make later usage of the network analysis tools that are available for electrical systems it is beneficial to transform the non-electrical system parts into analogue electrical ones. By introduction of suitable interfaces between different subsystems it will then be possible to find one single electrical network that describes the complete system. In many cases the inspection of the equivalent electrical network will already give fundamental insights into the principal operation of the system.

1.2 Mechanical systems

1.2.1 Quantities for the description of mechanical systems

At any point of a mechanical system the following two fundamental quantities can be identified:

- velocity $u = \frac{dx}{dt}$
- force F.

The velocity u can easily be transformed into other quantities that describe the movement, such as displacement $x = \int u\mathrm{d}t$ or acceleration $a = \frac{du}{dt}$.

1.2.2 Mechanical elements

Mass

The ideal mass is assumed to be incompressible, which means that each point on the mass body has identical velocity (Fig. 1.1). The fundamental physical principal is Newton's law (1.1).

$$\begin{aligned} u_1 = u_2 &= u \\ F_1 - F_2 &= m\frac{du}{dt} \end{aligned} \tag{1.1}$$

Spring

In the ideal spring the excursion is proportional to the applied force, according to Hook's law (1.2). In the idealized case the stiffness s is assumed independent of the excursion. It should be noted that there is no force drop off along the spring.

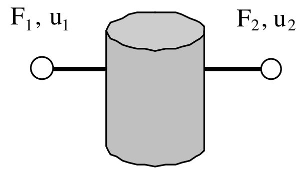


Figure 1.1: Symbol of the ideal mass.

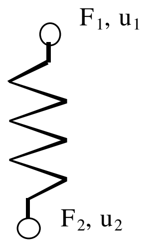


Figure 1.2: Symbol of the ideal spring.

$$
\begin{aligned}
F_1 &= F_2 = F \\
F &= s \int u_1 - u_2 \mathrm{d}t
\end{aligned} \tag{1.2}
$$

Friction

In the ideal friction element there is proportionality between friction force and velocity (1.3). Again, as in case of the spring, there is no force drop off over the element.

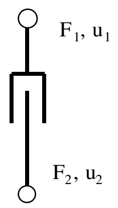


Figure 1.3: Symbol of an ideal friction element.

$$
\begin{aligned}
F_1 &= F_2 = F \\
F &= R(u_1 - u_2)
\end{aligned} \tag{1.3}
$$

Links

Links are connectors for mechanical elements. The idealization assumes that the links have no mass and are incompressible. In some cases it may be necessary to model the mass of a real link by a separate mass element.

Levers

Levers are mechanical transformers. In the ideal case they have no mass and can rotate without friction losses. For the lever in Fig. 1.4 the relations between forces and velocities obey Eq. 1.4.

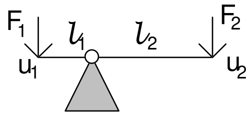


Figure 1.4: Geometry of a lever.

$$
\begin{aligned}
F_1 l_1 - F_2 l_2 &= 0 \\
u_1 l_2 + u_2 l_1 &= 0
\end{aligned}
\tag{1.4}
$$

Mechanical sources

Mechanical systems can be excited by sources that impress a force or a velocity or a mixture, depending on the load. An example of a pure force source is a conductor in a magnetic field. The force acting on the conductor depends only on the strength of the magnetic field and the current in the conductor but *not* on its velocity. A pure velocity source would excite a mechanical structure with a certain velocity, independent of the necessary force.

1.2.3 Mechanical resonance - spring pendulum

The spring pendulum is a prototype mechanical resonance system that is found in many electro acoustical transducers. It consists of a mass, a spring and a damping element as shown in Fig. 1.5.

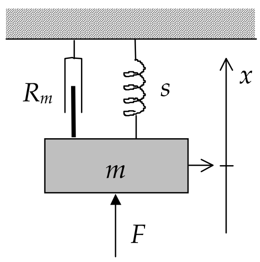


Figure 1.5: Spring pendulum acting as a mechanical resonance system. A mass m is suspended from a ceiling by a spring with stiffness s. In addition there is a damping element R_m with velocity dependent friction.

We are interested in the movement of the mass for a given exterior harmonic force F. The movement shall be described as the time dependent excursion $x(t)$, the velocity $u(t) = dx/dt$ and the acceleration $a(t) = d^2x/dt^2$.

As a first step the equilibrium of the forces is established:

$$F_{\text{acceleration}} + F_{\text{friction}} + F_{\text{spring}} = F \tag{1.5}$$

With the harmonic excitation in complex writing, Eq. 1.5 yields

$$m\frac{d^2x}{dt^2} + R\frac{dx}{dt} + sx = \hat{F}e^{j\omega t} \tag{1.6}$$

The general solution for this differential equation (1.6) is

$$x(t) = \frac{\hat{F}e^{j\omega t}}{(j\omega)^2 m + j\omega R + s} = \frac{\hat{F}e^{j\omega t}}{j\omega\left(j\omega m + R + \frac{s}{j\omega}\right)} \tag{1.7}$$

The expressions simplify significantly if a new variable Z_m is introduced:

$$Z_m = j\omega m + R + \frac{s}{j\omega} \tag{1.8}$$

With Z_m, the solution (1.7) reads as

$$x(t) = \frac{\hat{F}e^{j\omega t}}{j\omega Z_m} \tag{1.9}$$

The velocity $u(t)$ is

$$u(t) = \frac{\hat{F}e^{j\omega t}}{Z_m} \tag{1.10}$$

For the acceleration $a(t)$ we get

$$a(t) = \frac{j\omega\hat{F}e^{j\omega t}}{Z_m} \tag{1.11}$$

The behavior of the spring pendulum is summarized in Table 1.1. Figure 1.6 finally shows the frequency responses of the excursion x, the velocity u and the acceleration a. Of special interest is the fact that the three quantities show a frequency independent amplitude for a certain frequency range.

ω	condition	Z_m	$x(t)$	$u(t)$	$a(t)$
low	$\frac{s}{\omega} \gg \omega m, \frac{s}{\omega} \gg R$	$\approx \frac{s}{j\omega}$	$\boldsymbol{\frac{\hat{F}e^{j\omega t}}{s}}$	$\frac{j\omega\hat{F}e^{j\omega t}}{s}$	$\frac{-\omega^2\hat{F}e^{j\omega t}}{s}$
at resonance	$R \gg \omega m, R \gg \frac{s}{\omega}$	$\approx R$	$\frac{\hat{F}e^{j\omega t}}{j\omega R}$	$\boldsymbol{\frac{\hat{F}e^{j\omega t}}{R}}$	$\frac{j\omega\hat{F}e^{j\omega t}}{R}$
high	$\omega m \gg R, \omega m \gg \frac{s}{\omega}$	$\approx j\omega m$	$\frac{-\hat{F}e^{j\omega t}}{\omega^2 m}$	$\frac{-j\hat{F}e^{j\omega t}}{\omega m}$	$\boldsymbol{\frac{\hat{F}e^{j\omega t}}{m}}$

Table 1.1: Behavior of the spring pendulum for the three frequency ranges *low* (below resonance), *at resonance* and *high* (above resonance). In bold are the cases with frequency independent behavior.

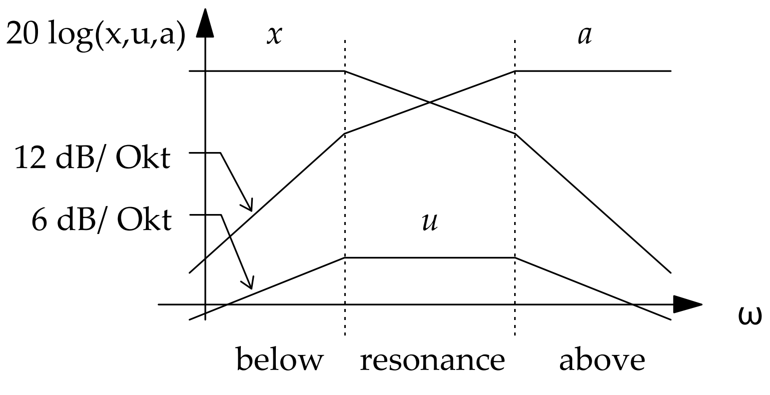


Figure 1.6: Frequency response of the excursion x, the velocity v and the acceleration a for the spring pendulum with force excitation.

The majority of acoustical transducers contain a mechanical resonance system as discussed above. In case of microphones, the electrical output is either proportional to the excursion or the velocity of the membrane. From above follows that the membrane system has to operate:

- below resonance in case of microphones with proportionality between electrical output and membrane excursion
- at resonance in case of microphones with proportionality between electrical output and membrane velocity

1.3 Acoustical systems

1.3.1 Quantities for the description of acoustical systems

The two fundamental acoustical quantities are

- sound pressure p
- volume flow or volume velocity Q which is in relation to sound particle velocity v by $Q = \int_S v \mathrm{d}S$ where S is the cross sectional area.

1.3.2 Acoustical elements

Acoustical mass

The ideal acoustical mass is accelerated but not compressed air. An acoustical mass can be realized by a small tube (Figure 1.7) where it is assumed that the sound wave length is much larger than the dimensions of the tube.

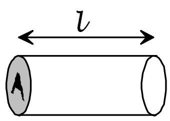


Figure 1.7: Realization of an acoustical mass by a thin tube of length l and cross sectional area A.

The properties of an acoustical mass can be derived from Newton's law. The force difference ΔF on both sides of the tube leads to an acceleration of the air mass in the tube:

$$\Delta F = \rho A l \frac{dv}{dt} \tag{1.12}$$

where
ρ: density of air.

Acoustical compliance

The ideal acoustical compliance is compressed but not accelerated air. It can be realized by a cavity (Figure 1.8) whereby again it is assumed that all dimensions are small compared to the wave lengths of interest.

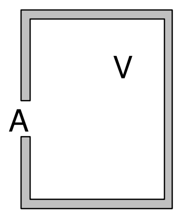


Figure 1.8: Realization of an acoustical compliance by a cavity of volume V and opening area A.

The properties of an acoustical compliance can be derived from Poisson's law (see Acoustics I) under the assumption of adiabatic behavior:

$$PV^{\kappa} = const. \tag{1.13}$$

where
P: absolute pressure
V: volume
κ: adiabatic exponent, for air $\kappa = 1.4$

It is assumed that a force F is acting on a virtual piston in the opening A. By this force the piston sinks in by Δl (1.14).

$$\Delta l = \frac{F}{s} \tag{1.14}$$

where
s: stiffness of the compliance.

As derived in Acoustics I, the value of s can be determined as

$$s = c^2 \rho \frac{A^2}{V} \tag{1.15}$$

where
c: speed of sound

It should be noted that the position of the opening of the cavity is irrelevant. As an example both variants in Figure 1.9 are identical compliances.

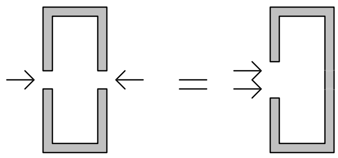


Figure 1.9: The behavior of the cavity with distributed openings is identical to the cavity with just one opening as long as the total opening area is equal.

Acoustical resistance

Acoustical resistances convert sound energy into heat. They are typically realized by porous materials. Usually the value of acoustical resistances has to be determined experimentally. However for a very small tube the resistance can be determined analytically [1]:

$$\Delta p = v \frac{8 l \eta}{r^2} \tag{1.16}$$

where
Δp: sound pressure difference on both sides of the small tube
v: sound particle velocity
l: length of the tube
r: radius of the tube ($r \ll l$)
η: dynamic viscosity of air: $= 1.82 \times 10^{-5}$Nsm^{-2}

The acoustical resistance of a small tube is always accompanied by an acoustical mass. However for small radii and low frequencies the effect of the mass can usually be neglected.

1.4 Analogies

1.4.1 Quantities

The first step in establishing the analogies is the definition of equivalent quantities in each system [2]. This can be done in different ways. Here the analogy is chosen that preserves the topology of the

[1] H. F. Olson, Elements of Acoustical Engineering, 1947
[2] A. Lenk et al., Elektromechanische Systeme, Mechanische und akustische Netzwerke, deren Wechselwirkungen und Anwendungen, Springer (2001)

network. In nodes in electrical networks the sum of all incoming currents equals zero ($\sum I = 0$). An identical condition can be formulated for the force in mechanical systems ($\sum F = 0$) and for the volume flow in acoustical systems ($\sum Q = 0$). Consequently, the analogy shown in Table 1.2 is established.

electrical	mechanical	acoustical
voltage U	velocity u	sound pressure p
current I	force F	volume flow Q

Table 1.2: Equivalent electrical, mechanical and acoustical quantities in the *FIQ* analogy.

During the transformation of mechanical and acoustical quantities into the corresponding analogue electrical ones, scaling factors have to be introduced that adjust the amplitudes and convert the units appropriately. It is often convenient to work with amplitude scaling factors of 1, however arbitrary amplitude scaling is possible.

The transformation of mechanical systems requires (1.17):

$$\begin{aligned} U &= G_1 u \quad \text{with} \quad G_1 = 1\frac{\mathsf{Vs}}{\mathsf{m}} \\ I &= \frac{1}{G_2} F \quad \text{with} \quad G_2 = 1\frac{\mathsf{N}}{\mathsf{A}} \end{aligned} \tag{1.17}$$

The transformation of acoustical systems requires (1.18):

$$\begin{aligned} U &= G_3 p \quad \text{with} \quad G_3 = 1\frac{\mathsf{Vm}^2}{\mathsf{N}} \\ I &= \frac{1}{G_4} Q \quad \text{with} \quad G_4 = 1\frac{\mathsf{m}^3}{\mathsf{As}} \end{aligned} \tag{1.18}$$

1.4.2 Impedances

Following the impedance definition for electrical elements

$$\text{impedance} = \frac{\text{voltage}}{\text{curent}} \tag{1.19}$$

corresponding relations can be found for the mechanical and acoustical elements identified above. For sinusoidal time dependencies, complex writing $x(t) = x_0 e^{j\omega t}$ is used.

1.4.3 Mechanical elements

Mass

With Eq. 1.1 the impedance of a mechanical mass m corresponds to

$$Z = \frac{u}{F} = \frac{u}{m\frac{du}{dt}} \tag{1.20}$$

With $u = u_0 e^{j\omega t}$ follows $\frac{du}{dt} = u_0 j\omega e^{j\omega t}$ and finally

$$Z = \frac{1}{j\omega m} \tag{1.21}$$

Translated into electrotechnics, the impedance of a mechanical mass corresponds to the impedance of a capacitor. Consequently the symbol for a mechanical mass is a capacitor. It should be noted that the inertia of the mechanical mass has to be understood relative to a reference system at rest. Thus the second terminal of the mass or the equivalent electrical capacitor is connected to velocity potential 0 (ground).

Spring

With Eq. 1.2 the impedance of a mechanical spring with stiffness s is found as

$$Z = \frac{u}{F} = \frac{u}{s \int u\mathrm{d}t} \tag{1.22}$$

With $u = u_0 e^{j\omega t}$ follows $\int u\mathrm{d}t = u_0 \frac{1}{j\omega} e^{j\omega t}$ and finally

$$Z = j\omega \frac{1}{s} \tag{1.23}$$

Translated into electrotechnics the impedance of a mechanical spring corresponds to the impedance of an inductance. Consequently the symbol for a mechanical spring is an inductance.

Friction

With Eq. 1.3 the impedance of a friction element R is

$$Z = \frac{u}{F} = \frac{u}{Ru} = \frac{1}{R} \tag{1.24}$$

The electrical equivalent is a resistance. Consequently the symbol for a friction element is an electrical resistance.

1.4.4 Acoustical elements

Acoustical mass

With Eq. 1.12 the impedance of the acoustical mass is found as

$$Z = \frac{p}{Q} = \frac{\frac{\Delta F}{A}}{Av} = \frac{\rho A l \frac{dv}{dt}}{AAv} \tag{1.25}$$

With $v = v_0 e^{j\omega t}$ follows $\frac{dv}{dt} = v_0 j\omega e^{j\omega t}$ and finally

$$Z = j\omega \frac{\rho l}{A} \tag{1.26}$$

Translated into electrotechnics the impedance of an acoustical mass corresponds to the impedance of an inductance. Consequently the symbol for an acoustical is an inductance.

Acoustical compliance

With Eq. 1.14 and 1.15 follows for the impedance of an acoustical compliance :

$$Z = \frac{p}{Q} = \frac{\frac{\Delta F}{A}}{Av} = \frac{c^2 \rho \frac{A^2}{V} \Delta l}{AAv} = \frac{c^2 \rho \frac{A^2}{V} \int v\mathrm{d}t}{AAv} \tag{1.27}$$

With $v = v_0 e^{j\omega t}$ follows $\int v\mathrm{d}t = v_0 \frac{1}{j\omega} e^{j\omega t}$ and finally

$$Z = \frac{c^2 \rho}{j\omega V} \tag{1.28}$$

Translated into electrotechnics the impedance of an acoustical compliance corresponds to the impedance of a capacitor. Consequently the symbol for an acoustical compliance is a capacitor. It should be noted that one terminal of these capacitors is usually connected to ground potential. This represents the fact that any net force is carried by the mounting support.

Acoustical resistance

The impedance of an acoustical resistance is frequency independent. The corresponding electrical element is thus the resistance. For a small tube the impedance can be determined with Eq. 1.16 as

$$Z = \frac{p}{Q} = \frac{8l\eta}{\pi r^4} \tag{1.29}$$

where
l: Length of the tube
r: radius of the tube ($r \ll l$)
η: dynamic viscosity of air = 1.82 $\times 10^{-5}$ Nsm^{-2}

Overview

Table 1.3 shows the compilation of the electrical, mechanical and acoustical impedances.

	electrical	mechanical	acoustical
$Z = R$	resistance R	friction resistance $R = \frac{1}{R_m} G_1 G_2$	resistance $R = R_m G_3 G_4$
$Z = \frac{1}{j\omega C}$	capacitor C	mass $C = m \frac{1}{G_1 G_2}$	compliance $C = \frac{V}{\rho c^2} \frac{1}{G_3 G_4}$
$Z = j\omega L$	inductance L	spring $L = \frac{1}{s} G_1 G_2$	mass $L = \frac{\rho l}{A} G_3 G_4$

Table 1.3: Equivalence of mechanical, acoustical and electrical elements ($G_1 \ldots G_4$ see Eq. 1.17 and 1.18).

1.4.5 Measurement of acoustical impedances

In many fields of applications it is desirable to measure acoustical impedances. One application is for example the quality control of musical instruments. Figure 1.10 shows one strategy to measure impedances. The central element is an acoustical resistance that is arranged in series to the impedance of interest. While the configuration is excited by a loudspeaker, the sound pressure (p_1, p_2) is measured on both sides of the acoustical resistance R.

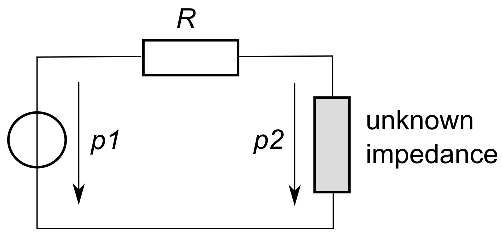


Figure 1.10: Schematics for the measurement of acoustical impedances.

If the difference of the pressures $p_1 - p_2$ is kept constant, the volume flow through the unknown impedance is constant and thus the pressure p_2 is proportional to the unknown impedance. The physical construction of the measurement arrangement is shown in Figure 1.11, Figure 1.12 gives some typical frequency response curves.

1.4.6 Examples

Combination of acoustical mass and acoustical compliance in series and in parallel

An acoustical mass and a compliance can be combined in two ways. The corresponding equivalent electrical circuits can be found by checking whether sound pressure or volume flow is constant over the first element.

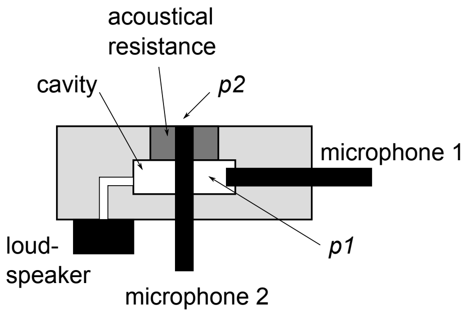


Figure 1.11: Construction for the measurement of acoustical impedances. The unknown impedance is put on top of the device. Microphone 1 measures the sound pressure p_1, microphone 2 measures p_2. Within the cavity the absolute pressure is $P_0 + p_1$, at the entrance of the unknown impedance the pressure is $P_0 + p_2$. The pressure outside is P_0.

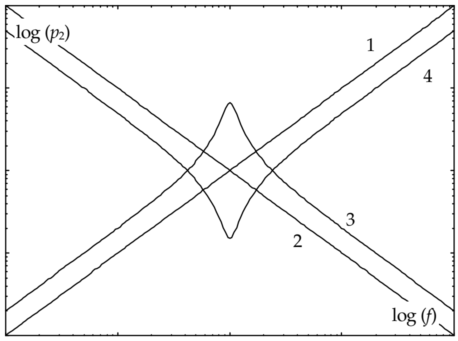


Figure 1.12: Typical frequency response curves of measured impedances: 1: acoustical mass, 2: acoustical compliance, 3: parallel resonance circuit of a mass and a compliance, 4: series resonant circuit of a mass and a compliance.

In the arrangement according to Figure 1.13 the first element is an acoustical compliance. This corresponds to compressed but not accelerated air which means that there is no pressure drop over the element. The mass as second element "sees" the same pressure (or voltage in the electrical analogue) and thus the two elements are in parallel.

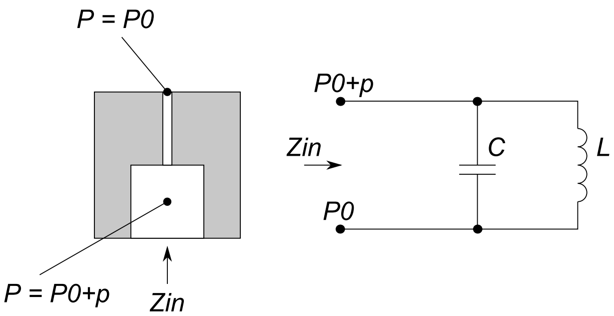


Figure 1.13: First possible arrangement of an acoustical mass and an acoustical compliance (section view) with the analogue electrical network. P is absolute pressure.

In the arrangement according to Figure 1.15 the first element is an acoustical mass. This corresponds to accelerated but not compressed air which means that there is no volume flow drop over the element. The compliance as second element "sees" the same volume flow (or current in the electrical analogue) and thus the two elements are in series.

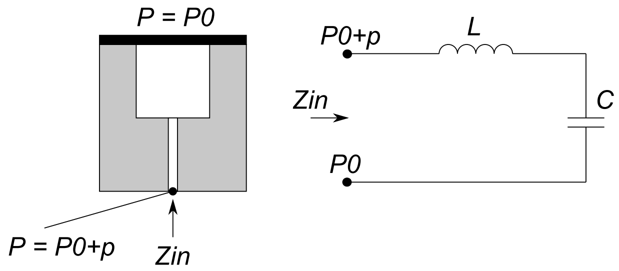


Figure 1.14: Second arrangement of an acoustical mass and an acoustical compliance (section view) with the analogue electrical network. P is absolute pressure. It should be noted that any force due to the pressure difference in the cavity and P_0 is provided by the mounting support.

Combination of mechanical mass and mechanical spring in series and in parallel

Similarly to the acoustical example above the two possible arrangements of a mechanical mass and spring are investigated. Again with help of considering which quantity remains constant over the first element, the corresponding electrical network can be identified.

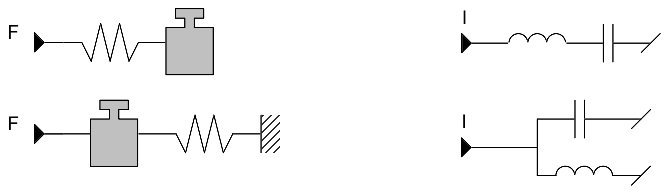


Figure 1.15: Equivalent electrical networks for the two possible arrangements of a mechanical mass and spring.

Mufflers or silencers

Mufflers or silencers are acoustical low pass filters that let pass low frequency air flow but attenuate high frequency noise. Figure 1.16 shows a possible realization with three cavities that are connected to each other by tubes and the corresponding electrical network.

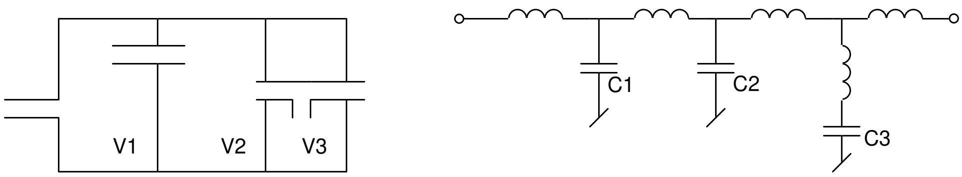


Figure 1.16: Example of the physical realization of a muffler and the corresponding electrical network.

Spring pendulum

Here the mechanical resonance system *spring pendulum* (Figure 1.17) shall be discussed by investigating the behavior of the corresponding electrical network (Figure 1.18).

The network in Figure 1.18 demonstrates that the exterior force F (current in the electrical analogue) is divided to the three elements *mass*, *spring* and *friction*. The excursion of the mass is proportional to the force of the spring and thus to the current F_F ($\Delta x = F_F/s$). The velocity is proportional to the force in the friction element and thus to the current F_R ($u = F_R/R$). The acceleration finally is proportional to the inertia and thus to the current F_M ($a = F_M/m$). The discussion of the frequency responses of the different quantities reduces to a discussion of the corresponding currents F_F, F_R and F_M in the electrical network.

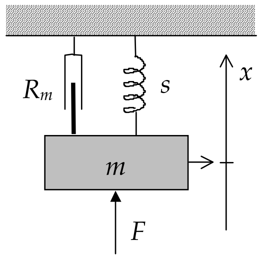


Figure 1.17: Spring pendulum as mechanical resonance system. A mass m is suspended from a ceiling by a spring with stiffness s. In addition there is a damping element R_m with velocity dependent friction.

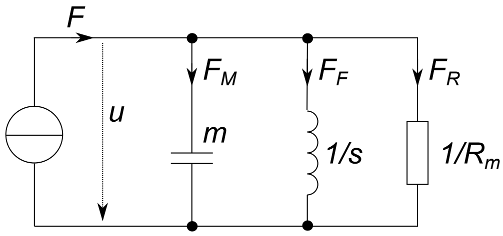


Figure 1.18: Equivalent electrical network that represents the spring pendulum. All elements are in parallel as they all have identical velocity u.

Damping of mechanical vibrations

Vibration dampers reduce the oscillations of a structure. If the oscillations of the structure are limited to a narrow frequency band, this can be done by a tuned resonance system consisting of a spring and a mass (Figure 1.19).

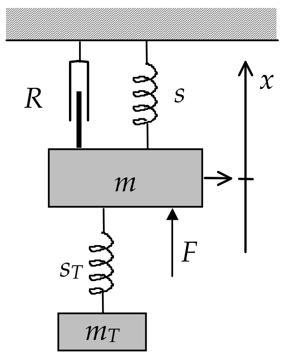


Figure 1.19: Vibration damper consisting of a spring s_T and a mass m_T to reduce the oscillations of the mass m (see Figure 1.17).

The corresponding electrical network is shown in Figure 1.20. The additional L-C series resonance circuit ($1/s_T$ and m_T) is tuned in such a way that the impedance vanishes at the frequency for which maximum damping is needed. The force F is short circuited and thus the velocity u over the mass element m tends to 0. However it has to be noted that the additional LC network introduces additional resonances of the total system which may amplify the oscillation at other frequencies.

1.4.7 Long tubes represented by a series of lumped elements

Often encountered structures in acoustical systems are tubes of small diameter but considerable length. The description of such tubes with lumped elements is possible even if their longitudinal extension is large compared to the shortest wave length of interest. The trick lies in a subdivision in slices that

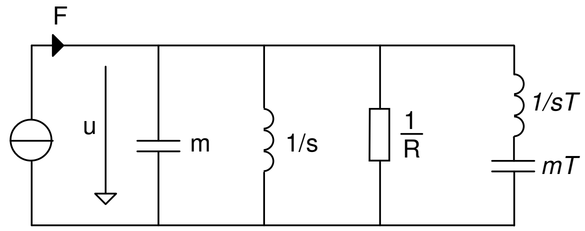


Figure 1.20: Equivalent electrical network for the mechanical spring pendulum with a vibration damper consisting of a spring s_T and a mass m_T.

fulfill the requirement regarding the ratio of dimension and wave length.

In the following a tube with cross sectional area A is considered. It is assumed that the tube is subdivided into sections of length l where $l \ll \lambda$. In the lossless case such a section element represents a mass and a compliance property. The obvious representation as an analogue electrical network is a T-element with two longitudinal inductances and a capacitor across. The two inductances both share the length l of the slice and thus have a value $L = \frac{\rho l}{2A}$ each. The compliance corresponds to a capacitance of $C = \frac{Al}{\rho c^2}$. While putting these T-elements in series two adjacent inductances can summarized and replaced by one inductance. The resulting network corresponds to a lossless electrical transmission line.

In Figure 1.21 the analogue electrical network for a hard terminated tube of length 0.24 m and cross sectional area $A = 10^{-4}$ m^2 is shown. The boundary condition "hard termination" corresponds to vanishing sound particle velocity. This is fulfilled by an open termination in the electrical network (current = 0). For the discretization a length of the slices of 0.04 m was assumed. The result of a numerical simulation with Spice is shown in Figure 1.22. It can be seen that the lowest resonance ($\lambda/4$) perfectly matches the expected frequency. However for the higher resonances the error increases for increasing frequency due to the fact that the chosen discretization is no longer fine enough.

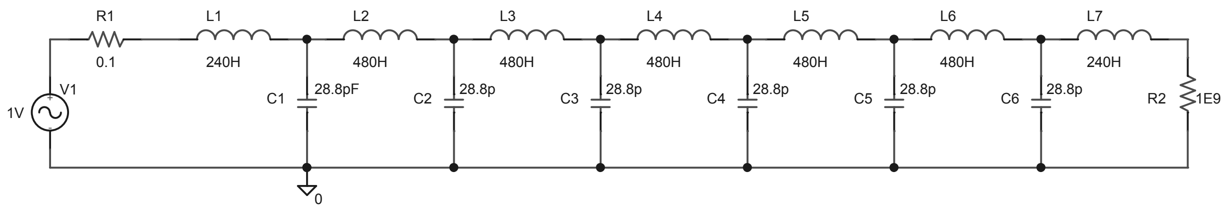


Figure 1.21: Equivalent electrical network of a tube or acoustical transmission line.

1.4.8 Long tubes represented by distributed elements

As an alternative to the transmission line representation from above, long tubes can be represented by distributed elements. They are more convenient for an analytical treatment. Distributed elements are introduced as four-poles in equivalent networks (Fig. 1.23).

The four-pole characteristics of the long tube of length d can be derived as follows. A tube allows two plane waves with pressure amplitudes A and B to travel in opposite directions. The sound field $\check{p}(x)$ and $\check{v}(x)$ as complex amplitude functions can thus be written as

$$\check{p}(x) = Ae^{-jkx} + Be^{jkx} \tag{1.30}$$

$$\check{v}(x) = \frac{A}{\rho c}e^{-jkx} - \frac{B}{\rho c}e^{jkx} \tag{1.31}$$

where x is the coordinate along the tube axis and k is the wave number. Setting $x = 0$ at the left port of the four-pole with index 1 and $x = d$ at the right port with index 2 allows to eliminate the pressure amplitudes A and B and express the relations of the four-pole as:

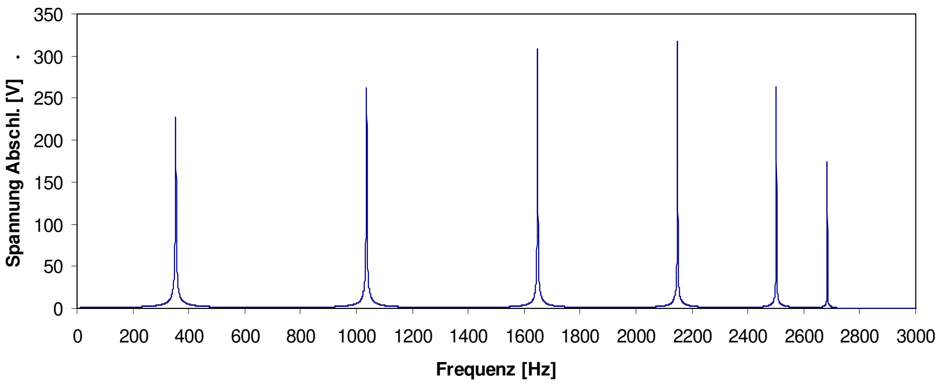


Figure 1.22: Result of a Spice simulation performed for the network in Figure 1.21 showing the voltage (sound pressure) at the end of the tube. In the simulation the resistor $R2$ had to be introduced with a very large but finite value.

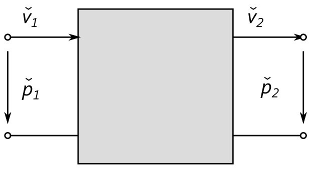


Figure 1.23: General four-pole that establishes the relations between pressure and velocity at the two ports 1 and 2.

$$\check{p}_2 = \frac{\check{p}_1}{2}e^{-jkd} + \frac{\check{p}_1}{2}e^{jkd} + \frac{\check{v}_1\rho c}{2}e^{-jkd} - \frac{\check{v}_1\rho c}{2}e^{jkd} = \check{p}_1\cosh(jkd) - \check{v}_1\rho c\sinh(jkd) \tag{1.32}$$

$$\check{v}_2 = \frac{1}{\rho c}\frac{\check{p}_1}{2}e^{-jkd} - \frac{1}{\rho c}\frac{\check{p}_1}{2}e^{jkd} + \frac{\check{v}_1}{2}e^{-jkd} + \frac{\check{v}_1}{2}e^{jkd} = -\check{p}_1\frac{1}{\rho c}\sinh(jkd) + \check{v}_1\cosh(jkd) \tag{1.33}$$

It can be shown [3], that the four-pole with the behavior described by Eq. 1.32 and 1.33 can be represented by a T-type circuit with generalized impedances (Fig. 1.24).

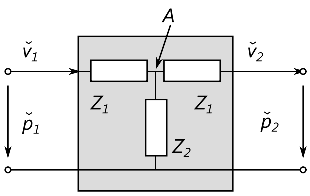


Figure 1.24: Four-pole as T-type circuit of generalized impedances Z_1 and Z_2.

