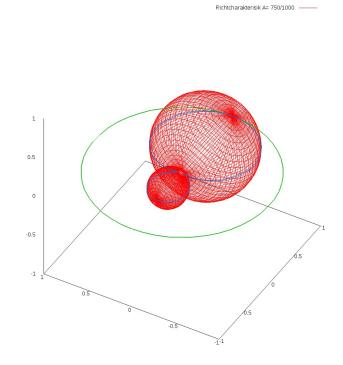
Introduction to Audio Technology



Wolfram Wagner

April 19, 2021

License: Creative Commons, CC by-sa, Version 4 An up to date version is distributed at http://www.baggy-trousers.de/AudioTechnology.pdf A german version is distributed at http://www.baggy-trousers.de/Tontechnik.pdf The author is grateful for remarks and improvements. E-Mail: mrthehowlingwolf@gmail.com

Contents

| 1 | Ove | rview | | 11 |
|---|------|----------------|--|-----------------|
| 2 | The | Sound | l Field | 13 |
| | 2.1 | Oscilla | ations and Waves | 13 |
| | | 2.1.1 | Oscillations | 13 |
| | | 2.1.2 | Waves | 13 |
| | 2.2 | Level, | dB | 13 |
| | 2.3 | Sine W | Vave, Fourier Transformation | |
| | | 2.3.1 | Sine Wave | 16 |
| | | 2.3.2 | Fourier Transformation | 16 |
| | 2.4 | The P | lane Sound Wave | 18 |
| | 2.5 | The S | pherical Sound Wave | 18 |
| | 2.6 | Interfe | erence | 18 |
| | 2.7 | Plane | Reflection, Comb Filter | 19 |
| | 2.8 | The R | | 21 |
| | | 2.8.1 | Room Modes | 21 |
| | | 2.8.2 | Reverberation, Diffuse Sound Field | 21 |
| | 2.9 | Size of | f Sound Sources | 23 |
| | 2.10 | Size of | f Microphones | 23 |
| | | 2.10.1 | Interference | 23 |
| | | 2.10.2 | Reflection | 23 |
| 3 | Hea | ring | | 25 |
| Ū | 3.1 | 0 | ar | - |
| | 3.2 | | of Auditory Perception | |
| | 0.2 | 3.2.1 | Curves of Equal Loudness | |
| | | 3.2.2 | Masking | |
| | 3.3 | | zation | |
| | | 3.3.1 | Angle to the Mid Plane | |
| | | 3.3.2 | Angle in the Symmetry Plane: Front, Back, Above, Below | |
| | | 3.3.3 | Distance | |
| | 3.4 | Effects | s to the Subconscious | 28 |
| 4 | Tho | Micro | nhono | 29 |
| - | 4.1 | | $\widehat{\operatorname{Prone}}$ | 29 29 |
| | 4.1 | 4.1.1 | Sensitivity | $\frac{29}{30}$ |
| | | 4.1.1 | Frequency Response | 30 |
| | | 4.1.2 | Directional Characteristic | $30 \\ 31$ |
| | | 4.1.3 | Noise Level | 33 |
| | | 4.1.4 4.1.5 | Maximum Sound Pressure Level | зэ 35 |
| | | 4.1.5 4.1.6 | Output Impedance | $\frac{35}{35}$ |
| | | -1.1.U | | 00 |

| | 4.2 | Directi | onal Characteristics | 36 |
|---|----------|----------------|--|------------|
| | | | Pressure Transducer - Spherical (Omnidirectional) Characteristic . | 36 |
| | | | Pressure Gradient Transducer - Figure of Eight, Proximity Effect . | 37 |
| | | 4.2.3 | Combinations - Cardioid, Super cardioid, Hyper cardioid, Broad | |
| | | | Cardioid \ldots | 44 |
| | | 4.2.4 | Combination with Interference Effects: Shotgun, Arrays | 53 |
| | | 4.2.5 | Geometrical Acoustics: Reflection, Refraction | 54 |
| | 4.3 | | bles of Transducers | 54 |
| | 1.0 | 4.3.1 | The (LF-)Condenser Microphone | 55 |
| | | 4.3.2 | Microphones Using Electromagnetic Induction | 59 |
| | | | Special Microphones | 64 |
| | | 4.3.4 | Active Transducers | 65 |
| | | 4.0.4 | | 05 |
| 5 | Mul | tiple M | icrophones/Signals, Stereophony | 67 |
| - | 5.1 | | bhony - not necessarily exactly two channels | 67 |
| | 5.2 | | hannel Stereophony | 67 |
| | 0.2 | | Reproduction | 67 |
| | | | Level Stereophony | 67 |
| | | 5.2.2 | Time Difference Stereophony | 71 |
| | | 5.2.3 5.2.4 | Combinations of Level and Time Difference | 71 |
| | | 5.2.4 5.2.5 | Alternative Microphone Assemblies | $72 \\ 73$ |
| | | | Main Microphones and Support Microphones | 73 74 |
| | | 5.2.0 5.2.7 | Ear Related Signals | 74 75 |
| | 5.3 | | lized Stereo: Surround | 75 76 |
| | 0.5 | | | |
| | | 5.3.1 5.2.0 | 5.1 | 76 |
| | | 5.3.2 | More Channels | 80 |
| | | 5.3.3 | Virtual Sound Sources with Direction Information | 81 |
| | | 5.3.4 | Ambisonics | 81 |
| | | 5.3.5 | Wavefield Synthesis | 82 |
| 6 | Flor | trical T | ransmission | 83 |
| U | 6.1 | | e Adaption | 83 |
| | 6.2 | | netrical transmission | 83 |
| | 6.3 | | etrical Transmission | 84 |
| | 6.4 | | , Connectors | 85 |
| | 6.5 | , | I-Box | 86 |
| | 0.0 | | Passive DI-Box | 86 |
| | | | Active DI-Box | 86 |
| | 6.6 | | | 86 |
| | | | m Power | 87 |
| | 6.7 | 0 | to Digital Conversion and Back | |
| | 6.8 | - | Transmission | 88 |
| | | 6.8.1 | S/PDIF | 88 |
| | | | AES/EBU | 89 |
| | | 6.8.3 | AES42 | 89 |
| | | | Integration into Video Signals, HDMI, DisplayPort | 89 |
| | <u> </u> | 6.8.5 | USB | 89 |
| | 6.9 | | ss Transmission | 90 |
| | | 6.9.1 | Analog, FM | 90 |
| | | 6.9.2 | Digital | 90 |

| | 6.10 | Digital | Storage - Formats | . 90 |
|---|-----------------|----------------|--|---------|
| | | | Compression - Lossless | |
| | 6.12 | Digital | Compression - With Losses | . 91 |
| 7 | Elee | twicel e | and Digital Dragoning of Signals | 93 |
| 7 | Elec 7.1 | | Ind Digital Processing of Signals | |
| | 1.1 | 7.1.1 | Microphone Preamplifier | |
| | | 7.1.1 7.1.2 | Power Amplifier | |
| | 7.2 | - | atic Level Control: Limiter, Expander, Gate | |
| | 1.2 | 7.2.1 | Compressor | |
| | | 7.2.1 | | - |
| | | 7.2.3 | Expander | |
| | | 7.2.4 | Gate | |
| | 7.3 | | ζ | |
| | 7.4 | - | | |
| | 7.5 | | zer | |
| | 1.0 | 7.5.1 | Shelving Filter | |
| | | 7.5.2 | Graphical Equalizer | |
| | | 7.5.3 | Parametric Equalizer | |
| | 7.6 | | Manipulation Effects | |
| | | 7.6.1 | Delay, Echo | |
| | | 7.6.2 | Reverb, Room, Early Reflection, Delayed Onset, Plate, Spring | |
| | 7.7 | Modula | | |
| | | 7.7.1 | Tremolo | |
| | | 7.7.2 | Vibrato | |
| | | 7.7.3 | Phasing | |
| | | 7.7.4 | Flanging | |
| | | 7.7.5 | Chorus | |
| | | 7.7.6 | Rotary | . 102 |
| | 7.8 | Freque | ncy Change | |
| | | 7.8.1 | Harmonizer | . 102 |
| | | 7.8.2 | Autotune | . 103 |
| | | 7.8.3 | Double Tracking | |
| | 7.9 | Nonlin | ear Characteristic: Saturation, Distortion | . 103 |
| | 7.10 | Analog | g Processing | . 104 |
| | 7.11 | | Processing | |
| | | 7.11.1 | Finite Impulse Response, FIR | . 104 |
| | | 7.11.2 | Infinite Impulse Response, IIR | . 105 |
| | 7.12 | | ms for Audio Processing | |
| | | | Wave Editors | |
| | | 7.12.2 | Digital Audio Workstation, DAW | . 106 |
| 8 | Ren | roducti | on | 107 |
| 0 | 8.1 | | \mathbf{D} eakers | |
| | 0.1 | 8.1.1 | Quality of Reproduction | |
| | | 8.1.2 | Types of Transducers | |
| | | 8.1.3 | Enclosures | |
| | | 8.1.4 | Sound Guides | |
| | | 8.1.5 | Passive Loudspeakers and Crossover Networks | |
| | | 0.1.0 | | • • • • |

| | | 8.1.6 | Active Loudspeakers and Crossovers | . 117 |
|----|------|----------|--|-------|
| | | 8.1.7 | DSP Based Loudspeakers | . 118 |
| | 8.2 | Room | | . 118 |
| | 8.3 | Headp | hones | . 119 |
| 9 | Prac | ctical R | Recording | 121 |
| | 9.1 | Requir | rements | |
| | | 9.1.1 | Artists, Music/Text, Instruments | . 121 |
| | | 9.1.2 | Together, One by One? | . 121 |
| | | 9.1.3 | Psychology of the Recording | |
| | | 9.1.4 | Perfection - Genie - "Magic" | |
| | | 9.1.5 | Live - Studio | |
| | 9.2 | - | phone Technique | |
| | | 9.2.1 | Close Miking | |
| | | 9.2.2 | Multiple Microphones | |
| | | 9.2.3 | Some Concrete Sound Sources | |
| | | 9.2.4 | The Large Distributed Sound Source - Orchestra, Choir, Organ . | |
| | | 9.2.5 | The Drum Set | |
| | 9.3 | | lix of the Recording | |
| | | 9.3.1 | The Goal: A Simulated Perfect Event, Not a Documentation | |
| | | 9.3.2 | Cleaning Up | |
| | | 9.3.3 | Balance of the Sound Sources | |
| | | 9.3.4 | The Sound of the Individual Tracks | |
| | | 9.3.5 | Localization | |
| | | 9.3.6 | Room Information | |
| | | 9.3.7 | Combination of Effects: Creating a Depth Scaling | |
| | | 9.3.8 | Focus | |
| | | 9.3.9 | Dynamics | . 134 |
| 10 | Prac | ctical S | ound Reinforcement | 137 |
| | | | Time | |
| | 10.2 | The Fe | eedback | . 137 |
| | | | Limitation of Volume or Amplification | |
| | | | Useful Position and Angle of Directional Loudspeakers, Isolation . | |
| | | | Close Miking | |
| | | | Directional Microphones, at Suitable Places and Angles | |
| | | | Take Care when Boosting Frequency Regions | |
| | | | DI Boxes, Wherever Suitable | |
| | | | Transducers Without Strong Resonances | |
| | | | Little Reverb in the Monitor | |
| | | | Careful Usage of Compressors & Co | |
| | | |) Equalizers | |
| | | | l Feedback Killers | |
| | | | ions, Wind and Pop Sounds | |
| | | | er Conditions | |
| | 10.5 | | bound for the Audience | |
| | | | Minimal Solution, no Bass | |
| | | | Full Spectrum, Subwoofer | |
| | | 10.5.3 | Classical Stack | . 141 |

| | 10.5.4 | Vertical Array | | 141 |
|------|--------|----------------------|------|-----|
| 10.6 | Monite | Dr | | 142 |
| | 10.6.1 | Monitor Loudspeakers | | 142 |
| | 10.6.2 | In Ear | | 142 |

List of Figures

| 2.1 2.2 | Comb filter, distance 4.3 mm | 19 19 |
|--------------|--|------------|
| $2.3 \\ 2.4$ | Comb filter, distance 100 mm | 20 22 |
| 3.1 | Schematic cross section of ear, take from Wikipedia, License CC by-sa Version 4, Created by Geo-Science-International, accessed 2016-06-07, translations by author | 26 |
| 3.2 | Curves of equal loudness, take from Wikipedia accessed 2016-06-07, trans- lated by author | 27 |
| 4.1 | Example of a frequency curve, from [18], page 253. It is a measurement of a ribbon microphone | 30 |
| 4.2 | Example for a plot of a directional charactersistic, from [18], page 248 | 32 |
| 4.3 | Maximum of sensitivity in front | 32 |
| 4.4 | Maximum of sensitivity on the side | 33 |
| 4.5 | Omnidirectional | 36 |
| 4.6 | Figure Of Eight | 38 |
| 4.7 | Gradient receiver frequency response, linear and logarithmic | 40 |
| 4.8 | Proximity effect | 45 |
| 4.9 | Random Energy Efficiency REE , its reciprocal value Directivity Factor | |
| | Gamma and the Distance Factor DSF plotted versus parameter A | 47 |
| 4.10 | Cardioid | 48 |
| 4.11 | Direction of cancellation by super cardioid | 49 |
| 4.12 | Super cardioid | 50 |
| 4.13 | Hyper cardioid | 51 |
| 4.14 | Broad cardioid | 52 |
| | All characteristics compared | 53 |
| | Example of an omnidirectional microphone | 55 |
| 4.17 | Noise characteristics of a condensor microphone compared to the noise of | |
| | typical environments, taken from Peus [20], page 7 | 57 |
| | Induction on a moving wire | 59 |
| | Schematic construction of a dynamic microphone | 60 |
| | Schematic frequency response of an omnidirectional dynamic microphone . | 60 |
| | Schematic construction of a dynamic microphone with gradient characteristic | 61 |
| 4.22 | Schematic frequency response of a gradient type dynamic microphone (di- | |
| | rectional) | 62 |
| 4.23 | Attenuation of bass or midrange response for a dynamic microphone. The | <i>c</i> - |
| | black rectangles are meant to be inductors/coils | 62 |
| 4.24 | Schematic construction of a ribbon microphone, right side: cut throug a - b | 63 |

| 5.1 | XY setup: Recording area plotted above the angle of two cardioid micro- |
|-----|---|
| 5.2 | phones |
| | phones |
| 5.3 | XY setup: recording areaplotted above the angle of two hyper cardioid |
| | microphones |
| 5.4 | 5.1 loudspeaker setup in a cinema |
| 5.5 | 5.1 loudspeaker setup in studio or home cinema |
| 5.6 | The IRT cross |
| 5.7 | The ORTF surround setup |
| 6.1 | The jack and cinch connector |
| 6.2 | The XLR connector |
| 6.3 | Stereo jack |
| 6.4 | Schema of phantom powering |
| 7.1 | The paramters of a compressor |
| 7.2 | Compressor: prolonging a delay |
| 7.3 | Compressor: shortening a delay |
| 7.4 | Basic analog mixer circuit |
| 9.1 | Piano recording: end position |
| 9.2 | Piano recording: side position |
| 9.3 | Piano recording: hammer position, AB and XY setup |
| 9.4 | Piano recording: hammer position, microphone array |
| 9.5 | Example: creating a two dimensional sound stage |

1 Overview

Never before the technology to handle sound was as cheap as today. If you pick the right components you may easily convert sound into electrical or digital signals, edit them, replay them or store them digitally for a long term. The inevitable degradation over time by multiple generations of analog storage technology lost their grip. Once digitized, every signal may be treated without losses with perfect linearity in the frequency and amplitude domain. The requirements for the computers to do these tasks are trivial these days.

On the other hand this means: there is one less excuse if the result is less than optimal.

Another requirement for success will not disappear anytime soon: the necessary knowledge and skills to use the technology sensibly and goal oriented.

With this script I want to help finding the necessary steps in the right direction. If I succeed, this means: yet another excuse gone. My fault this time.

If you need only one track at a time, for first steps in recording audio all you need is a microphone with built-in A/D-converter and USB port. By using electret capsules they may already deliver a reasonable sound quality. The downside of this minimalist approach is that you may not combine several of these devices for multiple channels e.g. for stereo recordings or even more simultaneous input channels (whole band, orchestra, drum set...) because they are not running on the same clock source and the signals will run out of sync in no time, destroying at least the phase correlations immediately. You can, however record lots of tracks (multitrack) one at a time and mix and edit them later.

Even if you want to work real cheap, you have to be very careful in choosing your listening reference equipment. A typical problem is trying to fix a weakness of your device in the mix. You might for example exaggerate the bass region because your loudspeaker has too little bass. The loudspeakers that are tailored for this task are called studio monitors. Do not rely on headphones for this task. This can lead to a very narrow stereo image as result, because the complete isolation of the signals on each ear exaggerates the breadth of the stereo base. Additionally there is no room reverberation in what you hear which makes it hard to decide correctly how much reverberation should be used in the mix.

The logical next step up is using at least two microphones of the same type – best small diaphragm condenser microphones with omnidirectional and unidirectional capsules or two microphones of each characteristic – and an interface with two microphone inputs. This opens up the whole world of stereophonic microphone configurations. As before, you can still add more channels with multitracking, but not simultaneous. One or two DI-boxes complement this and allow for unproblematic interfacing of directly electrical sound sources (instruments) if needed.