The relations between $\check{p}_1$, $\check{p}_2$, $\check{v}_1$ and $\check{v}_2$ in the T-type circuit from Fig. 1.24 are found with help of Kirchhoff's circuit laws and impedance relations:

$$\check{p}_1 = \check{v}_1 Z_1 + (\check{v}_1 - \check{v}_2)Z_2 \tag{1.34}$$

$$\check{p}_2 = -\check{v}_2 Z_1 + (\check{v}_1 - \check{v}_2)Z_2 \tag{1.35}$$

To determine the impedances Z_1 and Z_2, Eq. 1.34 and 1.35 are converted into the form of Eq. 1.32 and 1.33. From the comparison of the equations follows then

[3] F.P. Mechel, Formulas of Acoustics, Springer, 2nd edition, 2008.

$$Z_1 = \frac{\rho c \sinh(jkd)}{1+\cosh(jkd)} = \rho c \frac{\cosh(jkd)-1}{\sinh(jkd)} \tag{1.36}$$

$$Z_2 = \frac{\rho c}{\sinh(jkd)} \tag{1.37}$$

For real-valued wave numbers k which is the case for air ($k = \omega/c$), Z_1 and Z_2 can be simplified further

$$Z_1 = j\rho c \frac{1-\cos(kd)}{\sin(kd)} \tag{1.38}$$

$$Z_2 = \frac{-j\rho c}{\sin(kd)} \tag{1.39}$$

It should be noted that the pressure and the velocities related to node A in the T-type circuit in Fig. 1.24 are virtual quantities and do not correspond to any field point. As an alternative to the T-type circuit discussed above, a Π-type circuit can be used instead [4].

Application to multi-layer absorbers

Multi-layer absorbers consisting of thin sheets separated by large air cavities can be successfully modeled by an equivalent electrical network using the above introduced distributed element for the cavity in form of a T-type circuit. Fig. 1.25 shows an example with two absorber sheets. Z_a and Z_b represent the properties of the sheets, Z_t is the termination impedance. $\check{p}_1$ is the sound pressure on the rear side of sheet 1 (Z_a) and $\check{p}_2$ is the sound pressure on the front side of sheet 2 (Z_b). $\check{p}_3$ and $\check{p}_4$ are the sound pressures at the rear side of sheet 2 and on the surface of the backing.

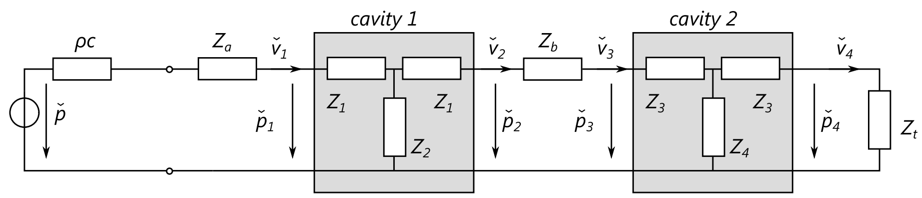


Figure 1.25: Equivalent electrical circuit for an absorber consisting of two sheets (represented by the two impedances Z_a and Z_b) separated by an air cavity (1) and mounted with an air cavity (2) in front of a termination (represented by Z_t).

For a perfectly rigid backing, Z_t becomes infinitely large. In this case $\check{v}_4 = 0$ and the T-type circuit of cavity 2 reduces to a series arrangement of Z_3 and Z_4. The equivalent impedance Z_{eq} of cavity 2 and the rigid termination (Fig. 1.26) is given as the sum of the two impedances:

$$Z_{\mathrm{eq}} = j\rho c \frac{1-\cos(kd_2)}{\sin(kd_2)} - \frac{j\rho c}{\sin(kd_2)} = -j\rho c \cot(kd_2) \tag{1.40}$$

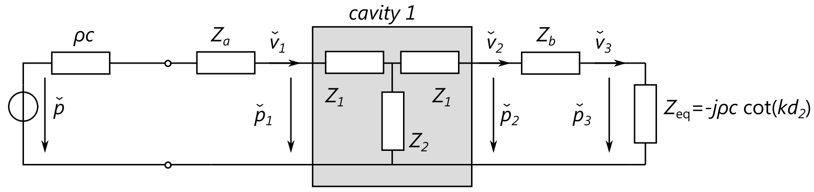


Figure 1.26: Equivalent electrical circuit for the absorber example from above in case of rigid backing.

[4] F.P. Mechel, Formulas of Acoustics, Springer, 2nd edition, 2008.

1.5 Coupling of mechanical, acoustical and electrical systems

1.5.1 Interfaces

The representation of coupled mechanical, acoustical and electrical systems by one equivalent network makes it necessary to introduce interfaces that fulfill the underlying physical conditions. Two kinds of interfaces are possible:

- Conversion of a potential into a potential quantity and a flow into a flow quantity
- Conversion of a potential into a flow quantity and vice versa

The first type of interface can be realized by a transformer (Figure 1.27).

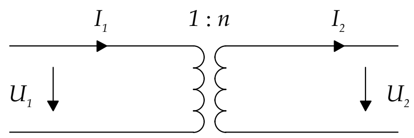


Figure 1.27: Network symbol of a transformer with conversion factor n.

The relations for a transformer with conversion factor n are

$$U_2 = nU_1 \tag{1.41}$$

$$I_2 = \frac{1}{n} I_1 \tag{1.42}$$

An example for coupled mechanical - electrical system with an interface of the first kind is a conductor that moves in a magnetic field. The induced voltage in the electrical system (potential quantity) is proportional to velocity in the mechanical system (potential quantity). The coupling can thus be represented by a transformer with $n = Bl$ (B: magnetic induction, l: length of the conductor).

The second type of interface is realized by a gyrator (Figure 1.28).

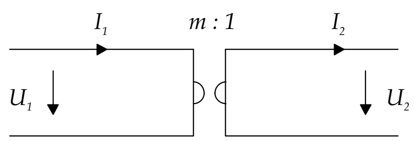


Figure 1.28: Network symbol of a gyrator with conversion constant m.

The properties of a gyrator are described by

$$U_2 = I_1 m \tag{1.43}$$

$$I_2 = U_1 \frac{1}{m} \tag{1.44}$$

An example for a coupled acoustical - mechanical system of the second kind is a stretched membrane. The mechanical force (flow quantity) acting on the membrane is proportional to the difference of the acoustical sound pressure (potential quantity) on both sides. The coupling can thus be described by a gyrator with constant $m = 1/A$ (A: area of the membrane).

1.5.2 Dual conversion

The analysis of a network containing gyrators is cumbersome. In many cases it is more convenient to get rid of gyrators by applying a dual conversion to a suitable portion of the network - the gyrators can then be replaced by transformers. To identify the part of the network that has to be replaced by it's dual, a closed curve is drawn that cuts in half all gyrators. The curve may not traverse any other element or connection of the network. In addition, an arbitrary conversion constant r has to be chosen. The following rules can be applied to determine the dual of a network:

1. A point is drawn in the middle of each mesh. An additional dot is placed outside all meshes.

2. All points introduced in step (1) are connected by auxiliary lines so that there is a line through each element and no line passes through more than one element.

3. The dual network is found by replacing all auxiliary lines be the corresponding dual element.

Figure 1.32 shows the dual conversion of all possible network elements for a conversion constant r. Figures 1.29 to 1.31 show an example of a conversion of a network in it's dual.

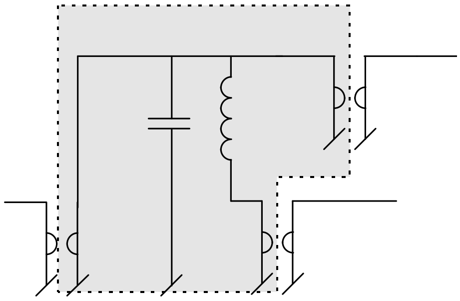


Figure 1.29: Original network. To eliminate the three gyrators a curve is drawn that cuts them in half. The region inside the curve is then converted.

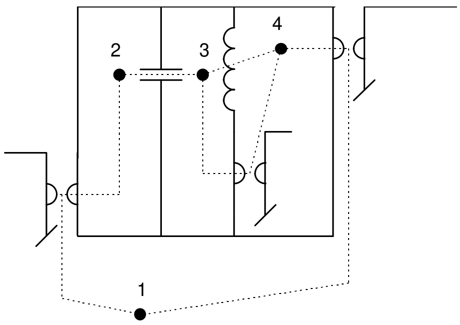


Figure 1.30: Redrawn original network with introduced mesh points that are connected by auxiliary lines.

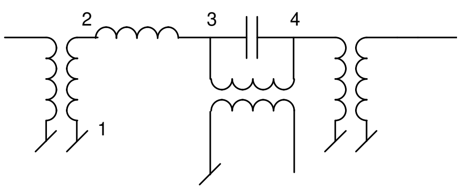


Figure 1.31: Dual network.

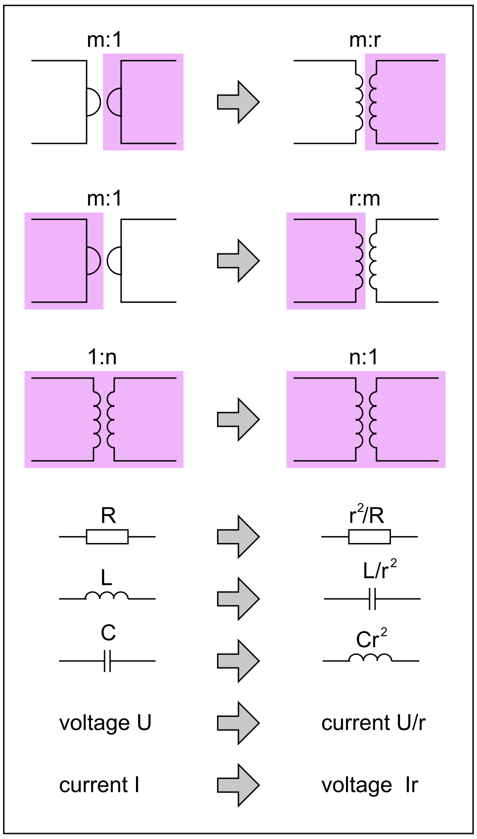


Figure 1.32: Original (left) and their dual network elements (right) for a conversion constant r.

1.5.3 Examples

Equivalent electrical network of double wall construction

Double wall constructions consist of two plates that are separated by a cavity of air. They have improved sound insulation characteristics compared to single layer constructions. A disadvantage is the occurrence of a coupling resonance for the double mass - compliance system.

For low frequencies it can be assumed that the two plates vibrate in their fundamental mode as pistons. It is further assumed that the thickness of the air layer is small compared to the smallest wave length of interest. In this case the system can be described by lumped elements. With the following quantities:

A: area of the plates
m_1: area specific weight of the first plate
m_3: area specific weight of the second plate
d: distance between the two plates (thickness of the air layer)

the analogue electrical network can be drawn as shown in Figure 1.33.

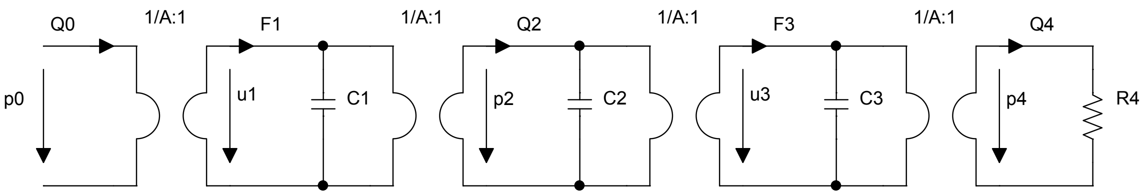


Figure 1.33: Analogue electrical network for a double wall construction at low frequencies. p_0 and p_4 represent sound pressure on both sides of the construction.

The network in Figure 1.33 has to be read from left to right. The sound pressure p_0 in the acoustical system is translated into a mechanical force F_1 that acts on the first plate by help of a $1/A : 1$ gyrator. F_1 excites the mass of the first plate - represented by the capacitance $C_1 = m_1 A$ - to a vibration of velocity u_1. On the rear side the sound particle velocity has to be equal to the plate velocity u_1. Consequently the translation from the mechanical into the acoustical system between the two plates is realized by a $1/A : 1$ gyrator. The cavity of air between the plates corresponds to an acoustical compliance and is represented by the electrical capacitance $C_2 = \frac{dA}{\rho c^2}$. The transition from the acoustical to the mechanical system of the second plate is again realized by $1/A : 1$ gyrator. The second plate is described by the capacitance $C_3 = m_3 A$. Finally the velocity u_3 of the second plate is transformed by a $1/A : 1$ gyrator into the volume flow Q_4. Under the assumption that the rear side of the construction radiates plane waves, the ratio of sound pressure and sound particle velocity is given as ρc. In the electrical network this condition transforms to the resistance $R_4 = \frac{\rho c}{A}$.

With a dual conversion (conversion factor $r = \frac{1}{A}$) of the most inner part of the network in Figure 1.33 the network in Figure 1.34 is found.

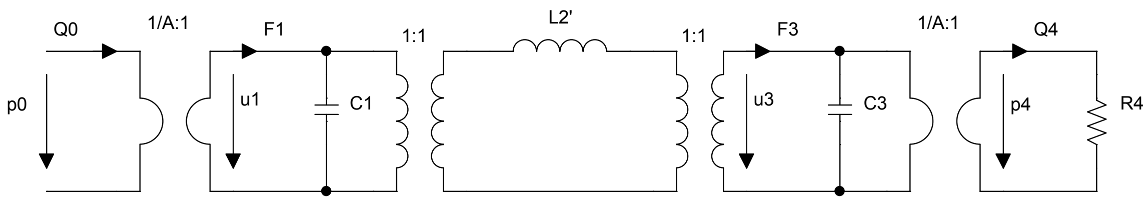


Figure 1.34: Analogue electrical network for a double wall construction at low frequencies after the first dual conversion of the most inner part ($L_2' = C_2 \frac{1}{A^2}$).

Once again a dual conversion with the same conversion factor can be applied to eliminate the last two gyrators (Figure 1.35).

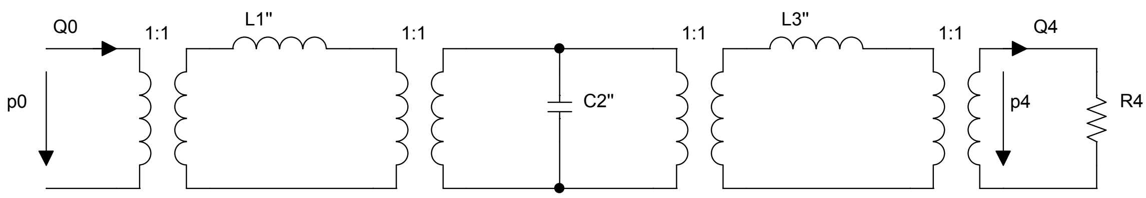


Figure 1.35: Analogue electrical network for a double wall construction at low frequencies after the second dual conversion ($L_1'' = C_1 \frac{1}{A^2} = \frac{m_1}{A}, C_2'' = L_2' A^2 = \frac{dA}{\rho c^2}, L_3'' = C_3 \frac{1}{A^2} = \frac{m_3}{A}$).

As a last step the 1:1 transformers in Figure 1.35 are omitted (Figure 1.36).

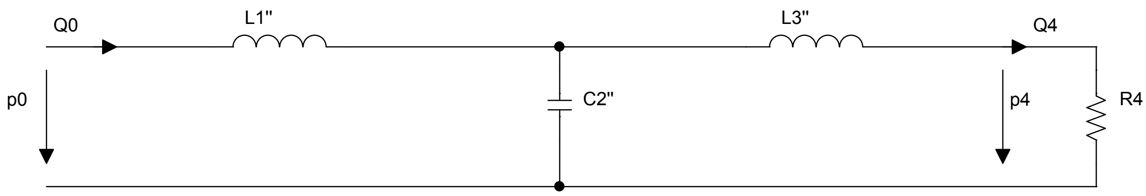


Figure 1.36: Analogue electrical network for a double wall construction at low frequencies after elimination of the 1:1 transformers.

The network in Figure 1.36 can be calculated by hand or discussed with help of network analysis tools such as Spice.

Chapter 2

Microphones

Microphones convert acoustical signals into electrical ones. The majority of microphones use a thin lightweight membrane that is set into motion by the sound field [1]. The membrane motion is converted into electrical signals based on different physical principles. The fundamental characterization of a membrane based microphone has to specify

- the membrane configuration, and
- the conversion principle.

Membrane configuration

The membrane configuration characterizes the exposition of the membrane with respect to the sound field. The motion of the membrane is the consequence of the resulting force on both sides of the membrane. If the sound field has access to both sides of the membrane, the resulting force vanishes for sound incidence parallel to the membrane. However, for perpendicular incidence, the force is maximal. On the other hand, if the rear side of the membrane is fully shielded from the sound field, the resulting force is no longer dependent on the angle of incidence.

Conversion principle

The conversion principle describes how the motion of the membrane is transformed into an electrical signal. The most important examples are the electrodynamic and the electrostatic conversion principle.

2.1 Electrostatic microphone

2.1.1 Principle of operation

Figure 2.1 shows the section of an electrostatic microphone. As the principle of operation is based on a capacitance, this type of microphone is often called condenser microphone.

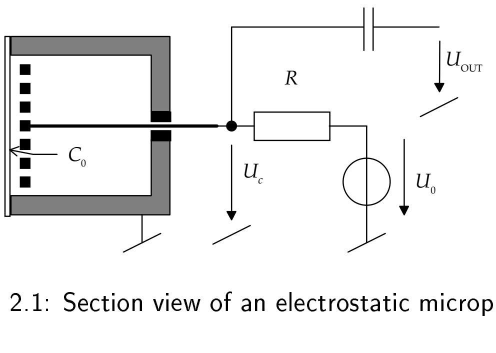


Figure 2.1: Section view of an electrostatic microphone.

[1]Alternative microphone principles could e.g. evaluate the temperature variation that is caused by a sound wave (adiabatic behavior) or observe the refraction index that is modulated by the density variation due to a sound wave.

A conductive, stretched membrane is placed in front of a back plate. This configuration forms a plate capacitor C_0 with varying capacitance depending of the position of the membrane. In series to a high-impedance resistance a high voltage (polarisation voltage) is applied to the capacitor to keep the charge on the capacitance constant.

The charge Q that is stored in a capacitor is

$$Q = CU \tag{2.1}$$

where
C: capacitance of the capacitor
U: voltage across the capacitor

The capacitance of a plate capacitor is

$$C = \frac{\epsilon_0 A}{x} \tag{2.2}$$

where
ϵ_0: electric constant = 8.85×10^{-12} $\mathrm{AsV^{-1}m^{-1}}$
A: area of one plate
x: distance between the plates

In the reference position (at rest) the charge Q_0 stored in the capacitor is

$$Q_0 = \frac{\epsilon_0 A}{x_0} U_0 \tag{2.3}$$

where
x_0: distance between the plates in the reference position

If the distance between the plates varies, the voltage change ΔU_c across the capacitor is

$$\Delta U_c = -\frac{Q_0}{\epsilon_0 A}\Delta x = -\frac{U_0}{x_0}\Delta x \tag{2.4}$$

where
Δx: displacement of one of the plates relative to the other

With Eq. 2.4 the relation between the membrane displacementΔx and the microphone signal ΔU_c is found. As the output signal is proportional to the excursion of the membrane, the system has to be operated below the spring-mass resonance.

2.1.2 Complete equivalent network

To determine the equivalent electrical network, the relevant acoustical and mechanical elements have to be identified first. Figure 2.2 shows again a section view of the electrostatic microphone with the corresponding elements.

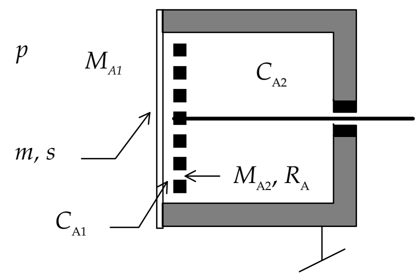


Figure 2.2: Section through an electrostatic microphone with the relevant acoustical and mechanical elements.

The following list describes the elements and quantities:

p sound pressure

m, s mechanical mass and stiffness of the membrane

M_{A1} acoustical mass in front of the membrane

C_{A1} acoustical compliance of the air between membrane and back plate

M_{A2}, R_A acoustical mass and resistance of the air in the holes of the back plate

C_{A2} acoustical compliance of the air in the rear cavity

Transition: sound field - membrane

p is the sound pressure in the undistorted sound field. This pressure takes effect on the membrane across a layer of air in front of the membrane M_{A1}. The mechanical force F on the membrane is

$$F = p_1 A \tag{2.5}$$

where
p_1: sound pressure at the membrane position
A: area of the membrane

Sound pressure is a potential quantity, the mechanical force on the other hand is a flow quantity. The interface to link the acoustical and mechanical subsystem is therefore a gyrator (Figure 2.3).

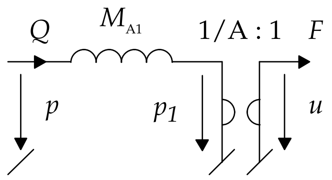


Figure 2.3: Equivalent network for the interface between sound pressure p_1 and force F acting on the membrane.

Membrane

The behavior of the membrane is determined by its mass and stiffness. Both elements move with the same velocity which is the potential quantity. Therefore they have to be arranged in parallel in the equivalent network (Figure 2.4).

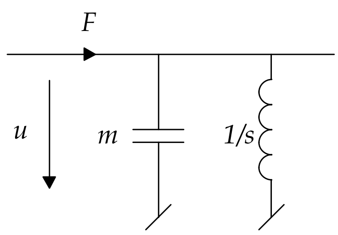


Figure 2.4: Network representing the mechanical aspects of the membrane.

Transition: membrane - interior of the microphone

The membrane moves with velocity u. This creates on the rear side an acoustical volume flow Q:

$$Q = uA \tag{2.6}$$

where
u: velocity of the membrane

A: area of the membrane

The membrane velocity u is a potential quantity while the volume flow Q is a flow quantity. Again a gyrator is needed for the translation of one quantity into the other (Figure 2.5).

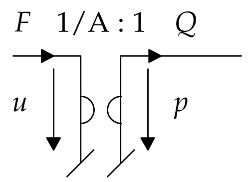


Figure 2.5: Network of the interface between membrane velocity and volume flow.

Interior of the microphone

The interior of the microphone is an acoustical system that consists of the compliance C_{A1} of the air between membrane and back plate, the acoustical mass and resistance M_{A2}, R_A of the air in the holes in the back plate and the compliance C_{A2} of the air in the rear cavity. C_{A2} is fed by volume flow through R_A and M_{A2}, resulting in a network according to Figure 2.6.

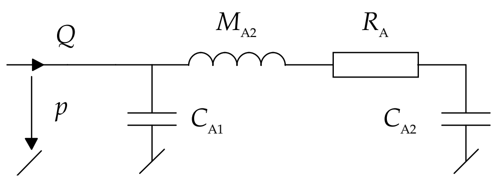


Figure 2.6: Equivalent network for the interior of the microphone.

Electrical output

The voltage at the output of the microphone is proportional to the variation of the capacity (Eq. 2.4:

$$\Delta U_c = -U_0 \frac{\Delta x}{x_0} \tag{2.7}$$

The excursion Δx of the membrane depends on the stiffness s of the spring and the force F_s acting on the spring:

$$\Delta x = \frac{F_s}{s} \tag{2.8}$$

Therefore

$$\Delta U_c = -F_s \frac{U_0}{s x_0} \tag{2.9}$$

The mechanical force is a flow quantity while the output voltage is a potential quantity. The conversion needs therefore again a gyrator (Figure 2.7). The capacitor in series to the electrical output represents the electrical capacitance of the membrane system.

The linkage of all parts finally leads to the equivalent network of the electrostatic microphone as shown in Figure 2.8.

The network in Figure 2.8 contains three gyrators. They can be eliminated by replacement of the grey network part by it's dual arrangement. As conversion constant $r = 1/A$ is chosen where A is the area of the membrane. The corresponding network is shown in Figure 2.9.

With the selection of $r = 1/A$ as conversion constant, two of the three transformers get a conversion factor of 1:1. As they are connected to ground they can simply be removed (Figure 2.10).

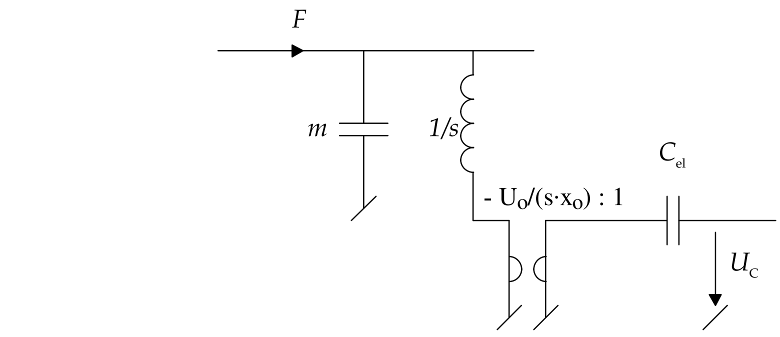


Figure 2.7: Equivalent network of the electrical output of the condenser microphone.

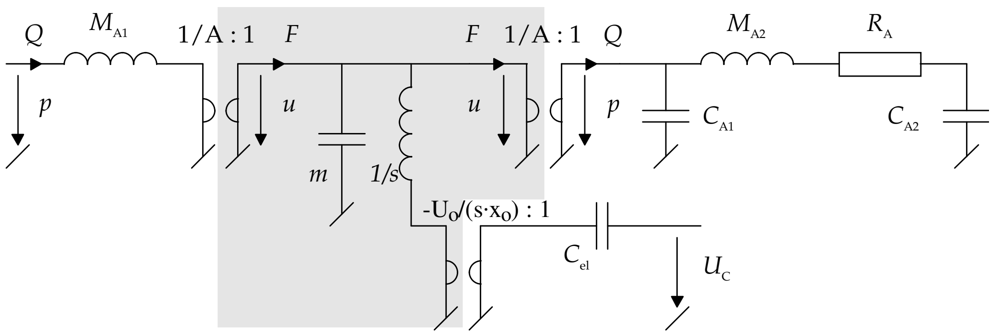


Figure 2.8: Complete network of the electrostatic microphone.

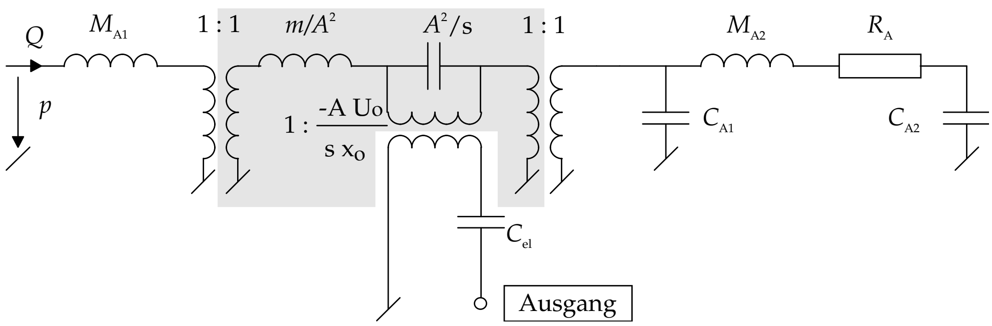


Figure 2.9: Equivalent network of the electrostatic microphone after dual conversion to get rid of the gyrators.

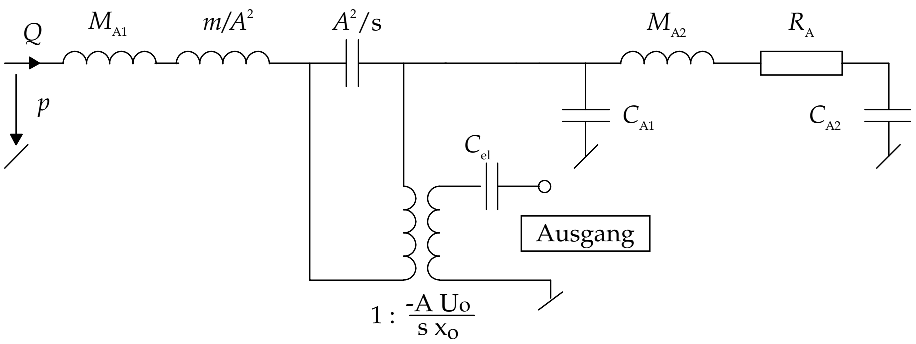


Figure 2.10: Complete equivalent network of the electrostatic microphone with the 1:1 transformers removed.

2.1.3 Simplified network

The electrostatic microphone senses the displacement of the membrane. To obtain a flat frequency response the system has to be operated below resonance. The principal behavior of the microphone can thus be discussed for low frequencies. In this case the network in Figure 2.10 can be simplified considerably:

- All inductances can be short-circuited.
- In the series configuration R_A and C_{A2}, the capacitance C_{A2} dominates. Thus R_A can be removed.
- The parallel arrangement C_{A1} and C_{A2} can be replaced by C_{A2}, as $C_{A2} \gg C_{A1}$ (compare the volume of the corresponding cavities).

With the above, a simplified low frequency approximation of the network can be drawn (Figure 2.11).

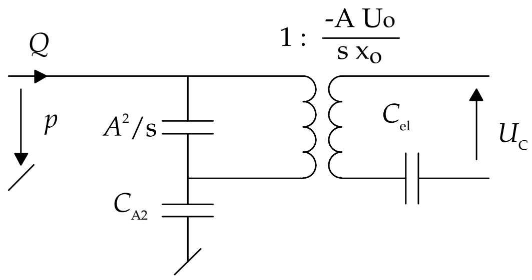


Figure 2.11: Simplified equivalent network of the electrostatic microphone for low frequencies.

Finally from the simplified network (Figure 2.11) the transfer function of the microphone can be determined:

$$\Delta U_c = p \frac{C_{A2}}{C_{A2} + A^2/s} \frac{-AU_0}{sx_0} \tag{2.10}$$

where
p: sound pressure
ΔU_c: microphone output voltage

Eq. 2.10 reveals the strategy to maximize the microphone sensitivity ($\frac{\Delta U_c}{p}$):

- The capacitance C_{A2} representing the rear side volume of the microphone body has to be chosen large enough so that $C_{A2} \gg A^2/s$.
- The polarization voltage U_0 should be made as large as possible while the membrane - back plate distance x_0 at rest should be made as small as possible. The corresponding limits are given by the isolation capability of the air. Typical values are U_0 = 200 V and x_0 = 20 μm.
- The area A of the membrane should be made as large as possible. However with increasing area of the membrane the sound field distortion by the microphone increases.
- The stiffness of the membrane should be minimized. At the same time this lowers the resonance of the system and thus brings down the upper end of the frequency range of operation.

The need for a polarization voltage can be avoided by making usage of electret materials. Certain synthetic materials can be loaded with a permanent charge. For that purpose they have to be exposed simultaneously to high temperatures and high electric fields. In cheap microphones the membrane is made from such electret materials. In this case compromises are necessary regarding the electrical and mechanical properties. The better solution is to make the back plate from electret material. Electret condenser microphones are often called prepolarized microphones.

If the microphone shall sense sound pressure, the membrane has to be arranged in such a way that standard absolute air pressure acts on the rear side. This is achieved by a tiny opening in the microphone body, allowing a slow pressure equalization. This measure introduces a lower limiting frequency for the microphone operation. However, if the opening is small enough this frequency can be made as low as 1 or 2 Hz.

2.1.4 Free-field and pressure microphones

Microphones for measuring purposes are usually built as pressure sensors with omnidirectional characteristics. This is reached by exposing only one side of the membrane to the sound field. However, at high frequencies (wave length < diameter of the membrane) the microphone body itself produces an increase of pressure for normal sound incidence. This corresponds to a signal amplification up to 13 dB (at 13 kHz) for a 1 inch membrane. For sound incidence parallel to the membrane no amplification occurs. There are two strategies to overcome this problem:

- The microphone is used for sound incidence parallel to the membrane only. In this case no measures are necessary. Such microphones are called pressure microphones.

- The microphone is used for normal sound incidence only. The amplification due to the sound field distortion at high frequencies is compensated by an acoustical filter in the microphone capsule itself or by an electrical filter in the amplifier. These microphones are called free-field microphones.

2.1.5 Power supply of the microphone amplifier

The capacitance at the output of the microphone is very small (5...50 pF). In order to minimize the current, a microphone amplifier has to be attached directly to the capsule. In audio applications two circuits are in use to feed the amplifier in case of a symmetrical cable with two signal wires and a ground shield.

T-powering

In case of T-powering (Tonaderspeisung) the power supply current flows over the two signal wires (Figure 2.12). This powering concept is no longer used for new microphones as it has several disadvantages compared to phantom powering.

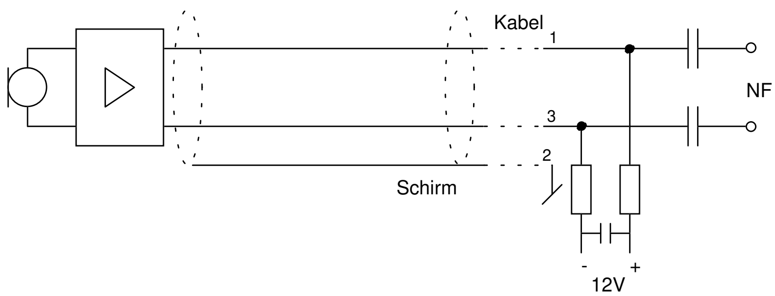


Figure 2.12: T-powering.

Phantom powering

The supply voltage of 12 V or - more common - 48 V is applied between the middle of the two signal wires and the ground shield (Figure 2.13). The current thus flows symmetrically on the two signal wires and back by the ground shield [2].
Compared to T-powering, phantom powering has several advantages:

- dynamic microphones can be plugged without switching off the power supply (security aspect).

- the two signal wires can be reversed in polarity (security aspect).

[2] DIN 45596

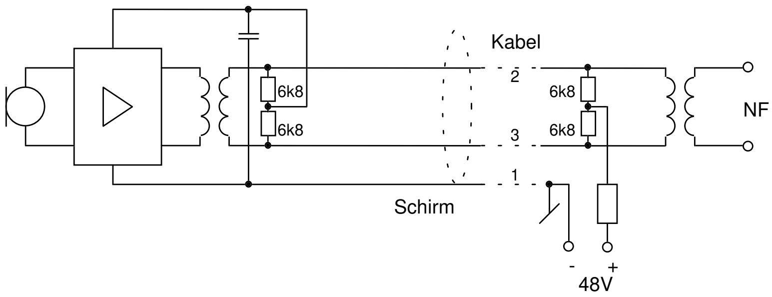


Figure 2.13: Phantom powering of microphones. The numbering of the wires corresponds to the standard configuration of XLR connectors.