If the main microphone configuration is to be complemented by auxiliary microphones that shall be recorded on additional channels, it is necessary to use an interface with enough inputs (eight, or better sixteen, you can't have too many) and the corresponding number of microphones that should be as universal (neutral) as possible.

If the task is not to create a recording but to directly amplify and distribute sound in real time, you need the corresponding microphones and/or DI-boxes, a mixing desk, amplifier(s) and the corresponding loudspeakers (or as combination of the last two items, active loudspeakers). At least in the case of musical performance you typically also need monitor speakers for the artists. With the right set of effect devices you may suppress acoustical feedback oscillation and optimize the sound.

If you have to amplify for an audience of non-trivial number, the speakers (PA = Public Access) should be put on stands that elevate them beyond ear height. That way the audience in the rear does not lose a large part of the high frequencies because the heads of the front rows are in the way. Except for cases with very little bass (speech) there should always be a sub woofer that is located on the floor. In the location on the stand, the first reflection from the floor typically leads to a cancellation in the bass region, even if these frequencies get reproduced by the speakers.

No matter what you want to achieve, it is absolutely necessary to have a clear target and the knowledge that allows you to systematically work towards this goal. Always keep in mind that in case of doubt the whole result is far more important than details like the sound of one of the instruments for example.

I will concentrate on concepts and not on ready-to-use recipes. That might make it harder to understand everything and read the book to the end. On the other hand this helps to find a straight way to get things done, solve problems and to find your own way and style. Recipes can keep you bound to the examples given and often do not really fit optimally to the situation at hand.

You learn best if you understand the concepts and get routine in applying them. This way you can really "grok" them.

2 The Sound Field

2.1 Oscillations and Waves

2.1.1 Oscillations

In a state of oscillation, a mechanical system reacts to an initial exciting impulse by cyclically exchanging between potential energy and kinetic energy in the form of movement after the inducing impulse has stopped. If the potential field is square to the distance from the equilibrium point, the result is movement in sinusoidal form. The system is then called a harmonic oscillator.

For electrical systems the corresponding exchange happens between the electrical field in a condenser and the magnetic field in a coil. Alternatively an analogous reaction is created by means of an amplifier with positive feedback loop.

2.1.2 Waves

If the excitation, in the form of a deviation from the equilibrium case, moves away from the origin, you have a wave. The typical equations that describe this phenomenon allow for any kind of wave form, as long as its square stays integrable over limited regions of time and space. This corresponds to a limited power an is typically the case in real life.

The typical equations that describe this phenomenon allow for any kind of wave form, as long as its square stays integrable over limited regions of time and space. This corresponds to a limited power as is typically the case in real life.

2.2 Level, dB

At a given position in a sound field, there is a density of power. It is the product of sound pressure and the speed of air movement by the sound. Similarly, in the electrical case this corresponds to the product of voltage and current.

In both cases there is a constant quotient of both variables whose product delivers the power or density of power. This quotient is a resistance, either mechanical (the wave resistance) or electrical. The square of one of the variable is therefore proportional to the power density or power. Depending on which variable is used, you have to multiply with or divide by the resistance to get the actual power density or power.

In the case of acoustics, the power density can be anywhere in the range of twelve orders of magnitude for a audible signal (between perceived silence and ear pain). The easiest way to handle this large range is to use a logarithmic view. This corresponds roughly to the way our ear perceives loudness.

You must not forget that the logarithm is only defined on pure (and positive) numbers, not on anything that has a unit. Therefore you always need a reference unit. The logarithm is then taken from the quotient of the measured value and the reference value. The unit cancels out this way. It the natural logarithm (base e) is used, the values are given in *Neper*. In technology this is rarely used

If the logarithm to the base 10 is used, the values are given in Bel^1 . For many practical uses the numbers are too small, therefore the most often form is a tenth of the unit, the *Decibel*, abbreviated as dB. The corresponding value is called the level, depicted as L in formulas.

This results in the electrical case for a given power P (and analogous for the power density in the sound field) in the definition:

$$L = 10 \log_{10}(P/P_{\rm ref}) \tag{2.1}$$

Because of the square law for the equation between the sound pressure p in the sound field and the power density, this leads to

$$L = 10 \log_{10}(p^2/p_{\text{ref}}^2) = 10 \cdot 2 \log_{10}(p/p_{\text{ref}}) = 20 \log_{10}(p/p_{\text{ref}})$$
(2.2)

and similarly for the particle velocity by the sound field (not to be confused by the velocity of sound c), called vs

$$L = 10\log_{10}(vs^2/vs_{\text{ref}}^2) = 10 \cdot 2\log_{10}(vs/vs_{\text{ref}}) = 20\log_{10}(vs/vs_{\text{ref}})$$
(2.3)

in the electrical case the voltage ${\cal U}$

$$L = 10 \log_{10}(U^2/U_{\text{ref}}^2) = 10 \cdot 2 \log_{10}(U/U_{\text{ref}}) = 20 \log_{10}(U/U_{\text{ref}})$$
(2.4)

and the current I

$$L = 10 \log_{10}(I^2/I_{\text{ref}}^2) = 10 \cdot 2 \log_{10}(I/I_{\text{ref}}) = 20 \log_{10}(I/I_{\text{ref}})$$
(2.5)

The corresponding values for the resistance are located on both sides of the quotient and cancel out.

Example:

$$P = U \cdot I = U \cdot U/R = U^2/R \tag{2.6}$$

and for the power

$$\begin{split} L &= 10 \log_{10}(P/P_{\text{ref}}) \\ &= 10 \log_{10}(UI/(U_{\text{ref}}I_{\text{ref}})) \\ &= 10 \log_{10}(UU/R)/(U_{\text{ref}}U_{\text{ref}}/R) \\ &= 10 \log_{10}(U^2/U_{\text{ref}}^2) \\ &= 10 \log_{10}((U/U_{\text{ref}})^2) \\ &= 10 \cdot 2 \log_{10}(U/U_{\text{ref}}) \\ &= 20 \log_{10}(U/U_{\text{ref}}) \end{split}$$
(2.7)

In audio technology the most interesting values are the sound pressure p and the voltage U, because they are converted in each other.

The most used reference levels are:

Sound pressure level p_0 is defined as 20 μ Pa for L_p

¹Named after Alexander Graham Bell, the inventor of the telephone.

Voltage here you see several reference levels, partly for historical reasons. By adding a character to the dB symbol, you depict which one is used.

 $\begin{array}{ccc} U_0 & \text{Unit} \\ 1 \text{ mW at } 600 \ \Omega, \text{ this is about } 0,775 \text{ V} & \text{dBm (obsolete)} \\ 0,775 \text{ V} & \text{dBu} \\ & 1 \text{ V} & \text{dBV} \end{array}$

If these definitions are understood, you may even compute approximate level differences in your head. As always, in the logarithmic scale a product is transformed to a sum, a division to a difference, a power to a factor.

Some examples:

$$U_1/U_2 = 10 \tag{2.8}$$

leads to a level difference of 20 dB

$$U_1/U_2 = 1000 \tag{2.9}$$

results in a level difference of 60 dB, the power 3 of 10 gets multiplied. In many real life cases you may benefit from the approximation

$$1000 = 10^3 \simeq 2^{10} = 1024 \tag{2.10}$$

1000 is corresponds to 60 dB (as shown above), so this is approximately also true for 1024. This is

$$log_{10}(2) \cdot 10$$
 (2.11)

Therefore,

$$U_1/U_2 = 2 (2.12)$$

is at about 6 dB level difference. It corresponds to four times the power or a doubling of the voltage.

The square root of this halves the dB value, so that

$$U_1/U_2 = \sqrt{2} \tag{2.13}$$

corresponds to a level difference of close to 3 dB. The exact number is irrational, the next useful approximation is 3.01 dB. Because the remaining third of a percent does not really make a difference e.g. for perception, normally 3 dB are used in real life as value for a doubling of power.

Let's move on to an example that is slightly less trivial. It may also be approximated by simple computations.

A loudspeaker is able to create a level of 98 dB with 1 W input power at a distance of 1 m. In a distance of 8 m a level of 105 dB shall be produced. How much power has the amplifier to provide – and the the loudspeaker be able to handle?

The power distributes at each distance at the same part of a sphere surface area with a radius that is equal to the distance of the source (at least for point sources, let's say this is approximately the case). So we have to achieve 105 dB at a distance of 8 m, and by each halving the distance we gain 6 dB, because the area that the sound gets distributed to is now only a quarter of the original size. If we go from 8 m to 1 m, we have to halve the distance three times. Therefore, the loudspeaker has to provide a sound presser level at 1 m distance of

 $105 \text{ dB} + (3 \cdot 6) \text{ dB} = 123 \text{ dB}$

With 1 W of input power, it creates 98 dB, the difference is

123 dB-98 dB=25 dB.

Let us simplify the computation by taking 26 dB, which also brings us on the safe side of erring. This gives us a factor of 100 for 20 dB and the remaining 6 dB lead to another factor of 4. We need approximately 400 W.

2.3 Sine Wave, Fourier Transformation

2.3.1 Sine Wave

Almost every text about the mathematical foundations of acoustic and electro-acoustical processes start with the special case of the sine wave. There are good reasons for that.

The mathematically simplest case of a periodic movement is the harmonic oscillator. One example is a mass point in a square potential (the force is linear to the elongation) of a spring. For small amplitudes, each stable system fits into this view with growing precision (Taylor series).

What seems like just the simplest special case reaches into the realm of all cases in which a system is linear. On the other hand this is a necessary feature for almost all devices and interfaces in the electro-acoustic technology. It is necessary to allow a clear distinction of signal components and for detailed perception.

The nice thing about linear systems is the fact, that you may take their reaction to known input signals and combine them to find the reaction to an arbitrary linear combination (weighted sum) of them just by computing the linear combination of the original answer functions. If you go from finite sums to an integral, the same principle holds.

2.3.2 Fourier Transformation

The Fourier transformation is a mighty tool to compute lots of relevant information, it changes the viewpoint from a time based description to a frequency based description and vice verse.

If a function is discrete in time and defined on a finite time interval, it may be divided into a finite sum of weighted sine- and cosine functions. There are two alternate equivalent ways to express this, an pair of amplitude and phase for each frequency or a complex amplitude for each frequency. This situation is typically given for real world sound applications. If a signal is cyclical, whether defined in an interval or on the whole range of real numbers, it is sufficient to work on one cycle.

If go to smooth functions, the sums turn to infinite sums. If there is a maximal frequency, it is sufficient to get the function value at equidistant sample times twice for the cycle of the highest frequency. The description of the whole function is still complete, loss-less. That is the result of Shannon's sampling theorem, one of the foundations of digital signal processing. If the interval is extended to the realm of the real numbers, the sums turn most of the time into functions (for all cases generalized as distributions) over frequencies. Again, these exist distinctly for sine and cosine, amplitude and phase or complex amplitude. For infinite definition regions there appears one complication. Already one cyclical function, say sine wave, leads to a divergence of the result at this frequency, the value goes to infinity. Therefore there is no defined function value at this place, but the integral is defined. This is described mathematically by distributions. The most basic distribution has a value of zero over the whole region of \mathbf{R} , with the exception of zero, where the integral over a interval containing zero is defined to be one. This distribution is called delta function. This misleading name ("function") is taken as given for historical reasons.

The delta function has an important meaning. If you know the answer of a linear system to the delta function (technically: the impulse response), you may use the mathematical operation of convolution to compute the response to an arbitrary input. The answer to the delta function is called the Green's function of the system. Its Fourier transformation is the answer function to arbitrary sine wave inputs and vice versa.

It is extremely more efficient to compute the Fourier transformation of the input signal and the Green's function, multiply them and re-transform the result to get the answer function that to directly compute the convolution.

The Fourier transformation can be inverted. If you take a matching constant in the definition and work with complex numbers, it is even symmetrical. That is the reason why it is called a transformation: both descriptions are complete, with the whole information.

This all sounds very theoretical and I just denoted the results, but these have important practical implications.

It is sufficient to know the answer of a linear system to an arbitrary sine wave input to know how it will handle an arbitrary signal, including pulses.

If someone complains that this cannot be all you need to know, because e.g. music is claimed to be more that some sine wave signals, it only shows ignorance. Like any signal, music may be decomposed to sine wave signals (including phase information) without losing any detail. Otherwise it would not be possible to reverse the transformation.

If a pulse response is completely expressible in real numbers (which is in reality always the case, like the name suggests) and the is only defined for positive arguments (the system is unable to look into the future, which is also rather common), there are important consequences for the relations between the amplitude and phase functions, The Kramer-Kronig relations. With the exception of delays and all-pass filters you may compute the minimal phase functions if you have the amplitude function. Many systems are minimal phase. If you see resonances or dips in the amplitude function, it is immediately clear, that this implies large phase changes in the proximity of their frequencies, often a lingering sound in the case of a resonance.

If you exploit some facts from number theory, the Fourier transformation can be computed very efficiently if the number of sampling points can be factored to many small prime numbers, ideally if it is a power of two. The computation can then add lots of input values and multiply them once with their common factor. This algorithm is calls fast Fourier transformation (FFT).

Conclusion: The frequency response of a system is one of the most important characteristics. If the system is completely linear and the information is resolved as fine as necessary, including phase information, it contains the complete description.

2.4 The Plane Sound Wave

This is the simplest case of all. Everything only depends on the position along one direction, perpendicular to this everything remains constant.

If we choose the direction of propagation as x axis, it is sufficient to describe everything in one dimension.

For a sine wave running along the x axis we get the simple equation

$$p(x,t) = \sqrt{2} p_{\text{eff}} \sin\left(2\pi \left(\frac{x}{\lambda} - (t-t_0)f\right)\right) = \sqrt{2} p_{\text{eff}} \sin\left(2\pi \left(\frac{xf}{c} - (t-t_0)f\right)\right) \quad (2.14)$$

Here p is the difference between the actual pressure and the static mean pressure, p_{eff} is the effective (root mean square, RMS) value of the sound pressure, λ the wave length, f the frequency, c the velocity of sound. c depends on the temperature, but not on the static pressure.

2.5 The Spherical Sound Wave

A spherical sound wave ideally starts a single point. More realistic is a (small) sphere that "breathes". This prevents the singularity in the origin of the wave (infinite pressure, infinite power density).

This wave can also be described in one dimension, this time this is the radius r, the result is independent of other polar coordinates.

In the limit of large distance the wave approaches a plane wave locally.

The amplitude is not constant with varying distance r, because the power density which is proportional to p_{eff} , is distributed over the corresponding surface area. The power density falls proportional to $\frac{1}{r^2}$. As result, the sound pressure falls proportional to $\frac{1}{r}$.

$$p(r,t) = \frac{1}{r}\sqrt{2}p1_{\text{eff}}\sin\left(2\pi\left(\frac{r}{\lambda} - (t-t_0)f\right)\right) = \frac{1}{r}\sqrt{2}p1_{\text{eff}}\sin\left(2\pi\left(\frac{rf}{c} - (t-t_0)f\right)\right)$$
(2.15)

2.6 Interference

If there is a superposition of independent signals (they are uncorrelated), the power densities, which are proportional to the square of the sound pressure, add up. So, for effective sound pressures p_1 to p_N , the resulting total sound pressure p_{res} is given by

$$p_{\text{res}} = \sqrt{(p_1^2 + p_2^2 + \ldots + p_N^2)} = \sqrt{\left(\sum_{i=1}^N p_i^2\right)}$$
 (2.16)

If there is a constant phase correlation, this has consequences for the resulting sound pressure. The most extreme case is the addition of two signals of the same amplitude. Here the resulting sound pressure can be anywhere between zero and the double pressure of one signal (four times the power density). For two sine wave signals with equal sound pressure and phase difference $\delta\phi$ the result is:

$$p_{\text{res}} = (p_1 + p_2) \cdot \cos(\delta\phi) \tag{2.17}$$

2.7 Plane Reflection, Comb Filter

If a sound wave is reflected by a acoustically hard wall (the particle velocity is stopped, the wave is completely reflected back with the same law about the reflection angle as in the case of optical reflection) the reflected and the incident wave interfere drastically.

At least close to the wall the sound pressure of the incoming and the outgoing waves are practically identical. Therefore the two extremes of double sound pressure (+6 dB level) and cancellation do happen in this case.

If the incoming and reflected waves are plane and move perpendicular to the wall, there each sine component gets canceled at each distance of $\frac{\lambda}{4} \cdot N$ for odd N and for even N you get the maximal sound pressure. For each given distance δ the frequency response is modulated with a factor $V(\delta, c, f)$ that is given by the equation

$$V(\delta, c, f) = 1 + \cos\left(\frac{4\pi\delta}{\lambda}\right) = 1 + \cos\left(\frac{4\pi\delta f}{c}\right)$$
(2.18)

Here are plots of resulting frequency responses as examples, shown linearly to ease the understanding of the effect and doubly logarithmic as usual for comparison.