- fluctuations of the power supply voltage have no influence on the signal if the coupling resistors are perfectly symmetrical. In practice the difference should be < 0.4 %.

For the generation of 48 V from a 9 V battery, see e.g. [3]

2.1.6 Overview of measuring microphones

Table 2.1 shows tyoical specifications of different measuring microphones.

type	frequency range	sensitivity	capacitance	dynamic range
1" pressure	3 Hz - 8 kHz	50 mV/Pa	55 pF	15 - 140 dB
1" free-field	3 Hz - 16 kHz	50 mV/Pa	66 pF	15 - 140 dB
1/2" pressure	4 Hz - 20 kHz	12.5 mV/Pa	18 pF	20 - 160 dB
1/2" free-field	4 Hz - 40 kHz	12.5 mV/Pa	18 pF	20 - 160 dB
1/4" pressure	4 Hz - 70 kHz	1.6 mV/Pa	6 pF	35 - 170 dB
1/4" free-field	4 Hz - 100 kHz	1.6 mV/Pa	6 pF	35 - 170 dB
1/8" pressure	6 Hz - 140 kHz	1 mV/Pa	3 pF	50 - 175 dB

Table 2.1: Specifications of different measuring microphones.

2.2 Dynamic microphone

2.2.1 Principle of operation

Figure 2.14 shows in a section view the principal construction of a dynamic microphone.

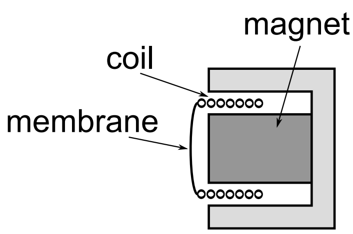


Figure 2.14: Principal building of a dynamic microphone.

A coil that is suspended in a magnetic field is attached to the membrane. The movement of the membrane and the coil induces a voltage U:

$$U = uBl \tag{2.11}$$

[3] T. Schärer, Phantom-Speisung für Kondensatormikrophone, MegaLink 21/1999

where
u: velocity of the coil and the membrane
B: magnetic field or magnetic induction
l: length of the wire of the coil

The electrical output (voltage U) of the microphone is thus proportional to the velocity of the membrane with the consequence that the system has to be operated at resonance. To realize a sufficient wide band of flat frequency response, several resonances have to implemented. This leads to much more complex structures of dynamic microphones compared to condenser microphones. The additional resonances can be realized in many different ways. In the following section one example is discussed in more detail.

2.2.2 Complete equivalent network

The relevant acoustical and mechanical components are shown in Figure 2.15.

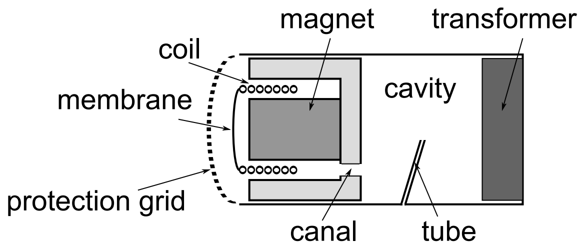


Figure 2.15: Section view of a dynamic microphone.

The following elements have to be considered in the equivalent electrical network:

p_1, p_2 sound pressure in front of the microphone and at the location of the exit of the tube

M_{A1} acoustical mass in front of the protection grid

M_{A2}, R_{A2} acoustical mass and resistance of the holes in the protection grid

C_{A3} acoustical compliance of the air between protection grid and membrane

m, s mass of the membrane and the coil, stiffness of the membrane and the suspension

C_{A4} acoustical compliance of the air between membrane and magnet

M_{A5}, R_{A5} acoustical mass and resistance of the air in the dampened canal through the magnet

C_{A6} acoustical compliance of the air in the rear cavity of the microphone

M_{A7}, R_{A7} acoustical mass and resistance of the air in the tube

M_{A8} acoustical mass of the air moving at the outer end of the tube

The equivalent network is shown in Figure 2.16.

The network in Figure 2.16 contains two gyrators. They can be eliminated by dual conversion of the network part shown in gray. It should be noted that the whole electrical section has to be included in the conversion as well. As conversion constant $r = 1/A$ (A: area of the membrane) is chosen. The converted part is shown in Figure 2.17.

The $1 : Bl$ transformer remains in the network but gets an modified conversion factor. Furthermore the original output voltage has now become a short circuited output current.

For low frequencies where the wave length is much larger than the dimensions of the microphone, the two sound pressures p_1 and p_2 are almost identical. For high frequencies this approximation is no longer valid. However in this case the impedance of the tube is so high that there is almost no air movement in the tube. For that reason it is possible to set generally $p_1 = p_2$ in the network. With insertion of the dually converted part, the final equivalent network is found as shown in Figure 2.18.

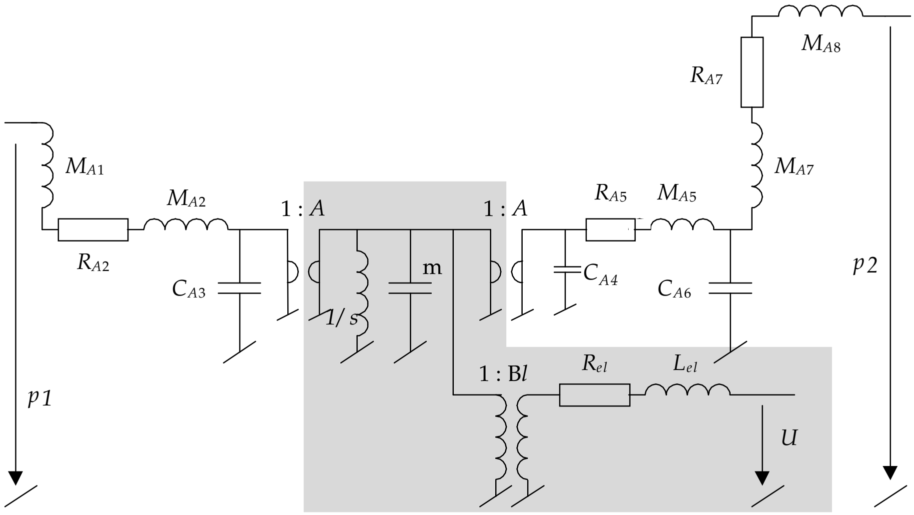


Figure 2.16: Complete network of the dynamic microphone.

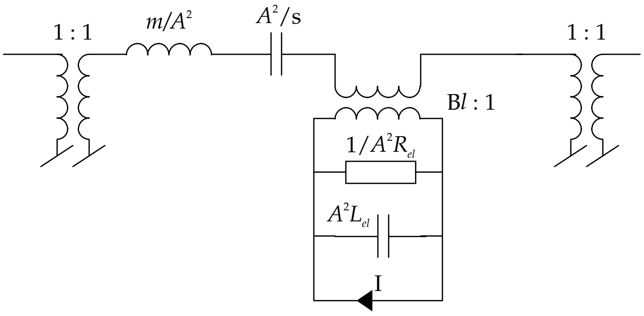


Figure 2.17: Dually converted part of the network of the dynamic microphone.

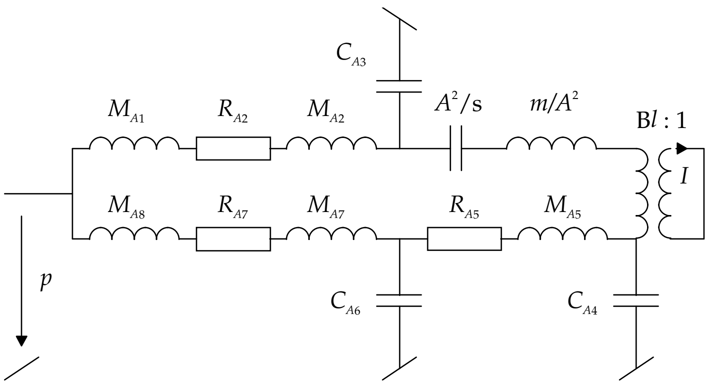


Figure 2.18: Complete equivalent network of the investigated dynamic microphone.

2.2.3 Simplified networks for different frequency ranges

In the above discussed example of a dynamic microphone five different frequency ranges can be identified. In each case certain elements dominate while the others can be neglected. The corresponding simplified networks are shown in Figures 2.19 to 2.23.

Very low frequencies

For very low frequencies the inductances can be replaced by a short-circuit. The only relevant capacitances are C_{A6} and A^2/s. The resulting frequency response shows an increase with 12 dB per octave.

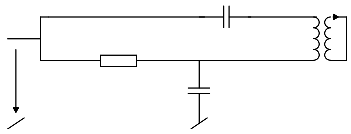


Figure 2.19: Simplified network of the dynamic microphone for very low frequencies.

First resonance (around 40 Hz. . .100 Hz)

The lowest resonance is determined by the tube and the rear cavity of the microphone. The relevant elements are the acoustical masses M_{A7} and M_{A8} and the acoustical compliance C_{A6}.

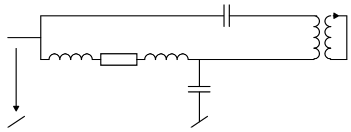


Figure 2.20: Equivalent network of the dynamic microphone for the lowest resonance.

Second resonance (around 100 Hz .. 2 kHz)

Above 100 Hz the impedance of the tube becomes so large that this path can be neglected. and the acoustical compliance C_{A6} can be replaced by a short-circuit. The relevant elements are the mechanical mass and stiffness of the membrane. This resonance is dampened by the acoustical resistance R_{A5} in the canal.

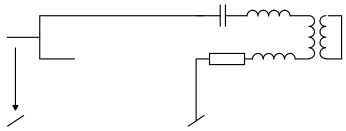


Figure 2.21: Equivalent network of the dynamic microphone for the second resonance.

Third resonance (around 2 kHz .. 5 kHz)

The behavior in this frequency range is determined by the mechanical mass of the membrane and the acoustical compliance C_{A4} of the air between the membrane and the magnet. The mechanical stiffness of the membrane can be neglected.

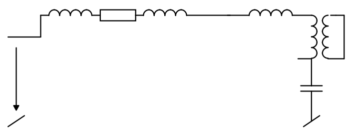


Figure 2.22: Equivalent network of the dynamic microphone for the third resonance.

Fourth resonance (around 5 kHz .. 8 kHz)

The highest frequency range is dominated by the acoustical mass of the air in front of the membrane and the acoustical mass of the air in the holes of the protection grid as well as by the acoustical compliance of the air between protection grid and membrane.

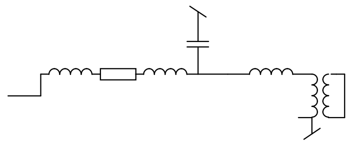


Figure 2.23: Equivalent network of the dynamic microphone for the fourth resonance.

By proper distribution of the resonances it is possible to realize an essentially flat frequency response over a wide frequency range as shown in Figure 2.24.

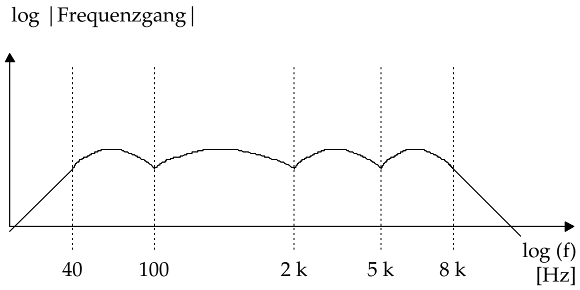


Figure 2.24: Frequency response of the discussed dynamic microphone.

2.3 Microphone directivity

As mentioned above it is possible to select the directivity (sensitivity regarding the angle of sound incidence) of a microphone independently of the conversion principle. The directivity is determined by the configuration of the membrane relative to the sound field.

2.3.1 Omnidirectional microphone

An omnidirectional sensitivity (Figure 2.25) is reached by a membrane configuration with only the front side exposed to the sound field. In this case the microphone senses the sound pressure which is a scalar quantity and thus contains no directional information. However for high frequencies the microphone body introduces a sound field distortion that depends of the angle of incidence. Measuring microphones are almost always omnidirectional microphones.

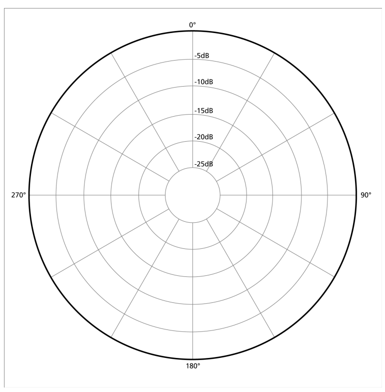


Figure 2.25: Polar plot of an omnidirectional directivity for a membrane that is exposed only one-sided to the sound field.

2.3.2 Figure of eight microphone

If both sides of a membrane are completely exposed to the sound field, the microphone gets a figure of eight directivity (Figure 2.26). The output signal is proportional to the cosine of the angle of incidence relative to the membrane normal direction.

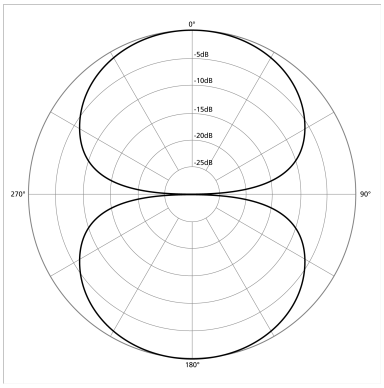


Figure 2.26: Polar plot of a figure of eight directivity for a membrane where both sides are exposed to the sound field.

The membrane movement is the result of the net force acting on it which is proportional to the sound pressured difference on both sides of the membrane. An incident sound wave with arbitrary direction will act on one side first and somewhat later on the second side. The delay corresponds to a path length difference of Δx whereby Δx depends on the cosine of the incident angle. For low and mid frequencies the relevant sound pressure difference on both sides of the membrane can be interpreted with good approximation as sound pressure gradient. Microphones with this directivity are thus sometimes called pressure gradient microphones.

From Newton's law a relation between the sound pressure gradient and the temporal derivative of the sound particle velocity can be deduced:

$$\text{grad}p = -\rho \frac{\partial v}{\partial t} \tag{2.12}$$

For harmonic time dependencies, Eq. 2.12 becomes in complex writing with $\underline{v} = \check{v}e^{j\omega t}$:

$$\text{grad}\underline{p} = -\rho j \omega \underline{v} \tag{2.13}$$

From Eq. 2.13 follows that pressure gradient microphones are sound particle velocity sensors. In order to obtain a flat frequency response, the ω proportionality has to be compensated for. This can be done e.g. with a dynamic microphone that operates above resonance. An example of such an implementation is the ribbon microphone.

2.3.3 Cardioid microphone

Cardioid microphones are a first form of directional microphones. Their sensitivity is significantly higher for frontal sound than for sound from the rear. The corresponding polar plot is shown in Figure 2.27. The output signal U of the cardioid microphone depends of the sound incident angle α according to

$$U(\alpha) \sim 0.5(1 + \cos\alpha) \tag{2.14}$$

As in case of the figure of eight microphone, both sides of the membrane of a cardioid microphone are exposed to the sound field. By introducing additional delay elements it can be reached that for sound incidence under $180°$ (from the rear), the sound pressure difference vanishes and thus the sensitivity of the microphone tends to 0. Besides the standard cardioid directivity, further variants such as supercardioid and hypercardioid are in use. Figure 2.28 shows the polar plots of the family of cardioid microphones.

A possible realization of a cardioid microphone [4] is shown in Figure 2.29. An air cavity with acoustical

[4] Leo L. Beranek, Acoustics, Acoustical Society of America, 1954/1986.

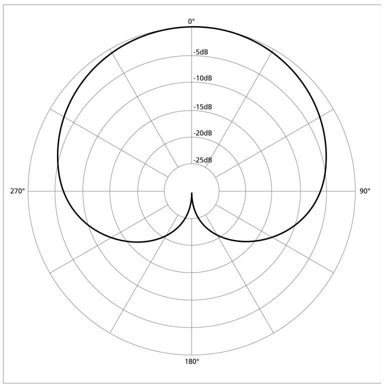


Figure 2.27: Polar diagram of a cardioid microphone.

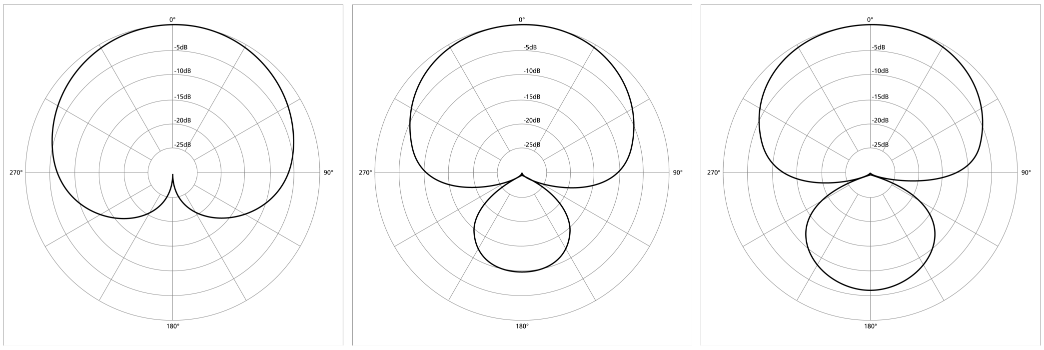


Figure 2.28: Comparison of the directivity of cardioid, supercardioid and hypercardioid microphones (from left to right).

compliance C_A is bounded on one side by the membrane. At the other end an opening with an acoustical resistance R_A is inserted. The sound pressures on both sides of the microphone is labeled as p_1 and p_2. The path length difference Δx from the front to the rear side of the membrane depends of the sound incidence direction α and corresponds to

$$\Delta x = \Delta l \cos(\alpha) \tag{2.15}$$

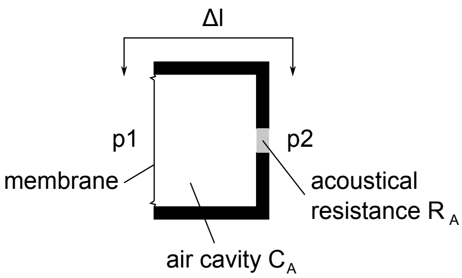


Figure 2.29: Possible realization of a cardioid microphone. The conversion elements are omitted.

A sound wave incident from direction α is assumed. With the x axis pointing in the propagation direction it can be written:

$$p_1 = \hat{p}e^{j(\omega t - kx)} \tag{2.16}$$

With this, p_2 can be expressed as:

$$p_2 = p_1 + \frac{\partial\left(\hat{p}e^{j(\omega t - kx)}\right)}{\partial x}\Delta l \cos(\alpha) = p_1\left(1 - j\frac{\omega}{c}\Delta l \cos(\alpha)\right) \tag{2.17}$$

The equivalent electrical network can then be developed as shown in Figure 2.30 where the impedance of the membrane is denoted as Z_{AD}.

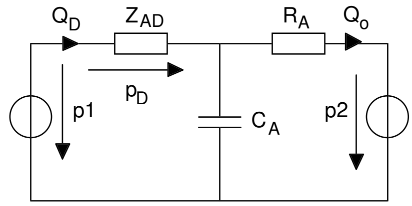


Figure 2.30: Equivalent network of the cardioid microphone. Z_{AD} denotes the impedance of the membrane.

For the network in Figure 2.30 follows:

$$p_1 = Q_D \left(Z_{AD} + \frac{1}{j\omega C_A} \right) - \frac{Q_0}{j\omega C_A} \tag{2.18}$$

$$p_2 = -Q_0 \left(R_A + \frac{1}{j\omega C_A} \right) + \frac{Q_D}{j\omega C_A} \tag{2.19}$$

Solving both equations (2.18) and (2.19) for Q_D yields the sound pressure difference across the membrane

$$p_D = Q_D Z_{AD} = \frac{Z_{AD} \left(p_1 R_A + \frac{p_1 - p_2}{j\omega C_A} \right)}{Z_{AD} R_A - j \frac{R_A + Z_{AD}}{\omega C_A}} \tag{2.20}$$

Insertion of Eq. (2.17) in (2.20) results in:

$$p_D = p_1 \frac{Z_{AD} \left(R_A + \frac{\Delta l \cos(\alpha)}{c C_A} \right)}{Z_{AD} R_A - j \frac{R_A + Z_{AD}}{\omega C_A}} \tag{2.21}$$

For suitable dimensioning

$$Z_{AD} \gg R_A \tag{2.22}$$

and

$$\frac{1}{\omega C_A R_A} \gg 1 \tag{2.23}$$

Eq. 2.21 can be simplified as:

$$p_D \approx p_1 j\omega C_A R_A \left(1 + \frac{\Delta l \cos(\alpha)}{c C_A R_A} \right) \tag{2.24}$$

With selection of the constant B, where

$$B = \frac{\Delta l}{c C_A R_A} \tag{2.25}$$

the directivity of the microphone can be adjusted. For the classical cardioid microphone B has to be set to 1. It should be noted that a $1/\omega$ correction is necessary in order to realize a flat frequency response.

2.3.4 Microphones with switchable directivity

Probably the first microphone with switchable directivity was the legendary U47 from Neumann, dating from 1947. The U47 used two cardioid capsules that were mounted back to back. The directivity of each of the two capsules can be expressed as function of sound incidence angle α as follows

$$\begin{aligned} &\text{capsule 1} \rightarrow 1 + \cos\alpha \\ &\text{capsule 2} \rightarrow 1 - \cos\alpha \end{aligned} \tag{2.26}$$

If only the capsule facing the source is used, the microphone has a cardioid characteristics. If the signals of the two capsules are summed up, an omnidirectional microphone results. The difference yields a figure of eight characteristics (see Table 2.2). By suitable scaling of one of the two signals, further characteristics can be realized.

operation	signal	directivity
capsule 1	$1 + \cos\alpha$	cardioid
capsule 1 + capsule 2	$1 + \cos\alpha + 1 - \cos\alpha = 2$	omnidirectional
capsule 1 - capsule 2	$1 + \cos\alpha - (1 - \cos\alpha) = 2\cos\alpha$	figure of eight

Table 2.2: Possible microphone directivities with two cardioid capsules that are mounted back to back.

Usually only one common back plate is used in the configuration of two back to back cardioid capsules (Figure 2.31). The holes in the back plate are of small diameter in order to realize an acoustical resistance (Figure 2.29). These microphones are usually built with large membranes.

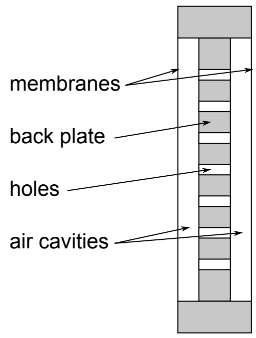


Figure 2.31: Section view of an electrostatic microphone in form of two cardioid capsules with one common back plate.

2.3.5 Directional microphone

In some applications microphones are needed that exhibit a more pronounced directional selectivity than a cardioid characteristics. The solution lies in the sampling and phase sensitive accumulation of the sound field over a region that is large compared to the wave length. The summation is organized in such a way that for wanted directions the contributions add up constructively while unwanted directions are attenuated by destructive interference.

A first possibility is a parabolic mirror [5]. Hereby the microphone capsule is mounted at the focal point of the parabola. This arrangement yields strong amplification of on-axis sound incidence, however only if the dimension of the mirror is large compared to the sound wave length.

A second variant is a line microphone. A line microphones consists of a long and narrow tube with openings in equidistant spacing. If a sound wave hits the tube end and the openings, secondary waves are emitted. These waves propagate inside the tube to the closed end where the microphone capsule is located. For on-axis sound incidence all contributions add up in phase to maximal amplitude. For lateral sound incidence, the secondary waves have all different phase and will thus interfere more or less destructively. Figure 2.32 shows the frequency response of a tube of length 50 cm for sound incidence $0°$ (on axis) and $90°$. As can be seen, the directional sensitivity shows up only for frequencies with wave lengths that are larger than the length of the tube. In concrete realizations of tube microphones, variable acoustical resistances are put in the openings in order to compensate for the damping inside the tube.

[5] Sten Wahlstrom, The Parabolic Reflector as an Acoustical Amplifier, Journal of the Audio Engineering Society, vol. 33, p.418-429 (1985)

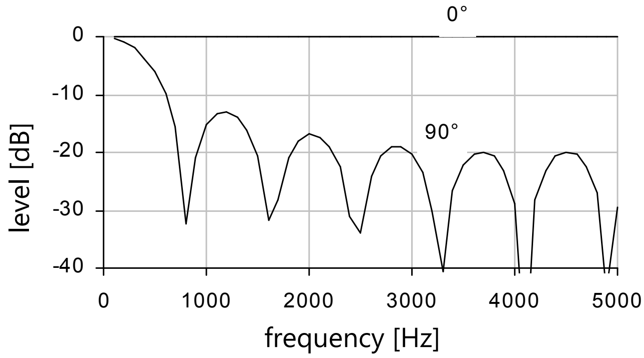


Figure 2.32: Example of a frequency response of a tube microphone of length 50 cm for $0°$ and $90°$ sound incidence.

Instead of having holes distributed along a tube, several microphone capsules can be placed in a linear arrangement. By adjusting the delays before summing up the microphone signals, the sound incidence direction for maximal sensitivity can be adjusted without manipulation of the geometry.

2.3.6 Proximity effect

Microphones with a directional selectivity (such as figure of eight or cardioid) produce an output signal that depends on the sound pressure difference at two points in the sound field. This pressure gradient is related to the sound particle velocity (Eq. 2.27). Therefore these microphones can be regarded as a sort of velocity sensors.

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial v_x}{\partial t} \tag{2.27}$$

In it's near-field and at low frequencies, a point source produces large sound particle velocity amplitudes compared to the sound pressure amplitude. In this case the impedance is significantly smaller than in the sound field of a plane wave. Thus if the microphone is sensitive to sound particle velocity, it's output shows strong amplification for low frequencies and close to the source. This phenomenon is called *proximity effect*. Some microphones have built in filters to compensate for this effect (note that this can be perfectly adjusted for one specific source-microphone distance only). However in some cases the low frequency enhancement is desirable as e.g. a voice may sound warmer and more voluminous.

2.4 Exotic transducers

Besides electrostatic and dynamic microphones there exist further transducer types that are used for special applications.

2.4.1 Microflown

The microflown transducer was invented by the university of Twente in Enschede, the Netherlands. However the basic ideas and concepts are already quite old. In collaboration with Sennheiser the company Microflown Technologies B.V. [6] developed a commercial product [7],[8]. The microflown transducer consists of two hot wires (heated up to 200...400°) and evaluates the cooling effect by microscopical wind which is produced by the sound particle velocity (Figure 2.33).

If the sound particle velocity vector points from one wire to the other, the cooling effect is more pronounced for the first wire than for the second one. The resulting temperature difference is proportional

[6] http://www.microflown.com/

[7] Jörg Sennheiser, MICRO-MINIATURIZED MICROPHONE FOR COMMUNICATION APPLICATIONS, 2nd Convention of the EAA, Berlin, 1999.

[8] W.F. Druyvesteyn, H.E. de Bree, A Novel Sound Intensity Probe Comparison with the Pair of Pressure Microphones Intensity, Journal of the Audio Engineering Society, vol. 48, p.49-56 (2000)

to the sound particle velocity and can be evaluated electronically. The sensitivity of the transducer decreases for higher frequencies due to the fact that the mass of the wires prevents to follow the quick temperature variation. In general, the sensitivity of the microflown sensor is rather low which manifests in a relatively high self noise. The directional sensitivity corresponds to a figure of eight directivity. This directivity is preserved up to high frequencies as the dimensions of the sensor are small.

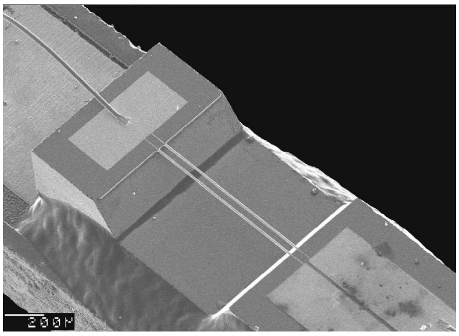


Figure 2.33: Construction of a microflown sensor. The two wires of approximately 1 mm length serve at the same time as heating elements and temperature sensors.

2.4.2 Soundfield Microphone

The soundfield microphone is a commercially available microphone system that captures sound by four nearly conicident cardioid capsules. Signals obtained by a soundfield microphone are - after some processing - suitable for reproduction by an ambisonics system.

The four cardioid capsules are arranged at the corners of a tiny tetrahedron. By electronic filtering, the capsule signals are interpolated to the center of the tetrahedron[9]. The conicidence can be achieved up to a frequency of about 10 kHz. The filtered capsule signals (LF, RF, LB, RB) form the so called A-format. By linear-combination, the corresponding B-format signals are obtained:

$$\begin{aligned} W &= LF + LB + RF + RB \\ X &= LF - LB + RF - RB \\ Y &= LF + LB - RF - RB \\ Z &= LF - LB - RF + RB \end{aligned} \tag{2.28}$$

2.5 Microphone measurements

The requirements to measure microphones are rather high. Usually measurements have to be performed under free field conditions which makes a reflection free chamber necessary. In addition, to investigate the lower end of the dynamic range, an extremely quiet environment is needed.

2.5.1 Frequency response measurements

For the excitation of the microphones usually loudspeakers are used. A difficulty is that the frequency response of the loudspeaker is in almost all cases worse than the expected response of the microphone. For that reason frequency response measurements of microphones need a reference microphone for comparison purposes. Of course the properties of the reference microphone have to be superior compared to the microphone under test. A complete documentation of the frequency response is the plot of the curve. If a specification by a few figures is needed, the lower and upper limiting frequencies are indicated for a deviation of $\pm x$ dB relative to the reference sensitivity at 1 kHz. x is usually understood as 3 dB or in some cases as 6 dB.

[9] M. Gerzon, The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound, Paper No. L-20, AES Convention, March 1975.

2.5.2 Microphone directivity

The directivity of a microphone is evaluated at a few discrete frequencies (e.g. in octave steps). The microphone under test is mounted on a turntable. For a constant excitation by a loudspeaker, the directivity is obtained by recording the microphone level as a function of sound incidence angle.

2.5.3 Non-linear distortion - maximum sound pressure level

The evaluation of non-linear distortions of a microphone makes it necessary to generate sound fields with very high undistorted sound pressure (140 up to 160 dB). As direct radiating loudspeakers would introduce exorbitant distortions, tricks such as Helmholtz resonators (specifically tuned bass reflex cabinets) are needed. The measurements are usually performed at 1 kHz. The maximum sound pressure level of a microphone is defined as the level that leads to a certain limit of the distortion factor (0.5%, 1%). The maximum sound pressure level of microphones with small membranes can easily reach values above 140 dB.

2.5.4 Self-noise

Self-noise describes the fact that any microphone generates a certain output signal even if there is no sound excitation at all. This self-noise is expressed as equivalent sound pressure level that corresponds to the output voltage of the microphone without excitation. Usually the A-weighting is applied to account for the frequency dependent sensitivity of the ear. In this case the specifications use the term " according to IEC 651". Large membrane microphones reach values as low as 10 dB(A). Alternatively a frequency weighting according to CCIR bzw. ITU-R 468 may be used. This weighting gives typically 11. . .14 dB higher levels.

Chapter 3

Loudspeakers

3.1 Transducer principles

3.1.1 Electrodynamic loudspeaker

The electrodynamic loudspeaker gets it's excitation in form of the force that acts on a conductor in a magnetic field. This force F is given as

$$F = BlI \tag{3.1}$$

where
B: magnetic field or magnetic induction
l: length of the conductor
I: current

On the other hand a movement of the conductor induces a voltage U with

$$U = Blu \tag{3.2}$$

where
B: magnetic field or magnetic induction
l: length of the conductor
u: velocity of the conductor

3.1.2 Electrostatic loudspeaker

The electrostatic loudspeaker gets it's excitation in form of the electrostatic force F that occurs between plates of different potential where

$$F = \frac{\epsilon_0 S U^2}{2x^2} \tag{3.3}$$

with
ϵ_0: electric field constant = 8.85×10^{-12} AsV^{-1}m^{-1}
S: area of each of the two plates
U: voltage between the plates
x: distance between the plates

The relation between voltage and force in Eq. 3.3 is not linear. To approximate linear behavior, a bias voltage is needed. To guarantee proper operation, this bias voltage has to be chosen much larger than the highest signal voltage.

3.2 Sound radiation of a membrane

3.2.1 Radiated power

The acoustically radiated power of a moving membrane in its piston mode depends on the membrane velocity, the area of the membrane and the medium in which sound is radiated. The loading of the membrane by the medium can be described by the radiation impedance Z_{Ro} where

$$Z_{Ro} = \frac{p}{v} \tag{3.4}$$

with
p: average sound pressure on the membrane surface
v: normal component of the sound particle velocity on the membrane surface (equals the mechanical velocity of the membrane)

The radiation impedance describes the sound pressure amplitude that establishes for a given membrane velocity. The calculation of the radiation impedance requires wave theoretical approaches. In the case where the membrane is surrounded by a wall of infinite extension, the Rayleigh integral can be applied.

An alternative definition for the radiation impedance is based on the volume flow instead of the sound particle velocity:

$$Z_R = \frac{p}{Q} \tag{3.5}$$

with
p: average sound pressure on the membrane surface
Q: volume flow (product of particle velocity and area of the membrane)

The definition (3.5) for Z_R differs from (3.4) for Z_{Ro} in a scaling by the membrane area only. The definition with the volume flow is very well suited for usage in the context of equivalent electrical circuits.

Usually the radiation impedance is a complex quantity indicating that there is an active and a reactive component. The active component leads to effective sound radiation while the reactive component corresponds to air that is moved back and forth in front of the membrane.