The first cancellation in the hearing range (20 KHz) happens at a distance of only $4.25 \ mm$

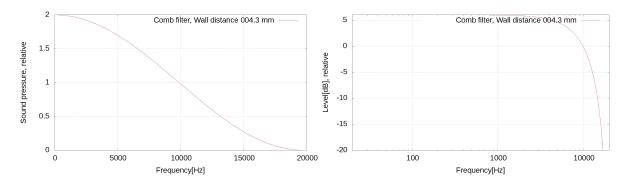


Figure 2.1: Comb filter, distance 4.3 mm

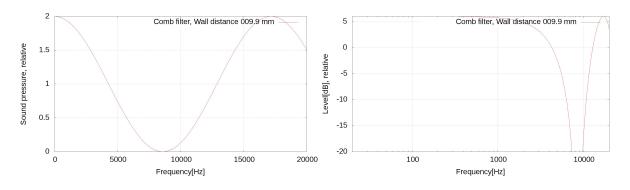


Figure 2.2: Comb filter, distance 10 mm

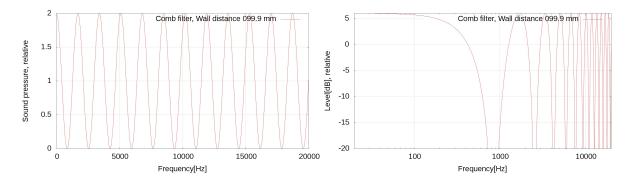


Figure 2.3: Comb filter, distance 100 mm

The sound changes even by small variations in the distance to the wall are a lot larger than that resulting from typical devices not being ideal.

An animation of the results of distance changes may be found on YouTube playlist Audio Visualisation[41] as linear plot and logarithmic, in dB.

Aside from absorption the rules for the path of sound work the same in both directions. Therefore the effect is the same if a omnidirectional source and the microphone are switching spaces.

As result the same effect applies to the reproduction of sound. If the sound source transmits in all directions, the retarded echo from the wall adds to the direct sound. Close to the source the effect is small, because the original signal is louder than the echo. The echo may also be shadowed by a large sound source. This has consequences for an optimal placement of the microphone. Depending on the characteristics of the source and the geometry of the situation it is important to know about the effect and to try to recognize if it happens. It may be a good idea to consider counter tactics.

A special case that is not included in the graphics, but that may be easily extrapolated: at a distance of 1 m the lowest frequency that is canceled out is a bit lower that 100 Hz, this is in the bass region. If a woofer is located at that place and the listener is not close to it, the power and efficiency of the woofer do not matter. You do not get a reasonable sound pressure around this frequency.

If the distance is greater that 5 cm, there are so many frequencies cancelled out that it is not possible to correct the frequency response. Even worse: in the worst cases, these are cancellations, these frequencies are missing in the resulting signal. There is no way to counter this effect by linear measures, a multiplication by zero does not give you any chance.

It is important to note that for diverging waves (finite distance of the source) and partial reflection the amplitudes of the incoming has a greater amplitude than the reflected sound. Instead of a total cancellation the effect leads to a dip in the frequence response. In a room the effect is also less prominent because of the incoherent addition of the diffuse room reverberation that can partially fill the gaps.

To avoid or at least minimize the comb filter effect, the following precautions should be used:

- Neither source nor microphone should be very close to the wall(s). In its vicinity direct and reflected sound have the most similar amplitude.
- If a single sound source is to be picked up, the microphone and the sound source should be much closer to each other than either wall to one of them. This maxi-

mizes the amplitude difference between the direct and the reflected sound. It also minimizes the effect of the room reverberation field, adding clarity(close miking).

- Surfaces that are close should be treated acouscically to minimize reflections.
- Several microphones that pick up the same source should maximize the difference in level by being at distances differing as much as possible. A rule of thumb is that the distances should differ by a factor of three (the three-to-one rule).

2.8 The Room

2.8.1 Room Modes

On all walls surrounding a room there are reflections along with the corresponding interferences. Additionally this leads to paths in which the sound wave is moving in closed loops in the geometrical-acoustical limit (wavelength is short compared to all relevant geometrical sizes) in which the sound is reflected in two directions. This results in so called "stationary waves" that oscillate in phase in the whole room (at least in the case of no or no relevant damping). Strictly speaking, this is not a wave phenomenon any more, but rather an oscillation that is distributed in space, a so called mode. They do not move (like waves do) but have fixed pressure maxima and minima, called nodes.

At frequencies matching a mode the energy is stored over time. The room as whole has a resonance on that frequency. With constant excitation the sound pressure is raised on these frequencies, sometimes drastically, and the pressure remains after the end of the excitation for times that can be impressive. Both effects lead to great problems for a clean reproduction of bass frequencies. This effect can be minimized by adding a large damping for the bass frequencies.

This effect complicates the task of recording bass signals acoustically in a room. It the room is not really huge (concert hall) and the modes are not damped, it is often necessary to record them in direct proximity, where the direct signal is much larger than that of the room modes.

Damping elements that favor the damping of bass frequencies (e.g. by having interfaces that reflect higher frequencies but are transparent for bass) are called bass traps.

2.8.2 Reverberation, Diffuse Sound Field

For middle and high frequencies the density of resonance frequencies gets so high, that they can't be heard distinctively. The transition frequency, over which the bandwidth of the modes gets broad enough that the modes overlap is called the Schrödinger frequency of the room.

For these frequencies the propagation of the sound waves can be seen as sound beams with the tools of geometric acoustics (analogous to geometric optics). In case of a room this is combined with statistics.

If the sound waves reach each point in the room from any possible angle due to chaotic multiple reflections, the resulting sound field is called a diffuse field. This is well approximated in a reverberation room. It is build with rigid, acoustically hard walls that omit plan parallel walls, sometimes including curved surfaces, e.g. cylinders.

This field is the prototype to describe the room part of an acoustical field, often viewed in contrast to the spherical field of a sound source.

2 The Sound Field

The diffuse field is taken to be evenly distributed (independent of the location) and incoherent to the source that excites it. For signals with extremely long coherency times (e.g. organ flute, stationary repeating signal) this is not really the case.

In real life there is always some damping (even by the air itself), so the sound pressure can never rise to arbitrary levels from a finite excitation. The sound pressure falls off exponentially after the end of the excitation. The time after which the reverberation falls off by 60 dB is called the reverberation time t + 60. After that time it is normally not possible to hear the remaining reverberation signal.

In a real measurement of the reverberation time it is often difficult to collect the data until the -60 dB are reached, especially if there is ambient noise at the same time. To get a meaningful result one may measure the time until e.g. -20 dB are reached and multiply the value with three to extrapolate to the theoretical time value for -60 dB. This theoretical time is called t_{20} . Ideally therefore you get $t_{20} = t_{60}!$

The direct and diffuse signal are assumed to be uncorrelated. This means that the resulting sound pressure is

$$p_{\rm res} = \sqrt{p_{\rm sphere}^2 + p_{\rm diffuse}^2} \tag{2.19}$$

The following plots depict the sound pressure of all components and the resulting total sound pressure. Please not: the origin is not included because it leads to a divergence and does not have a practical relevance, because no real sound source is point like. The absolute values are arbitrarily chosen, but the +3 dB at the crossing of both curves have a real meaning. The distance between this point and the origin defines the reverberation radius.

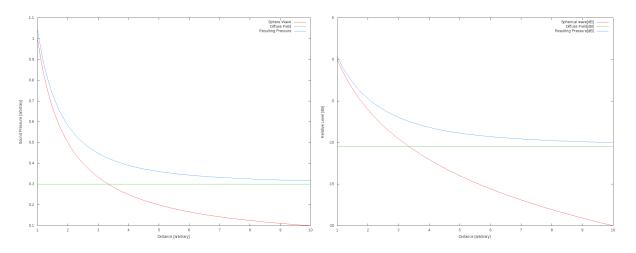


Figure 2.4: Levels of direct sound and room reverberation

By damping the sound, mostly by inserting porous substances like glass wool, polymer sponge, foils with micro holes. Care is advised not to overdo this with thin dampers, because they work mostly exclusively for high frequencies. This may lead to a reverberation that is extremely colored with excessive (undamped) bass.

If the structure of the reverberation is too regular (mostly because of parallel walls), this may be countered by means of diffusors. With curved or cleverly distributed depth of columns they distribute incoming sound into all directions without lowering the amount of reverberation.

2.9 Size of Sound Sources

All sound sources that are larger than $\frac{1}{4}\lambda^2$ (this meas for 20 KHz and a wavelength of 17 mm as a rule all of them) and that at least partly vibrate in phase over this range display a directivity of sound transmission. As a rule the directivity pattern is strongly frequency dependent, with high frequencies being most directive or irregular.

As a result of this, for recording it is necessary to find (at least...) a place that receives a sound that matches best the perception and/or desired ideal.

2.10 Size of Microphones

2.10.1 Interference

Except for very tiny diaphragms there is an interference for sound that comes from the side, because not all parts of the diaphragm get excited with the same phase. This leads to a reduced sensitivity for treble for sound that comes from the side. The larger the diameter, the larger is the effect.

2.10.2 Reflection

If the diameter comes into the same order of magnitude as the wavelength, the capsule (and body) of the microphone change the sound field. One effect is the reflection of the incoming sound waves. For diameters that are large compared to the wavelength the pressure of the reflected wave is as high as that of the incoming wave. Because the phase on the place of reflection is almost exactly the same as that of the incoming wave, the pressure doubles and the level is raised by 6 dB. The transition happens in a frequency interval of roughly one octave.

The same amount of level raising but with a smoother and broader transition happens on the surface of a sphere. To achieve this effect one may use a (mostly wood) sphere with a diameter of typically 5 cm with a hole that matches the microphone body for insertion³. The Neumann models M50 and M150 have such a construction internally. They look like large diaphragm microphones because of their size, but they are not. The output of such microphones can be used without any further processing for AB settings (see section 5.2.3, page 72) or Decca Trees (see section 5.2.5, page 73). The smoothly raised treble level matches approximately the typical loss by distance and room reverberation.

²That is the minimum size that leads to regions where part of the sound is transmitted out of polarity into some directions for geometric reasons.

³This has to be an omnidirectional pressure transducer. In all other cases the necessary additional holes for the back side of the capsule would be blocked by the sphere.

3 Hearing

3.1 The Ear

The outer ear, especially the pinna, does not simply collect and concentrate sound energy. It has several entry points and (open) wave guides that lead part of the acoustic signal over a bypass with the corresponding delay into the inner ear parts. This leads to resonances and interferences and as a consequence to changes in the frequency response that are indications for the angel from which the sound arrives, especially in the directions up and down.

The next part leads through the ear canal to the ear drum. The ear canal is a pipe with one open and one closed end. The resulting resonance raises the sensitivity of the ear close to the thermal noise of the air itself.

The ear drum gets excited by the sound pressure and moves the small bones malleus, incus and stapes. They transform the movement mechanically and match its mechanical impedance to that of the cochlea input. In the cochlea the frequencies get coarsely separated and detected by the hair cells. The resulting signal is transmitted to the brain via the vestibucochlear nerve. There the separation of frequencies is extremely sharpened.

3.2 Limits of Auditory Perception

The limits of perception for hearing lie between at about 0 to 130 dB of sound pressure (this is the origin of the definition for 0 dB at 1 KHz) and 16-20000 Hz for the frequency. For older persons the hearing threshold raises and the upper frequency limit gets lower. Recently there were measurements that indicate that sound with frequencies even below 16 Hz may be perceived. This does not have immediate impacts on practical work with audio technology, because with the current state of the art it is not really possible to play back such low frequencies and they are are not relevant to music, even less so for speech.

The limits of perception do not simply span a rectangle in a plot of sound pressure level versus frequency.

3.2.1 Curves of Equal Loudness

The comparison of perceived loudness and actual sound pressure level has been pioneered by Fletcher and Munson. Later it was refined with improved equipment. The plots show curves of sound pressure levels with respect to constant perceived loudness. The loudness is given in Phon which corresponds to a perceived loudness that is equal to that of a 1 KHz tone with the sound pressure level that has the same numerical value in dB.

The ear is most sensitive between 2 and 4 KHz, matching the open pipe resonance of the 2.5 cm long ear canal.

For the bass region, there are two remarkable effects.

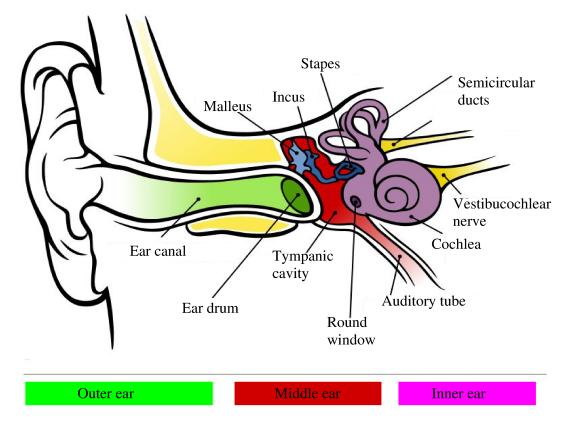


Figure 3.1: Schematic cross section of ear, take from Wikipedia, License CC by-sa Version 4, Created by Geo-Science-International, accessed 2016-06-07, translations by author

- A level of 70 dB is necessary to perceive the full frequency range. For 60 dB the lowest perceived frequency is 30 Hz, for 20 dB it is even 150 Hz (red curve).
- The perceived loudness curves above 20 Phon are closer to each other than at higher frequencies. This is extreme for the frequency of 20 Hz, where the perceived loudness difference of 20 Phon corresponds to just 10 dB of sound pressure level difference. A deviation in frequency response in for bass frequencies is more prominent that for higher frequencies.

This has interesting consequences for sound reproduction. For low levels there is no need for low frequencies, they can't be perceived anyway. That is the reason why it is possible to stand the sound of miniature loudspeakers if the signal does not have to be loud.

The other way round: a certain minimal sound pressure level is absolutely necessary (often assumed to be 85 dB to have space for soft passages) to judge the results of a mix. Otherwise there is no way to know how much bass is right. The correct level of bass is also dependent of the sound pressure level. Therefore it would be wise to agree on a norm level, which is now standard for movie mixes.

If the target level and the actual reproduction level are both known, which is the case for calibrated AV-receivers or cinemas, the known error in perception of the bass level can be correctly computed an corrected by a tailored gain change at these frequencies. In coarse form the old fashioned loudness or contour controls or switches tries to do the same.

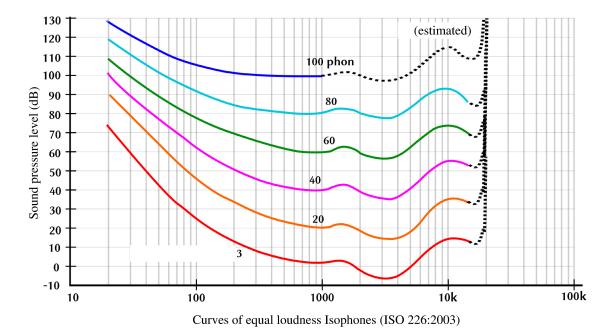


Figure 3.2: Curves of equal loudness, take from Wikipedia accessed 2016-06-07, translated by author

3.2.2 Masking

The perception of a tone leads to a raise of the perception level for tones of similar frequencies, most prominently for higher frequencies.

The lossy digital (data) compression methods use this fact to filter out or at least lessen the resolution for the coding of signals in this frequency regions.

This masking leads to the effect that high levels in the bass region leave little room for detail in higher frequency ranges. This is problematic for higher bass or low mid sounds. Therefore an exaggerated bass must be avoided, and it is very problematic to have several sources of strong bass signals. The act of searching and removing these conflicts is sometimes called tidying up the bass.

3.3 Localization

Additional to the recognizing of the signal structure of a sound, one very important aspect of hearing is the localization. For this end, several aspects of a sound field get evaluated.

3.3.1 Angle to the Mid Plane

To get the angle of the sound source to the mid plane of the head, two parameters of the two ear signals get compared.

Time difference of arrival at both ears. The signal reaches the ear that is closer before the other one. The additional way around the head geometry must be taken into account. This time difference can be one millisecond maximum. For signals that do not change at all over time or where a change happens only very slowly, this difference vanishes with the exception of a static phase difference. As result, it is hard to localize stationary sound sources and the localization is not very accurate. **Difference in level and frequency response between both ears.** By the shadow of the head, the treble and a bit the midrange are lower in level on the ear on the far side of the source.

Both effects get less prominent in the bass region. Most of the time it is assumed that under 150-200 Hz effective localization is not possible.

The hearer constantly moves the head a bit to each side unconsciously. The corresponding changes to the ear signals get processed in the brain. This improves the precision of the localization. If these changes are suppressed by using ear related recordings ¹, the localization is disturbed, mostly in the important region in front of the listener. The movements are still happening but the corresponding effects on the ear signals are missing, this leads to conflicts in the perception.

3.3.2 Angle in the Symmetry Plane: Front, Back, Above, Below

By the coloration of the sound in the pinna (which is specific to each individual) the hearing system can extract the elevation angle if the original sound or movement is known. The precision is far less than that of the angle towards the side. It is very low in the forward direction. enddocument The relevant frequency domains were identified by Jens Blauert and are called Blauert's bands.².

A raise in level in the region of 1 KHz and 10-15 KHz leads to a localization in the back, in the region of 8 KHz above and 400 Hz and 4 KHz in the front.

All in all a lowering at 1 KHz and 7-15 Hz and a raise of 400 Hz and 4 KHz leads to the localization at very small distance as opposed to a diffuse and distant localization. Eberhard Sengpiel recommends to use these frequencies with Q-factors of 2–4 and 6 dB of level change maximally.