The sound power W that is effectively radiated by a membrane is given as

$$W = Q^2 Re[Z_R] \tag{3.6}$$

with
Q: volume flow
$Re[Z_R]$: real part of the radiation impedance

The radiation impedance Z_{Ro} of a membrane mounted in a infinitely extended wall has Real and Imaginary parts according to Figure 3.1 (top). The parameter ka is the product of the wave number $k = 2\pi/\lambda$ and the radius a of the membrane. For small ka values the Real part of the radiation impedance is proportional to k^2. For $ka > 2$ the Real part is almost constant and approximates the value of the impedance of plane waves. The bottom part of Figure 3.1 shows the radiation impedance for the free membrane. Here for small ka values the Real part increases proportional to k^4. From this follows that the free membrane is very inefficient in radiating low frequencies.
In equivalent electrical networks, the radiation impedance is usually inserted as a frequency dependent resistor (Real part) and an inductance (Imaginary part) arranged in series. If a deviation of about $\pm$ 30% can be accepted, the radiation impedance can be represented by a frequency independent resistance $R = \rho c$ and an inductance $L = \rho a/\sqrt{2}$ in parallel [1].

3.2.2 Radiation directivity

The radiation directivity of a vibrating membrane mounted in an infinitely extended wall can be determined by evaluating the Rayleigh integral. It expresses the sound pressure p at an arbitrary position in space by the integration of the membrane velocity $\check{v}_n$ over the surface of the membrane S:

[1] B. B. Bauer, Notes on Radiation Impedance, J. Acoustical Society of America, v. 15, p. 223-224 (1944).

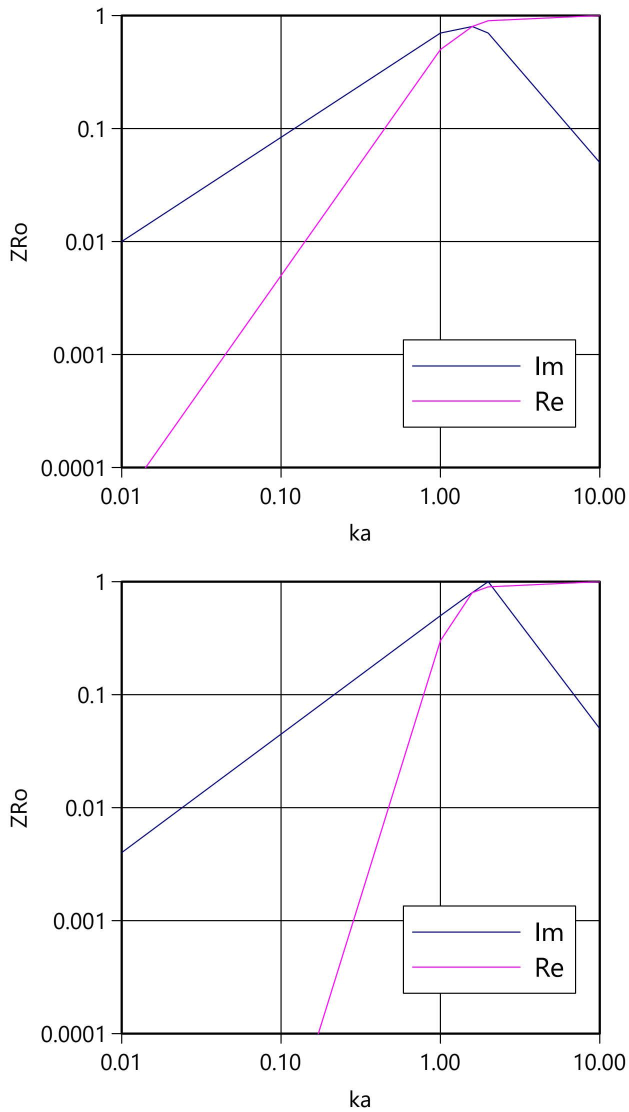


Figure 3.1: Real- and Imaginary part of the radiation impedance Z_{Ro} as a function of ka (k: wave number, a: radius of the membrane). The impedance is shown relative to the free field impedance of the plane wave ρc. Top: membrane mounted in an infinitely extended wall, bottom: free membrane.

$$\check{p} = \frac{j\omega\rho}{2\pi}\int_S \check{v}_n \frac{1}{r} e^{-jkr} \mathrm{d}S \tag{3.7}$$

Under the assumption that the receiver position is not too close to the membrane, the $1/r$ term in the integral (3.7) is in a first order approximation constant. The integration has to consider the phase term only. The solution can finally be expressed with help of Bessel functions:

$$\check{p}(\phi) = \frac{\check{v}_n}{r_0} jka^2 \rho c \frac{J_1(ka\sin\theta)}{ka\sin\theta} \tag{3.8}$$

with
$\check{v}_n$: velocity of the membrane
r_0: reference distance
k: wave number $= 2\pi/\lambda$
a: radius of the membrane
θ: angle between the membrane normal vector and the direction to the receiver point

Figure 3.2 shows examples of resulting directivities.

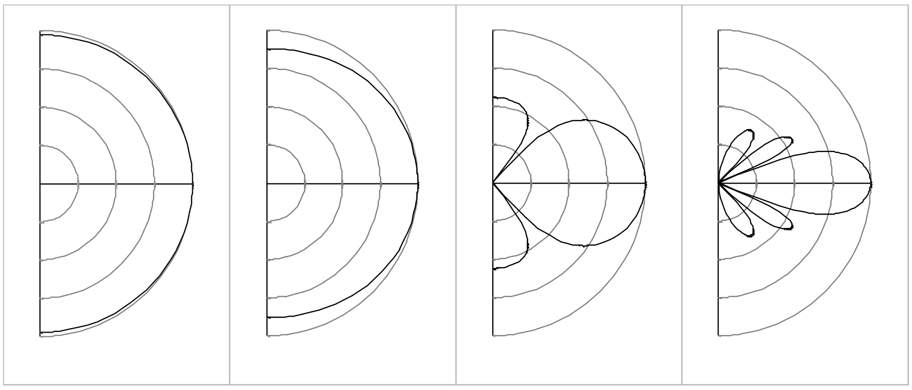


Figure 3.2: Radiation directivity for a membrane in an infinitely extended wall for $ka = 1$, $ka = 2$, $ka = 5$, $ka = 10$ (from left to right), shown as polar diagrams with a scaling of 10 dB per division.

3.3 Electrodynamic loudspeaker

3.3.1 Principle of operation

Figure 3.3 shows the section view of the construction of an electrodynamic loudspeaker. The cone-like membrane is driven by a coil that can move back and forth in a magnetic field. At the outer circumference the membrane is suspended to a ring by a soft spring. The inner part of the membrane is centered by an inner suspension (spider) to restrict the motion of the loudspeaker to one direction.

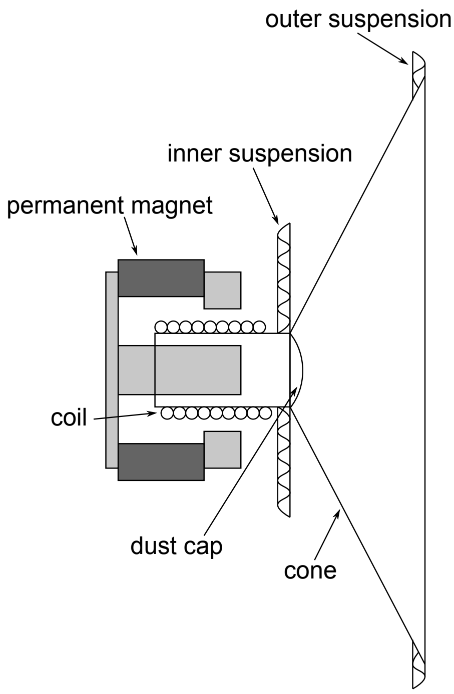


Figure 3.3: Section view of the construction of an electrodynamic loudspeaker.

3.3.2 Equivalent network

Referring to Figure 3.3, the following elements for the equivalent circuit can be identified:

M_{AR}, R_{AR} acoustical mass and resistance of the air at the rear side of the membrane (corresponding to real and imaginary part of the radiation impedance)

m, s mechanical mass of the membrane and the coil and stiffness of the membrane and the suspension

R_m mechanical friction losses of the suspension of the membrane

M_{AV}, R_{AV} acoustical mass and resistance of the air at the front side of the membrane (corresponding to real and imaginary part of the radiation impedance)

R_E, L_E electrical resistance and inductance of the coil

The above elements are arranged to an electrical equivalent network as shown in Figure 3.4. The interface to the connectors of the loudspeaker is realized by a $1 : Bl$ transformer with coil resistance R_E and inductance L_E. This represents in a correct way the driving force that depends on the current (3.1) and on the other hand the induced voltage that depends on the coil velocity (3.2).

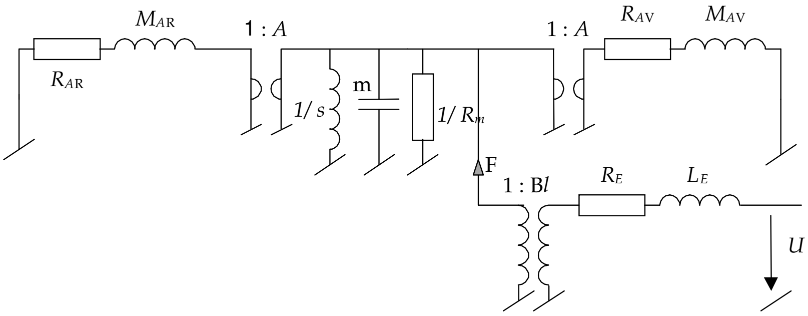


Figure 3.4: Equivalent electrical network of the electrodynamic loudspeaker. F represents the driving force.

The two gyrators in Figure 3.4 can be eliminated by dual conversion with $r = 1/A$ (Figure 3.5):

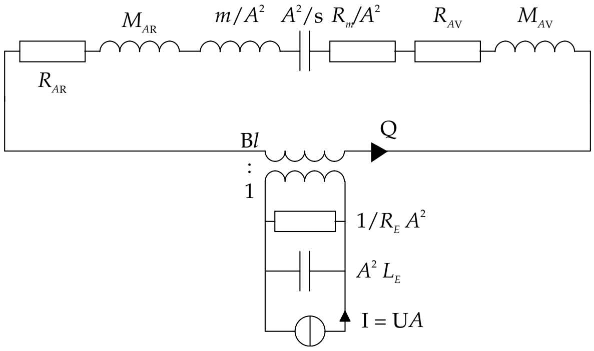


Figure 3.5: Equivalent electrical network of the electrodynamic loudspeaker after dual conversion and elimination of the gyrators.

The radiated power is given by the product of the square of the volume flow and the real part of the radiation impedance. The volume flow corresponds to the current in the RLC series connected circuit. Three frequency domains can be distinguished:

- At low frequencies the capacitance is dominating. Therefore Q increases proportionally to ω.
- Around resonance, Q is maximal and only limited by the resistances.
- Above resonance, the inductances are dominating. The volume flow Q decreases proportionally to $1/\omega$.

To obtain a flat frequency response (radiated power independent of frequency) the loudspeaker has to be operated above resonance where $Q^2 \sim 1/\omega^2$ compensates for the ω^2 dependency of the real part of the radiation impedance. This is fulfilled by a membrane in an infinite wall for not too high frequencies ($ka < 1$). This high frequency limit can be overcome to a certain degree with a membrane construction that separates the membrane area into a number of rings. The membrane is then tuned in such a way that for high frequencies only the inner part is in motion, resulting in smaller ka values.

3.3.3 Nonlinearities

Up to now it was assumed that the elements in the equivalent circuit are constant. However for higher signal amplitudes this is no longer perfectly true. The mechanical spring elements representing the inner and outer membrane suspension produce restoring forces that are no longer linearly related to the excursion. In addition large excursion of the coil can lead to variations of the inductance of the coil L_E and the force factor $B \times l$.

3.3.4 Mounting in a cabinet

The usage of an infinitely extended wall for a loudspeaker is of course not a suitable approach. Indeed the fundamental effect of the wall is a separation of the front and rear side of the membrane. This can be realized more conveniently by a box or cabinet. However, the cabinet introduces an additional compliance depending on the box volume. At the same time the radiation impedance on the rear side is no longer relevant. Figure 3.6 shows the modified equivalent network for a loudspeaker mounted in a cabinet.

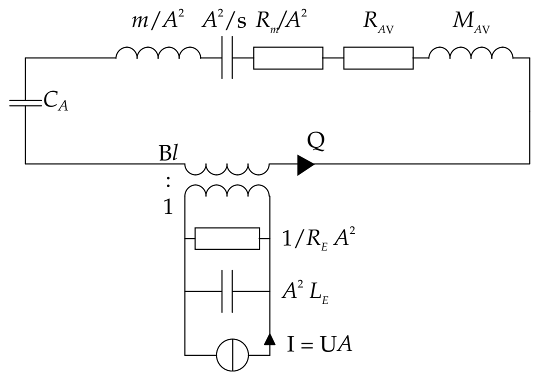


Figure 3.6: Equivalent electrical network of the electrodynamic loudspeaker mounted in a cabinet that introduces an additional compliance C_A.

The acoustical compliance that is introduced by the cabinet is a capacitance C_A that is in series to the existing capacitor representing the membrane stiffness. Due to this series arrangement the resulting capacitance is smaller than the smaller of the two. This leads to an increase of the resonance frequency and thus to an increase of the lower limiting frequency of the loudspeaker. This unwanted effect can be kept in limits if the volume of the cabinet is not too small. In addition, if the cabinet is filled with damping material, the processes in the box are no longer adiabatic but isothermal which increases the effective volume by 15 %.

An other approach to virtually simulate larger box volumes is the usage of the adsorption effect of certain porous materials [2]. These materials can accumulate or adsorb gas molecules at their surface. During the compression phase a certain portion of the air inside the cabinet is assimilated by the material. This reduces the density of the air and thus reduces the stiffness of the acoustical compliance. During the decompression phase this process is reversed. The effective volume seems larger by a factor of about 1.5 to 3 compared to the geometrical one.

Finally after combining corresponding elements and after dual conversion of the complete network with a conversion constant $r = 1/A$ the equivalent network as shown in Figure 3.7 is found. This the network that is usually presented in literature about dynamic loudspeakers. As a simplification the real part of the radiation impedance can be neglected for low frequencies.

The *volume flow* Q as quantity of interest has been transformed to the voltage U'_S where

$$U'_S = \frac{Q}{A} \tag{3.9}$$

The other network elements are

[2] J. R. Wright, The Virtual Loudspeaker Cabinet, Journal of the Audio Engineering Society, vol. 51, p.244-247 (2003).

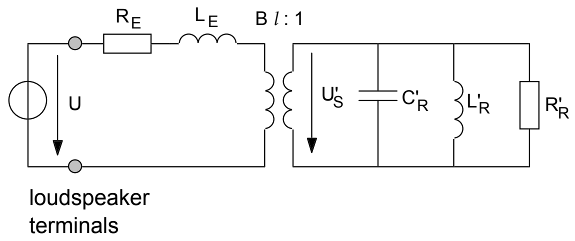


Figure 3.7: Simplified equivalent electrical network for the dynamic loudspeaker mounted in a cabinet.

$$C'_R = A^2 M_{AV} + m \tag{3.10}$$

$$L'_R = \frac{\frac{A^2}{s} C_A}{\frac{A^2}{s} + C_A} \frac{1}{A^2} \tag{3.11}$$

$$R'_R = \frac{1}{R_m} \tag{3.12}$$

The transformer in the above network (Figure 3.7) can be eliminated by scaling of the impedances on the right hand side by $B^2 l^2$ (Figure 3.8).

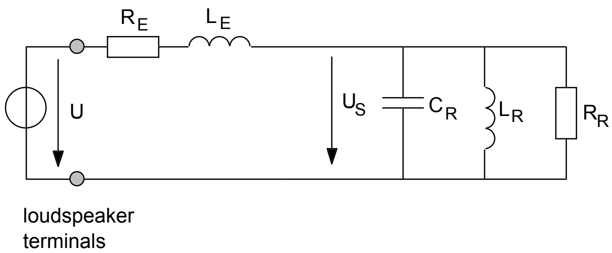


Figure 3.8: Network of the dynamic loudspeaker after elimination of the transformer.

In Figure 3.8, the voltage U_S corresponds to:

$$U_S = Bl U'_S = Bl \frac{Q}{A} \tag{3.13}$$

and the other elements are

$$C_R = \frac{1}{B^2 l^2} C'_R = \frac{A^2 M_{AV} + m}{B^2 l^2} \tag{3.14}$$

$$L_R = B^2 l^2 L'_R = \frac{\frac{C_A}{s} B^2 l^2}{\frac{A^2}{s} + C_A} \tag{3.15}$$

$$R_R = B^2 l^2 R'_R = \frac{B^2 l^2}{R_m} \tag{3.16}$$

With help of the network in Figure 3.8, the electrical impedance as seen from the loudspeaker terminals can be discussed. For very low frequencies the inductances represent short-circuits. The impedance is thus determined by R_E and consequently frequency independent. For an increase in frequency the impedances of the inductances become larger and thus relevant (note: $L_R \gg L_E$). C_R and L_R represent a parallel resonant circuit. At resonance it's impedance is limited by R_R. For further increase of the frequency above resonance the impedance becomes smaller again and reaches R_E as the capacitance will more and more represent a short-circuit. For very high frequencies the inductance L_E of the coil comes into play. The impedance will increase linearly with frequency. Figure 3.9 shows

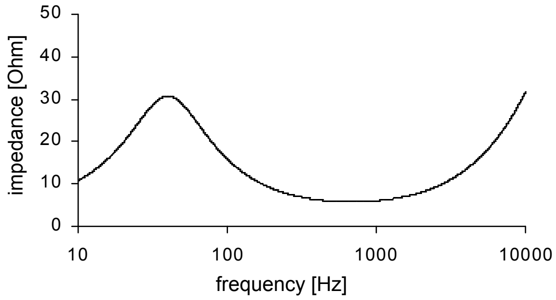


Figure 3.9: Frequency response (amplitude only) of the electrical impedance as seen at the terminals of a loudspeaker.

a typical frequency response of the electrical impedance at the loudspeaker terminals.

The characteristic impedance of a loudspeaker is given by the nominal impedance which corresponds to the minimum in the impedance curve above resonance. Typical values are 4 and 8 Ohms.

3.3.5 Frequency response of sound pressure

As mentioned above, the radiation of the loudspeaker is determined by the velocity of the membrane and the radiation impedance. For low frequencies an omnidirectional radiation with equal strength in all directions can be assumed. For a chassis mounted in a cabinet, the sound pressure p at distance d can be expressed as:

$$p(d) = \sqrt{\frac{W\rho_0 c}{4\pi d^2}} = \frac{Q\sqrt{\mathrm{Re}[Z_R]\rho_0 c}}{\sqrt{4\pi}d} \tag{3.17}$$

where
W: radiated sound power
Q: volume flow on the membrane (product of membrane velocity and membrane area)
$\mathrm{Re}[Z_R]$: real part of the radiation impedance

For low frequencies the real part of the radiation impedance of the membrane in a cabinet (Figure 3.1) can be approximated by:

$$\mathrm{Re}[Z_R] = \frac{\rho_0 c}{2}(ka)^2\frac{1}{a^2\pi} = \frac{\rho_0\omega^2}{2\pi c} \tag{3.18}$$

with
k: wave number
a: radius of the membrane
ω: angular frequency

Eq. 3.18 inserted in Eq. 3.17 yields

$$p(d) = Q\omega\frac{\rho_0}{d\sqrt{8\pi}} \tag{3.19}$$

With help of Eq. 3.13 the volume flow Q in Eq. 3.19 can be expressed by the voltage U_S from the network in Figure 3.8. Finally one gets:

$$p(d) = U_S\omega\frac{\rho_0 A}{d\sqrt{8\pi}Bl} = U_S\omega\frac{\rho_0 a^2}{d\sqrt{8}Bl} \tag{3.20}$$

With Eq. 3.20 and the network in Figure 3.8 the sound pressure frequency response $p(d)/U$ of the loudspeaker can be determined:

$$\frac{p(d)}{U} = \frac{U_S}{U}\omega\frac{\rho_0 a^2}{d\sqrt{8}Bl} \tag{3.21}$$

The voltage transfer function U_S/U in Eq. 3.21 can be found with the network in Figure 3.8 for low frequencies (L_E can be neglected) as

$$\frac{U_S}{U} = \frac{j\omega L_R R_R}{-\omega^2 L_R R_R C_R R_E + j\omega(L_R R_R + L_R R_E) + R_R R_E} \tag{3.22}$$

Eq. 3.22 inserted in Eq. 3.21 yields

$$\frac{p(d)}{U} = \frac{\rho_0 a^2}{d\sqrt{8}Bl}\frac{1}{j}\frac{1}{R_E C_R}\frac{-\omega^2 L_R C_R}{-\omega^2 L_R C_R + j\omega\frac{L_R R_R + L_R R_E}{R_R R_E} + 1} \tag{3.23}$$

The frequency dependent part of Eq. 3.23 corresponds to a second order high pass function and can be expressed in the form:

$$G(j\omega) = \frac{-\omega^2 T_c^2}{-\omega^2 T_c^2 + j\omega\frac{T_c}{Q_{TC}} + 1} \tag{3.24}$$

with
$T_c^2 = C_R L_R = \frac{1}{\omega_c^2}$, where ω_c is the lower limiting angular frequency of the high pass filter (corresponding to the resonance of the loudspeaker)
$Q_{TC} = \sqrt{L_R C_R}\frac{R_R R_E}{L_R R_R + L_R R_E}$ (quality factor of the high pass filter)

Insertion of the element values yields as quality factor:

$$Q_{TC} = \frac{1}{\omega_c \frac{\frac{C_A}{s}}{\frac{A^2}{s} + C_A}\left(R_m + \frac{B^2 l^2}{R_E}\right)} \tag{3.25}$$

Figure 3.10 shows the resulting frequency response $p(d)/U$ for different values of the quality factor Q_{TC}

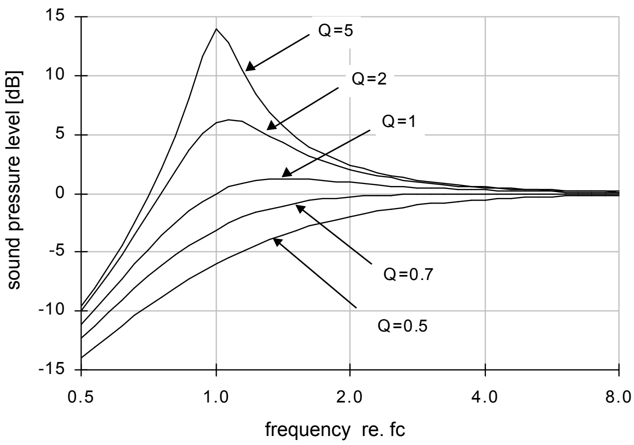


Figure 3.10: Sound pressure frequency response of the dynamic loudspeaker for different values of the quality factor Q_{TC} (abscissa: frequency relative to the resonance f_c).

3.3.6 Thiele-Small parameter

The specification of a loudspeaker chassis is usually based on the so called Thiele-Small parameters:

Resonance frequency f_s [Hz]:
The resonance frequency of the chassis indicates the lower end of the frequency range of operation.

The resonance frequency can be determined most easily from the peak in an electrical impedance measurement. The relation between the corresponding angular frequency and the elements introduced above is:

$$\omega_s = \sqrt{\frac{s}{A^2 M_{AV} + m}} \tag{3.26}$$

Compliance equivalent volume V_{AS} [l]:
The stiffness s of the membrane and the supporting device is expressed by an equivalent volume V_{AS} of equal compliance. For that purpose the mechanical spring is transformed into an acoustical system. As shown in Figure 3.5 the spring becomes a capacitance with $C_F = A^2/s$. The acoustical compliance of the air volume V_{AS} corresponds to a capacitance $C_{AS} = V_{AS}/\rho_0 c^2$. From the postulation of equal capacitances follows

$$\frac{1}{s} A^2 = \frac{V_{AS}}{\rho_0 c^2} \tag{3.27}$$

Quality factors:
For an RLC parallel resonance circuit the quality factor Q of the resonance in the impedance response is given as

$$Q = RC\omega_s \tag{3.28}$$

where
Q: Quality factor corresponding to the ratio of the resonance frequency to the -3 dB bandwidth
ω_s: angular frequency at resonance

The LC parallel resonance circuit in Figure 3.8 is damped in two ways. The resistance R_R represents mechanical damping, while R_E is responsible for the damping on the electrical side (assuming an ideal driving voltage source). Consequently two quality factors can be specified:

Mechanical Q factor Q_{MS}:

$$Q_{MS} = \omega_s C_R R_R = \omega_s \frac{A^2 M_{AV} + m}{R_m} \tag{3.29}$$

Electrical Q factor Q_{ES}:

$$Q_{ES} = \omega_s C_R R_E = \omega_s \frac{A^2 M_{AV} + m}{B^2 l^2} R_E \tag{3.30}$$

Total Q factor Q_{TS}: The total Q factor is then given as

$$Q_{TS} = \frac{Q_{MS} Q_{ES}}{Q_{MS} + Q_{ES}} = \frac{1}{\frac{\omega_s}{s}\left(R_m + \frac{B^2 l^2}{R_E}\right)} \tag{3.31}$$

Area of the membrane A [m2]

DC resistance R_E [Ohm]

Force factor $B \times l$ [T·m]

The Thiele-Small parameters Q, V_{AS}, f_s and R_E can be determined from the amplitude response of the electrical impedance at the chassis terminals. This measurement has to be performed once for the chassis in free air and once built in in a cabinet of known volume [3].

As mentioned above, the chassis has to be mounted in a cabinet to get a flat frequency response. The behavior at the lower end of the operating range is determined by the resulting resonance frequency f_c and the total quality factor Q_{TC}. For a cabinet of effective volume V_B follows:

$$f_c = f_s \sqrt{\frac{V_{AS}}{V_B} + 1} \tag{3.32}$$

[3] R. H. Small, Direct-Radiator Loudspeaker System Analysis, J. Audio Eng. Soc., vol. 20, n.5, p.383-395 (1972).

and

$$Q_{TC} = Q_{TS}\sqrt{\frac{V_{AS}}{V_B}+1} \tag{3.33}$$

3.3.7 Velocity and displacement of the loudspeaker membrane

As seen above the dynamic loudspeaker has to be operated above it's resonance frequency. Consequently the membrane velocity shows in the frequency range of interest an amplitude response that is inversely proportional to frequency. The membrane displacement follows thus a $1/f^2$ behavior. The question of maximum excursion of the membrane has to be discussed at the lower end of the operating range.

3.3.8 Bass reflex cabinet

The reproduction of low frequencies is one of the fundamental challenges in the construction of loudspeakers. A trick to improve the low frequency behavior is the usage of so called bass reflex cabinets (Figure 3.11). This cabinet type is equipped with an additional opening that is connected to the interior with a tube of distinct length and cross section. The mass of the air in the tube acts together with the compliance of the air in the cabinet as a spring-mass resonator. By appropriate tuning of this resonance, the range of operation of the loudspeaker can be extended somewhat towards lower frequencies. At resonance the oscillation of the air in the tube is maximal while the membrane motion is minimal. The sound radiation is generated by the tube alone.

A serious draw back of the bass reflex cabinet is a worsened transient response. In addition, if narrow tubes are used, flow induced noise can occur as the sound particle velocity in the tube may become very large. This noise has broad band character, however shows a maximum around the tube resonance (tube length = $\lambda/2$). An improvement in this respect is the widening of the tube diameter towards the opening [4]. On the other hand the shape of the tube shall not be too conical as no longer an oscillating air column is formed.

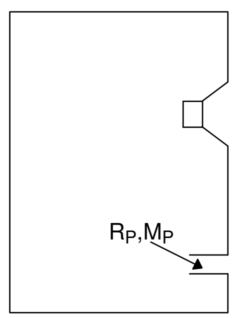


Figure 3.11: Bass reflex cabinet. An additional resonance is introduced by the interaction of the air mass M_p (and resistance R_p) in the tube and the compliance of the cabinet volume.

An alternative construction principle is the passive membrane system. Hereby the mass is realized mechanically by an additional membrane that has no coil and is therefore not driven actively. This membrane introduces an additional stiffness that adds to the compliance of the air in the cabinet.

3.3.9 Bandpass cabinet

A modification of the bass reflex cabinet is the bandpass cabinet (Fig. 3.12). It is sometimes used for sub-woofers that have to radiate low frequencies only. Hereby the driving chassis is mounted inside the cabinet. It acts on the air volume inside the cabinet which itself makes oscillate the air in a tube that is responsible for the exterior radiation. By this acoustical filter the radiation of higher frequencies is suppressed as desired for a sub-woofer.

[4] N. B. Roozen et. al. Vortex sound in bass-reflex ports of loudspeakers, Journal of the Acoustical Society of America, vol. 104, p.1914-1924 (1998).

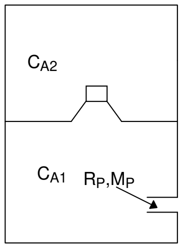


Figure 3.12: Bandpass cabinet. The acoustical compliance and the acoustical mass form an acoustical filter that attenuates the radiation at higher frequencies.

3.3.10 Feed-back strategies

Especially near resonance, loudspeaker chassis show non-optimal transient behavior. In addition, for large membrane excursions non-linearities may occur that produce harmonic distortions. Both aspects can be significantly improved by the introduction of a feed-back loop. A first idea is to install a second coil that induces a voltage proportional to the membrane velocity. This voltage can be used as a feed-back signal. Other methods make usage of information about the momentary displacement of the membrane [5].

A very elegant method is the combination of a loudspeaker chassis with a driving power amplifier that has a negative output impedance that compensates the resistance of the coil of the chassis [6]. From the equivalent electrical network of the dynamic loudspeaker (Figure 3.8) can be seen that such a compensation allows for direct control of the velocity of the membrane by the driving voltage. In this case the necessary $1/f$ dependency of the velocity has to be realized by an electrical filter that acts on the signal voltage. For high frequencies the inductance of the coil becomes relevant and makes the compensating effect less effective. The negative output impedance is realized in a practical application by superposition of a voltage signal that is deduced from the measured output current. This has the effect that an increase of the output current produces an increase of the output voltage. To avoid instability of the system it has to be guaranteed that the sum of the negative output impedance and the resistance of the coil is greater than 0 at any time.

3.3.11 Digital equalizing

An other method to improve the non-ideal behavior of loudspeakers is to use digital filters [7]. A promising strategy is to measure the impulse response $h(t)$ of the loudspeaker and then to calculate the inverse filter $g(t)$ where $h(t) * g(t) = \delta(t)$. By pre-filtering of the signal with $g(t)$ - at least theoretically - a system with perfect sound pressure impulse response and frequency response is obtained. Of course there is a limitation by physical constraints. If the radiation of the chassis itself is weak at certain frequencies, the inverse filter will introduce very high amplifications and thus very high voltages that can destroy the chassis. Furthermore as the radiation characteristics of a loudspeaker varies with angle, a perfect compensation is possible for one distinct direction only. And finally it should be noted that non-linear behavior can not be corrected by digital equalizing.

3.4 Horn loudspeaker

The radiation of a piston mounted in a wall or a cabinet is very inefficient as the real part of the radiation impedance is low in the typical frequency range of operation. The introduction of a horn (Figure 3.13) improves the efficiency significantly due to a more optimal impedance matching.

Most common are exponential horn shapes. The cross sectional area S can be described as

[5] W. Geiger, Servo Control of Loudspeaker Cone Motion Using an Optical Linear Displacement Sensor, J. Audio Engineering Society, vol. 53, p.518-524 (2005).

[6] K. E. Stahl, Synthesis of Loudspeaker Mechanical Parameters by Electrical Means, J. Audio Eng. Soc., v.29, p. 587-596 (1981).

[7] Matti Karjalainen et. al., Comparison of Loudspeaker Equalization Methods Based on DSP Techniques, J. Audio Eng. Soc., v.47, p. 14-31 (1999).

Figure 3.13: Example of a horn loudspeaker. The horn is folded to achieve a smaller overall length of the speaker.

$$S(x) = S_T e^{mx} \tag{3.34}$$

where
$S(x)$: cross sectional area at position x
S_T: cross sectional area at the throat $(x = 0)$
m: geometrical constant

If the length of the horn and the circumference at the mouth that are larger than the wavelength, the horn can be regarded as infinitely extended and no significant reflections occur at the mouth. Assuming that the cross sectional area in the horn changes only slowly, the sound propagation can be regarded as plane. In this case the concerning differential equation is [8]

$$\frac{\partial^2 p}{\partial t^2} - c^2 m \frac{\partial p}{\partial x} - c^2 \frac{\partial^2 p}{\partial x^2} = 0 \tag{3.35}$$

The effect of the horn becomes visible in the second term in Eq. 3.35. If the horn constant m is set to 0 (the horn degenerates to a tube with constant cross section), Eq. 3.35 just becomes the ordinary one-dimensional wave equation. As solution of Eq. 3.35, a plane wave with general amplitude term can be assumed:

$$\underline{p}(t) = \hat{p} e^{ax} e^{j\omega t} \tag{3.36}$$

By insertion of Eq. 3.36 into Eq. 3.35, the coefficient a is found as

$$a = -\left(\frac{m}{2} + j\frac{\sqrt{4(\omega/c)^2 - m^2}}{2}\right) \tag{3.37}$$

From Newton's law follows a relation between the local derivative of sound pressure and the temporal derivative of sound particle velocity in the corresponding direction:

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial v_x}{\partial t} \tag{3.38}$$

In complex writing, Eq. 3.38 translates into

$$\underline{v} = -\frac{1}{j\omega\rho}\frac{\partial \underline{p}}{\partial x} \tag{3.39}$$

By insertion of the derivative of sound pressure, the sound particle velocity is found as

$$\underline{v} = \frac{1}{j\omega\rho}\left(\frac{m}{2} + j\frac{\sqrt{4(\omega/c)^2 - m^2}}{2}\right)\underline{p} \tag{3.40}$$

[8] Leo L. Beranek, Acoustics, 1954 and 1986

For the radiation impedance follows

$$\underline{Z}_{Ro} = \frac{\underline{p}}{\underline{v}} = \frac{2j\omega\rho}{m + j\sqrt{4(\omega/c)^2 - m^2}} \tag{3.41}$$

Depending on the parameter setting of m, two frequency regions for $\underline{Z}_{Ro}$ can be distinguished:

$2(\omega/c) < m$ $\quad$ $\underline{Z}_{Ro}$ is pure imaginary (Re[$\underline{Z}_{Ro}$] = 0)

$2(\omega/c) \geq m$ $\quad$ $\underline{Z}_{Ro}$ has a real and an imaginary part with Re[$\underline{Z}_{Ro}$] = $\rho c\sqrt{1 - \frac{m^2}{4(\omega/c)^2}}$

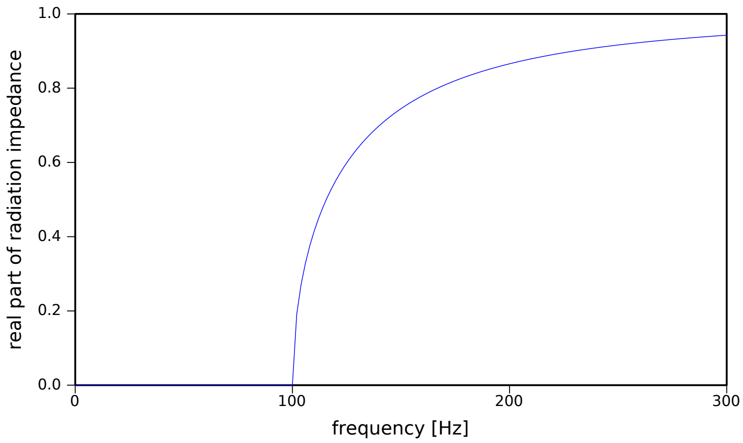


Figure 3.14: Frequency response of the real part of the radiation impedance (rel. to ρc) of an exponential horn with m = 3.7.