3.3.3 Distance

The distance information is inferred by the sound pressure level in case of known source signals, the structure of echoes at early reflections, the relative amount of reverberation (if present) and by the damping of treble in air for large path lengths (Which is also the reason why an extremely broadband bang of e.g. a flash gets transformed to a thunder growl).

3.4 Effects to the Subconscious

Hearing is universally an extremely important sense for warning. Therefore it has fast and strong effects to the subconscious and the mood. No other sense can lead to a similarly fast push of adrenaline compared to the reaction to an unanticipated extremely loud sound signal.

If taken for positive effects this is one of the reasons why music can touch so deeply or excite so strongly, especially with rhythmic bass.

 $^{^1\}mathrm{See}$ dummy head, chapter 5.2.7, page 75

²For a description, see http://www.sengpielaudio.com/DieBedeutungDerBlauertschenBaender. pdf, on [24]

4 The Microphone

The Microphone is the device that captures the sound and converts it to an electronic or digital signal that can be processed and/or stored.

Any information about the sound that gets lost by the limitations of the microphone cannot be regained later. Therefore it is extremely important to pay attention to its characteristics and use it in a suitable way, starting with the optimal place.

The chapter starts in section 4.1 by defining the the parameters that describe the characteristics of the transducer in general, independent of its type.

Most microphones use a diaphragm to be moved by the sound pressure. It can be coupled to a transducer that reacts to its position (static, e.g. condenser, piezo, coal microphones) or to a transmitter that reacts to its movement, the momentary speed (dynamic in a general sense, dynamic in the narrow sense or voicecoil, ribbon).

Diaphragms react to pressure differences. The microphone capsule can be build to create a

- **pressure transducer**. One side (called back) is ideally closed to deliver a static reference pressure. This ideal is not suited for practical devices, because the air pressure slowly changes with height or weather conditions. Those variations exceed those by typical sound pressure levels. Therefore the back is not really sealed but a capillary allows the back to reach an equilibrium to the mean ambient pressure but still keeps the back side isolated from sound.
- **gradient or pressure difference transducer.** The back is left open to the sound, the mechanical construction has to be symmetrical to a plane normal to the axis of maximum sensitivity.
- a combination of gradient and pressure transducer. The back is partially open. Most microphones for music are from this category.

This distinction and interferences result in the different directional charactersistics which are described in section 4.2, page 36.

This part is followed by section 4.3, page 54 which provides an overview of the principles and types used. It includes recommendations about which types of microphones are best suited for different tasks.

4.1 Features

The function of microphones can be characterized by several important parameters that can be measured objectively. They are important for determining the fitness of the microphone for the different use cases.

4.1.1 Sensitivity

The sensitivity Sens is the output voltage for a sound pressure level of 1 Pa, which is 94 dB. It may be stated as Sens in mV or as G in dBV.

$$G = 20 \cdot \log_{10}(Sens/1 \text{ V}) \tag{4.1}$$

Because acoustical sound pressure levels are given relative to the approximate hearing threshold of 0 dB, the electrical output level L_{el} for a sound pressure level of L_{ac} can be obtained by adding the sensitivity G and subtracting 94 dB.

$$L_{el} = L_{ac} + G - 94 \text{ dB}$$
(4.2)

The sensitivity normally depends on frequency and angle of incidence (see below). The specs secify the sensitivity at 1 KHz for plane sound waves coming from the direction of maximum sensitivity. This direction will be called *the axis* from now on.

To get as little noise from the preamplifier as possible, this value shoud be high. Typical values are at about -40 dB for condenser microphones and at about -50 dB for dynamic microphones.

To be fair, this value is not the most important one. Its variations over incident angle and frequency are important, however (see below).

4.1.2 Frequency Response

The sensitivity with respect to the frequency has the most prominent effect to the sound that the microphone can achieve. The plot depicts in double logarithmic scale the output level in dB against the frequency with constant sound pressure level, see figure 4.1.2, page 30 for example.

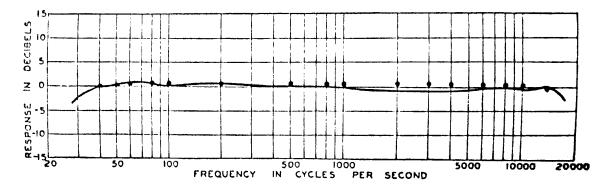


Figure 4.1: Example of a frequency curve, from [18], page 253. It is a measurement of a ribbon microphone

If there is only one plot, it is valid for a (hopefully) plane sound wave arriving from the main axis. Seldomly you may see a second plot that is measured in an angel of 180° or alternatively a plot for diffuse sound. For all usual non-unidirectional microphones that use first order spherical functions (almost all of them), the frequency response depends on the distance or, more precisely, on the curvature radius of the incoming sound waves. Very few microphones come with additional plots for several distances. Even more problematic is the absence of any kind of norm at which distance the measured curve was obtained.

Except for very small (measurement-) microphones and some broad cardioid capsules the frequency response is always dependent on the angle of the incoming sound. In coarse form this may be recalculated from the plot of the directional characteristic.

Sadly, many manufacturers, especially in the low cost market, provide plots that are strongly smoothed.

The ideal curve is flat (constant) inside the bandwidth of the microphone.

Normally the microphone should not change the sound itseld but be as precise as possible. Sound changes further up in the chain by EQ can be tuned precisely as needed and are not fixed by the hardware. Why should a certain curve be optimal for e.g. *all* voices?

Aside from this, the following deviations from a flat line can be unproblematic or even beneficial, depending on the use case:

- A rise in sensitivity before the treble cut-off. It is useful for omnidirectional microphones to counter the narrowing of the directional characteristic in the top octave if the ambience is to be recorded. Otherwise the sound is too dull without the full off-axis treble. The same effect can be achieved by an EQ, but this will slightly increase the noise level.
- A rise in sensitivity in the upper mid range by reflected sound. This may be used if this frequencies will be boosted anyway. However, if the onset of the rise in sensitivity does not happen at the optimal frequency, it is more complicated to reach the optimal overall curve by an equalizer than setting it directly from a neutral signal. The EQ must produce the possibly complex difference curve to the optimum.
- A graual (ramped) rise in high frequency response. This can counter the damping of high frequencies by a large distance. Again, contrary to a later EQ the microphone characteristic leads to less noise. This can be achieved by using a sphere around an omnidirectional microphone.
- A controlled sloped attenuation of low frequencies for large distances . On directional microphones this can counter the proximity effect (see below) for one distance, if it is clear that at bout this distance will be used. Sometimes this is switchable. If this is done by the microphone it may prevent overloading of preamps or tansformers before any EQ can be used, notably for wind noise of popp signals.

4.1.3 Directional Characteristic

The plot of the directional characteristic shows the sensitivity depending on the angle of incidence, typically using polar coordinates, an example is given in figure 4.1.3, page 32.

It is important for the user to know the direction of maximal sensitivity, which is taken as origin. With most thin cylindrical microphones this is the axis of symmetry, see figure 4.3.

For some other microphones, usually all large diaphragm microphones, the direction of maximal sensitivity is at one side of the microphone, which must be marked, see figure 4.4.

For a directional characteristic of a figure-of-eight¹ microphone, both directions of maximum sensitivity (front and rear) must be on the side, because otherwise the symmetry of front and back is not possible.

¹See section 4.2.2 page 37

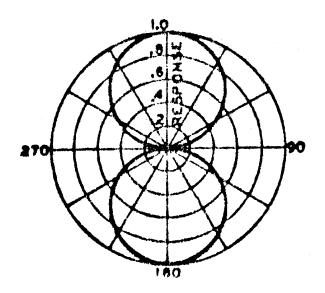


Figure 4.2: Example for a plot of a directional charactersistic, from [18], page 248

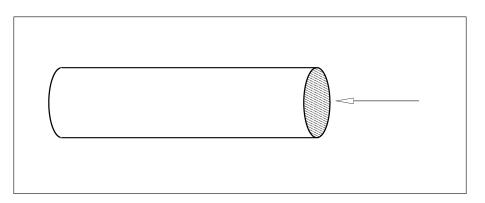


Figure 4.3: Maximum of sensitivity in front

Sometimes the sensitivity is plotted logarithmic in dB, sometimes linearly as absolute value. Care must be taken to read the plot correctly.

As a rule the directional characteristic depends on the frequency. Therefore the sound captured by different microphones may be different even if the frequency response for the main direction has been equalized to be the same. Microphones do not only pick up the direct sound from the front, but also possibly other sounds, for example, a diffuse room reverberation sound, which is colored differently by the unequal directional characteristics for all frequencies. The accompanying plot should contain the characteristic for many frequencies to be meaningful. Sometimes the plot presents only one curve for 1 KHz, which is typically the one most closely matching the nominal form.

A strong effect of frequency on the directional characteristic is very problematic for the XY stereo geometry ², because the localization becomes dependent on the frequency and thus inconsistent. It also affects the usage for live sound reinforcement. If the directivity is less prominent than the ideal for some frequency, the danger for a feedback oscillation gets worse.

Most microphones get built to fit a directional characteristic that can be derived as a combination of the ideal spherical functions of zeroth and first order, the sphere³ and

²See Chapter 5.2.2, page 68

 $^{^{3}}$ See section 4.2.1, page 36

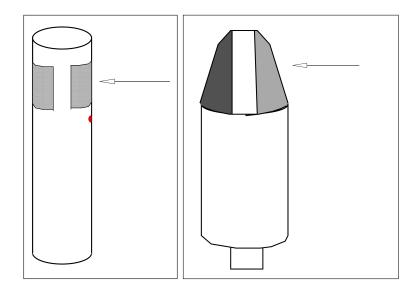


Figure 4.4: Maximum of sensitivity on the side

figure of eight⁴. See chapter 4.2, page 36. As a rule, the directional characteristic should be as constant as possible for all frequencies. Otherwise the abient sound will be colored even if the frequency response on axis is neutral.

Deviations from this uniformity are the reason why it is generally impossible to create the sound characteristic of a differenct microphone by equalization.

4.1.4 Noise Level

No microphone is able to create a noise-free output signal. Even the nominally constant air pressure is a mean pressure of extremely many particles that collide with the diaphragm because of the thermal movement. At least this pressure variation will always show up in the output signal.

The voltage of this component of the noise signal is proportional to the square root of the diaphragm radius and consequently proportional to the fourth root of the area. The voltage that is created by the input sound signal is proportional to the the diameter of the diaphragm and thus proportional to the square root of the diameter, assuming that the input is coherent over the area of the diaphragm. This results from the fact that the sound power is proportional to the square of the diameter and proportional to the area. Therefore the signal to noise ratio is proportional to the square root of the diameter of the diaphragm.

Additionally there is thermal noise from the resistance of the impedance of the microphone. For condenser microphones this impedance is proportional to $\frac{1}{f}$. Therefore this noise is most prominent in the bass. Its voltage is proportional to $\frac{1}{\sqrt{f}}$. Luckily the ear is very insensitive to low bass levels. The electric circuit of condenser microphones also adds to the total noise. In the mid and treble region the total noise of a good condenser microphone is really close to the theoretical optimum.

For dynamic microphones the electrical noise results from the resistance of the coil or the ribbon, respectively. Because of the relatively low sensitivity (compared to condenser microphones) it is very important to use microphone preamplifiers with as low as possible

 $^{^{4}}$ See section 4.2.2, page 37

4 The Microphone

inherent noise, especially for ribbon microphones.

Because the sensitivity of the ear to low level signals varies largely over differnt frequencies, the measured noise spectrum is equalized and integrated to obtain an approximation of the perceived overall level.

Two different norms are in use for the weighting of the noise level.

- You may weight the noise with the ANSI A weighting curve. The bass and the treble (to a lesser degree) are each attenuated.
- More closely matching the psycho-acoustic knowledge of the hearing process is the CCIR weighted number. It gives more weight to pulse noise and does not attenuate bass and treble as much as the other norm. Because the A-curve and CCIR weightings are different, there is no possibility to compute one of the two numbers from the other. The CCIR numbers are a lot higher than the A-curve weighted ones, however; typically at about 10 dB higher. Therefore this number is not the friend of the marketing department.

To get acoustically meaningful and comparable values, it is necessary set the noise leven in relation to acoustical levels. To do so, subtract the electrical noise level from the logarithmic sensitivity G. The result is called the signal-to-noise ratio SNR.⁵

$$SNR = G - L_{el} \tag{4.3}$$

To compare the inherent microphone noise to the level of acoustical signals one computes the level that an acoustical noise signal would have that leads to the measured electrical noise level under the assumption that the microphone is noise free. This is called the equivalent noise level L_{eq} . You get its value subtracting the SNR from 94 dB.⁶

$$L_{eq} = 94 \text{ dB} - SNR = L_{el} - G + 94 \text{ dB}$$
(4.4)

An example:

A microphone has a senitivity of -40 dBV (10 mV/Pa). The measured weighted electrical noise in silence is -116 dBV. The equivalent signal-to-noise ratio and equicalent acoustic noise level are

$$SNR = G - L_{el} = -40 \text{ dB} - (-116 \text{ dB}) = 76 \text{ dB}$$
 (4.5)

$$L_{eq} = L_{el} - G + 94 \text{ dB} = -116 \text{ dB} - (-40 \text{ dB}) + 94 \text{ dB} = 18 \text{ dB}$$
(4.6)

A typical value of the A weighted effective noise level for a 20 mm condenser microphone is in the range of 15 to 20 dB.

Needless to say: the noise level should be low as possible, the SNR as high as possible.

⁵Notice: The dynamic range that the microphone can cover is much larger, because just about any microphone can work with sound levels exceeding 94 dB. The signal-to-noise ratio and the dynamic range must not be confused.

 $^{^{6}}$ The noise is to be compared to the approximate hearing threshold at 0 dB SPL and not to 1 Pa, equivalent to +94 dB acoustical.

4.1.5 Maximum Sound Pressure Level

At a given maximum sound pressure level each microphone becomes non-linear and creates additional harmonics and intermodulation. This is not really a problem with dynamic microphones in typical use cases. The Author assumes that the high pressure level limit of dynamic microphones stems from the fact that the mechanics of a voice coil does not imply hard mechanical limitations. The author was told that in a discussion a Sennheiser developer stated that some of their microphones can deliver output signals with peaks over 200 V. Guessing that maybe it was the MD421 with a Sensitivity Sens of 2 mV/Pa, this means it can handle (however linearly) 100.000 Pa, 100 dB above 94 dB. This is about the normal athmospheric pressure. If this is correct, there is no practical limit in general operation. Condenser microphones have a buffer preamplifier in the microphone itself which has a limited maximum output voltage..

Modern condenser microphones of good quality can reach values of 130 dB, and with attenuation sometimes even 10 dB more. It is a bad idea always to use the attenuator in an attempt to be on the safe side of clipping at all circumstances. The signal gets attenuated and the noise in the amplification stays the same, diminishing the signal-to-noise ratio.

Cheap condenser microphones often have this limit much lower. Some begin to clip at 110 dB. This value is easily transgressed if a source is really close.

For comparisons it is necessary to pay attention to the stated distortion factor for the limit. 1% or 3% are both sometimes used.

With the value of MSPL and the SNR the dynamic range DR can be computed. The difference between 94 dB and the maximum pressure limit MPL has to be added, because the microphone can handle this level range, too.

$$DR = SNR + MSPL - 94 \text{ dB} \tag{4.7}$$

Let the MSPL of the microphone used as example in the noise level subsection be 130 dB, then is has a DR of

$$DR = 76 \text{ dB} + 130 \text{ dB} - 94 \text{ dB} = 112 \text{ dB}$$
(4.8)

The Maximum Sound Pressure Level MSPL should be as high as possible, and subsequently the Dynamic Range DR, too.

4.1.6 Output Impedance

Because the connection in the electro-acoustic technology is usually done by voltage matching, the output impedance should be as low as possible. This allows for long cables, because there is enough signal current not to lose audibly high frequencies through the cable capacitance, and electrical external noise is short circuited.

Condenser microphones may reach values down to 50 Ohms because of their internal amplifiers.

For dynamic microphones, low values are typically in the range between 200 and 600 Ohms.

Microphones with high output impedance for direct connection to typical valve inputs are mostly obsolete. They are only used in special cases, e.g. to feed the signal of a blues harp into a valve guitar amp, to create a typical overdriven blues or rock sound.

4.2 Directional Characteristics

In the following discussion, I will show the directional characteristic with three plots each, in plane Cartesian coordinates ⁷, in plane polar coordinates ⁸ and in three dimensional polar coordinates. I will only use linear plots for this comparison, no logarithm (dB).

All plots are normalized to one for on-axis incidence (0°) .

4.2.1 Pressure Transducer - Spherical (Omnidirectional) Characteristic

The simplest directional characteristic of all is the ideal small pressure transducer, it is independent of the angel of incidence.

The three dimensional plot shows a sphere, its cross-sectional area cut with the plane is a circle.

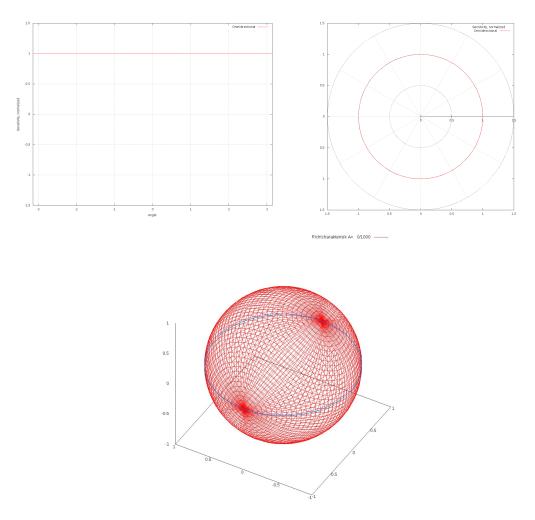


Figure 4.5: Omnidirectional

This characteristic has one very positive property, its frequency plot is independent of

⁷This plot is the clearest one to show the different rates of change of the sensitivity over the angle.

 $^{^8 {\}rm The\ most\ commonly\ used\ plot}.$

the distance to the source.⁹

As the size of the transducer is finite, there is an interaction with the sound field for high frequencies. There, the dimensions get comparable to the wave length.

The main effects are

- **Reflection:** The sound that reaches the front of the diaphragm gets reflected more and more for diminishing wave lenghts. The compliance of the diaphragm is much smaller than that of the air. Therefore it has the effect of a rigid surface, stopping the air movement of the sound wave (almost...) completely. This is mathematically the same as if a wave is added to the sound field in front of the diaphragm that has the same properties as the incoming sound but moves in the opposite direction. The velocity of the air movement at the place of the diaphragm cancels (as needed) and the momentary pressure adds, the pressure doubles (ca. +6 dB). For long wavelenghts this effect does not show because the air has time to move to the side, where it is not stopped. This is not possible if the phase changes noticably before the air can move to the circumference of the diaphragm. Any movement to the sides arrives to late.
- **Interference:** Sound waves that arrive from a non-vanishing angle to the axis reach different places of the diaphragm at different times, because they have different path lenghts. This means that different places on the diaphragm see different phases of the wave. The total signal will be a weighted moving average over the wave form, which flattens the wave form out. This diminishes the sensitivity high frequencies. As result the directional characteristic changes for high frequencies. It gets narrower and the backside is more and more suppressed (shadow effect).

For common small diaphragm condenser microphones with a diameter of at about 20 mm, both effects are fully developed above 17 KHz. This leaves a choice to the developer.

- **Free field equalization**: The microphone may be built in a way to make the frequency response from 0° as linear as possible. The treble response in all other directions is lowered, also the mean frequency response for all directions, the diffuse frequency response.
- **Diffuse field equalization:** The microphone is built in a way to raise the sensitivity for the treble range to linearize the the diffuse frequency response.

Both equalizations do not change the relative response for all angles - they are mainly determined by geometry. Therefore each microphone can be changed in this parameter even after a recording solely by using an equalizer in the treble range.

The free-field equalized microphone gets an optimal signal for a sound source that is closely located in front of it, the diffuse-field equalization is optimal to pick up a diffusely distributed sound field, such as room reverberation.

4.2.2 Pressure Gradient Transducer - Figure of Eight, Proximity Effect

The first order of directional characteristic in the theory of spherical functions is created by a transceiver that reacts to the gradient of air pressure. The gradient of air pressure is

⁹Strictly speaking – it is independent of the curvature of the phase front of incoming sound. For small sources this in turn is equal to the reciprocal distance to the sound source.

the derivative of the pressure function in relation to a given direction. It is proportional to the projection of the incidence angle α on this direction (say, the x axis), which leads to a factor of its cosine to the sensitivity relative to that of the main direction.

$$E_{\rm rel}(\alpha) = \cos(\alpha) \tag{4.9}$$

If the result is plotted in a spherical coordinate system, it results in a circle of radius $\frac{1}{2}$ around $(\frac{1}{2}, 0, 0)$, the negative part is plotted exactly over the positive one. If you plot the absolute value instead to show that in the reverse direction the sensitivity has the same absolute value, the plot results in two circles around $(-\frac{1}{2}, 0, 0)$ and $(\frac{1}{2}, 0, 0)$, the famous figure of eight in two dimensions.

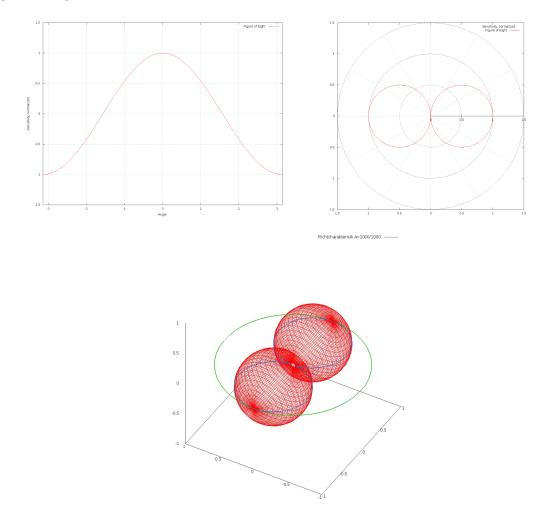


Figure 4.6: Figure Of Eight

A problem arises from the fact that the derivation to the location adds an inner derivative of $\frac{2\pi}{\lambda}$. λ is the wave length. If the speed of sound is constant with respect to the frequency, this factor can be written as $\frac{2\pi f}{c}$. If the construction were to really deliver a signal that is purely proportional to pressure derivative, it would have almost only treble. The level difference between 20 KHz and 20 Hz would be a linear factor of 1000 for the voltage, logarithmic this is 60 dB¹⁰.

¹⁰According to the rule of thumb, there are at about 10 octaves (1024 instead of 1000) with about 6 dB each.

Additionally this introduces a phase change of 90° to the signal compared to the pressure signal.

Both can be compensated by countering the spatial derivative with an integral over time. Mechanically this can be done for dynamic (including ribbon) microphones by lowering the resonance frequency of the diaphragm at the lower end of usable frequency range by making the suspension very weak.

For condenser microphone the resonance has to be in the middle of the usable frequency range and the resonance has to be broadened by mechanical or acoustical damping. A movement that is controlled by damping is proportional to the pressure in this case. The frequency region may be broadened by adding additional resonances at the limits.

As result of this measures each microphone that at least partly works as gradient transceiver is a lot more sensitive to air blows and mechanical shock than pure pressure transducers.

The principles of this kind of transceiver can also be seen by using differences on short distances instead of the derivative. In fact this is even closer to the real circumstances, because the finite size of the diaphragm causes sound coming from the forward direction to reach the back side after a detour of length between zero and the diameter of the diaphragm (mid to side in the front, side to mid in the back) with the corresponding delay. As approximation we assume one single effective detour and assume it to be determined by the radius of the diaphragm including frame. Additionally the diaphragm may be put into a short pipe to increase the length of the detour. Care must be taken regarding the resulting pipe resonances.

For low frequencies and large wave lengths the results are the same as with the derivative, with the possible exception of constant factors. If for high frequencies the detour gets in the order of magnitude of a quarter wave length, there are large differences in the results. The increase in sensitivity because of the inner derivative is decreased an with even larger relative detours the sensitivity is even diminished itself. Therefore the whole construction has an upper limit of frequency, determined by the size. The author is not aware of a widely used type of pure pressure gradient microphone whose frequency range extends above 15 KHz. enddocument A derivation of the results for difference transducers can be found in a book by Olson.¹¹ If you consider the pressure difference between two circular diaphragms with distance D, equivalent to a cylinder of length D that is freely movable in axial direction, with circular area of S with frequency f, wave length λ , wave number $k = 2\pi/\lambda$, pressure amplitude p_m , incidence angle Θ , you get for the two pressures p_1 , p_2 and their difference Δp

$$p_1 = p_m \sin\left(k\left(ct + \frac{D}{2}\cos\Theta\right)\right) \tag{4.10}$$

enddocument

$$p_2 = p_m \sin\left(k\left(ct - \frac{D}{2}\cos\Theta\right)\right) \tag{4.11}$$

enddocument an from that

$$\Delta p = p_1 - p_2 = 2p_m \cos(kct) \sin\left(\frac{kD}{2}\cos\Theta\right) = 2p_m \cos(2\pi ft) \sin\left(\frac{2\pi D}{2\lambda}\cos\Theta\right) \quad (4.12)$$

¹¹See Olson, [18], pp. 237-253

end
document The force on the cylinder / the diaphragm f_m
is found by multiplication with the area end
document

$$f_m = S\Delta p = 2Sp_m\cos(2\pi ft)\sin\left(\frac{2\pi D}{2\lambda}\cos\Theta\right) = 2Sp_m\cos(2\pi ft)\sin\left(\frac{\pi fD}{c}\cos\Theta\right)$$
(4.13)

end document The velocity v_m of a free mass m (or a mass above the resonance frequency) is given by end document

$$v_m = \frac{Sp_m}{m\pi f} \sin(2\pi ft) \sin\left(\frac{\pi fD}{c}\cos\Theta\right)$$
(4.14)

It can clearly be seen, that the phase is equal to that of a pure pressure transducers because of the time integration by the mass (sine instead if cosine).

If you plot the frequency response for the reference angle 0°, normalized at a starting amplitude $v_m(0)=1$, you get

$$v_m/v_m(0) = \frac{\sin(\frac{\pi fD}{c})}{f} = \frac{\sin(\frac{f}{f_0})}{f}$$
 (4.15)

with $f_0 = \frac{c}{\pi D}$.

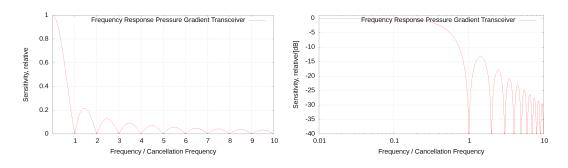


Figure 4.7: Gradient receiver frequency response, linear and logarithmic

For real pressure gradient transducers, things are a bit different. Instead of a cylinder of finite length, there is a diaphragm with its surrounding construction (for ribbon microphones including the magnet setup). The whole construction causes a path difference for incoming sound between front and back. Even for the ideal circular construction the correct computation of the effects needs sums over e.g. Bessel functions.¹² Interestingly, the resulting plots are extremely similar to those of the approximation shown above, and real measurements are reasonably close to the computed values. The first cancellation happens for $r = \lambda$, for the wavelength that matches the diameter. For 20 KHz this happens at 17 mm. The frequency limit in the sense of -6 dB relative sensitivity lies at a factor of 0.6 without pressure increase by reflection, this is 12 KHz in the given example. With every root the polarity changes. Additionally the directivity index is deformed including double peaks to the side with cancellation in the main direction. As result only the frequency region until the first root us usable at all.

¹²See Sivian et. al.[25], citation in Olson[18], p. 243

Proximity Effect

If the sound source is close the derivative or difference in sound pressure does not only stem from the phase difference but also from the change of sound pressure level because of the distance difference. If the source is far away compared to its size, the pressure at distance s is proportional to 1/s.¹³ The most important result of this additional component of the derivative is its frequency independence. The compensation in the frequency response is not necessary to linearize it. Because of this compensation this leads to a component of the signal whose frequency response in the bass region rises proportional to the reciprocal frequency, that is 6 dB per octave. This is the dreaded proximity effect.

To calculate the magnitude of the increase in sensitivity it is completely sufficient to take the derivative of the sound pressure. In the frequency region that shows the effect the wave lengths are extremely large compared to a typical microphone capsule, so the approximation does not lead to a significant error. Another positive side effect of this approach is that its result is universal, independent of the size of the microphone.

The pressure field of a sine sound wave emanating from the origin

$$p(r,t) = p_1 \frac{\sin(\omega(t-t_0) - kr)}{r}$$
(4.16)

The angular frequency is ω , the wave vector k, its absolute value is $\frac{2\pi}{\lambda}$, the wave length λ and the radius r.

The gradient transducer picks up the derivative. We assume it is directed at the source, therefore we can work with scalars with r in one dimension.

The derivative with respect to x is given as

$$p'(r,t) = p_1 \frac{-kr\cos(\omega(t-t_0) - kr) - \sin(\omega(t-t_0) - kr))}{r^2}$$
(4.17)

Without the term that vanishes for large r you have the sensitivity that corresponds to the frequency response without proximity effect:

$$p'_{\text{distant}}(r,t) = p_1 \frac{-kr\cos(\omega(t-t_0)-kr)}{r^2}$$
 (4.18)

The quotient is exactly the sensitivity rise by proximity. Here we are interested in the absolute value. Because the sin and cos components differ in phase by 90°, their squares add (Pythagoras), which leads to the quotient

$$\frac{|p'|(r)}{|p'_{\text{distant}}|(r)} = \frac{\sqrt{(kr)^2 + 1}}{kr} = \frac{\sqrt{(\frac{2\pi r}{\lambda})^2 + 1}}{\frac{2\pi r}{\lambda}}$$
(4.19)

The level difference is given by

$$\Delta L_{\text{proxFigureOfEight}} = 20 \log_{10} \left(\frac{\sqrt{(kr)^2 + 1)}}{kr} \right) = 20 \log_{10} \left(\frac{\sqrt{(\frac{2\pi r}{\lambda})^2 + 1}}{\frac{2\pi r}{\lambda}} \right)$$
(4.20)

A plot is given below, chapter 4.2.3, Page 45.

¹³The power gets distributed to a part of the area of a sphere, that leads to a factor proportional to $1/s^2$. The pressure is proportional to its square root.

4 The Microphone

Apart from this theoretical view it is really hard to forecast how large the proximity effect will be in a given situation. The deviation rises from the fact that many real sound sources are far away from the ideal point like source. The deciding parameter is the curvature of the phase front at the location of the diaphragm. This can be very complicated, just think about the situation close to a large diaphragm e.g. a drum. The mouth is also very different from a point source in the proximity region, so the proximity effect can hardly be reliably computed for the typical case of a vocal microphone for on stage use. In this special case the size of the source is even changing all the time, and the distance remains rarely constant and all this affects the frequency response.

Those effects are all caused only by the gradient transducer; a pressure transducer (Omnidirectional) is free of such effects. For this reason, one should consider whether the use of omnidirectional microphones is not the better approach in some cases. One example is a miniature omnidirectional vocal microphone that is fixed at the head in direct proximity beside the mouth.

Not to consider the bass boost by the proximity effect and not to compensate it, even to develop a taste that likes this special sound is one of the most grave errors (most of the time, but not always from newcomers) in the use of audio technology. If it affects several sound sources it leads to massive occlution of signals and destroys any chance to achieve a transparent overall sound.

On the other hand the effect is not always bad. It is an interesting means to counter noise with large amounts of low bass in the surrounding (Machines,...). They get not only lowered by the directivity characteristic, but the bass cut that counters the proximity effect lowers them even more without cutting away the bass in the signal proper.

If a breathing mouth is really close or in open air with wind these microphones display yet another problematic effect. Vortexes and pressure peaks by plosives there is a turbulent air motion directly at the diaphragm. They result in pressure peaks at more or less nonexistent distance and only on one side of the diaphragm. This results in (very) low frequency noise with exceptional high level. They can easily overdrive transformers or preamplifiers, long before any kind of equalizer can do something to prevent this. In all cases where this danger is present, it is absolutely necessary to use a wind screen. Most typical microphones for close voice pick up on stage have this component built in either as open plastic foam or wire meshes. This may put limits on the sensitivity for extreme treble. The protection must not be too thin, because otherwise the remaining vortexes are still too close to the diaphragm and get amplified by the remaining proximity effect. An ideal pop screen is therefore ideally rather thick and empty in direct proximity of the diaphragm to allow for a undisturbed propagation of pressure directly at the location of the microphone.¹⁴ Ribbon microphones are easily destroyed by large pressure shocks.

Not many pressure gradient transducers put the lowest usable frequency at the technically possible limit. Microphone for on stage vocals often cut off far above 100 Hz to tame the proximity effect, be less mechanically sensitive and lower the sensitivity for plosives. Models that transmit below 50 Hz are the exception from the rule.

Because the frequency characteristic is dependent on the distance, the manufacturers should deliver plots for several distances or at the very least give the distance at which the plot was obtained. Both is almost never done.

For pure pressure gradient transducers, the interference for sound from the side does not play an important role compared to other directional characteristics, because sound from that direction is canceled. In the plot the figure of eight stays rather stable with

¹⁴Mikrophonaufsätze, [36]

varying frequencies, the lobes just get a bit thinner for high frequencies.

If this kind of microphones is used to cut out noise sources you have to keep in mind the the cancellation happens at a very small interval of angles. The (unusual) Cartesian plot of the directional characteristic shows that the cancellation happens exactly a the point of maximal slope. This means that even for small deviations of angle the cancellation becomes weak. You have to control the angle very precisely and the source must not cover a wide angle.

The pure pressure gradient transceiver is not used very often, but a thorough understanding is necessary to understand the characteristics of the other common directional characteristics. The main use is the Blumlein setup ¹⁵ and the MS-setup¹⁶.