The real part of the radiation impedance $\underline{Z}_{Ro}$ is 0 for low frequencies. For frequencies above the threshold $2(\omega/c) \geq m$, Re[$\underline{Z}_{Ro}$] increases quickly (Figure 3.14) and reaches asymptotically the value of the free field impedance ρc. Frequency independent radiated power can be obtained in the range where Re[$\underline{Z}_{Ro}$] and the volume flow are both almost constant. The electrodynamic driving unit has therefore to be operated at resonance. By strong damping, a flat frequency range can be obtained within a bandwidth of about two octaves. With electronic filters the low end high end decay can be compensated for to a certain degree.

Compared to the direct radiator speaker, horn loudspeakers show a significantly higher efficiency. However a possible problem arises due to the high sound particle velocities in the throat of the horn. They may introduce non linear distortions.

3.5 Electrostatic loudspeaker

Modern electrostatic loudspeakers usually work on basis of the push-pull principle. Figure 3.15 shows the principal construction. A very lightweight membrane (0.2 g/dm^2) is mounted between two perforated and thus acoustically transparent electrodes.

The polarization voltage U_p adjusts the point of operation in order to linearize the relation between input voltage U_{signal} and force acting on the membrane. The resulting force and thus the excursion of the membrane remains relatively small. To obtain a reasonable amount of volume flow at low frequencies, the membranes of electrostatic loudspeakers have to be relatively large. However without counter measures this introduces high directivity at higher frequencies which is not desirable. This can be avoided either by curved membranes or by separating the membrane into several segments that are driven individually depending on frequency.

Electrostatic loudspeakers are dipole radiators. In order to avoid interference with sound that is radiated from the rear side and reflected at the wall, they have to be set up with a certain distance to the walls.

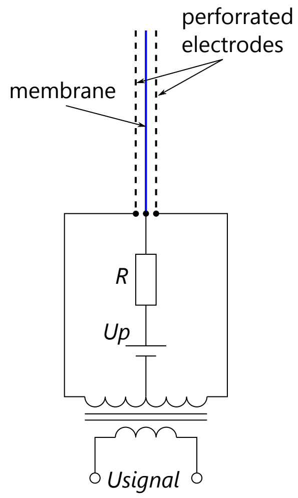


Figure 3.15: Construction principle of an electrostatic loudspeaker in push-pull arrangement.

3.6 Loudspeaker systems

The ideal loudspeaker covers the audible frequency range from 20 Hz to 20 kHz. However this is very difficult to reach. The chassis resonance limits radiation at the lower end, the high frequency limit is given by the condition $ka < 2$.

An other difficulty with a broadband chassis is the occurrence of Doppler distortions. These may become audible if a low and a high frequency signal component are radiated simultaneously. The membrane movement due to the low frequency component will then modulate the frequency of the high frequency component.

Due to the reasons mentioned above, sound generation of a loudspeaker is usually split to several specialized chassis. The lowest frequencies are reproduced by a large membrane woofer, the high frequencies are radiated by a small tweeter. In some cases additional mid-range speakers are used. The signal separation is achieved by crossover networks that filter the corresponding frequency ranges for each chassis.

Loudspeaker systems can either be passive or active. Passive systems contain no amplifier, they are fed by a low impedance voltage of an external power amplifier and filtered by a crossover network with passive elements. Active systems have a power amplifier built-in and can be fed with a high impedance line signal. Active systems do the crossover filtering before the power amplification takes place. Therefore active filters with OpAmps can be used. They can be more complex than passive filters.
The crossover networks are often second or third order filters (12 or 18 dB/octave). The dimensioning of passive filters is somewhat complicated as the chassis do not represent pure resistive loads. It is often necessary to implement a circuit to linearize the impedance frequency response. Active systems allow for the implementation of Lipshitz-Vanderkooy filters [9] that realize an optimal amplitude and phase response. The phase response is of special relevance in the transition frequency range where two chassis contribute to the total sound pressure.
Possible phase differences are the consequence of the chassis properties, the phase response of the crossover filters and the path length difference from the acoustical centers of the chassis to the receiver point. The path length difference can be made zero just for one specific listening position. As the listener position varies mainly in the horizontal plane it is advantageous to stack the chassis vertically on top of each other. It should be noted that the path length difference problem with different chassis can be avoided completely by using coaxial systems where woofer and tweeter have an identical acoustical center.

[9] S. Lipshitz, J. Vanderkooy, A Family of Linear-Phase Crossover Networks of High Slope Derived by Time Delay, J. Audio Eng. Soc. 31, p.2-19 (1983).

The maximum power that can be handled by a loudspeaker system is usually specified as electrical power at the terminals of the loudspeaker. The limitations come in on one hand by the maximal allowable excursion of the membrane (given by the peak power) and by the heating of the moving coil (given by the average power). The maximum allowable average power is determined by usage of a weighted noise signal according to IEC 268 (Figure 3.16). This signal represents the frequency content of typical audio material (speech and different kinds of music). The noise weighting is crucial as it determines the portions of the total power that has to be handled by each chassis in a loudspeaker system.

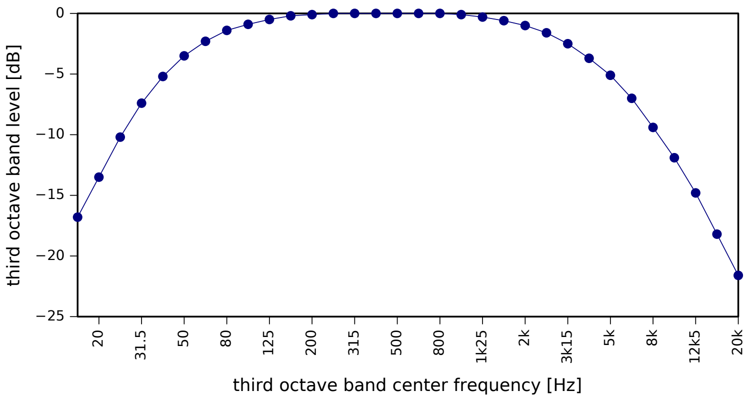


Figure 3.16: Third octave band spectrum of average audio material according to IEC 268.

A signal with third octave spectrum according to Figure 3.16 can be generated by using pink noise and a weighting filter according to Figure 3.17.

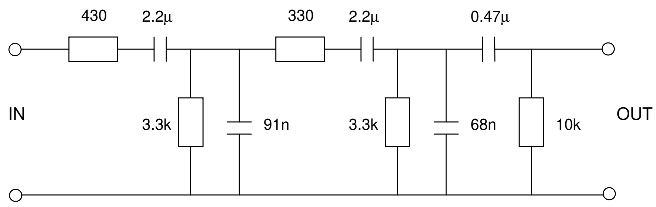


Figure 3.17: Weighting filter to transform pink noise into a noise signal with a spectrum according to IEC 268.

Figure 3.18 and Table 3.1 show the cumulated power of the IEC 268 noise signal. For a given frequency range of operation, the fraction of the total power that has to be handled by a single chassis can easily be determined. It is common praxis to specify the maximum power of a chassis by the total power of the system. Thus a typical specification for a tweeter may be: maximum power = 100 W with a crossover network of 12 dB per octave and a limiting frequency of at least 3 kHz. As can be seen from Table 3.1, the effective power at the tweeter is in this case 7 W only.

3.7 Loudspeaker measurements

The document IEC 268 [10] describes standard procedures for loudspeaker measurements. The most relevant quantities to describe loudspeakers are [11]:

- frequency response of the electrical impedance at the terminals
- frequency response of the sound pressure on axis
- impulse response of the sound pressure on axis
- cumulative decay spectrum of the sound pressure on axis
- directivity of the sound pressure for various discrete frequencies

[10] IEC 268, Part 1 and 5, Electroacoustic devices, General and Loudspeakers.
[11] Joseph D'Appolito, Testing Loudspeakers, Audio Amateur Press (1998).

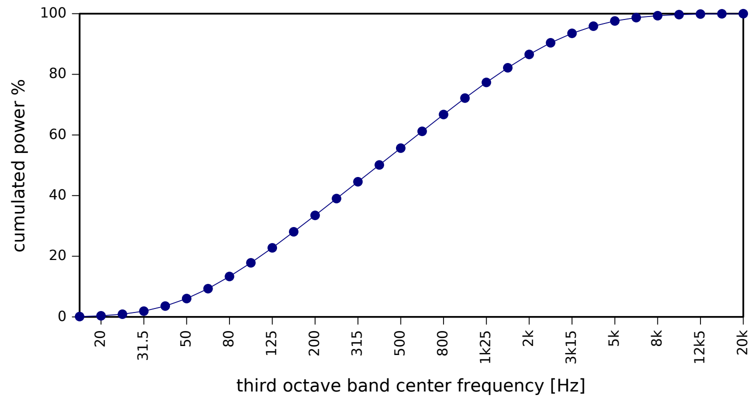


Figure 3.18: Cumulated power from 0 Hz up to the specified frequency for the noise IEC 268 signal.

frequency [Hz]	cumulated power [%]	frequency [Hz]	cumulated power [%]
16	0.12	630	61.20
20	0.36	800	66.74
25	0.89	1k	72.16
32	1.90	1k25	77.33
40	3.58	1k6	82.16
50	6.05	2k	86.56
63	9.32	2k5	90.40
80	13.33	3k15	93.52
100	17.84	4k	95.88
125	22.78	5k	97.59
160	28.07	6k3	98.70
200	33.49	8k	99.34
250	39.03	10k	99.69
315	44.57	12k5	99.88
400	50.12	16k	99.96
500	55.66	20k	100.00

Table 3.1: Cumulated power from 0 Hz up to the specified frequency for the noise IEC 268 signal.

3.7.1 Cumulative decay spectrum

The cumulative decay spectrum of the sound pressure describes the transient behavior of the loudspeaker. In contrast to the impulse response the cumulative decay spectrum contains frequency dependent transient information.

A direct method to obtain a decay spectrum is to excite the loudspeaker with a sinusoidal signal that is discontinued in a zero crossing (tone burst) as shown in Figure 3.19.

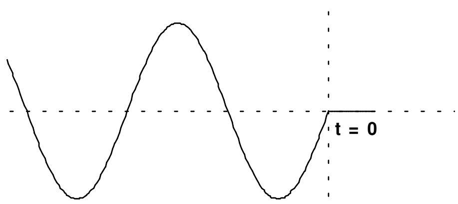


Figure 3.19: At time $t = 0$ discontinued sinusoidal signal of frequency f for the evaluation of the transient behavior at frequency f.

The loudspeaker will respond to the signal of Figure 3.19 with an oscillation of decaying amplitude. This can be represented as $s_\omega(t)$ where

$$s_\omega(t) = H_\omega(t)\sin(\omega t) \tag{3.42}$$

with
$H_\omega(t)$: modulation of the amplitude
ω: angular frequency of the signal

$H_\omega(t)$ is a function that usually decays to zero very quickly. These modulation functions can now be determined for many different frequencies and be displayed in a waterfall representation (Figure 3.20). At time $t = 0$ the amplitude response for stationary excitation appears.

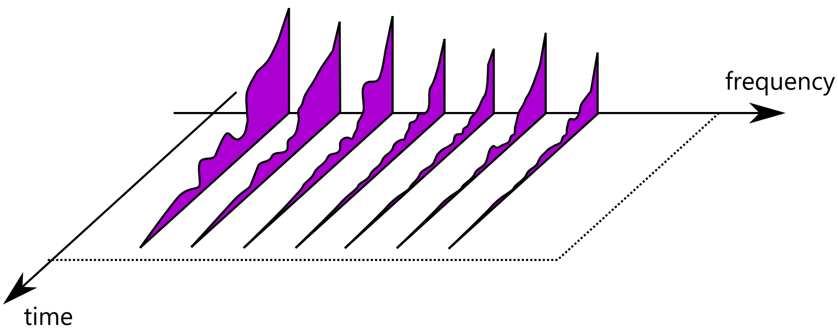


Figure 3.20: Waterfall representation of the amplitude modulation as a function of time for differen frequencies (decay spectrum).

The determination of the decay spectrum with single tone bursts is very time consuming. It is much more efficient and elegant to evaluate the impulse response first. The cumulative decay spectrum can then be calculated without additional information [12].

In complex writing the tone burst signal from Figure 3.19 can be represented as

$$f_\omega(t) = e^{j\omega t}u(-t) \tag{3.43}$$

where
ω: angular frequency of the sinusoidal signal
$u(t)$: step function $u(t) = 1$ für $t \geq 0, u(t) = 0$ für $t < 0$

The signal of interest corresponds to the imaginary part of $f_\omega(t)$. With $h(t)$ as the impulse response of the loudspeaker, the response $g_\omega(t)$ to the tone burst can be described as convolution where again the imaginary part is used:

$$g_\omega(t) = \text{Im}\left[\int_{-\infty}^{\infty} h(\tau)e^{j\omega(t-\tau)}u(\tau - t)\mathrm{d}\tau\right] = \text{Im}\left[e^{j\omega t}\int_{-\infty}^{\infty} h(\tau)e^{-j\omega\tau}u(\tau - t)\mathrm{d}\tau\right] \tag{3.44}$$

with
Im[ ]: imaginary part of [ ]

The integral in the expression on the right hand side of Eq. 3.44 corresponds to the Fourier transformation of the impulse response h(t) that is time windowed with u($\tau - t$). As we are only interested in the amplitude, the phase term $e^{j\omega t}$ can be ignored. For discrete time steps, the set of all functions g$_\omega$ can be represented as series of Fourier transforms of the appropriately time windowed impulse response functions (Figure 3.21). Figure 3.22 shows as an example the cumulative decay spectrum of a midrange loudspeaker.

[12] T. Suzuki et. al, Three-dimensional displays for demonstrating transient characteristics of loudspeakers. Journal of the Audio Engineering Society, vol 26, p.946-955 (1978).

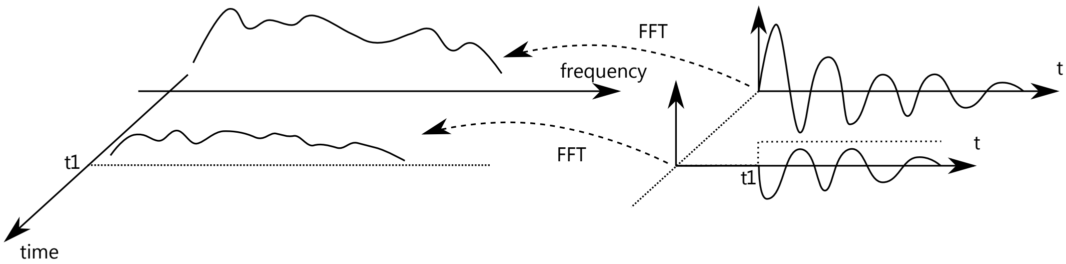


Figure 3.21: Generation of the cumulative decay spectrum by putting together the amplitude of Fourier transforms of the suitably time windowed impulse response functions.

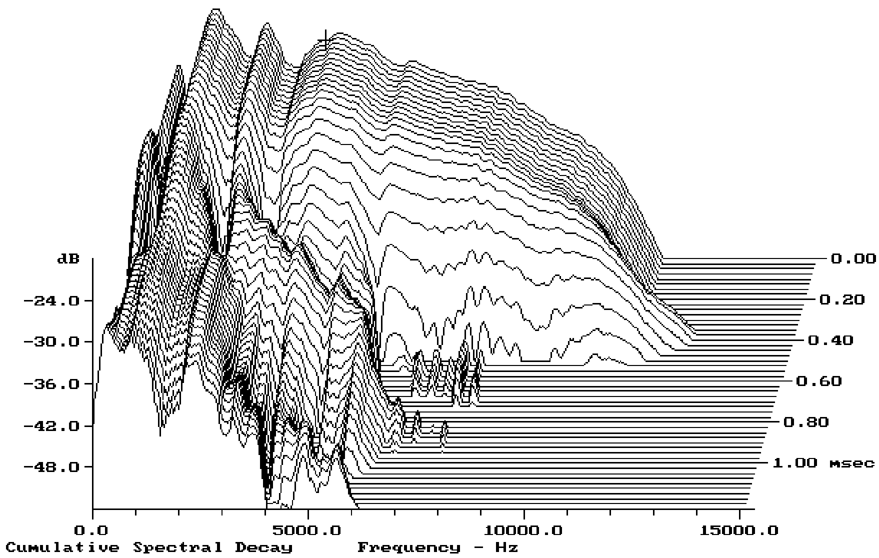


Figure 3.22: Cumulative decay spectrum of a midrange loudspeaker (Dynaudio D-76).

Chapter 4

Sound storage media

Devices for storage and reproduction of acoustical signals are available for about one hundred years. Figure 4.1 shows the main building blocks of a sound storage device.

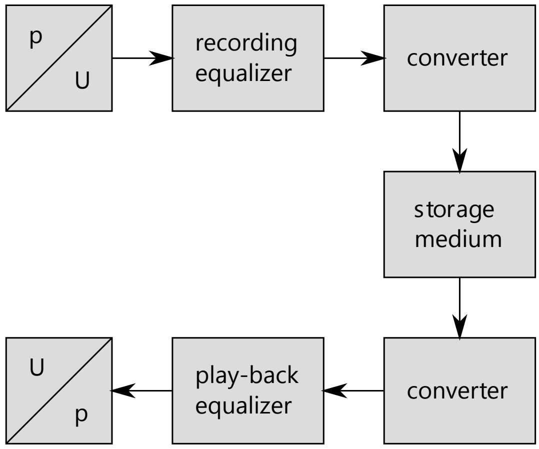


Figure 4.1: Principal construction of a storage device for acoustical signals. The input converter p/U is usually a microphone, the output converter U/p is a loudspeaker. The recording and play-back equalizers realize optimal adaption of the signal to the storage medium.

4.1 Vinyl records

The history of vinyl records starts with the invention of the phonograph by Edison in 1877. In his arrangement a graver that was coupled to a membrane transformed the acoustical signal into a groove of varying depth. At the beginning tin cylinders were used as media, later the cylinders got a layer of wax. During play back a needle followed the groove and transformed the depth variations into vibrations of a membrane. In 1887 Berliner came up with a disk as medium. Thereby the acoustical signal was stored in form of lateral excursion of a groove. The main advantage of this principle was the possibility of simple duplication by a press process. Westrex launched in 1957 the 45-45 system that allowed for storage of two channels (stereo signal). The two channels modulate the two flanks of the groove. As the two flanks form an angle of 90°, the signals are orthogonal and can thus vary independently of each other - at least in theory. In practice a channel separation in the order of 30 dB can be reached. Figure 4.2 shows a section view through a groove.

Both transducers for recording and play-back operate as electrodynamic converters with the consequence that the voltage is proportional to the velocity of the graver or the needle. The excursion shows a $1/f$ proportionality relative to the velocity. Together with the fact that typical audio signals show largest amplitudes at low frequencies it becomes obvious that the excursions at high frequencies are very small. Without further measures, inaccuracies in the production process and dirt in the groove would affect in an unbalanced way the signal/noise ratio at high frequencies only. To distribute the noise more

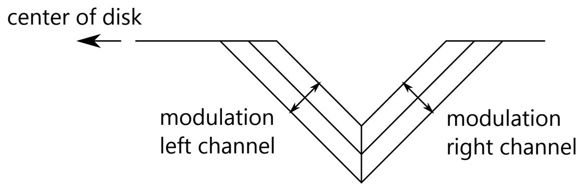


Figure 4.2: Section view of a groove in the 45-45 recording format.

equally over the whole frequency range, the RIAA [1] equalization was introduced. At recording, the high frequencies are amplified with a filter slope of about 15 dB per decade. This equalization is inverted at play-back (Figure 4.3).

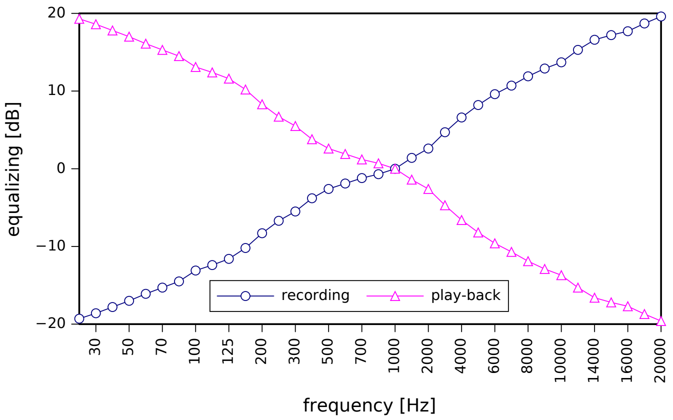


Figure 4.3: RIAA equalization used for vinyl records.

4.2 Analog tape recorder

The principle to store an audio signal in form of magnetization of a wire or tape was invented by Oberlin Smith [2]. The first practical implementation of the concept was presented by Valdemar Poulsen around 1898 [3]. Modern forms of tape recorders use as storage medium a synthetic tape coated with material that can be magnetized. By help of a winding system the tape is moved in longitudinal direction along three heads. Figure 4.4 shows the principle construction.

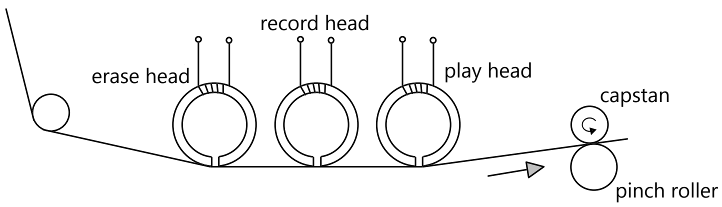


Figure 4.4: Principle of operation of an analogue tape recorder.

During recording the record head produces a magnetic field that leaves behind a permanent magnetization on the tape. During play-back the time varying magnetic field stored on the tape induces a voltage in the play head that corresponds to the signal that was originally recorded.

Materials that can be magnetized show a hysteresis as shown in Figure 4.5.

[1] Recording Industry Association of America.
[2] Oberlin Smith, Some Possible Forms of Phonograph, The Electrical World, September 8, 1888, p.116-117.
[3] Heinz H.K. Thiele, The 25 Essential Development Steps of Analogue Magnetic Technology, 104th AES Convention 1998, Amsterdam

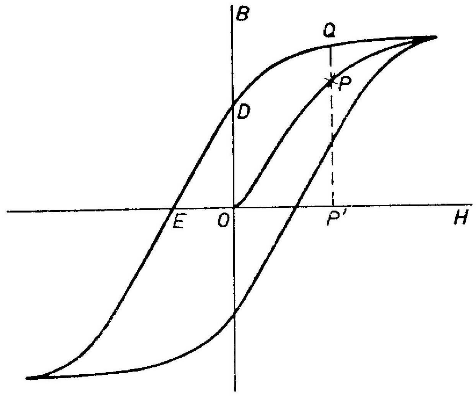


Figure 4.5: Typical hysteresis curve of the relation between magnetic field H (horizontal axis) and induction B (vertical axis).

The exposure of a magnetic tape to a field H_1 leads to an induction B_1. After discontinuation of the magnetic field, the induction decreases until the remanence B_{R1} is reached. This remanence represents the information that is stored on the tape. A larger initial magnetic field H_2 yields a higher remanence B_{R2}. However the relation between H and B_R is strongly nonlinear with the consequence of high distortions especially for low amplitude signals (Figure 4.6).

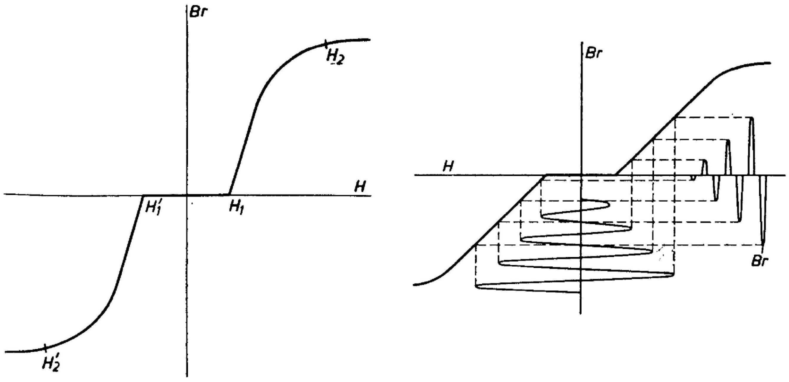


Figure 4.6: Curve of remanence showing the relation between the magnetic field H and the remanence B_R (left) and resulting nonlinear distortions for a sinusoidal signal of varying amplitude (right).

The distortions shown in Figure 4.6 can be lowered dramatically by introducing a bias current. The most obvious strategy is to add a DC component to the signal (dc bias). However for the following reasons this is not the best solution:

- The remaining distortions are still significant due to the nonlinear and asymmetric remanence curve. The asymmetry produces odd and even order harmonics.
- The signal-to-noise ratio is relatively small as only one half of the possible magnetization is used.

More advantageous is the introduction of ac bias. Hereby a high frequency signal (f = 50. . .150 kHz) is added to the audio signal in such a way that the amplitude of the high frequency signal spans the region of zero-remanence. The magnetization is now symmetric and thus no even order harmonic components occur. As the positive and negative branch of the magnetization curve are used, the signal-to-noise ratio is improved significantly. Figure 4.7 shows the dc- and ac biasing of magnetic recording.
The record on a magnetic tape is in longitudinal writing. With tape speed v and signal frequency f, the wave length λ of the magnetic induction is given as

$$\lambda = \frac{v}{f} \tag{4.1}$$

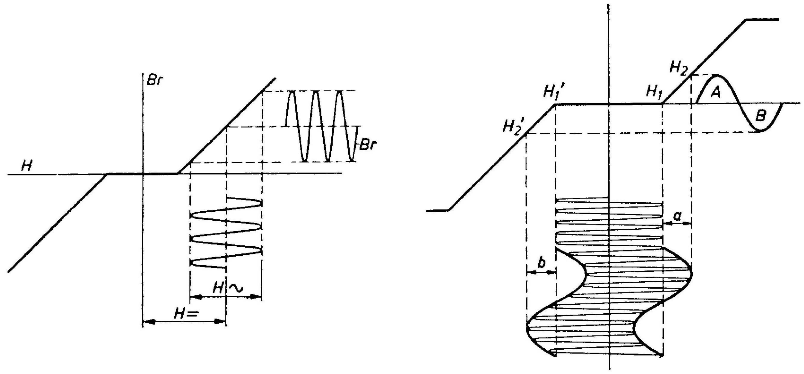


Figure 4.7: Reduction of non-linear distortions by dc (left) and ac (right) biasing.

The tape speeds are standardized. For professional applications 38.1 cm/s (15 inch/s) and 19.05 cm/s are used while consumer devices usually offer 9.5 cm/s and 4.75 cm/s. If the wave length λ equals the width of the tape head, the resulting induction is 0. This limit defines an the upper end of the frequency range of operation. For a typical tape head width of 5 μm and a tape speed of 19.05 cm/s an upper limiting frequency clearly higher than 20 kHz is possible.

For high record and play-back quality the tape heads have to be adjusted optimally regarding their position relative to the tape. A misalignment leads to an enlargement of the effective head width and thus to a lowering of the high frequency sensitivity.

The amplitude of the remanence B_r that is stored during recording is almost proportional to the magnetic field H and thus to the signal current. During play back a voltage is induced that is proportional to the change of the magnetic flux. This implies a ω-proportional frequency response that has to be compensated for in a play back equalizer.

Dolby noise reduction

Analog recordings typically suffer from an insufficient signal-to-noise ratio. Early attempts to improve this situation go back to Scott [4] and Olson [5]. A promising idea is the implementation of compander-expander systems. Hereby low level segments of the audio signal are amplified during recording and attenuated during play back. Since the sixties of the last century, Dolby is the most important manufacturer of such systems. Figure 4.8 shows a typical Dolby noise reduction block diagram.

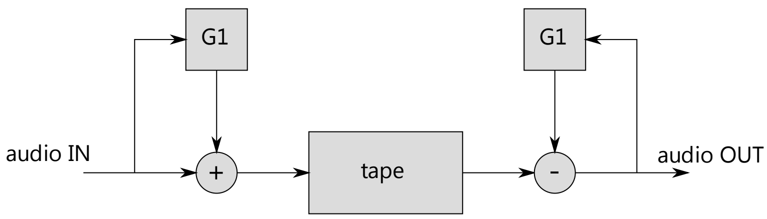


Figure 4.8: General block diagram of a Dolby noise reduction system.

The networks $G1$ and $G2$ in Figure 4.8 are frequency dependent amplifiers that are steered by the input signal. It can be shown easily that *Audio Out* = *Audio In* if $G1 = G2$. The different Dolby variants differ from each other regarding the complexity of these networks and in the number of individually processed frequency bands.

In the following the Dolby B system is discussed in more detail [6]. Dolby B uses only one single frequency

[4] H. H. Scott, Dynamic Noise Suppressor, Electronics 20, 96 (1947).
[5] H. F. Olson, Audio Noise Reduction Circuits, Electronics 20, 118 (1947).
[6] R. Berkovitz, K. Gundry, Dolby B-Type Noise Reduction System, Audio, September/October 1973.

band. During recording low level mid and high frequencies are amplified up to 10 dB. At play back this amplification is reversed with the consequence of significantly lowering the noise from the tape. The principal block diagram of a Dolby B system is shown in Figure 4.9. It corresponds to the general diagram from Figure 4.8 with the trick, that the same network can be used for recording and play back (but not simultaneously). This minimizes errors due to parameter variations.

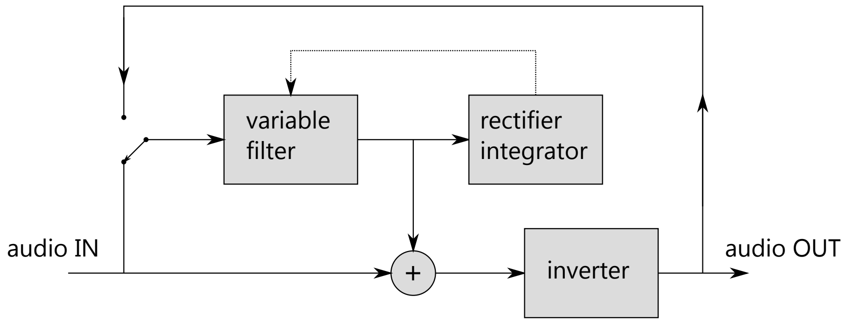


Figure 4.9: Block diagram of a Dolby B noise reduction system. The switch decides whether the network is used for recording or play back.

At recording, the mid and high frequency components are amplified by a filter with variable amplification and added to the original signal. The amplification depends on the average mid and high frequency signal level. The lower this level, the higher the amplification with a maximum of 10 dB (Figure 4.10). During play back the feedback loop adjusts itself in such a way that the original signal is reconstructed. As with all compander-expander systems a serious difficulty is the temporal adjustment of the time varying amplification factor. Dolby B uses a time constant in the range of 1. . .100 ms, depending on the signal variation. The signal-to-noise improvement with Dolby B is about 10 dB. For professional applications Dolby A and Dolby SR were common [7]. Both systems operated in several frequency bands to minimize the difficulties with the tracking of the amplification of the filter. Together with professional tape recorders Dolby SR reached a dynamic range of more than 90 dB.

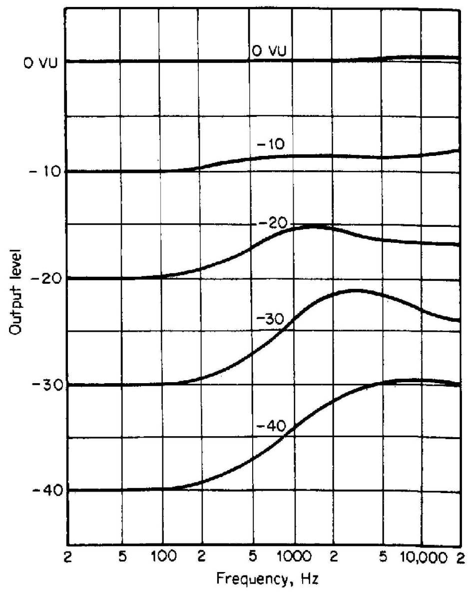


Figure 4.10: Amplification of the variable filter in Dolby B as a function of the mid- and high frequency signal level.

[7] R. Dolby, The Spectral Recording Process, J. Audio Eng. Soc. 35, 99-118 (1987).

4.3 Compact Disc

Besides online downloads the Compact Disc (CD) is the most important medium for the distribution of sound. The CD stores sound in digital form with a sampling rate of 44.1 kHz and a resolution of 16 Bit. This yields a frequency response from 20 Hz to 20 kHz and a dynamic range of more than 90 dB. The CD is a plastic disc of 12 cm in diameter and 1.2 mm thickness. The CD is produced by a press process and consists of several layers (from top to bottom):

- protection layer with imprint
- metal layer of 50...100 nm thickness
- transparent layer

The digital information is stored in form of grooves of about 120 nm depth. The metal layer follows these grooves and acts as reflective termination. The grooves are arranged on a spiral with the starting point at the center of the disc. The width of the grooves is 0.5 μm, their length is 1 or 3 μm, depending on the value of the Bit. The distance between two spiral courses is 1.6 μm, the total length of the spiral is about 5 km. The grooves are sensed optically by a laser diode from the bottom side of the disc. The wavelength of the laser light is 500 nm. The reflection from the bottom of a groove leads to a path length difference of about half a wavelength and thus together with the reflection from the surrounding area of the groove a pronounced destructive interference occurs. The reflected light is detected with a photodiode. By usage of an array of several sensors a control signal can be deduced to track the spiral course. Depending on the disc type the speed of reading lies between 1.2...1.4 m/s. Depending on the reading position this results in a rotational speed between 500 U/min at the beginning (center of the disc) and 200 U/min at the end. The bitstream presents a constant bit rate of 4.3218 Mbit/s. For a maximum play length of 74 minutes this corresponds to a capacity of 2.4 GByte.

Although the information on a CD is stored in a robust way, one has to assume a error rate of typically 10^{-6}. To correct for these errors, additional redundancy has to be introduced. For that purpose the CD uses the CIRC (cross-interleave Reed-Solomon code). The CIRC is very efficient as only one fourth of the total data is redundant. With this error correcting system missing data due to gaps of about 1 mm can be reconstructed.

The CIRC coded audio signal can not be written directly to CD. An appropriate modulation is needed to avoid that:

- there are too high frequency components in the Bit stream (inter-symbol interference)
- there are too low frequency components in the Bit stream (inaccurate reconstruction of the clock signal)

These requirements are fulfilled by FEM (Eight to Fourteen Modulation). Thereby each Byte (8 Bits) is mapped onto a symbol consisting of 14 Bits. Additionally the linking of two adjacent symbols has to be controlled by insertion of three transition Bits. The audio information is then completed by a subcode. The subcode contains information about the number of tracks, the start and end times of the tracks and so on. In total, the original audio data stream (16 Bit $\times$ 2 channels $\times$ 44'100 samples per second) of 1.4 Mbit/s grows to a volume that is three times as high.

4.4 DVD Audio and Super Audio CD

With increasing density of storable data on optical discs the appetit for more audio data grows as well. A higher capacity compared to the CD would allow for larger play lengths, an increased resolution with more Bits, a higher sampling rate and more than two channels. While the quality improvement due to an increased sampling rate and a higher resolution (44.1 kHz, 16 Bits $\rightarrow$ 192 kHz, 24 Bits) is usually inaudible [8], the step to more than two channels is a very convincing change in the audible experience.

With DVD-Audio and Super Audio CD two concurring formats were launched. DVD-Audio will disappear soon, there are no longer new recordings published in this format. The Super Audio CD will probably endure the same destiny.

[8] E. B. Meyer, D. R. Moran, Audibility of a CD-Standard A/D/A Loop Inserted into High-Resolution Audio Playback, J. Audio Engineering, vol. 55, no. 9, 775-779 (2007).

4.4.1 DVD Audio

The DVD Audio standard was published in 1999. One layer of a DVD has a net capacity of 4.7 GByte, which corresponds to the capacity of about 7 CD's. DVD Audio uses 16, 20 or 24 bit quantization and sample frequencies between 44.1 and 192 kHz. The audio data are usually stored in uncompressed form. However MLP (Meridian Lossless Packing) or perceptual coding such as Dolby Digital (AC3) or MPEG are possible as well. Without any data reduction a single layer DVD can store 258 minutes of a stereo signal sampled at 48 kHz. 43 minutes are possible for a 5.1 surround signal sampled at 96 kHz.

4.4.2 Super Audio CD

The Super Audio CD format was developed by Sony and Philips. The capacity is identical to the one of a single layer DVD. With Super Audio CD, standard stereo signals as well as surround formats are possible. Some Super Audio CD's have an extra layer that can be read by standard CD players.

4.5 Harddisk recorder

Harddisk recorder store digital audio information on common digital media such as harddisks or memory cards. Computer based solutions use the computer hardware and an audio interface with amplifiers and A/D and D/A converters for the connection to the analog world. Moreover there exist stand-alone recorders that combine all necessary blocks in one device. Maximum sampling frequencies are typically 192 kHz, maximal resolution is 24 Bit.

Figure 4.11: Example of a harddisk recorder as a stand-alone device.

4.6 Compressed audio formats

Digitized audio material is quite demanding regarding memory space and necessary channel capacity for transmission. Consequently solutions were sought to reduce the bit-rate of audio signals. Two categories of compression strategies have to be distinguished:

lossless compression hereby only true redundancy is eliminated which means that the original signal can be reconstructed perfectly, sample by sample.

lossy compression hereby the original signal will *not* be reconstructed perfectly, the aim is that the reconstruction *sounds* like the original signal without relevant audible artifacts.

4.6.1 Lossless compression

Lossless compression eliminates true redundant information. If a new sample can be predicted completely from the past samples, the new sample carries no new information and is thus redundant. An extreme example is a pure sinusoidal signal. As soon as one full period is recorded, the value of the signal can be predicted at any time in the future. Such a signal would thus be describable very compactly. The other extreme is white noise. Hereby all past information about the signal is of no help to predict the new sample. The signal is not redundant at all - white noise can not be compressed in a lossless manner.

Real audio signals lie somewhere between these two extremes, however closer to noise than the sinusoid. The compression potential is thus rather low in the order of 50 %. The strategies for lossless compression are based on similarities between different channels in multichannel signals and the partial predictability of future samples from the past of a signal. Different formats are in use such as:

Apple Lossless ALAC: proprietary format of Apple for the compression of WAV or AIFF files.

Free Lossless Audio FLAC: freely available format from Xiph.Org

Meridian Lossless Packing MLP: proprietary format of Meridian Audio, also known as Dolby Lossless.

MPEG-4 Audio Lossless MPEG-4 ALS: compression format specified by the Moving Picture Expert Group under ISO/IEC 14496-3:2001/AMD 4, Audio Lossless Coding (ALS).

4.6.2 Lossy compression

Much higher compression factors can be reached if lossy compression is accepted that only reproduces a signal that *sounds like* the original. Basis for lossy compression is the fact that the human ear only evaluates certain attributes of an audio signal. In addition some frequency components of the signal are not heard due to masking by other components. In consequence, it is in principal possible to modify an audio signal without any change of the audible impression. With suitable coders, compression factors in the order of 10 are possible, which means that 90 % of the information can be eliminated. Coders that are based on perceptual effects of the ear are called *Perceptual Coders*.

Some actual examples are:

Adaptive Transform Acoustic Coding ATRAC: proprietary format of Sony, used e.g. in MiniDisc.

MPEG-1 Layer 2,3 MP2, MP3: format defined by the MPEG (Moving Picture Experts Group) MP2 mainly in digital broadcast applications, MP3 for internet and portable music players.

Advanced Audio Coding AAC: developed by the MPEG and launched as successor of MP3.

Windows Media Audio WMA: proprietary format defined by Microsoft.

Lossy compression is suited for end products that don't have to fulfill highest demands. Difficulties arise as soon as compressed audio material is edited or if an encoded signal is decoded and encoded again.

Principal structure of perceptual coders

Figure 4.12 shows the principal building blocks of a perceptual coder. The digital audio signal $x[n]$ is split into N frequency bands by help of a filter bank. At the outputs of the filter bank, the filtered time signals x_1 to x_N are available. Simultaneously a psychoacoustic model is operated to simulate the fundamental properties of the human ear. This psychoacoustic model defines the necessary quantization of each time signal x_1 to x_N. Finally the set of coded signals is collected to a data stream x_c.

Compared to the coder, the decoder is much simpler (Figure 4.13) and needs substantially less computational effort. The decoder only has to demultiplex and decode the data stream x_c into the filtered time signals x'_1 to x'_N and then to resynthesize the audio signal $y[n]$.

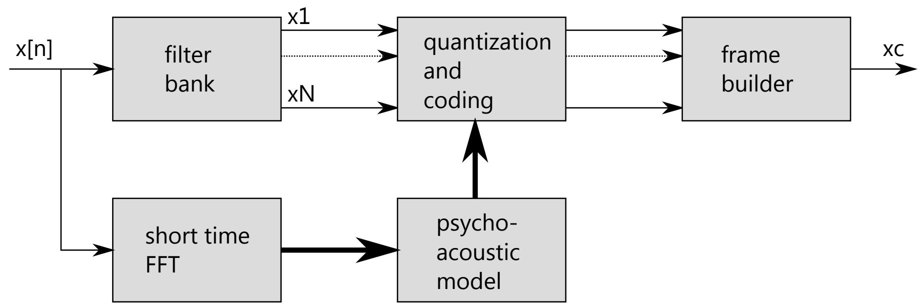


Figure 4.12: Principial building blocks of a perceptual coder based on subband coding.

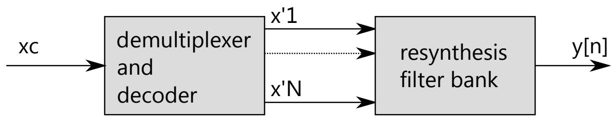


Figure 4.13: Principal building blocks of a perceptual decoder based on subband coding.

ISO-MPEG-1 Coding

In the following the perceptual coding scheme MPEG-1 is illustrated in more detail [9]. For MPEG-1 different variants - labeled as layer 1, 2 and 3 - were developed. The MPEG-1 specifications define the decoders, the encoders can vary within the corresponding compatibility limits. Since 1999 the layer 3 variant (MP3) is very popular. A driving force behind the development of the format was the Fraunhofer Institut für Integrierte Schaltungen[10].

The analysis filter bank in MPEG encoding splits the signal into 32 subbands of constant width. According to the layer 3 specification each of the 32 subbands splits then into another 6 or 18 bands. The quantization in each band depends on a psychoacoustic model. The main task of this model is to simulate the effect of frequency masking. For that purpose in each band i the signal power level $L_S(i)$ is determined. The presence of a signal component in band i lifts the hearing threshold in this band. For noise like signals this new threshold $L_T(i)$ lies 5.5 dB below $L_S(i)$. For tonal signals the difference is somewhat larger. The quantization in band i is then chosen in such a way that the quantization noise just lies below the shifted threshold (Figure 4.14).

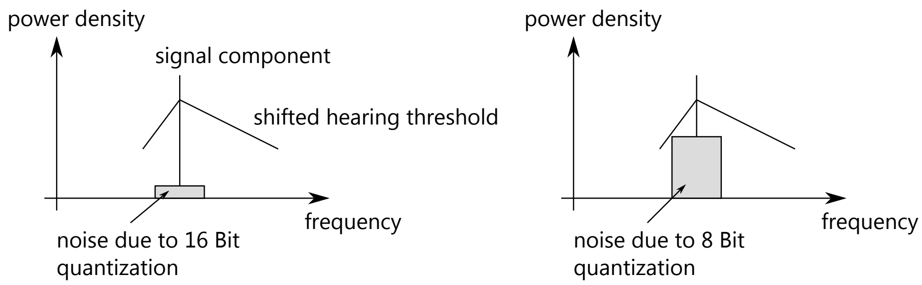


Figure 4.14: Inaudible increase of the quantization noise due to coarser quantization.

The effect of frequency masking is used not only within the band of the masker but in the neighbor bands as well. With the frequency response L_{T_m} of the shifted threshold a level difference $SMR(i) = L_S(i) - L_{T_m}(i)$ between the signal power level $L_S(i)$ and L_{T_m} is calculated for each band. $SMR(i)$ determines then the necessary quantization of each subband. The bit allocation is dynamic which means that the number of bits per subband varies with the signal.

[9] Karlheinz Brandenburg, Marina Bossi, Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding, J. Audio Engineering Society, v.45, n.1/2, 4-21 (1997).
[10] http://www.iis.fhg.de/amm/

In case of multichannel signals further data reduction is possible by using the fact that the differences between the channels are usually small. Therefore it can be beneficial not to code the channel signals but only the differences.

With bit rates of about 128 kBit/s for a stereo signal - corresponding to a compression of 1:12 - MP3 yields very good results with barely audible differences between original and compressed signal. For bit rates of 160 kBits/s and higher most people can not distinguish between original and compressed signals [11].

[11] Gilbert A. Soulodre et al., Subjective Evaluation of State-of-the-Art Two-channel Audio Codecs, J. Audio Engineering Society, v.46, no. 3, 164-177 (1998).

Chapter 5

Recording technique

For a more thorough introduction, see e.g. the book by the Bartletts [1].

5.1 Stereo recordings

Stereo was patented in 1931 by the British Alan D. Blumlein. He stressed the fact that directional information is necessary for a plausible reproduction of direct sound and reflections that occur in a room. The human auditory system can localize sounds with help of information about intensity and time of arrival differences at the two ears. Consequently stereo recording configurations can be categorized according to the directional information they provide in the stereo signal:

- intensity stereophony: the two microphone signals differ only in their intensities
- time-of-arrival stereophony: the two microphone signals differ only in their phase
- mixed stereophony: the two microphone signals differ in their intensities and in their phase.

Although intensity stereophony reproduces directional information quite well, spaciousness is usually more convincing in time of arrival stereophony. However an advantage of intensity stereophony is it's mono compatibility.

5.1.1 Intensity stereophony

XY arrangement

The XY arrangement uses two cardioid microphones that are installed at the same position but with different orientation (Figure 5.2). As the two capsules lie very close to each other, almost no phase differences occur. Due to the cardioid directivity one microphone will record a sound wave incident from a lateral direction with higher amplitude. The microphones are usually oriented with an angle ϕ of $65°$ relative to the frontal direction. This corresponds to the -3 dB point for the directivity $20\log(1+\cos(\phi))$.

A serious disadvantage of the XY arrangement is the fact that the most important incidence direction *front* is detected with $65°$ off-axis. As microphones are tuned for best performance on-axis, certain deviations in the amplitude response are probable.

As the two microphone membranes can't be installed exactly at the same location, they have to be arranged in such a way that lateral sound arrives at the corresponding microphone firstly (Figure 5.2).

MS arrangement

For the MS arrangement two microphones with different directivities are installed as close as possible to each other. The first capsule is usually an omni or cardioid microphone, the second capsule is always a figure of eight microphone. The membranes of the two microphones form an angle of $90°$. The name *MS* expresses that one microphone records a *M*id signal and the other a *S*ide signal. The two stereo

[1] On-Location Recording Techniques, Bruce and Jenny Bartlett, Focal Press, 1999.

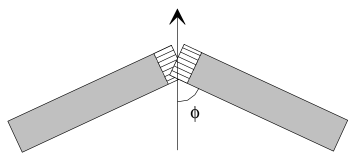


Figure 5.1: Stereo microphone configuration in the XY arrangement. The angle ϕ is usually chosen as $65°$.

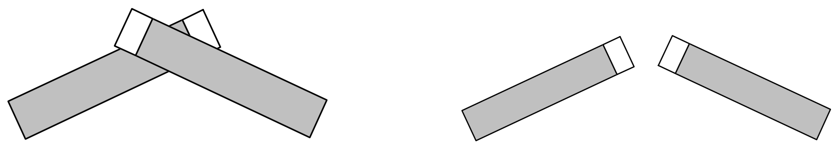


Figure 5.2: left: correct XY arrangement (sound from the left hits the left microphone firstly), right: wrong XY arrangement (sound from the left hits the right microphone firstly).

channels left and right correspond to the sum and difference of the M- and the S-signal. By scaling the M- and S-signal differently, the opening angle can be varied. If a high quality recording device is at hand, it is beneficial to record the M- and S-signal and do the left/right mixing later. This allows for an adjustment of the opening angle at a later time.

Blumlein arrangement

The Blumlein arrangement is one of the oldest stereo recording technique. It is based on two figure of eight microphones that form an angle of $90°$. The microphone signals correspond directly to the left and right stereo channels. The difficulties with the Blumlein arrangement is the high amount of room signal due to the high sensitivity towards the rear direction and the weakness in the low frequencies of the figure of eight microphones.

5.1.2 Time of arrival stereophony

AB arrangement

The AB arrangement uses two omnidirectional microphones that are installed with a certain lateral distance. The stereo effect results solely due to time of arrival differences. A typical lateral separation is 20 cm. However, if the microphones are placed further away from the source, the spacing can be made larger. If the spacing is made too large, the localization will be lost due to the ambiguity of the phase information at high frequencies.

A big advantage of AB stereophony lies in the fact that pressure microphones can be used that come closer to the ideal microphone compared to microphones with a cardioid or figure of eight directivity. The AB arrangement is often used for the recording of classical music in churches and in concert halls.

5.1.3 Mixed stereophony

ORTF arrangement

The ORTF arrangement was developed by the French Radio. Thereby two cardioid microphones are used that are installed with a lateral spacing of 17 cm. Each microphone is oriented with an angle of $55°$ relative to the frontal direction, resulting in a total opening angle of $110°$. This arrangement produces intensity differences as well as time of arrival differences.

Jecklin disc (=OSS)

An alternative method to introduce intensity and time of arrival differences is the use of two laterally separated omnidirectional microphones with a massive disc in between. This arrangement is named after it's inventor Jürg Jecklin[2]. The lateral spacing is responsible for time of arrival differences, the disc leads to intensity differences. A difficulty of the Jecklin disc are reflections at the disc surface that lead to interferences with the direct sound and thus to comb filter effects. Therefore the disc has to be made as absorptive as possible.

Binaural stereophony

Instead of a disc as used by Jecklin, an artificial head can be installed as separation of the two microphones. Consequently the microphone membranes are then placed at the position of the eardrums. Such a binaural stereophony arrangement produces intensity and time of arrival differences and in addition frequency filtering by the auricle and the ear canal. Instead of an artificial head, the own head can be used with small microphones placed in the ear canals. Binaural stereo recordings are suited for reproduction by headphones. For best quality, the transfer function from the headphone membrane to the eardrum has to be compensated for. Binaural stereo produces an impressive auditory spaciousness, however the front-back differentiation is often difficult. A good overview of the location accuracy of binaural recordings can be found in the paper by Möller [3]. The reproduction of binaural recordings by loudspeakers is not satisfying as the head related transfer function is involved two times.

Summary of stereo microphone configurations

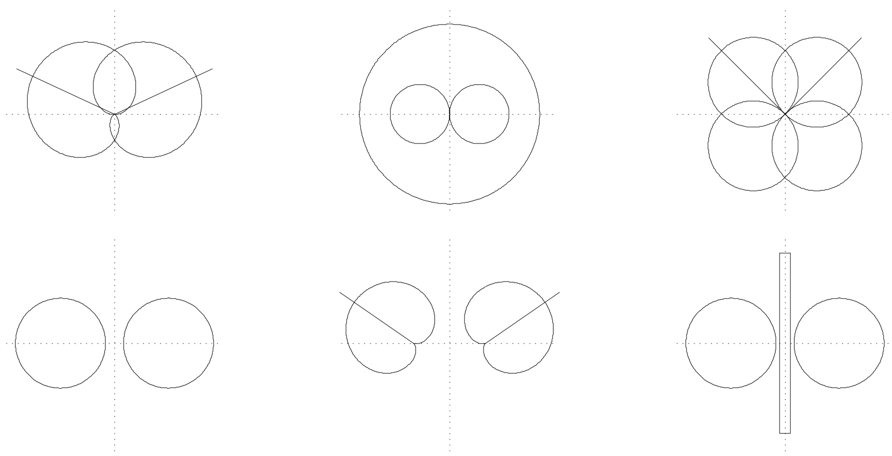


Figure 5.3: Polar diagrams of the microphones used in different stereo recording configurations. Top row: XY, MS and Blumlein, bottom row: AB, ORTF and Jecklin disc.

5.2 Positioning of microphones

Usually a recording shall provide information about the source and the room. If only a stereo microphone is used, the distance to the source determines the ratio of direct and diffuse sound that carries information about the room. If the microphone is installed closer to the source than the *critical distance*, the direct sound dominates. The recording is very transparent but lacks of spaciousness. If the distance of the microphone to the source is larger than the *critical distance*, the recording is too diffuse and blurred. The critical distance is thus often a good compromise and the ideal position for a stereo microphone. In concert halls the critical distance is in the order of a few meters. If directional microphones are used, the optimal distance to the source increases.

[2] Jürg Jecklin, A Different Way to Record Classical Music, Journal of the Audio Engineering Society, vol. 29, pp. 329-332 (1981).

[3] Henrik Möller et al., Evaluation of Artificial Heads in Listening Tests, Journal of the Audio Engineering Society, vol. 47, pp. 83-100 (1999).

The above mentioned necessary compromise between direct and diffuse sound can be dismissed, if both contributions are recorded separately. For that purpose often a stereo microphone is installed close to the source to record the direct sound and additionally a distant omnidirectional microphone records the diffuse sound.

The amount of diffuse sound in a recording can be used to adjust the apparent distance to a source. If the signal of an instrument is mixed with a large portion of diffuse sound, the instrument appears far away. On the other hand, if the signal contains mostly direct sound, the instrument seems to be very close to the listener.

Maximal flexibility in the mixing process is obtained, if each instrument or group of instruments is recorded separately with its own microphone. However the placement of the microphones needs special care. In any case it has to be avoided that the same instrument is recorded with similar strength by two different microphones as this would lead to possible interference effects while summing up the microphone signals. To avoid this problem, small walls can be installed between the instruments. If the microphones are then placed suitably, the walls shield the unwanted contributions. However special care has to be taken that the walls don't introduce comb filter effects due to reflections. Pressure zone microphones that are installed directly on the reflecting surface can help to overcome this difficulty.

5.3 Surround sound

With stereo a sound image can be produced that seems to origin from a frontal direction within an opening angle of about $60°$. This represents more or less the direct sound of an extended orchestra on stage. However stereo is unable to reproduce appropriate directional information of room reflections and diffuse reverberation. An extension to a perspective of $360°$ are surround sound formats. For movie sound the 5.1 format is nowadays a quasi standard[4]. With newer sound storage media surround sound becomes interesting for pure audio productions. First ideas for multichannel audio date back to the 1930's. Essential experiments were conducted by Disney for the movie Fantasia. In the 1960's the four channel system quadrophony was launched, however without commercial success.

The 5.1 format stores the audio information digitally, usually in a compressed format [5]. 5.1 offers two stereo channels *left* and *right* and in addition a *center* channel, a *left surround* and a *right surround* channel and (indicated by .1) a *low frequency effects LFE* channel. The first five channels cover the full audio range while the LFE channel is responsible for frequencies between 20 and 120 Hz only. The motivation for the LFE channel is a relief of the 5 main channels as they don't need to reproduce the highest low frequency amplitudes. The center channel is important for sound that should be heard from the middle in case of a listener that is not exactly seated on the symmetry axis between the left and right loudspeakers.

Regarding recording techniques, surround sound introduces additional degrees of freedom compared to stereo [6]. Firstly it has to be clarified which perspective should be reproduced. Besides special effects recordings, two main concepts are common:

- The *direct/ambient* perspective seeks to create a realistic listening experience as it would be heard in a real environment. For that purpose the three front channels *left*, *right* and *center* are used to reproduce the direct sound of the source. The two surround channels cover the lateral and rear sided room reflections and reverberation. For the front channels usually a stereo microphone is used while the surround channels are produced by omnidirectional room microphones.
- The *inside the band* perspective creates the illusion of a listening position within the source, e.g. in the middle of an orchestra. For maximal flexibility usually distributed microphone arrangements are used.

[4] Tomlinson Holman, 5.1 Surround Sound - Up and Running, Focal Press, 2000.
[5] Marina Bosi, High-Quality Multichannel Audio Coding: Trends and Challenges, Journal of the Audio Engineering Society, vol. 48, no. 6, p.588-595 (2000)
[6] Francis Rumsey, Microphone and Mixing Techniques for Multichannel Surround Sound, Journal of the Audio Engineering Society, vol 46, n. 4, p. 354-358 (1998).

Chapter 6

Reproduction of audio signals

6.1 Stereo loudspeakers

Probably still the most common reproduction arrangement is a stereo loudspeaker pair. As a rule of thumb the listening position should form an equilateral triangle together with the two loudspeakers (Figure 6.1). The line between the two loudspeakers is called stereo basis. Due to intensity and time of arrival differences between the loudspeaker signals the ear creates the illusion of phantom sources that seem to be located on the stereo basis. A level difference of 25 dB or a delay of 2 ms between the loudspeaker signals results in completely lateral localization.

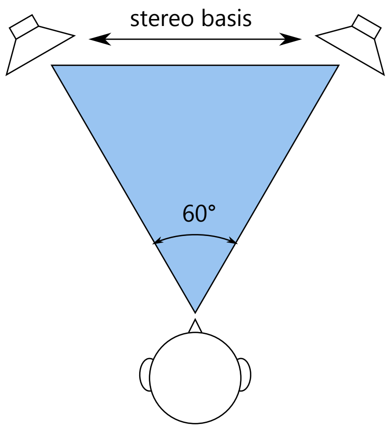


Figure 6.1: Standard stereo loudspeaker arrangement. The listener and the two loudspeakers form an equilateral triangle.

6.2 Headphones

At the reproduction by headphones the transducers are located directly in front of the outer ear. In contrast to the loudspeaker case, with headphones there is no crosstalk from the left channel to the right ear and vice versa. An other advantage is that the acoustical condition of the listening room is irrelevant. However headphone reproduction often leads to localization inside the head.

In a typical live listening situation an extended source is located in front of the listener. The direct sound signals stem from an azimuthal angle of maybe $\pm 30°$. As a consequence of the head related transfer function from the source to the ear drums a significant linear frequency distortion occurs. If one listens with headphones to a stereo recording that was performed with help of a normal microphone arrangement such as XY or AB, the head related transfer function is by-passed. For that reason, headphones are usually equipped with a free field equalization.

Headphones can be built in open or closed form. The speciality of the closed variant is that the headphone and the ear canal form a pressure chamber. The sound pressure at the ear drum just follows the excursion of the headphone membrane.

Most headphone transducers are of the electro-dynamic type. However some headphones use electro-static systems.

6.3 Loudspeaker reproduction with crosstalk cancelation

The sound pressure at a listener's ear drums can be controlled almost perfectly with loudspeaker reproduction and crosstalk cancelation [1],[2]. For that purpose a conventional stereo listening arrangement is used. Prior to listening the transfer functions from both loudspeakers to both ear drums have to measured. From these the corresponding crosstalk cancelation filters can be derived. They make sure that the signals from the left speaker to the right ear and from the right speaker to the left ear are nullified. In contrast to headphone reproduction this arrangement produces no localization inside the head. However a serious disadvantage of the method is that the crosstalk cancelation transfer functions are very sensitive to changes in geometry. This makes it necessary that the head of the listener is kept at rest at the predefined position.

6.4 Room simulators

Two channel stereo reproduction can create phantom sources along the stereo basis. A step further go room simulators that create a room impression by adding reflections and reverberation from non frontal directions. A difficulty with normal recordings and reproduction in not specifically treated listening rooms is the fact that information of three rooms overlays:

- the stereo recording contains reverberation and thus information about the recording room
- the room simulator adds the acoustics of the artificial room
- at reproduction the listening room adds reflections and reverberation.

Best results with room simulators can be expected if anechoic recordings are used and the listening room is heavily dampened with absorptive material at the room surfaces.

The room simulator adds two components to the original signal:

discrete reflections: delayed copies of the original signal from specific directions and possibly equalized according to the absorption coefficient of the corresponding reflecting surface

diffuse reverberation: reflections with high temporal density (reverberation) from all directions, specified by the parameters *decay time* and *frequency response*

While the generation of discrete reflections is quite straight forward, the simulation of diffuse reverberation is more elaborate. Before digital solutions [3] were available, people used tape recorders with feedback loops [4], spring reverb devices [5], hall plates [6] or echo chambers. Today's systems work exclusively digital.

The simplest structure to produce reverberation is shown in Figure 6.2.

The condition for stability for the structure shown in Figure 6.2 is $g < 1$. The resulting impulse response $h(t)$ has the form:

$$h(t) = \delta(t-\tau) + g\delta(t-2\tau) + g^2\delta(t-3\tau) + g^3\delta(t-4\tau) + \ldots \tag{6.1}$$

[1] M. R. Schroeder, B. S. Atal, Computer Simulation of Sound Transmission in Rooms, IEEE International Convention Record, Part 7, p.150-155 (1963).

[2] O. Kirkeby, P. Nelson, H. Hamada, Local sound field reproduction using two closely spaced loudspeakers, J. Acoustical Society of America, vol. 104, p.1973-1981 (1998).

[3] Barry Blesser, An Interdisciplinary Synthesis of Reverberation Viewpoints, Journal of the Audio Engineering Society, vol. 49, p.867-903 (2001).

[4] G. R. Crane, G. A. Brookes, Artificial Reverberation Facilities for Auditoriums and Audio Systems. J. Audio Engineering Soc. n.3, vol. 9, p.198-204 (1961).

[5] H. E. Meinema et. al. A New Reverberation Device for High Fidelity Systems. J. Audio Engineering Soc. n.4, vol. 9, p.284-326 (1961).

[6] K. Arcas, A. Chaigne, On the quality of plate reverberation, Applied Acoustics, vol. 71, p.147-156 (2010).

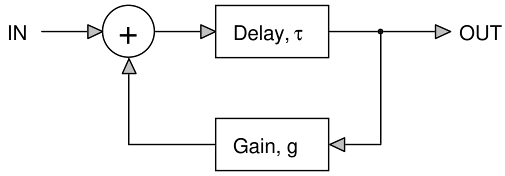


Figure 6.2: Simplest structure of a reverberation generator.

with the spectrum

$$H(\omega) = e^{-j\omega\tau} + ge^{-j\omega 2\tau} + g^2 e^{-j\omega 3\tau} + g^3 e^{-j\omega 4\tau} + \ldots \tag{6.2}$$

Eq. 6.2 represents a geometric series. The sum is given by

$$H(\omega) = \frac{e^{-j\omega\tau}}{1 - ge^{-j\omega\tau}} \tag{6.3}$$

The amplitude response of (6.3) is:

$$|H(\omega)| = \frac{1}{\sqrt{1 + g^2 - 2g\cos(\omega\tau)}} \tag{6.4}$$

The function (6.4) corresponds to a comb filter. The amplitude differences between maxima and minima increase for growing g. This non-flat amplitude response is a drawback of this simplest reverberation structure. Schroeder [7] has shown that by the subtraction of a suitable amount of the original signal (Figure 6.3) a flat amplitude response can be obtained.

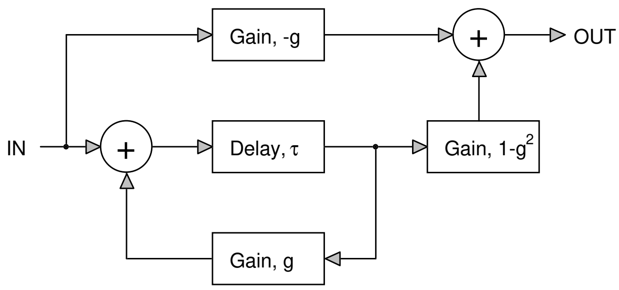


Figure 6.3: Modified structure of a reverberation generator with a flat amplitude response.

The output of the reverberation generator decays by the factor g per delay time τ. This corresponds to a level drop of $20\log(g)$. Consequently the resulting reverberation time for a decay of 60 dB is

$$RT = \tau \frac{60}{-20\log(g)} \tag{6.5}$$

The reverberation generator produces a constant echo density $1/\tau$. For a realistic auralization of a diffuse field the echo density should be in the order of 1000 per second. To reach such a value, usually several reverberation units have to be put in series.

6.5 Surround sound

Surround sound recordings can reproduce sound events not only from within the $60°$ of the stereo angle but from backward directions as well. Thus, surround sound allows for a more realistic reproduction of the acoustics of rooms. Earlier systems such as Dolby Surround used matrix-coded formats with four channels mapped onto two physical tracks - however with the disadvantage of a limited channel separation. Actual surround sound such as Dolby Surround Digital (5.1) is based on fully separated channels.

[7] M. R. Schroeder, Natural Sounding Artificial Reverberation, J. Audio Engineering Soc. n.3, vol.10, p.219-223 (1962).

In 1992 Dolby launched the format Dolby Surround Digital for movie sound. It is often labeled as *SR.D* or *5.1*. The format uses 5+1 digitally coded independent audio channels. The channels represent the following directions:

- left (20...20'000 Hz)
- center (20...20'000 Hz)
- right (20...20'000 Hz)
- surround left (20...20'000 Hz)
- surround right (20...20'000 Hz)
- LFE (low frequency effects) (20...120 Hz)

The LFE channel is optional. The main purpose is to relieve the 5 other channels from the reproduction of the lowest frequencies with high amplitude. The audio signals are usually coded in a compressed format such as Dolby AC-3. The ideal loudspeaker configuration for the 5.1 sound is shown in Figure 6.4.

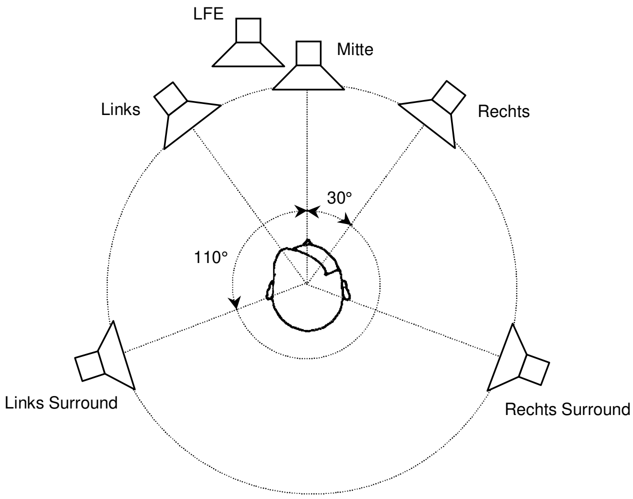


Figure 6.4: Ideal listening configuration for 5.1 surround sound with the five main channels *left*, *right*, *center*, *left surround* and *right surround* and the optional low frequency effects channel *LFE*.

The reproduction of surround sound material is possible by headphones as well. The headphone signals are generated by processing of each loudspeaker channel with the left and right head related transfer functions HRTF for the corresponding sound incidence direction. Advanced headphone systems are equipped with a head tracker to detect head movements. In this case the HRTF's are adjusted accordingly with the result that the sound field remains at it's original position in the virtual room. These systems usually allow for the addition of artificial reverberation to compensate for the missing reverberation of the listening room.