The mean senitivity over all angles in space for a directive microphone in a diffuse sound field can be compared to that of a frontally incident plane sound wave of the same sound pressure level, which is normalized to one. To achieve this, the square of the relative (voltage-) sensitivity is integrated over all directions in space and the result is divided by the total space angle 4π . The resulting value is called Random Energy Efficiency, abbreviated as REE.

$$REE = \frac{1}{4\pi} \int_0^{\pi} E_{\text{rel}}^2(\alpha) \sin(\alpha) \int_0^{2\pi} d\theta \, d\alpha = \frac{1}{4\pi} \int_0^{\pi} E_{\text{rel}}^2(\alpha) \sin(\alpha) 2\pi \, d\alpha \tag{4.21}$$

and thus

$$REE = \frac{1}{2} \int_0^{\pi} E_{\text{rel}}^2(\alpha) \sin(\alpha) \,\mathrm{d}\alpha \tag{4.22}$$

The reciprocal value is sometimes given the symbol γ and called directivity factor.

$$\gamma = \frac{1}{REE} \tag{4.23}$$

Its square root has an important practical meaning. In a mixture of direct and diffuse sound the directional microphone can be at a distance from the source that is exactly that factor larger than the distance of an omnidirectional microphone if both have the same ratio of direct and diffuse signal, hence its name, distance factor DSF.

$$DSF = \sqrt{\gamma} \tag{4.24}$$

For the pressure gradient transducer you get

$$REE_{\text{grad}} = \frac{1}{2} \int_0^{\pi} \cos^2(\alpha) \sin(\alpha) \, \mathrm{d}\alpha \tag{4.25}$$

$$= \frac{1}{2} \int_0^{\pi} (\frac{1}{2} + \frac{1}{2}\cos(2\alpha))\sin(\alpha) \,\mathrm{d}\alpha$$
 (4.26)

$$= \frac{1}{2} \int_0^{\pi} \frac{1}{2} \sin(\alpha) \,\mathrm{d}\alpha + \frac{1}{2} \int_0^{\pi} \cos(2\alpha)) \sin(\alpha) \,\mathrm{d}\alpha \tag{4.27}$$

$$= \frac{1}{2} + \frac{1}{2} \int_0^{\pi} (-\frac{1}{2}) \sin(\alpha) + \frac{1}{2} \sin(3\alpha) \, \mathrm{d}\alpha \tag{4.28}$$

$$= \frac{1}{2} - \frac{1}{4} \int_0^{\pi} \sin(\alpha) \, \mathrm{d}\alpha + \int_0^{\pi} \frac{1}{2} \sin(3\alpha) \, \mathrm{d}\alpha \tag{4.29}$$

 15 See section 5.2.2, page 68

 $^{^{16}}$ See section 5.2.2, page 71

$$=\frac{1}{2} - \frac{1}{2} + \frac{1}{3} \tag{4.30}$$

$$REE_{\text{grad}} = \frac{1}{3} \tag{4.31}$$

This leads to $\gamma_{\text{grad}} = 3$ and the distance factor DSF of a pure gradient transducer is $\sqrt{3}$, at about 1.73.

4.2.3 Combinations - Cardioid, Super cardioid, Hyper cardioid, Broad Cardioid

Mathematically speaking, the directional characteristics create a vector space and you can create linear combinations of them. In short: you may add and subtract them and multiply them with weighting factors. This can be done in real life by putting two microphones at one place as good as possible. If the sources are in a plane, you may put them in the normal vector one over the other, so that they have the same distance and consequently phase from each source. Now you may do the calculations with the output signals. To subtract, you can reverse the polarity on one signal. If you change the levels, you can obtain all sums and differences, all linear combinations in the mathematical view. With this method you can even change effective directional characteristics after a recording, this is called matrixing.

All classical directivity characteristics that can be obtained without interference effects can be built by linear combinations of omnidirectional and figure-of-eight signals. The resulting characteristic can be described by the parameter A, which is between zero and one by the equation

$$E_{\text{rel}}(\alpha) = (1 - A) + A\cos(\alpha) \tag{4.32}$$

All of these characteristics have one thing in common: Because the figure-of-eight has a root (value of 0) at 90°, only the omnidirectional component is active at this angle. In this angle there is no proximity effect. Its relative level corresponds to the factor of the omnidirectional component (1 - A). In the case of the cardioid (see below) it is exactly 0.5 or -6 dB. If you want to take advantage of this fact, keep in mind that the frequency response for this angle is less than optimal because of interference, especially in the treble region, as explained in detail for omnidirectional charactereistics above.

A microphone with a single capsule, and in most cases a fixed directional characteristic in the region of these linear combinations, is constructed by having a diaphragm open at the front side and closing it partially on the back side. This partial closing is done by small holes in a back chamber combined with components for damping and retardation. The whole setup is optimized to get the target directional characteristic as close as possible and, usually, as independent from the frequency as possible.

The component of the figure-of-eight characteristic shows its proximity effect (increased sensitivity for bass) also in the combination corresponding to its weight A in the resulting characteristic. This increased value of A results in a change of the characteristic for low frequencies and close distances, because it shifts more and more in the direction of the figure-of-eight. As example, also an ideal cardioid picks up bass from the back side for close sources, because the cancellation in this direction is no longer effective. The

cancellation happens for a value of A of 0.5, but the proximity effect increases A for low frequencies.

Here is an equation for the general proximity effect.

$$\Delta L_{\text{nah}}(A) = 20 \log_{10} \left(1 - A + A \frac{\sqrt{(kr)^2 + 1}}{kr} \right) = 20 \log_{10} \left(1 - A + A \frac{\sqrt{(\frac{2\pi r}{\lambda})^2 + 1}}{\frac{2\pi r}{\lambda}} \right)$$
(4.33)

In figure 4.2.3 this is shown for the figure-of-eight and for the most common case of the cardioid (see below, A=0.5).

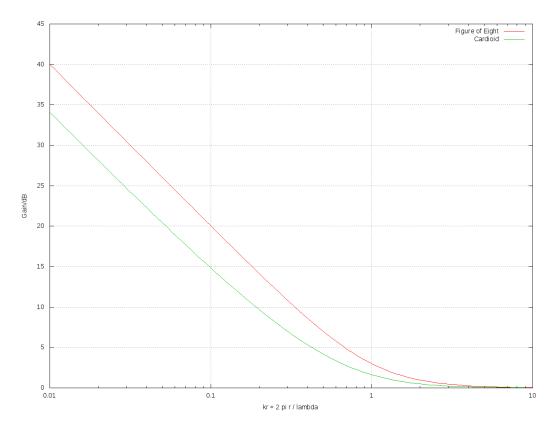


Figure 4.8: Proximity effect

For a distance of 5.4 cm the kr values have to be multiplied by 1000 to obtain the corresponding frequency in Hz, for 10.8 cm with a factor of 500, for 54 cm with a factor of 100, etc.

The other way round: To find the distances that belong to the given value of kr for typical frequencies, here is a tabular of the distance r in cm.

| | 0,01 | $0,\!02$ | $0,\!05$ | $0,\!1$ | $0,\!2$ | 0,5 | 1 | 2 | 5 | kr |
|-------------------|------------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----|
| 20 Hz | 2,706 | 5,411 | $13,\!53$ | 27,06 | 54,11 | 135,3 | $270,\!6$ | 541,1 | 1353 | |
| $50 \mathrm{~Hz}$ | 1,082 | $2,\!164$ | $5,\!411$ | $10,\!82$ | $21,\!64$ | $54,\!11$ | 108,2 | 216,4 | 541,1 | |
| 100 Hz | $0,\!5411$ | 1,082 | 2,706 | 5.411 | $10,\!82$ | 27,06 | $54,\!11$ | 108,2 | 270,5 | |
| $150~\mathrm{Hz}$ | 0,361 | 0,722 | $1,\!81$ | $3,\!61$ | $7,\!22$ | 18,1 | 36,1 | 72,2 | 181 | |
| $200~{\rm Hz}$ | $0,\!2706$ | $0,\!5411$ | $1,\!353$ | 2,706 | $5,\!411$ | $13,\!53$ | $27,\!06$ | $54,\!11$ | $135,\!3$ | |

With changing parameter A, the values for REE, γ and, of course, DFS get also changed.

$$REE(A) = \frac{1}{2} \int_0^{\pi} ((1 - A + A\cos(\alpha))^2 \sin(\alpha) \, \mathrm{d}\alpha$$
 (4.34)

$$= \frac{1}{2} \int_0^{\pi} ((1-A)^2 + 2(1-A)A\cos(\alpha) + A^2\cos^2(\alpha))\sin(\alpha) \,\mathrm{d}\alpha \tag{4.35}$$

$$= \frac{1}{2} \int_0^{\pi} ((1-A)^2 + 2(1-A)A\cos(\alpha) + A^2 - A^2\cos^2(\alpha))\sin(\alpha)\,\mathrm{d}\alpha \tag{4.36}$$

$$= \frac{1}{2} \int_0^{\pi} ((1-A)^2 + A^2 + 2(1-A)\cos(\alpha) + A^2\cos^2(\alpha))\sin(\alpha) \,\mathrm{d}\alpha \tag{4.37}$$

$$= \frac{1}{2} \int_0^{\pi} (1 - 2A + A^2 + A^2 + 2(1 - A)\cos(\alpha) + A^2\cos^2(\alpha))\sin(\alpha) \,\mathrm{d}\alpha \tag{4.38}$$

$$= \frac{1}{2}(1 - 2A + 2A^2) \int_0^{\pi} \sin(\alpha) \, \mathrm{d}\alpha + (1 - A) \int_0^{\pi} \cos(\alpha) \sin(\alpha) \, \mathrm{d}\alpha + \frac{A^2}{2} \int_0^{\pi} \cos^2(\alpha) \sin(\alpha) \, \mathrm{d}\alpha$$
(4.39)

$$= \frac{1}{2}(1-2A+2A^2)\int_0^{\pi}\sin(\alpha)\,\mathrm{d}\alpha + (1-A)\int_0^{\pi}\cos(\alpha)\sin(\alpha)\,\mathrm{d}\alpha + \frac{A^2}{2}\int_0^{\pi}\sin(\alpha)\,\mathrm{d}\alpha - \frac{A^2}{2}\int_0^{\pi}\sin^3(\alpha)\,\mathrm{d}\alpha + \frac{A^2}{2}$$

$$= 1 - 2A + 2A^2 - A^2 + A^2 \cdot \frac{1}{3}$$
(4.41)

$$REE = \frac{4}{3}A^2 - 2A + 1 \tag{4.42}$$

$$\gamma = \frac{1}{\frac{4}{3}A^2 - 2A + 1} \tag{4.43}$$

$$DSF = \frac{1}{\sqrt{\frac{4}{3}A^2 - 2A + 1}} \tag{4.44}$$

Here is a plot of these values depending on the parameter A. The most commonly used directional characteristics that will be explained in more detail below are marked.¹⁷

As discussed above, microphones with figure-of-eight characteristic that have usual diameters lead to a high frequency cutoff around 15 KHz. The combinations have one advantage, they can use the cutoff to blend over to an omnidirectional operation for high frequencies in which the microphone is directional by interference effects. The cutoff frequency of the gradient part as well as its contribution can be tailored by choosing the length of the detour that the backwards sound has to take. If its length is increased, the path difference and subsequently the pressure difference is increased, too. At the

http://www.sengpielaudio.com/TheoretischeMikrofondaten.pdf.

 $^{^{17}\}mathrm{A}$ detailed overview for all directional characteristics, including quotient of frontal and backwards sensitivity etc. can be found at

The definitions of the data is in

 $[\]verb+http://www.sengpielaudio.com/ErklaerungZuMikrofondaten.pdf.$

The equations in

http://www.sengpielaudio.com/FormelnZumPolardiagramm.pdf, everything at [24]

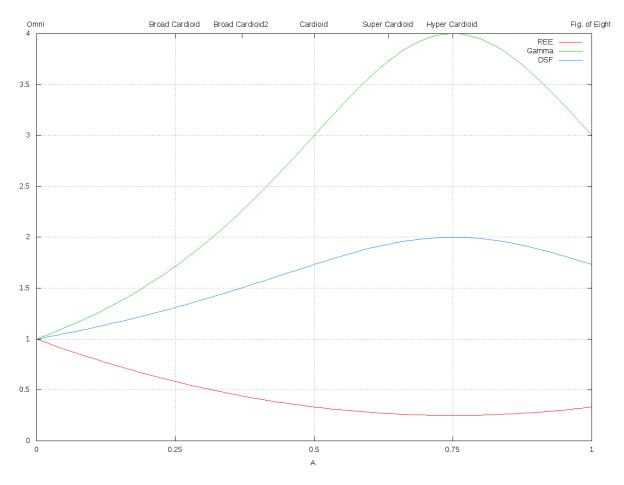


Figure 4.9: Random Energy Efficiency REE, its reciprocal value Directivity Factor Gamma and the Distance Factor DSF plotted versus parameter A

same time the cutoff frequency of the gradient part is lowered. Care must be taken that this cutoff frequency is not too low, otherwise a frequency region above the cutoff is not supressing sound from the back side as much as expected by the nominal charactersitic, it gets broader. This leads to a colored, uneven sound of ambient signals and it makes the microphone completely unusable for sound reinforcement, because this is a recipe to destroy the possible gain before feedback which will start in this transition region of frequencies.

The Cardioid results from equal weight for the omnidirectional and figure-of-eight characteristics. The equation becomes

$$E_{\rm rel}(\alpha) = 0.5 + 0.5\cos(\alpha)$$
 (4.45)

This characteristic cancels sound from 180°. Because this is the place where the cosine function has broad minimum, the angular region of cancellation is relatively broad.

The distance factor DSF is $\sqrt{3}$, ca. 1.73, the same value as for the figure-of-eight.

Most microphones for studio or stage use are built with this fixed characteristic.

The Super Cardioid is the result of optimizing the quotient of overall sensitivity from forward to backwards directions. The cardioid has a broad minimum, but only in one direction. If the figure-of-eight weight gets higher, the cancellation happens for a whole

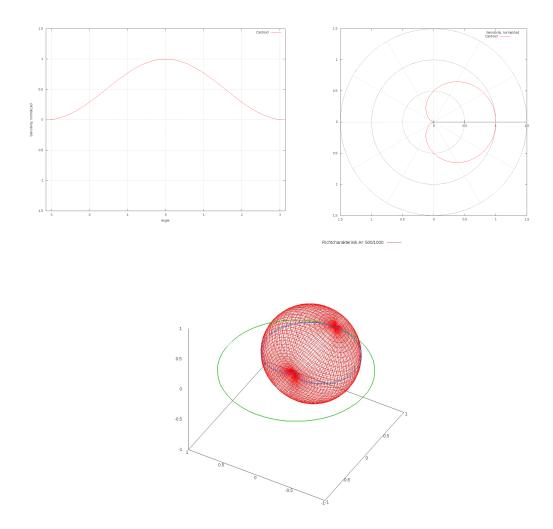


Figure 4.10: Cardioid

cone behind the microphone, but with a smaller region of very low sensitivity in the angle relative to the main direction. Geometrically this can be visualized at about as shown in figure 4.11.

This maximal quotient of sensitivities between front and back is achieved with a complicated value for A and the equation

$$E_{\rm rel}(\alpha) = 0.366025 + 0.633975\cos(\alpha) \tag{4.46}$$

The cancellation happens at ca. 120° .

Notice: a small angular region around 180° has a small secondary maximum. Like the figure-of-eight, the super cardioid picks up sound from this direction with the inverse polarity.

The distance factor DSF is at about 1,93.

The Hyper Cardioid optimizes the suppression of diffuse sound relative to direct sound from the main angle. The weight of the figure-of-eight characteristic is raised once more.

$$E_{\rm rel}(\alpha) = 0.25 + 0.75\cos(\alpha)$$
 (4.47)

The cancellation happens an angle of about $109,5^\circ.$

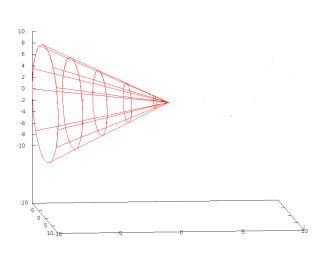


Figure 4.11: Direction of cancellation by super cardioid

The region of secondary maximum in the backwards direction is broader and more sensitive than in the super cardioid case, therefore the front to back quotient is less strong.

The distance factor DSF is 2, the maximum for all combinations of pressure and gradient transducers.

The Broad Cardioid lies between the cardioid and the omnidirectional characteristic. It shows the downsides of the gradient transducer in a very small amount. For some models it was possible to avoid the changes of the characteristic for high frequencies almost completely although the diaphragm has a diameter of ca. 20 mm. This characteristics lowers the sensitivity for backward sounds, but there is no direction with zero sensitivity.

Exactly between omnidirectional and figure-of-eight is the equation

$$E_{\rm rel}(\alpha) = 0.75 + 0.25\cos(\alpha) \tag{4.48}$$

Most of the times the constructions are closer to the cardioid with the equation

$$E_{\rm rel}(\alpha) = 0.63 + 0.37\cos(\alpha) \tag{4.49}$$

The distance factor DSF is still at about 1.5.

Comparison of all of these characteristics in the same visualization can be seen in figure 4.15.

There is also the matching 3D visualization as video[41]. Without the change in time this cannot really be shown very clearly.