6.6 Wave field synthesis

Stereo- and surround sound reproduction create the desired listening impression in a sweet spot by appropriate level- and time of arrival differences. The optimal reproduction system however would be capable to synthesize an arbitrary sound field in an extended region of a room. From a theoretical point of view this appears possible if sound pressure $\check{p}_S$ and the normal component of sound particle velocity $\check{v}_S$ on a closed surface around the volume of interest can be controlled perfectly. This principle of sound generation is called wave field synthesis [8]. The mathematical-theoretical principle is given in form of the Kirchhoff-Helmholtz integral:

[8] Günther Theile, Wave field synthesis - a promising spatial audio rendering concept, Proceedings of the 7th Int. Conference on Digital Audio Effects, 2004.

$$\check{p}(x,y,z,\omega) = \frac{1}{4\pi}\int\limits_S \left(j\omega\rho_0 \check{v}_S(\omega)\frac{e^{-j\omega r/c}}{r} + \check{p}_S(\omega)\frac{\partial}{\partial n}\frac{e^{-j\omega r/c}}{r}\right) \mathrm{d}S \tag{6.6}$$

or in a somewhat modified form:

$$\check{p}(x,y,z,\omega) = \frac{1}{4\pi}\int\limits_S \left(j\omega\rho_0 \check{v}_S(\omega)\frac{e^{-jkr}}{r} + \check{p}_S(\omega)\frac{1+jkr}{r^2}\cos\phi e^{-jkr}\right) \mathrm{d}S \tag{6.7}$$

The integrand in Eq. 6.7 is composed of an omnidirectional monopole contribution $\check{v}_S$ and a dipole contribution $\check{p}_S$ with a cos(ϕ)-directivity. For the control of sound pressure $\check{p}(x,y,z)$ at a point in space, the surface S has to be packed densely with monopole and dipole sources. These sources are then steered accordingly to create the desired sound field.

In a practical realization, the following limitations are encountered:

spacial discretization the continuous sound pressure and sound particle velocity distribution on the surface S has to be approximated by a limited number of discrete sources. Above a certain limiting frequency, this spacial discretization leads to aliasing with the consequence of errors in the sound field. This upper limiting frequency can be increased by installation of more loudspeakers on a finer mesh.

room reflections reflections at the boundary of the listening room introduce errors in the synthesized sound field. The relevance of these reflections can be reduced by covering walls and ceiling by absorptive material. If the reflection at the boundary surfaces can be described exactly, their presence can be taken into account and compensated for in the wave field synthesis calculation.

effort due to the enormous effort regarding the number of sources, practical wave field synthesis realizations are often restricted to two dimensions. A suitable loudspeaker arrangement is then a circle.

6.7 Ambisonics

Ambisonics is a multi-channel recording and reproduction technique, invented by Michael Gerzon in the early 1970s. Ambisonics represents sound from a certain direction with help of a multipole expansion. The multipole expansion is formulated as a sum of spherical harmonics. The distribution of signal directivity at a certain point in space is thus decomposed into

- 1 monopole part C_0^0
- 3 dipole parts C_1^{-1}, C_1^0, C_1^1
- higher order parts $C_n^{-j}, \ldots, C_n^j$

Ambisonics is mostly used in form of *first order Ambisonics* which utilizes the monopole (zeroth order) and dipole parts (first order) only. These four channels are labeled as W, X, Y and Z and form the so-called B-format. W is the monopole part while X, Y and Z correspond to the directional components. X can be thought of as a recording with a pressure microphone (omnidirectional). The X, Y and Z channel correspond to coincident recordings with figure-of-eight microphones pointing in x, y and z direction of a three dimensional coordinate system (Figure 6.5) [9].

The signal $s(t)$ of a source that is seen under azimuthal angle α (0 = frontal direction, counter-clockwise direction of rotation assumed) and elevation angle β (relative to the horizontal plane, positive angles represent points above the plane) (Figure 6.6) is consequently encoded in the four B-format channels as:

[9] An example of a commercial B-format sensor system is the Soundfield microphone.

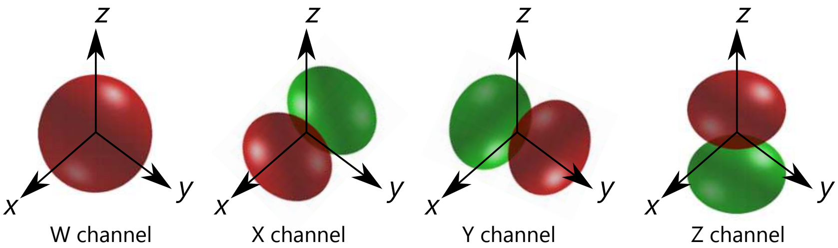


Figure 6.5: Zeroth and first-order spherical harmonics, corresponding to the directivities of the four B-format channels (green and red represent opposite signs).

$$\begin{aligned} W &= \frac{1}{\sqrt{2}} \cdot s(t) \\ X &= \cos(\alpha)\cos(\beta) \cdot s(t) \\ Y &= \sin(\alpha)\cos(\beta) \cdot s(t) \\ Z &= \sin(\beta) \cdot s(t) \end{aligned} \tag{6.8}$$

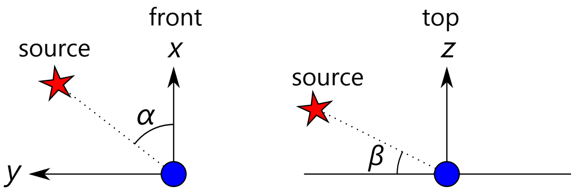


Figure 6.6: Azimuthal angle α and elevation β definition. Left: top view, right: section view.

If only directional information in the horizontal plane is needed, Eq. 6.8 reduces to the 2D-version in Eq. 6.9.

$$\begin{aligned} W &= \frac{1}{\sqrt{2}} \cdot s(t) \\ X &= \cos(\alpha) \cdot s(t) \\ Y &= \sin(\alpha) \cdot s(t) \\ Z &= 0 \end{aligned} \tag{6.9}$$

It should be noted that the scaling factor $1/\sqrt{2}$ in the W channel is introduced by convention. By linear-combination of the four B-format channels it is possible to create the audio signal that would have been captured by any first-order microphone (omni, cardioid, figure-of-eight) pointing in any direction.

Similarly it is easily possible to modify the orientation of a source coded in B-format. An acoustic scenery can be rotated around the x-axis (tilt), the y-axis (tumble) or the z-axis (rotate) or any combination of them. Starting with given W, X, Y, Z the new B-format signals W', X', Y', Z' are found by appropriate scaling:

$$\begin{aligned} W' &= k_1 W + k_2 X + k_3 Y + k_4 Z \\ X' &= k_5 W + k_6 X + k_7 Y + k_8 Z \\ Y' &= k_9 W + k_{10} X + k_{11} Y + k_{12} Z \\ Z' &= k_{13} W + k_{14} X + k_{15} Y + k_{16} Z \end{aligned} \tag{6.10}$$

The reproduction of a B-format sound can be done in a very flexible way by any number of loudspeakers. The creation of the loudspeaker signals from the B-format is called the Ambisonic decoding process.

Depending on the application, several decoders have been proposed [10],[11]. As localization of the human auditory system uses interaural time differences at low frequencies and interaural level differences at high frequencies, an optimal decoding scheme should behave differently at low and high frequencies.

In the following it is assumed that sound is reproduced in 2D only (horizontal plane) by a regular arrangement of loudspeakers. The N loudspeakers are placed on a circle around the listener and form azimuthal angles θ_i with respect to the frontal direction (Figure 6.7).

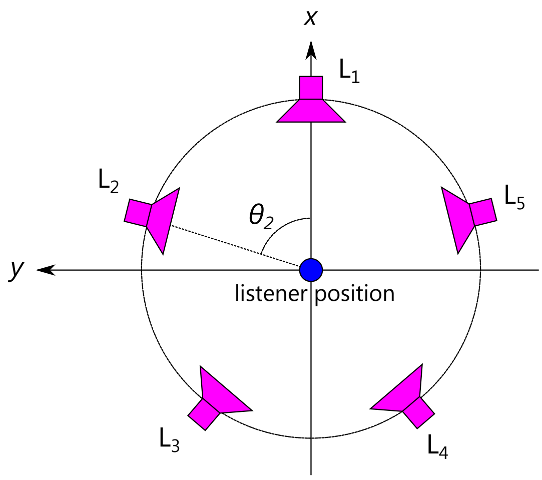


Figure 6.7: Example of a regular arrangement of loudspeakers for the reproduction of B-format sound.

The loudspeaker signals F_i can be expressed as

$$F_i = K_1 \cdot W + K_2 \cdot [X\cos(\theta_i) + Y\sin(\theta_i)] \tag{6.11}$$

The *cardioid decoder* or *in-phase decoder* uses a ratio

$$\frac{K_1}{K_2} = \sqrt{2} \tag{6.12}$$

and has the advantage that under no circumstances signal canceling (out-of-phase summation) occurs at the listener position. In comparison to other decoders, this aspect makes the the localization less sensitive with respect to changes in listener position. Consequently a larger listening area with proper localization can be expected. Figure 6.8 shows the loudspeaker signal amplitudes for the arrangement in Figure 6.7.

The *energy localization vector decoder* uses a ratio

$$\frac{K_1}{K_2} = 1 \tag{6.13}$$

and is believed to yield optimal localization results in the frequency range from 700 to 4000 Hz where inter-aural level differences are of importance (Figure 6.9).

Finally the *velocity localization vector decoder* with a ratio

$$\frac{K_1}{K_2} = \frac{1}{\sqrt{2}} \tag{6.14}$$

is considered as optimal in the low frequency range where inter-aural time differences are responsible for localization (Figure 6.10).

[10] M. A. Gerzon, Ambisonics in Multichannel Broadcasting and Video, Journal of the Audio Engineering Society, vol. 33, p.859-871 (1985).

[11] A. J. Heller et al., Is My Decoder Ambisonic? 125th Convention of the Audio Engineering Society, San Francisco, USA.

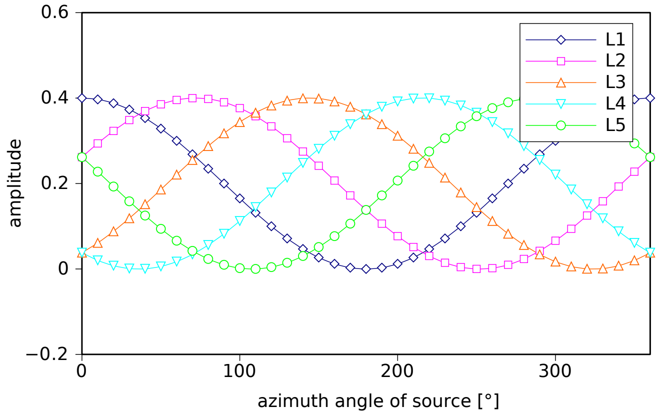


Figure 6.8: Loudspeaker signal amplitudes for the cardioid decoder as a function of azimuth angle of the source.

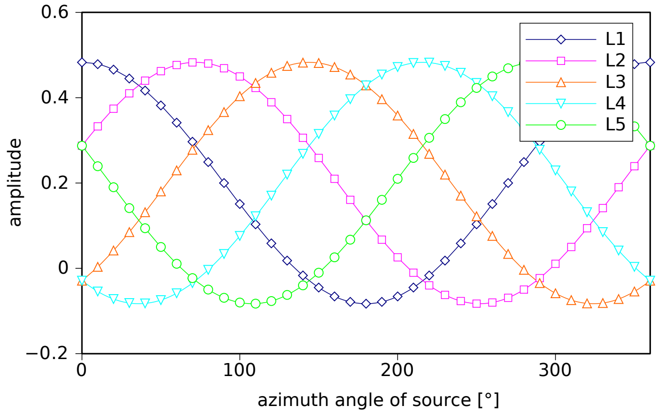


Figure 6.9: Loudspeaker signal amplitudes for the energy localization decoder as a function of azimuth angle of the source.

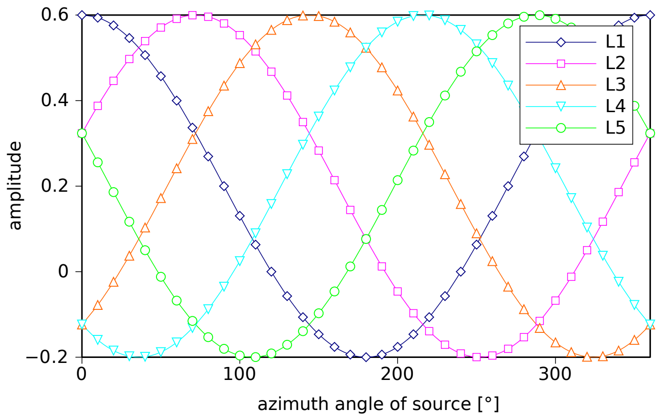


Figure 6.10: Loudspeaker signal amplitudes for the velocity localization decoder as a function of azimuth angle of the source.

As for different frequency bands different localization cues are of importance, the optimal decoder setting is expected to be frequency dependent. This can be realized by applying shelf-filters.

The ambisonics decoders assume coherent summation of the loudspeaker signals at the listener position. At least at higher frequencies this assumption is violated by the fact that the listener's head introduces a sound field distortion and that the two ears sample the sound field at slightly different points in space. At higher frequencies a more appropriate model for the superposition is thus *energy summation*. As in this case out-of-phase signals make no sense, the K_1/K_2 ratio should tend to the value of the cardioid decoder.

6.8 Auro 3D

6.9 Audibility of transmission errors

6.9.1 Distortion of the amplitude response

In the best case the amplitude of the transfer function of an audio chain (microphone - storage medium - sound generating element (headphones or loudspeaker and listening room) is frequency independent. This can never be reached perfectly. The resulting deviations from a flat frequency response curve are called linear distortions. The term "linear" is important, as no new frequencies occur as in case of "non-linear" distortions. The ear is most sensitive to amplitude errors, if the deviation is restricted to a relative narrow band of about one third-octave. In a direct A/B listening test, a deviation of about 1 dB can be heard. Without the possibility of a direct comparison, the just noticeable difference is about 3 dB. Exaggerations are heard more easily than dips.

6.9.2 Distortion of the phase response

Helmholtz and Ohm postulated more than 100 years ago that the audible coloration of a complex tone depends on the amplitude spectrum only whereas the phase spectrum has no influence. This finding is still recognized as correct. However there exist special signals that differ in perception if the phase is changed. Such signals are composed of low and mid frequency components. The masking effect of the low frequency components for the mid frequency component depends on the time course of the signal and thus is influenced by possible phase shifts.

6.9.3 Non-linear distortion

A non-linear transfer function of one element of the audio chain introduces "non-linear" distortion. As a consequence, new frequencies are generated that are not contained in the original signal. Non-linear transfer characteristics are usually described by a power series:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots \tag{6.15}$$

where
x: time response of the input signal
y: time response of the output signal
a_i: coefficients

In case of a weak nonlinearity, the following relation for the coefficients holds:

$$a_i \ll a_1 \text{ for } i = 2, 3, \ldots \tag{6.16}$$

The transmission of a sinusoidal signal $x(t) = X\cos(\omega t)$ yields

$$y(t) = a_0 + a_1 X\cos(\omega t) + a_2 X^2 \frac{1}{2}\left[1 + \cos(2\omega t)\right] + a_3 X^3 \frac{1}{4}\left[3\cos(\omega t) + \cos(3\omega t)\right] + \ldots \tag{6.17}$$

Ignoring the DC component and having the relation 6.16 in mind, the output $y(t)$ from Eq. 6.17 can be written as:

$$y(t) \approx a_1 X\cos(\omega t) + a_2 \frac{X^2}{2}\cos(2\omega t) + a_3 \frac{X^3}{4}\cos(3\omega t) + \ldots \tag{6.18}$$

The transmission of a sinusoidal signal of frequency f through a non-linear transfer characteristics produces harmonic components at the frequencies $2f, 3f, \ldots$. In case of sound of musical instruments these components usually cause no harm as they are masked by harmonic components already present in the original signal.

However, if the original signal contains two sinusoidal components of different frequencies, additional combination frequencies occur. To keep the analysis simple, the case of a quadratic distortion characteristics is investigated here. The input signal is composed of a first component with amplitude X_1 and angular frequency ω_1 and a second a second component with amplitude X_2 and angular frequency ω_2:

$$x(t) = X_1 \cos(\omega_1 t) + X_2 \cos(\omega_2 t) \tag{6.19}$$

The transfer characteristics of a weak quadratic non-linearity is assumed as

$$y = a_1 x + a_2 x^2 \tag{6.20}$$

with
$a_2 \ll a_1$

The output is found as

$$\begin{aligned} y(t) \approx a_1 \left(X_1 \cos(\omega_1 t) + X_2 \cos(\omega_2 t)\right) + \frac{1}{2} a_2 \left(X_1^2 \cos(2\omega_1 t) + X_2^2 \cos(2\omega_2 t)\right) \\ + a_2 X_1 X_2 \left[\cos\left((\omega_2 + \omega_1)t\right) + \cos\left((\omega_2 - \omega_1)t\right)\right] \end{aligned} \tag{6.21}$$

The transmission of to sinusoidal signal components with frequencies f_1 and f_2 yields harmonic distortion components $2f_1$ and $2f_2$ and in addition sum- and difference tones at frequencies $f_2 + f_1$ and $f_2 - f_1$. Often, the difference tone is audible as he lies below the lowest signal component and is thus not masked. Very critical in this respect is the performance of two instruments that are poor in harmonics such as e.g. flutes.

The degree of non-linearity can be described by different quantities, depending on the effect one is interested in. The occurrence of additional harmonic components for excitation with one pure tone is called harmonic distortion. The generation of sum- and difference tones for simultaneous excitation with two or more pure tones is described by the intermodulation distortion.

Harmonic distortion is quantified by the distortion factor k or by the *THD* (Total Harmonic Distortion). For sinusoidal excitation k is given as

$$k = \sqrt{\frac{\sum\limits_{n=2}^{\infty} X_n^2}{\sum\limits_{n=1}^{\infty} X_n^2}} \times 100\% \tag{6.22}$$

where
X_1: amplitude of the fundamental component
X_n: amplitude of the n-th harmonic produced by the non-linearity

If the measurement of the distortion factor contains the contribution of the noise as well, the quantity is labeled as *THD+N*.

6.9.4 Residual noise

Each component of an audio chain contributes to some extent to the residual noise. Residual noise can be separated in the categories *hum* and *noise*.

Hum describes tonal components that are related to the power supply frequency or its harmonics. With proper wiring of the components (in the best case with differential signaling) hum components can be easily kept as low as that they are inaudible.

Noise is produced by each individual electric component of the circuits involved. However modern low noise amplifier concepts achieve very good values in self-noise. For the specification of residual noise, different procedures are applied. They differ in the signal attributes that are evaluated and the usage of frequency weighting.

6.10 Differential signaling

In professional audio, analog components have balanced inputs and outputs and are connected to each other by differential signalling. Hereby two signal wires and a shield are used. The voltages of each of the two signal wires with respect to shield potential have opposite sign or $180°$ phase shift. At the input of a receiving component the difference between these two signals is determined. By this operation, possible unwanted electromagnetic interference is eliminated as it can be expected to occur on both signal wires in the same way. For that reason, long cables can be used without relevant interference.

Differential signalling uses often XLR-connectors (Figure 6.11). The wiring of the pins is shown in Figure 6.12. If a balanced and an unbalanced component are connected, the cold wire (b) is connected to the shield on the side of the unbalanced component (Figure 6.13).

Figure 6.11: XLR cable connectors, female on the left and male on the right.

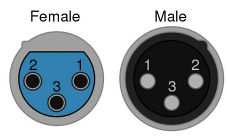


Figure 6.12: Wiring of XLR connectors. 1: shield, 2: positive signal (hot), 3: negative signal (cold).

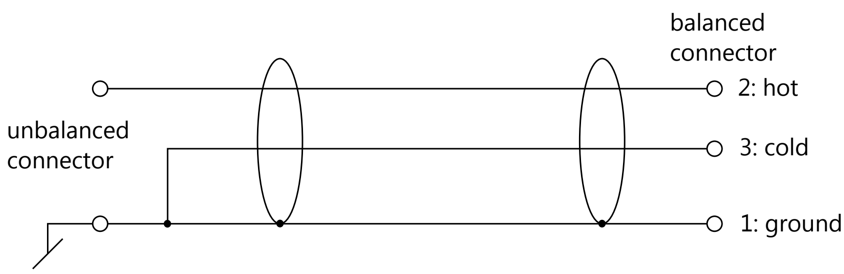


Figure 6.13: Connection of a component with an unbalanced connector and a component with a balanced connector.

6.11 Digital connections, AES/EBU, SPDIF

Digital audio components such as CD-players or hard-disc recorders usually offer - besides analog input/output - digital connections for the signal exchange. Professional devices use the serially operating AES/EBU interface for stereo signals. The signal is transmitted by balanced wiring and XLR connectors, the voltage lies between 2 and 7 V peak-peak. For moderate lengths, standard symmetrical audio cables can be used. The AES/EBU format uses 32 Bit per single channel audio sample. For the representation of the sample itself, 24 Bit are available. The remaining bits are used for *status information*, *parity check* and so on. With help of the *channel status bit* further information can be coded by distribution over several samples. The two audio channels are sent one after the other. With a sampling rate of 48 kHz, the volume of the data stream is $48'000 \times 2 \times 32 = 3.072$ MBit/sec.

Consumer components are usually equipped with an SPDIF-interface (Sony Philips Digital Interface). The format is closely related to the AES/EBU standard. In many cases a direct data transfer SPDIF $\leftrightarrow$ AES/EBU is possible. The main differences lie in the usage of the channel status information and in the electrical specifications of the connection. SPDIF uses unbalanced wiring with Chinch connectors and has a peak voltage of 0.5 V. Recent devices are often equipped with an additional optical interface such as TOSLINK introduced by Toshiba.

Chapter 7

Public address systems

7.1 Introduction

The primary goal of a public address system (PA system) is the suitable acoustical reproduction of an electrical signal by loudspeakers. Hereby a constant (+/-3 dB) direct sound pressure level in the whole audience region shall be produced. A signal-to-noise level difference of 10...25 dB shall be achieved. Typical fields of application for PA systems are:

- amplification of sources that are too weak
- amplification of sources with an insufficient directionality, with the consequence that the diffuse field is excited too strongly
- amplification of sources that are not present (reproduction of an audio signal that was captured elsewhere.

Historically seen, the demand for PA systems sprung up with the introduction of sound in movies. For the history of PA systems see e.g. [1].

An important area of application of PA systems is the amplification of speech in auditoria. In the following, PA systems for speech will be discussed in more detail.

To start with, some facts about speech signals are presented. A person produces with normal voice in 1 m distance about 65 dB(A) sound pressure level. The corresponding third-octave band spectrum is shown in Figure 7.1.

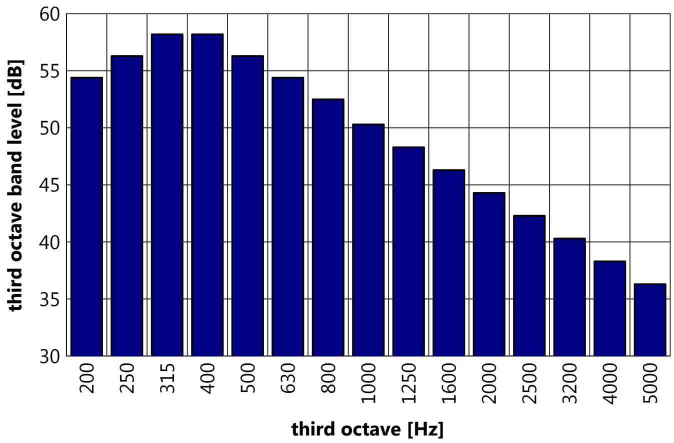


Figure 7.1: Third-octave band spectrum of average human voice according to ANSI.

[1] J. Eargle, M. Gander, Historical Perspectives and Technology Overview of Loudspeakers for Sound Reinforcement, Journal of the Audio Engineering Society, vol. 52, p.412-432 (2004).

Table 7.1 shows the relevance of each third-octave band for the speech intelligibility. Of great importance for the practice is the fact that frequencies below 300 Hz do not contribute significantly to speech intelligibility.

third-octave [Hz]	200	250	315	400	500	630	800	1000
weight	4	10	10	14	14	20	20	24
third-octave [Hz]	1250	1600	2000	2500	3150	4000	5000	
weight	30	37	37	34	34	24	20	

Table 7.1: Relevance (weight) of each third-octave band for speech intelligibility (source: determination of the Articulation Index).

7.1.1 Considerations of sound pressure levels

The building blocks of a PA system for speech are the *microphone*, *amplifier* and *loudspeaker* and the *room-acoustical conditions*. The goal is to produce by help of the loudspeakers a sound pressure in the audience area that is significantly amplified compared to the direct sound of the original source (the speaker).

A potential problem is the occurrence of feedback in the loop *loudspeaker - room - microphone*. Feedback may become audible already below the point of instability. It manifests as linear distortions of the frequency response and in stretching of transient signals. For that reason, as a rule of thumb, a margin of 10 dB to the point of instability should be guaranteed. This condition has to be fulfilled at all frequencies. It is therefore beneficial to flatten the frequency response of the loudspeaker system (in it's interaction with the room) by an equalizer.

If the PA system has to reproduce speech signals only, the frequency response of the system can be limited according to Figure 7.2.

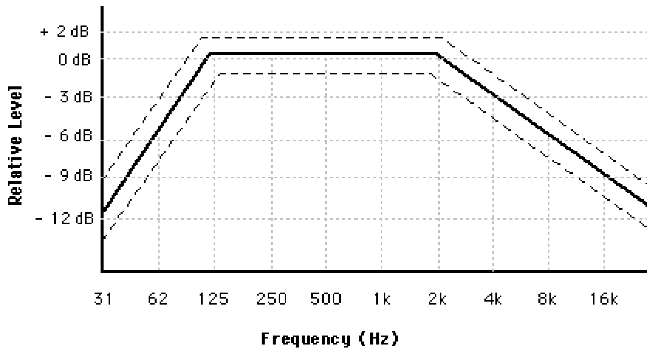


Figure 7.2: Sufficient frequency response for the reproduction of speech signals.

7.1.2 Maximal amplification

For a typical sound reinforcement situation in a room, as shown in Figure 7.3, the maximal possible amplification can be estimated from a few parameters. The key for the analysis is the separation of the sound field into direct and diffuse sound. The direct sound decreases with 6 dB per doubling of distance, the diffuse field is constant all over the room. It is assumed that the PA system consists of n identical loudspeakers, where the microphone is in the direct sound field produced by one speaker only. Similarly the listener receives direct sound from one speaker only.

The situation from figure 7.3 can be translated into a block diagram according to Figure 7.4.

The relevant quantities for the calculation are:

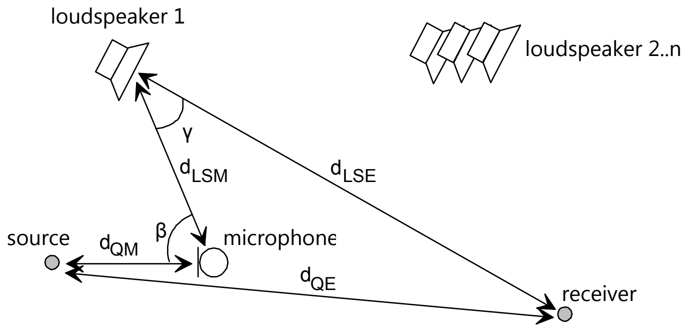


Figure 7.3: Situation of a PA system in a room with the source, one microphone, n identical loudspeakers and the receiver.

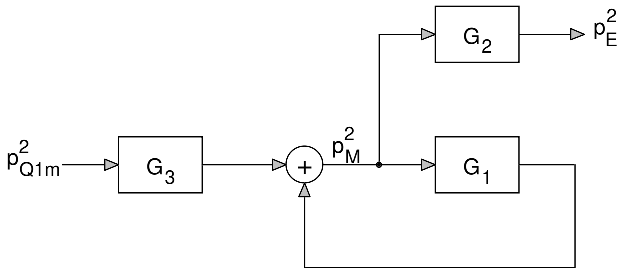


Figure 7.4: Block diagram of the PA system from Figure 7.3.

p_{Q1m}^2: sound pressure square produced by the source (speaker) in 1 m distance

p_M^2: sound pressure square acting on the microphone (taking into account the directional characteristics of the microphone

p_E^2: sound pressure square at the receiver position

p_{LS1m}^2: sound pressure square produced by one loudspeaker in 1 m distance in the main radiation direction

$\nu = \frac{p_{LS1m}^2}{p_M^2}$

$p_{LS1m}^2(\gamma)$: sound pressure square produced by one loudspeaker in 1 m distance in direction γ relative to the main radiation direction

$\overline{p_{LS1m}^2}$: sound pressure square produced by one loudspeaker in 1 m distance averaged over all directions

$RW_{LS}(\gamma)$: directivity of the loudspeaker $= \frac{p_{LS1m}^2(\gamma)}{p_{LS1m}^2}$

$RW_M(\beta)$: directivity of the microphone

Q: directivity factor of the loudspeaker $= \frac{p_{LS1m}^2}{\overline{p_{LS1m}^2}}$

A: total absorption in the room

The transfer function G_1 is given as superposition of the direct sound signal from loudspeaker 1 and the diffuse field contribution from all loudspeakers. The direct sound corresponds to

$$p_{M,direct}^2 = p_M^2 \nu RW_{LS}(\gamma) \frac{1}{d_{LSM}^2} RW_M(\beta) \tag{7.1}$$

The diffuse field contribution $p_{diffuse}^2$ from one loudspeaker with radiated sound power W is

$$p^2_{diffuse} = \frac{4\rho_0 cW}{A} \tag{7.2}$$

The radiated sound power W can be expressed by sound pressure square in 1 m distance and the directivity factor Q:

$$W = \frac{p^2_{LS1m} 4\pi}{\rho_0 cQ} \tag{7.3}$$

By insertion of (7.3) in (7.2), the diffuse field contribution of one loudspeaker is found as

$$p^2_{M,diffuse} = p^2_M \nu \frac{16\pi}{AQ} \tag{7.4}$$

G_1 is then found as

$$G_1 = \nu RW_{LS}(\gamma) \frac{1}{d^2_{LSM}} RW_M(\beta) + \nu \frac{16\pi}{AQ} n \tag{7.5}$$

The condition for G_1 for a safety margin of 10 dB to the point of instability is

$$G_1 < 0.1 \tag{7.6}$$

The transfer function G_2 is found analogously as superposition of direct and diffuse sound:

$$G_2 = \nu \frac{1}{d^2_{LSE}} + \nu \frac{16\pi}{AQ} n \tag{7.7}$$

G_3 finally describes the direct sound decrease from the source to the microphone:

$$G_3 = \frac{1}{d^2_{QM}} \tag{7.8}$$

The sound pressure square $\dot{p}^2_E$ produced by the source at the receiver without PA system is

$$\dot{p}^2_E = p^2_{Q1m} \left(\frac{1}{d^2_{QE}} + \frac{16\pi}{A} \right) \tag{7.9}$$

With the PA system in operation, the sound pressure square at the receiver becomes (ignoring the feed-back path):

$$\ddot{p}^2_E = p^2_{Q1m} G_3 G_2 + p^2_{Q1m} \left(\frac{1}{d^2_{QE}} + \frac{16\pi}{A} \right) \tag{7.10}$$

The pressure square amplification by the PA system is thus given as

$$\text{Gain} = \frac{\ddot{p}^2_E}{\dot{p}^2_E} = 1 + \frac{G_3 G_2}{\frac{1}{d^2_{QE}} + \frac{16\pi}{A}} \tag{7.11}$$

Maximal amplification is reached for $G_1 = 0.1$ (corresponding to 10 dB margin to the instability point). From this follows with Eq. 7.5

$$\nu = \frac{0.1}{RW_{LS}(\gamma) RW_M(\beta) \frac{1}{d^2_{LSM}} + \frac{16\pi}{AQ} n} \tag{7.12}$$

After insertion of (7.12) in (7.7) and (7.11) the maximal amplification is found as

$$\text{Gain}_{MAX} = 1 + \frac{0.1 \frac{1}{d^2_{QM}} \left(\frac{1}{d^2_{LSE}} + n \frac{16\pi}{AQ} \right)}{\left(\frac{1}{d^2_{QE}} + \frac{16\pi}{A} \right) \left(RW_{LS}(\gamma) RW_M(\beta) \frac{1}{d^2_{LSM}} + n \frac{16\pi}{AQ} \right)} \tag{7.13}$$

It should be noted that for speech transmission only the direct sound contribution is useful for speech intelligibility. This part is found by evaluating Eq. 7.13 with $A \to \infty$.

In the following, it is assumed that the PA system is well configured and produces significant amplification. In this case the following relations are valid:

- Gain$_{MAX} \gg 1$
- $\frac{1}{d^2_{LSE}} \gg n\frac{16\pi}{AQ}$ (the distance loudspeaker-receiver is smaller than the critical distance, meaning that the receiver is supplied mainly with direct sound)
- $n\frac{16\pi}{AQ} \gg RW_{LS}(\gamma)RW_M(\beta)\frac{1}{d^2_{LSM}}$ (feed-back is dominated by the diffuse field)

With the above assumptions it can be concluded that:

- Gain$_{MAX} \sim \frac{1}{d^2_{QM}} \rightarrow$ the maximal amplification is inversely proportional to the square of the distance of the speaker to the microphone
- Gain$_{MAX} \sim \frac{1}{n} \rightarrow$ the maximal amplification is inversely proportional to the number of loudspeakers
- Gain$_{MAX} \sim Q \rightarrow$ the maximal amplification is proportional to the directivity factor of the loudspeakers

7.1.3 Examples of sources with high directivity

As seen above high amplification can be reached with loudspeakers of high directivity only. This can be achieved by two strategies:

- horn speakers
- one- or two-dimensional arrangement of several speakers (with arrangement dimension $\approx$ wavelength

Horn loudspeakers

Horn loudspeakers achieve an improved efficiency due to better impedance matching. In addition they allow for a better control of directivity. Table 7.2 shows Q values and level reductions for $60°$ relative to the the main radiation direction for a typical mid-frequency horn with an opening area of 70 cm x 40 cm and a length of 40 cm.

	500 Hz	1 kHz	2 kHz	4 kHz
Q	14	12	14	12
$\Delta L(60°)$ [dB]	5	12	10	18

Table 7.2: Q values and level reductions ΔL at $60°$ relative to the main radiation direction for a typical mid-frequency horn loudspeaker.

Column speaker

Column speakers with many coherently driven chassis arranged in a line are a common solution to focus the radiated sound energy in one dimension [2]. Table 7.3 shows typical Q values and level reductions at $60°$ for a 1.3 m long column.

	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Q	8	11	18	22	18	45
$\Delta L(60°)$ [dB]	15	15	20	12	15	10

Table 7.3: Q values and level reductions ΔL at $60°$ relative to the main radiation direction for a 1.3 m long column speaker consisting of 8 chassis.

The focusing of column speakers depends strongly on frequency. Consequently the resulting amplitude response may be very uneven. The high directionality at high frequencies compared to the low frequencies can be reduced by dividing the column into shorter segments that are tilted relative to

[2] M. S. Ureda, Analysis of Loudspeaker Line Arrays, J. Audio Engineering Society, vol. 52, p.467-495 (2004).

each other (Figure 7.5). At low frequencies the directionality is still given by the total length while at higher frequencies the directionality is determined by the shorter segments.

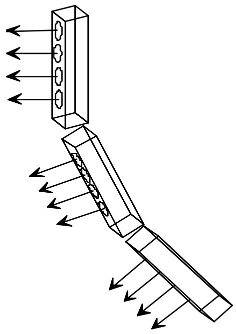


Figure 7.5: Splitting of a column speaker into several segments to reduce the strong frequency dependency of directionality.

A more versatile solution is to control the amplitude and phase of each chassis individually. Such an arrangement is called a line array. Within the limits given by the ratio of column length and wave length, almost any possible directivity can be realized [3] (of course only in one dimension). Of special interest is the possibility to steer the elevation of the beam independently of the physical orientation of the column. It is therefore possible to mount such columns flush to a wall which allows for highly aesthetic solutions. Furthermore the problem of wall reflections is eliminated which is common in case of inclined mounting (Figure 7.6). With suitable excitation it is even possible to produce two or more beams that point in different directions. Conventional columns focus at distance = ∞. Line arrays allow for an arbitrary focal point. By proper adjustment of this parameter, it possible to produce a sound field that varies only little with distance. This is a highly desirable effect as it reduces the danger of feed-back.

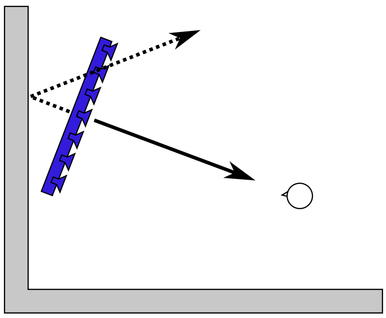


Figure 7.6: Conventional column speakers require tilted mounting to maximize radiation towards the audience. This leads to reflections at the vertical wall and produces a second unwanted beam.

7.1.4 Feed-back suppression

Wobbling

Feed-back can be significantly suppressed if the radiated loudspeaker signal is modulated in frequency. This results in a continuous alteration of the phase relation of the feed-back loop with the consequence that frequencies with positive feed-back at one time will experience negative feed-back the next moment. The idea of frequency modulation as feed-back suppressor dates back to 1928. The modulation was realized by mounting the loudspeaker and the microphone like a pendulum. In our times, Nishinomiya[4]

[3] M. Möser, Amplituden- und Phasen-gesteuerte akustische Sendezeilen mit gleichmässiger Horizontal-Richtwirkung. Acustica, vol. 60, p.91-104, (1986).

[4] Nishinomiya, G. Improvement of Acoustic Feedback Stability of P.A. Systems by Warbling. 6th International Congress of Acoustics, 1968, E-3-4, pp.E-93.

optimized the frequency modulation parameters for this application. He found a variation of 10 Hz with a modulation frequency of 4.5 Hz best suited. The improvement is about 5 dB, however a drawback is the audibility of the modulation at low frequencies.

Frequency shifting

The room transfer function between a source and a receiver shows isolated resonances at low frequencies. Towards higher frequencies the resonance density increases and becomes so high, that the transfer function gets a random character with a fast alteration of peaks and dips. By introducing a frequency shift before the signal is radiated by the loudspeaker, the uneven transfer function is smoothed to a certain extent as peaks may be mapped to dips and vice versa. In consequence an improved stability limit is reached. For practical applications a frequency shift of 5 Hz is typically used which results in a stability increase of about 6 dB.

Notch filter

For a given room and fixed microphone and loudspeaker positions, the number of frequencies that tend to positive feed-back is typically in the order of 10...20. It is therefore possible to suppress these frequencies by a relatively small set of notch filters. However if the microphone position is not fix, the feed-back frequencies vary. With adaptive filters these frequency shifts can be followed up to a certain extent.

Estimation and compensation for the feed-back loop

If the feed-back loop (G_1 in Figure 7.4) can be estimated with sufficient accuracy, it can be reproduced by a digital filter and thus compensated for. A method to do this with a subtraction filter [5] is shown in Figure 7.7. Hereby G_1' is the estimate of the feed-back loop G_1. By the subtraction operation it is achieved that signals that have passed the feed-back loop once will annulate. The principle of operation can be understood most easily by investigating a pulse like excitation signal. The pulse hits the microphone and runs through the filter G_1'. At the same time the pulse is radiated by the speaker. After the travel time from the speaker to the microphone (G_1), the feed-back portion of the pulse is captured again by the microphone. In the meantime the pulse has passed through filter G_1' as well. Both signal components annulate by the subtraction operation.

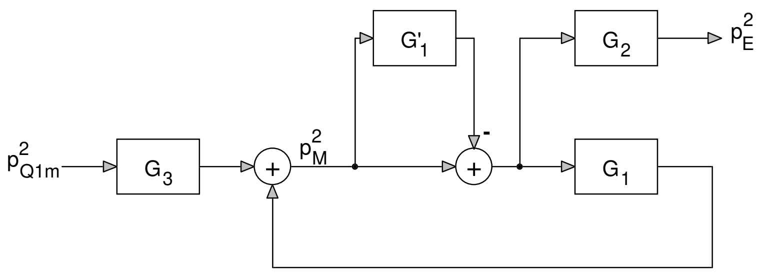


Figure 7.7: Block diagram of a PA system with a subtraction filter to suppress feed-back. G_1 represents the acoustic feed-back loop, G_1' is the estimate and electronic simulation of G_1.

The major difficulty is the accurate estimation of the feed-back loop. For fixed loudspeaker and microphone positions it may be possible to measure the system properties in advance. However it can't necessarily be assumed that such a system is time-invariant. The presence of audience for example or a change in temperature can alter the system transfer function. A way out is the continuous measurement of the system properties, for example with MLS at an inaudible level.

7.2 Speech intelligibility

The fundamental quantity to describe speech intelligibility is the intelligibility of syllables. It is determined by listening tests where test persons have to protocol the syllables they have understood.

[5] A. Goertz, Erprobung verschiedener Digitalfilter zur Unterdrückung der akustischen Rückkopplung in Beschallungsanlagen. Fortschritte der Akustik, DAGA 94, p.305-308 (1994).

The percentage of syllables that were understood correctly corresponds to the intelligibility of syllables. Good speech intelligibility requires a value of about 90%. With the redundancy inherent to speech, words and sentences can then be understood to 100%.

In rooms, speech intelligibility is influenced by the following parameters:

- signal/noise ratio
- reverberation time of the room
- ratio of direct to diffuse sound (determined by the distance source - receiver)
- early reflections (comb filter effects as a consequence of delays in the order of about 1 ms are detrimental)
- late reflections (echo effects for delay times above 50 ms)

For the objective measurement and prognosis of speech intelligibility, the following measures are in use:

- $\%AL_{cons}$[6] (percentage of lost consonants)
- STI (Speech Transmission Index or $RASTI$)
- $D50$ Deutlichkeit[7] (ratio of the early and thus beneficial sound energy to the total energy in the impulse response)

Investigations at a great many of rooms have shown that the three mentioned measures for speech intelligibility are highly correlated among each other. Consequently, formulas exist to transform one measure into an other [8]. For the objective evaluation of speech intelligibility an IEC standard exists [9].

7.2.1 Articulation Loss

Peutz found a semi-empirical relation between the percentage of lost consonants $\%AL_{cons}$ and acoustical and geometrical criteria of the room:

$$\%AL_{cons} = \frac{200 D^2 RT^2 N}{VQ} \qquad \text{for } D < 3.2 r_H$$
$$\%AL_{cons} = 9RT \qquad \text{for } D \geq 3.2 r_H$$

where
D: distance source - receiver
RT: reverberation time in the frequency range 500 Hz...2 kHz
V: room volume
Q: directivity of the source
N: ratio of the loudspeaker power that contributes to the diffuse field and the one that contributes to the direct sound (for a single loudspeaker $N = 1$)
A: total absorption
r_H: critical distance = $\sqrt{\frac{QA}{16\pi}}$

$\%AL_{cons}$ - values below 10 % correspond to a very good speech intelligibility, values between 10 and 15 % signify good speech intelligibility. With help of Eq. 7.14 a PA system can be designed for sufficiently high speech intelligibility. As a design target a value of $\%AL_{cons}$ of 15 % is usually chosen. For an overview of algorithms for the calculation of $\%AL_{cons}$ see [10].

[6] Peutz V.M.A. Articulation Loss of Consonants as a Criterion for Speech Transmission in a Room. Journal of the Audio Engineering Society, v.19 (1971).

[7] Ahnert, Reichardt, Grundlagen der Beschallungstechnik, Hirzel 1981.

[8] Bradley, J. S. Relationships among Measures of Speech Intelligibility in Rooms, Journal of the Audio Engineering Society, v. 46, p.396-405 (1998).

[9] IEC 60268-16:2011 Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index, 2011.

[10] Sylvio R. Bistafa, John S. Bradley, Revisiting Algorithms for Predicting the Articulation Loss of Consonants, Journal of the Audio Engineering Society, vol. 48, p. 531-544 (2000).

7.2.2 Speech Transmission Index

The *S*peech *T*ransmission *I*ndex STI is based on the modulation transfer function of the system source - room - receiver. The speech signal is simulated as noise that is slowly modulated in amplitude. The original modulation depth is reduced by effects that are detrimental for speech intelligibility such as reverberation and noise. The evaluation of the loss of modulation depth results in a measure for the speech intelligibility.

The STI is mostly used to assess existing systems by measurements. However in its original specification the measurement is very time consuming. A simplified and faster procedure, called $RASTI$ (*R*apid *S*peech *T*ransmission *I*ndex) was therefore introduced. Hereby the excitation signal is composed from 500 Hz and 2 kHz octave band noise. The levels are chosen to correspond to the typical source strength of a real human speaker. The noise signals are modulated in amplitude with 9 different frequencies between 0.5 and 12 Hz. At the receiver the corresponding reductions in modulation depth are evaluated and combined to a $RASTI$ value between 0 and 1. Values around 0.85 stand for very good speech intelligibility, 0.65 is good while values below 0.3 represent a bad intelligibility.

Alternatively, $RASTI$ and STI can be calculated from impulse response measurements. In this case $RASTI$ has no advantage over STI. Caution is advised if feed-back cancelers are part of the system as the impulse response measurement (e.g. MLS) may react very sensitive to nonlinear or non-stationary system properties.

7.2.3 Relation between Articulation Loss and RASTI

The quantity $\%AL_{cons}$ is mainly used for calculations while $RASTI$ and STI are suited for measurements. $\%AL_{cons}$ and $RASTI$ can be converted into each other with help of the equations 7.14 and 7.15.

$$\%AL_{cons} = 170.54e^{-5.42RASTI} \tag{7.14}$$

$$RASTI = 0.95 - 0.185\ln(\%AL_{cons}) \tag{7.15}$$

7.2.4 Deutlichkeit D50

The quantity Deutlichkeit $D50$ is based on the fact that signal contributions that arrive up to 50 ms after the direct sound are beneficial for speech intelligibility while the later part in the impulse response is detrimental. $D50$ is determined from an impulse response $h(t)$ between source and receiver:

$$D50 = \frac{\int_0^{50ms} h^2(t)\mathrm{d}t}{\int_0^{\infty} h^2(t)\mathrm{d}t} \times 100\% \tag{7.16}$$

$D50$ can be predicted easily with room acoustical simulation programs that provide energy impulse responses. Alternatively $D50$ can be determined from room impulse response measurements. A $D50$ value of 50 % corresponds to an intelligibility of syllables of about 90 %, which represents a very good speech intelligibility.

7.2.5 STI-PA

STI-PA is a new Speech Transmission Index variant. It can be understood as successor of $RASTI$. Similarly as with $RASTI$ an asynchronous operation of sender and receiver is possible. However $STI-PA$ covers a wider frequency range and uses more different modulation frequencies. Modern measurement instruments allow for numerical adaption of the results for different levels of background-noise.

7.3 Localization

Besides the amplification of the source signal, a high-end sound reinforcement system tries not the affect the localization of the original source. In the ideal case the PA system is even not heard consciously. This is possible thanks to the precedence effect of the ear. As long as the reinforced signal by the PA system is not more than 10 dB above the original signal level and if the delay of the loudspeaker signal is chosen appropriately, the ear will localize the original source. The reinforcement system increases the level and the intelligibility but as such is not audible. The correct delay of the loudspeaker signal can be achieved by suitable placement or by introducing a digital delay.

7.4 Types of PA systems

Depending on the local distribution of the loudspeakers, different types of PA systems are distinguished.

7.4.1 Centralized PA system

Centralized PA systems use one or several loudspeakers that are all located at the same position. In rooms with an axial orientation, horn speakers are often used. In non-directional rooms such as ice hockey stadiums, clusters of loudspeakers may be installed at a center position.

Centralized PA systems usually allow for a homogeneous sound pressure distribution in the audience region. In addition the localization is often good, if the loudspeakers are installed in the vicinity of the position of the original source. However if the reverberation times are too large (above about 2 seconds) centralized PA systems usually fail, as the distances from the speakers to the receivers are generally large.

7.4.2 Distributed PA system

In very reverberant rooms or rooms with a low ceiling, often distributed PA systems with several loudspeakers placed at different locations have to be employed. Problems may arise in transition zones that are supplied by different speakers. Here interference effects can play a role and may produce local sound pressure variations. As the speakers are relatively close to the listeners, the level distribution is in general not that homogeneous as in case of a centralized system. Localization can often be improved by inserting a delay to ensure that the direct sound of the original source arrives at the listeners ears earlier than the reinforced signal from the loudspeakers. In its extreme form, a distributed system provides a loudspeaker for each listener.

7.4.3 100 V systems

Analogous to the idea of high voltage transmission lines, some PA system use the so called 100 V technique. To minimize the necessary current and thus the necessary cross sectional area of the cables, the loudspeaker signal is transformed at the output of the power amplifier to a higher voltage of typically 100 V. At the loudspeaker end of the cable the voltage is then transformed down to the original voltage level appropriate for the speaker. Besides 100 V systems, voltage levels of 70 and 25 V are common. Compared to conventional cabling, the 100 V systems have the following advantages and disadvantages:

advantages:

- cables with smaller diameters can be used
- installations with a common central power amplifier unit and long cables to the speakers are possible

disadvantages:

- good sound quality requires expensive transformers
- the transformers increase the source impedance as seen by the loudspeaker

7.5 Tools for the simulation of PA systems

For the design of a PA system (selection, placement and orientation of appropriate loudspeakers) a variety of different software tools are available. They model the radiation characteristics of the loudspeaker systems and simulate the acoustics of the room. For a systematic collection of data and unified description of the loudspeaker properties, the CLF (*C*ommon *L*oudspeaker *F*ormat) was proposed. Meanwhile CLF is supported by many manufacturers and software developers. There is a free viewer for loudspeaker CLF data [11].

7.6 Electroacoustical adjustment of reverberation

Many today's rooms are used in a multi-functional way. As varying usage usually requires different reverberation times, the room acoustical properties have to be made adjustable. This can be reached by varying the absorption properties of the room surfaces. An even more elegant solution is to increase reverberation by electroacoustical means. Experience shows that the initial reverberation time can be prolonged by about 100 %.

Electroacoustical prolongation of reverberation can be realized in two different ways:

In-Line systems These systems pick up the source signal by microphones on stage, add reverberation by means of a reverberation generator and finally radiate the signals by several loudspeakers to the audience region.

Feed-back systems, Assisted Resonance Hereby autonomous microphone-loudspeaker units are installed on the room surfaces. Each loudspeaker radiates the digitally processed signal picked up by his own microphone. A major difficulty is the suppression of feed-back between loudspeaker and microphone. With one unit, a prolongation of the reverberation time of about 1...2 % can be obtained without relevant signal coloring. For a doubling of the reverberation time, typically about 25 to 40 units are needed.

[11] http://www.clfgroup.org/index.htm

Chapter 8

Audio signal processing

8.1 Effects

In electronic music, effects to modify the original audio signal are often used. Motivations can be to reduce the audibility of unwanted signal components or to create new sounds. Modern effect processors work digitally [1].

8.1.1 Equalization

Equalizers allow for an adjustment of the amplitude response of a signal. A typical field of application is the correction of the frequency response of a loudspeaker.

Parametric Equalizers

Parametric equalizers are composed of a series of bandpass and notch filters with freely adjustable filter properties such as center frequency, bandwidth and amplification of attenuation.

Graphic Equalizers

Graphic equalizers consist of a series of bandpass filters with fixed center frequencies and constant relative bandwidths. Usually they correspond to the standard octave or third-octave filter series. The user can adjust the amplification or attenuation in each band. The bandwidth and slope of the amplitude response is adjusted in such a way that two filters next to each other produce a flat response if they are set at identical level. The amplification and attenuation range is usually between +/- 10 and +/- 20 dB.

8.1.2 Reverbs

Reverbs are systems that add artificial reverberation to an audio signal. They are used e.g. to make single instruments sound more reverberant. In a multi-channel recording this is then perceived as if the instrument would be located more distant. An other application is to add reverberation to a recording that sounds too dry.

Today's computer power allows for real-time reverberation algorithms that perform a convolution with a specific room impulse response. Of course the corresponding calculation is done in the frequency domain.

8.1.3 Compressors / Limiters

Compressors and limiters are used to reduce the dynamics of an audio signal. As loudness of an audio signal is a very strong psychoacoustical attribute, sound engineers use compressors in many applications. This is of special importance if the audio signal has to attract the attention of people. Compressed audio signals are beneficial in listening situations with high environmental noise, e.g. in cars. Limiters ore often found in digital sound recorders. They prevent clipping if the audio signal

[1] Jon Dattorro, Effect Design Part 1, Journal of the Audio Engineering Society, vol. 45, p.660-684 (1997).

exceeds the maximum range of the A/D converter.

The principle of construction of a compressor or a limiter is shown in Figure 8.1. The key element is a voltage controlled amplifier (VCA) whereby the steering voltage depends on the audio signal itself. Whenever the audio signal level exceeds a certain limit, the amplification is lowered accordingly. In case of a limiter, the amplification is lowered in such a way that the output level remains constant.

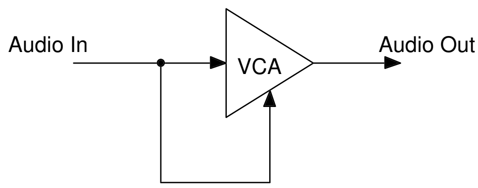


Figure 8.1: Principle of operation of a compressor / limiter.

Figure 8.2 shows the dependency of the VCA amplification from the input level or as an alternative representation the output level as a function of the input level for different compression ratios.

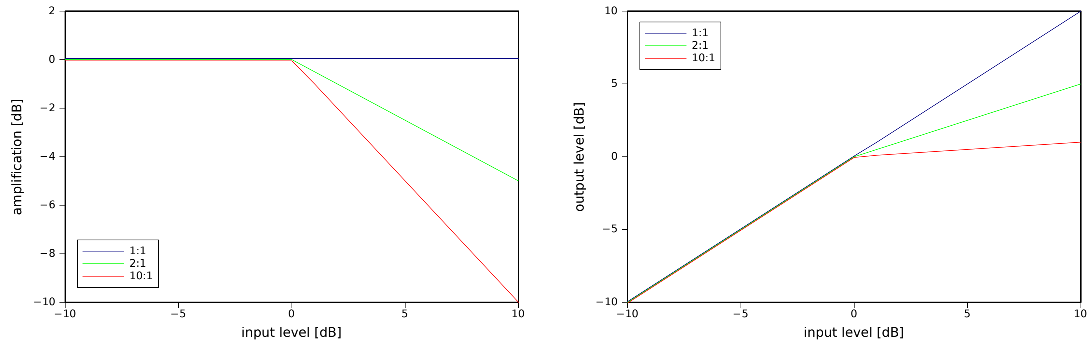


Figure 8.2: Compressor / limiter characteristics. *Left:* compressor VCA amplification in dependency of the level of the input signal for different compression rates and a threshold of 0 dB. *Right:* compressor output level in function of the level of the input signal.

To obtain the VCA steering voltage, the audio input signal has to be transformed into a time dependent squared average with a suitable time constant. Usually two separate values are used for the positive and negative slopes. For a signal increase the time constant is typically in the order of a few milliseconds. For a signal decrease, the time constant is in the range of 10 to 3000 milliseconds. The choice of these time time constants is crucial. If the positive slope time constant is too small, start transients of musical instruments may be distorted. If the negative slope time constants are too small, a modulation of weak signal components by dominant components may occur (pumping effect). On the other hand, a too large time constant prevents the necessary reaction after a loud signal part followed by a silent part.

As a last step before release, many music productions run through a so called *finalizer* process. These processors introduce limiting and compression in separate frequency bands to increase the perceived loudness. According to the formula: "louder sounds better", these practices have already been called *loudness war* [2]. It has been found in pop music from 1980 to 2000 that the typical average rms level increased by almost 20 dB. This increase in rms is accompanied by a corresponding reduction in dynamic range. There is increasing concern that this loss in dynamic range deteriorates the quality of modern music productions.

8.1.4 Noise Gates

Noise Gates are used to suppress unwanted background noise during silent phases. Therefore a threshold is defined, below which the signal is attenuated by a certain factor. Above the threshold the signals

[2]E. Vickers, The Loudness War: Background, Speculation and Recommendations, AES 129th Convention, San Francisco, 2010.

are transmitted unaltered. Figure 8.3 shows the relation between input and output of a noise gate. As with compressors and limiters the selection of an appropriate time constant is crucial.

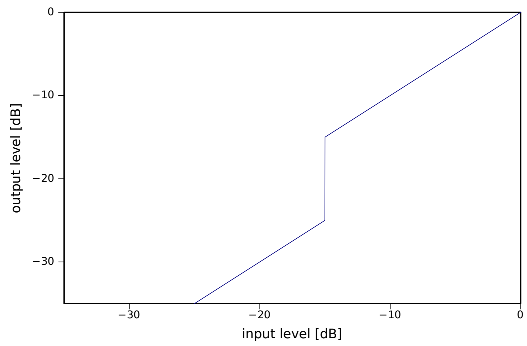


Figure 8.3: Output signal level dependency from the input signal level for a noise gate. The parameters threshold (here -15 dB) and step (here 10 dB) can be chosen freely.

8.1.5 Flanger

The flanging effect is created by summation of the original signal with a delayed copy, where the delay time varies slowly (Figure 8.4). The effect modifies the signal spectrum according to a comb filter with varying frequency positions of the minima and maxima. The delay times vary typically between 0 and 5 ms, the modulation frequencies lie in the range of a few tenth of a Hz.

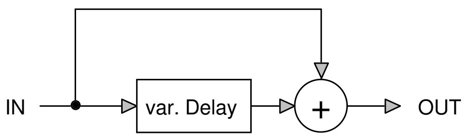


Figure 8.4: Block diagram of a flanger.

8.1.6 Chorus

If several flangers are put in parallel but with independently varying delay times in the range of about 10. . .25 ms, a chorus effect results that simulates the duplication of the original instrument or voice.

8.1.7 Phaser

The phasing effect is produced by summation of the original signal and a copy of it that is shifted in phase. The phase shift is varied over time. Similarly to the flanging effect, amplifications and attenuations in the spectrum occur. However with the phaser, the frequencies for constructive and destructive interference are not necessarily harmonics of a fundamental component.

8.1.8 Leslie

The Leslie effect introduces a Doppler frequency shift and an amplitude modulation by help of a rotating loudspeaker. If the Leslie cabinet is used in a room, the spacial distribution of the sound field varies with rotation due to varying distribution of the reflections.

8.1.9 Pitch-Shift and time-scaling

An important modification of an audio signal is the shift of its pitch without changing the time scale (*pitch-shift*). An equivalent task is to modify the signal duration without changing the pitch (*time-scaling*). The problem can be solved in the time domain [3] or in the frequency domain. In general

[3] R. Tan, A. Lin, A Time-Scale Modification Algorithm Based on the Subband Time-Domain Technique for Broad-Band Signal Applications, Journal of the Audio Engineering Society, vol. 48, p.437-449 (2000).

algorithms that work in the frequency domain are computationally more demanding but at the benefit of larger possible scaling factors.

8.2 Audio watermarking

8.2.1 Introduction

With *audio watermarking*[4],[5] additional information that has nothing to do with the audio data is hidden in the audio stream. This information can be used e.g.

- for copyright management
- to proof that an audio signal has not been modified
- to label the originator
- to monitor the usage by television or radio

Naturally, the music industry holds back information about implementations of audio watermarking. However it can be assumed that at least in parts, watermarking techniques are applied today.

The concept of hidden messages has a long tradition in history. A famous example is the usage of invisible ink. For an outsider, the message is not recognizable and the corresponding document of no use. On the other hand, a receiver that knows about the secret can treat the document in a suitable manner and then discover the written message.

8.2.2 Requirements for audio watermarking

The insertion of watermarking information has to consider the following requirements:

audio quality, transparency the watermarking process shall not seriously influence the quality of the audio signal. There is no clear limit regarding the allowable distortions. In the ideal case the watermarked signal is indistinguishable from the original. In this case that watermarking process can be regarded as totally transparent.

capacity the information density of the watermark shall be high.

robustness, reliability watermarks shall not be easily detectable for an attacker. In addition, the watermark shall be insensitive to possible manipulations of the audio signal. In the best case, the watermark information is preserved as long as the manipulated audio signal is of any value. Depending on the purpose of the audio watermark, a high reliability is essential. Such a sensitive application is the allowance of access based on a watermark.

8.2.3 Methods for audio watermarking

Although the watermarking methods already used commercially are unknown, a few fundamental principles can be deduced from literature. Hereby four categories can be distinguished.

The *time-domain methods* manipulate the signal in the time domain and use attributes that can be evaluated by an analysis of the time course of the signal. With *frequency-domain methods*, attributes are used that can be detected in the frequency domain. *wavelet-domain methods* are based on a signal representation as wavelets and consequently manipulate the corresponding coefficients. The *code-domain methods* finally are applicable in the context of MPEG coding and modify directly the code coefficients.

[4] Changsheng et al., Applications of Digital Watermarking Technology in Audio Signals, Journal of the Audio Engineering Society, vol. 47, p.805-812 (1999).

[5] R. Walker, Audio Watermarking, BBC R&D White Paper, WHP 057 (2004).

Time-domain methods

Superposition of 'noise'

In its simplest form, an audio watermark can be added to an audio signal by superposition of an additional signal that carries the information. However the level of this additional signal has to be sufficiently low in order not to seriously impair the original audio signal. From a perceptional point of view it is beneficial if the additional signal has a noise-like character.

If a 6 dB higher noise floor in a digital signal is acceptable, the *least significant bit* of each sample can be modulated according to the watermark information. This method yields a very high information density but is very fragile. Already simplest signal manipulations destroy the watermark completely.

Superposition of 'echoes'

According to the precedence effect, the human hearing is insensitive to echoes of that arrive shortly after the direct sound. By superposition of an artificial echo, watermark information can thus be added to an audio signal without influence on audible quality. The echoes have typical levels of -20...-40 dB relative to the direct sound. A coding method might be that an echo at 10 ms corresponds to a Bit value 0 while an echo at 30 ms signifies a Bit value 1. Detection is performed by an autocorrelation analysis. The superposition of echoes is a very robust method, however the watermark information density is relatively low.

Modulation of amplitude

Watermark information can be introduced by modifying the envelope of the audio signal. One application is to set very short sections of the audio signal to 0. Alternatively, the envelope can be manipulated on a larger time scale. The modulation of amplitude yields robust watermarks, however the information density is very small.

Frequency domain methods

Modification of the phase

The human auditory system is insensitive to phase information. With this in mind a watermark strategy can be developed that modifies explicitly the phase spectrum of an audio signal. For that purpose the audio signal is Fourier transformed on a frame-by-frame basis. In this representation the phase is modified according to the watermark information and a predefined coding strategy. By inverse Fourier transformation the signal is then retranslated into the time domain.

Superposition of a modulated tone

Watermark information can be added to an audio signal by superposition of one or several pure tone components. To make sure that these components are inaudible, their level is adjusted in such a way that they are masked by the original audio signal.

Suppression of frequency bands (notch filter)

Watermark information can be introduced by appropriate suppression of narrow frequency bands of the audio signal. These notch filters are usually realized in the frequency domain. In a spectral representation, one or a few frequency lines are removed. If this spectral gap is sufficiently narrow, the perceptual quality of the audio signal is unaltered. The disadvantage of this method is the rather low information density and the fact that the watermark is easily detectable.

Wavelet-domain methods

By application of a wavelet transformation, an audio signal can be translated into the wavelet-domain. In this domain the corresponding signal coefficients can be modified according to the watermark information [6].

[6] I. Daubechies, Orthonormal bases of compactly supported wavelets, Comm. Pure & Appl. Math. 41, pp. 909-996 (1988)

Code-domain methods

In audio signals that are already coded in a format such as e.g. MPEG or AC3, the watermark information can be added directly as a suitable bit stream.

8.2.4 Tests of audio watermarking

To test the robustness of watermarks, different classes of attacks (signal manipulations) are used. A well known collection is the StirMark Benchmark for Audio (SMBA) [7]. Some of the manipulations are listed in Table 8.1.

attack	description
Compressor	dynamics compression
Highpass	high-pass filtering
Lowpass	low-pass filtering
Echo	addition of an echo
Reverb	addition of reverberation
Flanger	addition of an echo with variable delay
Pitchshifter	pitch variation without influence on the time scale
Timestretcher	time stretching without influence on pitch
AddNoise	addition of pink noise
AddSinus	addition of a pure tone
CutSamples	deleting every x-th sample
CopySamples	duplicating every x-th sample
Exchange	exchanging neighbor samples for all samples
LSBZero	the least significant bit of each sample is set to 0
Normalize	amplification up to the maximal possible value
Resampling	changing the sampling rate
Smooth	averaging over neighbor samples
ZeroCross	samples below a certain threshold are set to 0
ZeroRemove	elimination of all samples with value 0

Table 8.1: Some of the attacks listed in the StirMark Benchmark for Audio.

[7] M. Steinebach et al., StirMark Benchmark: Audio watermarking attacks, Int. Conference on Information Technology: Coding and Computing (2001)

Chapter 9

Appendix

9.1 dB conventions in electroacoustics

name	type	definition	reference value
dBm	power ratio	$dBm = 10 \log \frac{W}{W_0}$	$W_0 = 1.0 \times 10^{-3}$ Watt
dBu	voltage ratio	$dBu = 20 \log \frac{U}{U_0}$	$U_0 = 0.775$ Volt
dBW	power ratio	$dBW = 10 \log \frac{W}{W_0}$	$W_0 = 1.0$ Watt
dBV	voltage ratio	$dBV = 20 \log \frac{U}{U_0}$	$U_0 = 1.0$ Volt

Table 9.1: dB conventions in electroacoustics.

dBm and dBu are related to each other with respect to a 600 Ohm line. The voltage reference $U_0 = 0.775$ Volt corresponds to the voltage that produces a power of $W_0 = 1.0 \times 10^{-3}$ Watt in a resistance of 600 Ohm.

Index

100 V-systems, 93
5.1 Surround Format, 71

A-format, 37
AB stereophony, 69
ac bias, 60
Adsorption, 44
AES/EBU interface, 82
Ambisonics, 76
analog tape recorder, 59
Articulation Loss, 91
audio data compression, 64
audio watermarking, 98

B-format, 37, 76
bandpass cabinet, 49
bass reflex cabinet, 49
Berliner, 58
Blumlein, 68
Blumlein stereophony, 69

cardioid microphones, 32
chorus, 97
CIRC, 63
clf format, 94
compact disc, 63
compliance acoustical, 5, 8
compliance equivalent volume, 48
compressors, 95
critical distance, 70
crossover networks, 53
cumulative decay spectrum, 55

D50, 92
dc bias, 60
DC resistance, 48
decay spectrum, 55
Deutlichkeit, 92
differential signalling, 82
distortion factor, 81
Dolby, 61
Dolby A, 62
Dolby SR, 62
Dolby Surround Digital, 75
Dolby-B, 61
dual conversion, 17
DVD Audio, 64

echo chamber, 73
Edison, 58
Equalizer, 95

feedback, 85
figure of eight microphone, 32
flanger, 97
Force factor, 48
force source, 3
free-field microphone, 26
friction mechanical, 2

gyrator, 16

hall plate, 73
Harddisk Recorder, 64
harmonic distortion, 81
headphones, 72
horn loudspeaker, 50
horn loudspeakers, 88

intensity stereophony, 68
intermodulation distortion, 81

Jecklin disc, 70

Leslie, 97
levers, 3
limiters, 95
line array, 89
links, 2
lossless compression, 65
lossy compression, 65
loudspeaker cabinet, 44
loudspeaker directivity, 40
loudspeaker electrodynamic, 39, 42
loudspeaker electrostatic, 52
loudspeaker nonlinearities, 44

mass acoustical, 5, 8
mass mechanical, 1, 7
Microflown, 36
microphone condenser, 20
microphone dynamic, 27
microphone electrostatic, 20
microphones, 20
mixed stereophony, 68
MPEG-1, 66
MS stereophony, 68
mufflers, 11

Noise Gates, 96

omnidirectional microphone, 31
ORTF arrangement, 69

passive membrane cabinet, 49
Perceptual Coding, 65
Phantom powering, 26
phaser, 97
pitch-shift, 97
pressure microphone, 26

Q factor electrical, 48
Q factor mechanical, 48

radiation impedance, 40
remanence, 60
resistance acoustical, 6, 9
resistance mechanical, 8
reverbs, 95

SMBA: Stirmark Benchmark for Audio, 100
soundfield microphone, 37
SPDIF, 83
speech intelligibility, 90
Speech Transmission Index, 92
spring mechanical, 1, 8
spring pendulum, 3, 11
spring reverb, 73
stereo basis, 72
stereo recordings, 68
STI, 92
STI-PA, 92
Super Audio CD, 64
surround sound, 71, 74

T-powering, 26
Thiele-Small parameters, 47
time-of-arrival stereophony, 68
time-scaling, 97
TOSLINK, 83
transmission line, 13

velocity sources, 3
vibration dampers, 12
vinyl records, 58

wave field synthesis, 75

XLR, 82
XY stereophony, 68