Variable Directivity Characteristic, Mechanical may be achieved by opening and closing of valves. The requirements for the mechanical construction are very hard. This

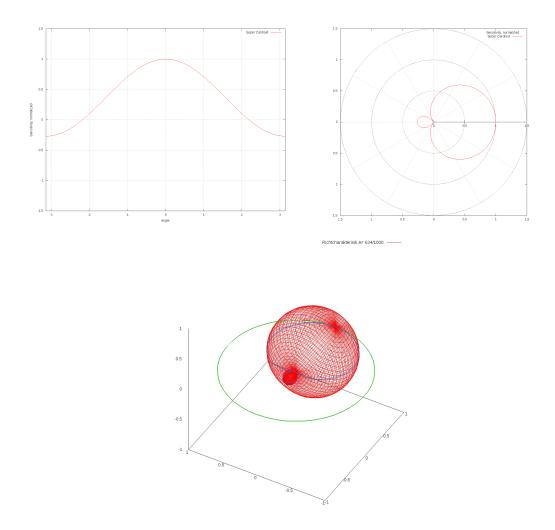


Figure 4.12: Super cardioid

is the reason why this method is not often used. The advantage is that the whole construction works with one capsule. This allows to keep the rotational symmetry and to avoid complications by interference between two distinct capsules.¹⁸.

Variable Directivity Characteristic, Electronical may be achieved with several capsules or at least diaphragms. As a rule the base is two cardioid capsules that are mounted in opposing directions.

You may achieve the same effect by mounting one cardioid microphone over another one in opposing direction. The polarity of one is used inverted.

Most practical realizations use one combined capsule with two diaphragms that are mounted very closely parallel to each other with a damping wall between both. If it is correctly done, each diaphragm delivers a cardioid pattern. In the deep bass region the separation of both diaphragms gets weak and as result the characteristics get changed into an omnidirectional one. Sometimes this is even preferred to get a weaker proximity effect.

¹⁸Schoeps builds a capsule that may be switched between omnidirectional and cardioid characteristic for the Colette system, the MK5. The MK6 goes even farther: it uses an input from the side and adds a figure-of-eight characteristic.

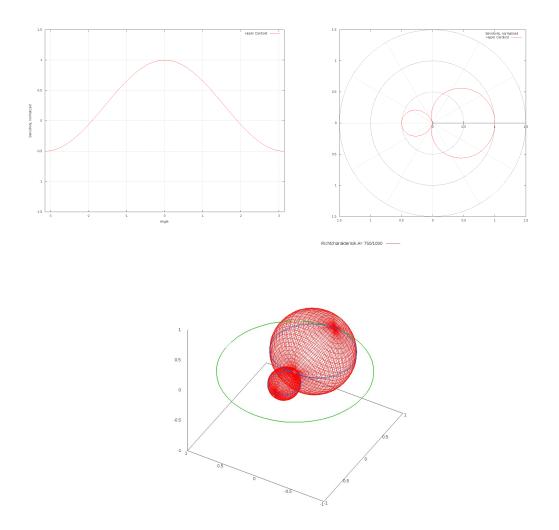


Figure 4.13: Hyper cardioid

Both diaphragms or capsules may be connected to a distinct impedance buffering amplifier and connected separately.¹⁹.

Normally the microphone creates the sum or difference of both signals. This can be done in the buffering circuit or by controlling the polarizing voltages of both diaphragms. For fixed voltages this allows the following combinations and directional characteristics:

$$signal_{Cardioid} = singal_{Cardioid1}$$
 (4.50)

$$signal_{Omnidirectional} = singal_{Cardioid1} + singal_{Cardioid2}$$
 (4.51)

$$signal_{Figure of Eight} = singal_{Cardioid1} - singal_{Cardioid2}$$
 (4.52)

The finite size of the combination capsule leads to possible cancellations of higher mid range frequencies in the case of the omnidirectional characteristic. Frontally incident waves get diffracted round the capsule and reach the back side diaphragm with a delay and resulting phase change. This is no problem in the bass region where the phase is effectively the same and also not in the treble range because the incoming wave is reflected and only a small signal reaches the back side. In between there is a frequency region where the phase difference may reach 180°, which weakens the sum signal. This

¹⁹e.g. Sennheiser MHK 800 Twin

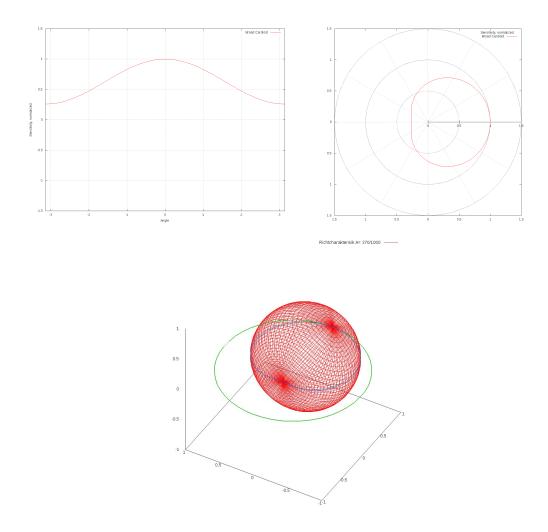


Figure 4.14: Broad cardioid

happens in the frequency range where our ear is very sensitive to deviations from the ideal. Before using such a combination microphone as omnidirectional device it is wise to be sure that the effect is not problematic for the task at hand.²⁰

If the weight of one cardioid may be smoothly changed between +1 and -1 you get all directivity characteristics that were described above, including all that lie in between.²¹

The Polarflex Concept created by Schoeps is the most flexible way to handle the directivity characteristics. Two microphones deliver the base to build any of the standard characteristics. One omnidirectional and one figure-of-eight microphone or two cardioid microphones are used, one above the other. In a matrix the figure-of-eight component is taken and can be weighted with a flexible frequency response. This allows the creation of frequency selective characteristics, for example the shift to omnidirectional characteristics in the bass only. Unlike the back to back capsules the transition may be tailored as needed.

²⁰Taken from a presentation by Jörg Wuttke.

²¹An example with 5 characteristics: Sennheiser MKM 800 P48, continually variable with large diaphragms: AKG C414, Røde NT2000

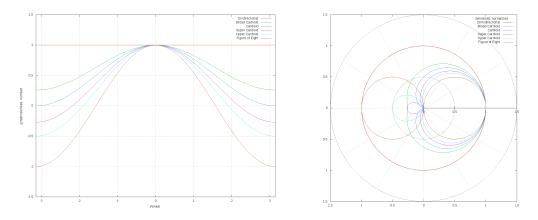


Figure 4.15: All characteristics compared

4.2.4 Combination with Interference Effects: Shotgun, Arrays

As noted before, for high frequencies the directivity changes for all capsules that are not extremely small because of interference effects. It gets more narrow. This is more so for large diaphragms. The term "large" is traditionally used for diameters larger than 2.5 cm or 1".

The interference may also be used explicitly to change the directional characteristic, as a rule to make it more narrow. Without precaution this leads to a narrowing of the characteristics with growing frequency. This means that sound from the sides is strongly colored (dull). Jörg Wuttke therefore recommends to consider the usage of super or hyper cardioids which do not necessarily show this effect.

Shotgun

A pipe with slits diffracts the sound waves on each obstacle in all directions. For sound arriving from an angle this means destructive interference. This lobe of maximum sensitivity gets therefore sharper for high frequencies than for any of the standard directional characteristics. For geometrical reasons the width of the lobe is inverse to the length of the pipe. Secondary lobes to the side are typical, as well as an extreme frequency dependence of the directivity.

Arrays

Another variant approach to achieve narrow directivity characteristics is the combination of several microphone capsules. If the signals are added, the base length increases and the main lobe of maximum sensitivity can get arbitrarily narrow. Like all interference effects it is important to keep in mind that the directivity characteristics results from the Fourier transformation of the local distribution of sensitivity. If the capsules are arranged in an uniform lattice, this leads at least for the treble region to a sensitivity characteristic that also has the form of a lattice. This effect is called lobing. This can be countered to a degree by using non-uniform distances between the capsules, the outer ones should be spread out more.

If you transgress the simple constant addition of the signals, you may vary the directional characteristic in wide ranges. You may decrease the treble for the outer capsules to approximate a geometry that is proportional to the wave length of the sound and by this a directional characteristic that is less dependent on the frequency. To compensate the inherent loss in overall treble, its level has to be raised for the inner capsule(s).

If DSP technology is used, you may go even further. By controlled delay of the signals of capsules the angle can be controlled at which for example the maximum of sensitivity shall be. This can be done without moving anything mechanically, you also compute several distinct signals for distinct angles. Taken to the extreme, this allows to create an acoustical image of the sources. This is extremely helpful when combined with a superimposed optical image to find sources of noise with high precision. [1]

The array may be one-dimensional, this means that the directional characteristic perpendicular to its axis is that of a single capsule. His construction is used vertically for speakers 22 and horizontally at conference desks 23 .

4.2.5 Geometrical Acoustics: Reflection, Refraction

Analogous to the geometrical optics for short wave lengths and comparably large setups you can use reflections or delays in porous materials or detours by curved ducts to build systems with mirrors and lenses.

The most common usage is the paraboloid mirror with a microphone in its focus. For mid and high frequencies this assembly can get the directivity of a theoretical diaphragm of the same diameter. This construction can help to pick up distant sound sources with a maximum of suppression op surrounding noise. This is not necessarily of good use for spies (the setup is a bit large not to attract attention) but it can ease the recording of bird song in free nature.

If the source is not extremely far away, the optimal form is a rotational ellipsoid with the microphone in one focus and the source in the other one. The paraboloid is the limit for large distances. On the other hand this is the tyical case where such a strong directivity is really needed.

4.3 Principles of Transducers

The large majority of microphone works with passive transducers. This means that the diaphragm reacts to a pressure difference by creating a mechanical oscillation whose energy is at least partially transferred in electrical energy. In principle it is possible to reverse that, but for an efficient reproduction the parameters have to be different to that of a microphone. The most important parameter is that loudspeakers have to be relatively large to provide the necessary volume change for low frequencies. This is absolutely of no concern for microphones.

The reversal is used for the absolute calibration of sensitivity, because the same factors apply in both directions. They can be eliminated by computation.

Because each diaphragm reacts to a pressure difference, both sides must be prepared in a way to get the requested directional characteristic, see 4.2, 36.

If the back side is connected to a closed air volume (strictly speaking it must be effectively sealed for frequencies above the lowest limit of the frequency characteristic), you measure the sound pressure on the front. In the limit of small capsules compared to the

 ²²e.g. KEM 970 from Microtech-Gefell, famous for its use in the German Bundestag.
 ²³Revoluto System, Beyerdynamic

wave length, this is the sound pressure at the place of the microphone, see section 4.2.1, Page 36.

If both sides of the diaphragm remain open, it reacts to the pressure difference. In the limit of a small diaphragm compared to the wave length this is the pressure gradient in the direction of the normal of the diaphragm, see section 4.2.2, Page 37.

If one side is open and the other connected to the air over acoustical resistances, you get the mixtures of both extremes, see section 4.2.3, Page 44.

Rule of thumb: If only the front of the capsule is connected to the air, the microphone has to be a pressure transducer, it is omnidirectional. All other directivity characteristics need holes on the back side, most of the times as slits or bore. To give an example, I repeat a graphic in figure 4.16.

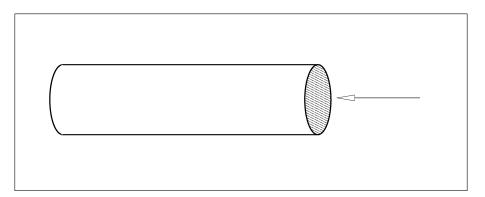


Figure 4.16: Example of an omnidirectional microphone

The microphone that is shown in the figure is necessarily omnidirectional because there are no openings on the back side.

4.3.1 The (LF-)Condenser Microphone

Of all practical constructions, the condenser principle is closest to the ideal microphone. This stems in part from the simplicity of the concept.

A diaphragm that is a conductor - or plated with a conductor - is one pole of a condenser. By its vibration the distance to the other pole changes and the capacitance with it.

Because no other components have to participate in the movement, the moving mass is low and it is relatively simple to achieve a high upper frequency limit and a broad bandwidth.

If you take the uniform movement of the diaphragm (which is close to the reality in its mid for good constructions), the capacitance is that of a plane condenser.

$$C = \mu_0 A / D \tag{4.53}$$

C is the capacitance in Farad, μ_0 the dielectric constant of the vacuum, A the area of the electrodes, D the distance.

In the LF circuit the capsule gets loaded to a equilibrium charge that is constant for time constants that are large against the longest period of time for transferred signals. The time constant that also determines the lowest working frequency is given as

$$\tau = RC \tag{4.54}$$

4 The Microphone

or, the other way round

$$\omega_0 = 1/(RC) \tag{4.55}$$

and thus

$$f_0 = \omega_0 / (2\pi) = 1 / (2\pi RC) \tag{4.56}$$

with τ as time constant, ω_0 as lower frequency limit of the transducer output, R as charging resistance and C as equilibrium capacitance of the capsule (no signal). The charging resistor typically has a value in the region of tens of gigaohms.

Q is the charge of the capsule. It is obtained from the polarization voltage U_0 and the equilibrium capacitance of the capsule C_0 as

$$Q = U_0 C_0 \tag{4.57}$$

The voltage at the condenser is

$$U_{\rm caps} = Q/C = D \frac{Q}{\mu_0 A} \tag{4.58}$$

It is clear that this voltage is directly proportional to the electrode distance and its alternating current component is thus directly proportional to its displacement.

To achieve a high sensitivity, a large charge and therefore polarization voltage is necessary. It can be created externally. In terms of design, care has to be taken not to create a discharge in the capsule and not to make it so high that the attractive electrostatic force results in a contact of the electrodes and short circuit.

For electret capsules the polarization voltage is kept in a material that is permanently polarized, the name giving electret. It can be part of the diaphragm – which requires a compromise in the mechanical characteristics – or better on the back electrode. These capsules are called back electret capsules. Currently the best available electret materials are so stable that there is no reason any more to consider them as inferior to externally polarized "real condenser" capsules.

The capsule is connected to an impedance converter amplifier that delivers an extremely high input impedance by using a J-FET transistor (or valve). The capsule must not be loaded in a way that interferes with its frequency characteristics.

For studio microphones there is normally a symmetrical output (at least symmetrical for the impedances, not necessarily the voltages) ²⁴ and the polarization voltage source, the impedance converter and the symmetry and buffer amplifier stage get their voltage vial phantom powering²⁵.

Looking at the noise level it is important to see that the received sound power increases proportionally to the area of the diaphragm and that the statistic thermal noise level increase is proportional to the diameter. Therefore very small diaphragms (3-10 mm diameter) are usually used when the typical sound pressure level is not very low (Lavalier/lapel, drums, measurement microphone). Very common are capsules of at about 20 mm in diameter. It is now possible to cover the whole listening frequency range without inherent limitations. Such microphones may be used almost universally. On the other hand they begin to show a change in directivity characteristic for the treble region because of diffraction.

²⁴Section 6.3, page 84

 $^{^{25}}$ Section 6.6, page 86

Capsules with a diameter of more that 25 mm or 1 inch are called large diaphragm capsules. A real advantage is only a possible extremely low noise level for uses that do not require a very neutral frequency characteristics, because their movement breaks up into partial oscillations for high frequencies. The author cannot agree with the widespread notion that these microphones with their uneven frequency characteristics and an upper frequency limit of typically 15 KHz are by principle the better studio microphones for vocal use, especially singers²⁶. By the way: the often cited trend that the directional characteristic morphs into omnidirectional in the bass is a result of double back-to-back diaphragm capsules. This is a typical construction for large diaphragm microphones with variable directional characteristics. It has nothing to do with the size of the diaphragm[23].

With state of the art amplification circuits a LF Condenser Microphone can be brought close to the theoretical limit in noise and sensitivity.

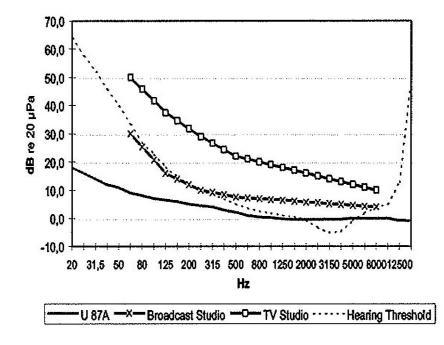


Figure 4.17: Noise characteristics of a condensor microphone compared to the noise of typical environments, taken from Peus [20], page 7

This plot shows the comparison of the frequency characteristics of the inherent noise of a large diaphragm studio microphone type U87A (12 dBA noise level) and the noise of typical studio rooms and the limit of perception. Modern microphones have a noise floor that is at about 5 dB lower, small diaphragm microphones have similar noise levels or a little bit higher, for example 17 dBA. This clearly demonstrates that the noise floor can be considered a solved problem with this class of microphones.

For a capsule that is small compared to the minimal wave length that is processed (e.g. less that 6 mm in diameter) it is relatively simple to achieve a directional characteristic that is close to independent from the frequency.²⁷

²⁶See Wuttke: Wissenswertes[38], Section 3

²⁷If the capsule wall does not have vanishing thermal conductivity and capacity, the compliance towards low frequencies may rise. The cycles get longer and there is more and more thermal energy transferred during the times of higher and lower momentary pressure with the surrounding material. This prevents the additional stiffness by the heating with compression that is typical for the adiabatic compression

4 The Microphone

• One side of the diaphragm is attached to a small closed volume of air (a small leak is left to equalize the internal and external static pressure) and the resonance frequency of the diaphragm is put on the upper limit of frequency region of operation with a suitable amount of damping or even higher, so that the damping is less relevant.

In this case the diaphragm works as spring, and the displacement is proportional to the momentary pressure, independent of the frequency. This is closely matched for small (diameter of at about 6 mm) measurement microphones.

• Both sides of the diaphragm remain open, the construction is symmetrical, the resonance frequency is put in the mid of the working region. The movement is heavily dampened to enlarge the bandwidth, which may be broadened even more by putting additional resonances at both ends.²⁸

The diaphragm gets excited by the pressure difference (in the limit to small geometry the derivative of pressure). Because of the inner derivative the gradient is proportional to the sound pressure and the frequency. Because the movement is controlled by friction, the speed is proportional to the pressure difference, and the dependency about the frequency is canceled as well as the phase shift of $+90^{\circ}$ (derivative of the pressure about the location) and -90° (integral of the velocity over time). The result is a capsule with a directional characteristic of a figure-of-eight with a linear frequency response in the ideal case. The output signal is in phase with the pressure form the front and in inverted phase for sound from the rear.

• By a partial opening of the back side, all characteristics between omnidirectional and figure-of-eight can be reached. In these constructions the resonance frequency must be relatively low, too.

Because the impedance buffer must be located directly near the capsule, there is the need for electronics at this place. This renders condenser microphones relatively sensitive to any kind of humidity, because they compare rather high polarization voltages and extremely low signal voltage very close together. This has to be taken into account for close miking in recording of vocals.

By the impedance buffer that amplifies the current and the relatively high output voltage of the capsule the sensitivity of these microphones is rather high. As a rule the impedance buffer does not amplify the voltage, for voltage symmetrical outputs, the difference voltage is however multiplied by two, resulting in 6 dB more sensitivity.

An interesting overview in circuitry concepts for LF condenser microphones and their development over time can be seen in the article by Meitz[11].

by sound, the system gets closer to isothermal behavior. The maximum quotient of the effect is dictated by the adiabatic power. For two atomic molecules like in air this is 1.4, corresponding to almost exactly 3 dB (because 1.4 is close to $\sqrt{2}$). In real life the effect is smaller because at least a part of the back driving force results from the tension of the diaphragm and is not affected by this effect. All in all this may lead to a raised sensitivity for low frequencies depending on the details of the construction and materials involved.

²⁸This is similar to the measures to increase the bandwidth for dynamic microphones that are pressure transducers. They work as small bandwidth transducers without these measures, too. See section 4.3.2, page 60.

(4.59)

4.3.2 Microphones Using Electromagnetic Induction

If a conducting wire W is moved in a magnetic field with induction \vec{B} of absolute value B which is constant over its length L, optimally in a direction perpendicular to both the length of the wire and the field with the velocity $\vec{v_M}$, a voltage U_a is induced according to

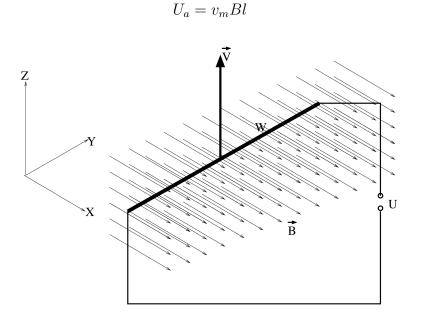


Figure 4.18: Induction on a moving wire

For a typical strength of induction in the magnetic field and a realistic speed of movement of the diaphragm the induced voltage per length unit is rather small. Therefore the shape of the inductor is chosen to be either a

- **coil**, in order to maximize the length L with several turns inside a pot magnet, or a short, conductive
- **ribbon**, which is the diaphragm in itsef. The extremely low but also low impedance voltage is subsequently transformed up or amplified right next to the converter.

The mechanical paramter that is converted into the voltage is the velocity and not a position as in e.g. the condensor microphone. The conversion is therefore not done statically but dynamic. As consequence, both variants described above are dynamic microphones. The common nomenclature is to call the first variant dynamic microphone and the second one ribbon microphone.

The Dynamic Microphone

The dynamic microphone (also, more precisely, called moving coil microphone) works as a reversal of the dynamic loudspeaker.

The schematic cross section figure 4.19 shows the principal setup. A coil C is fixed to a diaphragm D and is positioned inside a pot magnet with radial magnetic field (depicted by the poles N and S) via a suspension Su. The output voltage is typically directly connected to the output terminals. A few models add an output transformer and put less turns of wire to the coil.

4 The Microphone

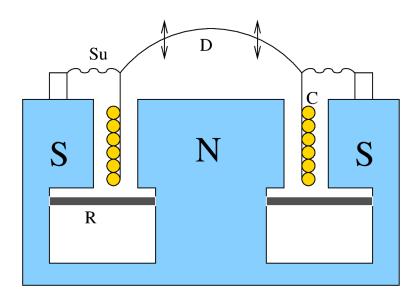


Figure 4.19: Schematic construction of a dynamic microphone

Without further design measures, the frequency response is rather resonant.

Without a strong mechanical resistance (e.g. fine fabric in the air path) the response falls with ca. 6 dB/octave below the resonance frequency because the mounting of the diaphragm works as a spring. This implies a constant amplitude of the movement compared to the pressure amplitude, so the temporal derivative (the speed) falls with frequency. Above the resonance the movement without damping is controlled by the combined mass of diaphragm and coil, which leads to a falling of 6 dB/octave to high frequencies. By mechanical resistance R and acoustical resonators (e.g. small holes or pipes: Helmholz resonators, not included to the figure) the frequency response gets optimized.²⁹

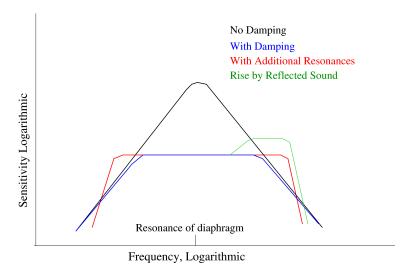
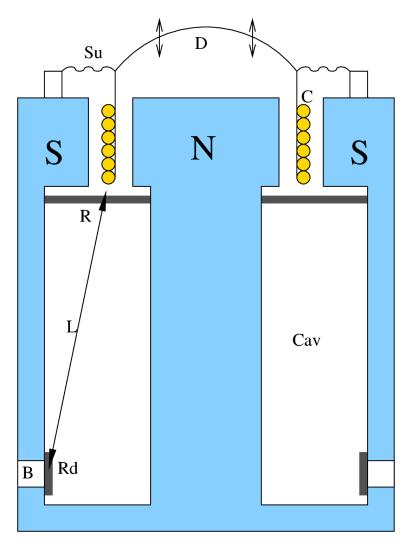


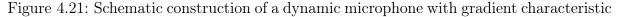
Figure 4.20: Schematic frequency response of an omnidirectional dynamic microphone

For a directional microphone, the back side of the capsule must be acoustically open in the frequency region required, see figure 4.21 on page 61. The exact dimensions of bores B, cavities Cav, delay paths L and mechanical resistors Rd define the directional

 $^{^{29}\}mathrm{For}$ a detailed explanation see Olson[18], pp. 226-229

characteristics and optimize the frequency response. Those require some space. Therefore high quality transducers are not very small as a rule.





As always, the frequency response of the pressure gradient receiver has an additional factor proportional to the frequency (or a quotient to the wave length). In logarithmic view this means an additional rise of ca. 6 dB/octave (exactly 20 /decade). The implication is that the resonance frequency of the moving parts (coil and diaphragm) define the low end of the bandwidth. The frequency response of the dynamic pressure gradient microphone above this frequency is linear in principle. Deviations result from reflections, resonances and compromises in the construction when a omnidirectional component is added to get one of the common directional characteristics and the upper cutoff of the gradient receiver.

A downside of the pot magnet construction is the more or less closed back side. It rules out possibility of a real figure-of-eight capsule, because at least the high frequencies are blocked from the back side. If needed, ribbon microphones (see 4.3.2, page 63) may be used instead,

Very few dynamic microphones have a frequency response up to 20 KHz. Because of the large diaphragms, the rise in sensitivity of the upper mids starts often for comparably low frequencies. Some microphone models counter this by fixed (Beyerdynamic M201) or switchable (Sennheiser MD441, Shure SM7B) parallel RLC circuits. The rise in bass re-

4 The Microphone

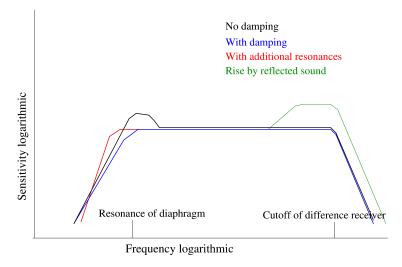


Figure 4.22: Schematic frequency response of a gradient type dynamic microphone (directional)

sponse by the proximity effect can be countered by parallel inductors (Sennheiser MD441, Shure SM7B). See figure 4.23.

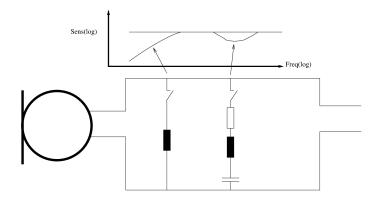


Figure 4.23: Attenuation of bass or midrange response for a dynamic microphone. The black rectangles are meant to be inductors/coils

As a rule the diaphragms are rather large to achieve a reasonable sensitivity. Even so the sensitivity is typically 10-20 dB below that of condensor microphones.

Because of the coil inherent to this construction, external varying magnetic fields generate an additional voltage that can degenerate the signal (mostly hum). Some of the better models therefore embed a second coil concentric to the moving coil which is dimensioned to compensate the noise voltage.

An advantage of this principle is the robustness of the construction. Some models can survive harsh handling. Dampness cannot lead to electrical problems, because there are no active components included.

This kind of microphone is very often used for live events for rock and pop music. For vocal pickup in extreme proximity there are specialized models with high lower limit of the frequency range, directional characteristics between cardioid or hyper cardioid to help against feedback loops, a stable protection metal mesh and pop protection by integrated plastic foam filters.

The Ribbon Microphone

The ribbon microphone is a special case of the dynamic microphone. The conductive diaphragm R itself is the inductive element in a strong magnetic field (between the magnetic poles N and S in the figure).

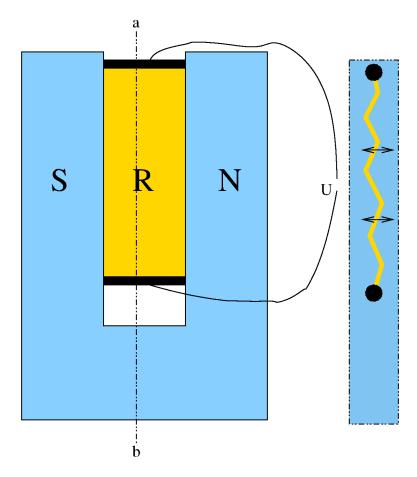


Figure 4.24: Schematic construction of a ribbon microphone, right side: cut throug a - b

Because this does not allow for multiple loops, its output voltage is very low, as is its output impedance. As a rule therefore an output transformer is switched before the output. Nevertheless the resulting sensitivity is typically less than that of a standard dynamic microphone. This means that for these microphones it is very important to use a microphone preamplifier with extremely low noise.

The extremely thin and loosely mounted diaphragm makes these microphones very sensitive to mechanical shock and even more so against pop and wind. The transformer renders them sensitive against current that is brought to the output connections, which can happen the case of applying unnecessary phantom power.

Most ribbon microphones are open at both sides and axis symmetrical, so their directivity characteristic is that of a figure-of-eight. Historically this was the first practical use of non-unidirectional microphones and got used to cancel the sound of the noisy film cameras of that time.

Some constructions have a back side that is partly closed to achieve other directional characteristics such as cardioid.

Because the ground principles work the same as for moving coil dynamic microphones, the same principles apply for the frequency response. Most ribbon microphones have a rather linear frequency characteristics in the mid range, but they do not reproduce frequencies over 15 KHz.³⁰ An example can be seen above in figure 4.1.2, page 30.

Because of the necessary magnetic setup these microphones are not rotational symmetrical, neither the construction nor the directional characteristics. In one direction the directivity is a lot more narrow than in the other one because of interference effects.

4.3.3 Special Microphones

The MEMS Microphone

It is built on a chip with micro electro mechanical technology by lithography. This allows a high repetition precision and allows the inclusion to a circuit directly on the circuit board. It is widely used inside of smart phones. Because of the miniaturization there is no noticeable interference in the hearing range. The frequency response is normally flat and the directional characteristic omnidirectional. How far this stays that way in the application depends on the details of the surrounding construction and the way the openings for sound are done.

The Piezo Microphone

This principle uses the piezo electrical effect of some substances. They are very resonant. It is easily possible to reach ultrasonic frequencies with this technology.

The Contact Microphone

Very often the piezo microphone is used in music because of its relatively high mechanical impedance as a bulk sound or contact microphone. They are fixed in some acoustic guitars. Because the electrical impedance is very high (the principle is ruled by the same equation as the LF condenser microphone), it needs an impedance-matching buffer amplifier in direct proximity, too.

Sometimes miniature condenser microphones are also used. Normally they get glued to an instrument.

The Boundary Layer Microphone

The boundary layer microphone is more or less just a common (typically condenser) microphone, that is placed directly on or even integrated in a solid rather large plane, sometimes in a specialized casing. This prevents interference effects by the reflections of these planes (no comb filter effect).

Most often pressure transducers with resulting hemispherical directivity characteristics are used, other characteristics (halved) are also possible. These lead to problems with vibrations of the plane, because they are directly coupled mechanically. Beware of table tops and bottoms.

A halved super cardioid is able to counter the decrease in level by increased distance to a degree, because the source simultaneously gets into the more sensitive angle. This works if the source stays in the same (non-zero) height and the microphone is positioned at the bottom. This is useful for theater recordings.[36]

 $^{^{30}\}mathrm{A}$ detailed discussion can be found in Olson[18], pp. 237-267

4.3.4 Active Transducers

Deviating from all concepts that were described above, some types of microphones are active transducers. They do not convert the sound or vibration energy to electrical signals but they use an external power source to create the signal. They work as amplifiers themselves. This should not be confused with the concept of amplification that is done behind the transceiver, as with LF condenser microphones.

The HF Condenser Microphone

The HF condenser microphone uses an HF generator in combination with phase controlled rectification in a bridge circuit to measure the capacitance change of the capsule. The power is delivered by the HF oscillator.

This microphone type can work without lower frequency limit. In the extreme case it can work as a barometer.

A very good realization of this principle is the MKH series by Sennheiser.

The condenser capsules must be optimized differently from those of LF condenser microphones.

The Carbon Microphone

This was the first microphone type that was used in technology. The resistance at the surface of carbon particles changes with pressure variations and this modulates the current if a voltage is applied. The inherent amplification (the voltage source is the power source) eased the construction of the first telephones quite a bit. The mechanical resonances of the carbon particle container and the typical metal diaphragms with their partial vibrations led to extremely uneven frequency characteristics. Additionally the device is not very linear in its characteristic, which leads to high distortion factors. In a typical use for one single voice this can be tolerated. With push-pull constructions the first studio microphones were constructed, with a vastly improved linearity. ³¹

The Tube Microphone

The tube microphone used the changing electrical characteristics of a tube by the vibration of its electrodes. 32

The Optical Microphone

The optical microphone construction modulates the amount of light that is typically passed between open fibers. It can be used in places with large electrical or magnetic fields, such as the inside of a NMR (Nuclear Magnetic Resonance, for tomography) device.

The Heat Flow Microphone

The heat flow microphone uses two closely positioned parallel wires. One wire is heated and the temperature of the other is measured by its resistance. Air moving from the heated wire to the measurement wire increases its temperature by transfer of additional

 $^{^{31}{\}rm see}$ Olson
[18], pp. 214-220

 $^{^{32}}$ see Olson[18], pp. 220-224

heat and vice versa. This device really measures the movement of the air by sound, which is proportional to the pressure gradient.

This principle is used to build a microphone that combines this setup for the three axes in space with a pressure transducer. This allows to simultaneously measure the complete set of the zeroth and first order of the spherical functions representing the sound field at a given point.

5 Multiple Microphones/Signals, Stereophony

5.1 Stereophony - not necessarily exactly two channels

Stereophony means spacial reproduction. Most see this as synonym for the typical two channel systems, but this is not correct.

If you stay to the original meaning, it encompasses all surround techniques, wave field synthesis and ambisonics. They all try to achieve a spatial distributed acoustical image. They use different means and more expense than the common two channel method but they achieve more realism.

5.2 Two Channel Stereophony

5.2.1 Reproduction

For the common two channel stereophony the well known stereo triangle defines the standard that also is the target for all mixes of records.

The two loud speakers and the recipient are located in the angels of a equilateral triangle. The room should be symmetrical around the line that connects the mid position of both loudspeakers and the recipient. This is necessary to provide the same parameters (amount of room sound, temporal characteristics, coloring, level) for both signals.

For larger distances to the loudspeakers the perceived angles get smaller and smaller. The perception gets closer to that of a monophonic (one channel) reproduction. For smaller distances the perceived angles to the side are exaggerated.

A lateral movement of the listener leads to a shifting of the perceived so called phantom sources towards the nearer loudspeaker. This effect gets stronger if the listener is closer to the loudspeakers.

For short distances and stereophonic positioning by time differences the dreaded "hole in the mid" can appear. The smallest deviation from the central hearing position leads to a localization off all but the most extreme positions towards the nearer speaker.

5.2.2 Level Stereophony

This method to achieve a given localization works with simultaneous signals on both channels and controls the respective levels. If the levels are equal the resulting phantom source appears in the mid of the loudspeakers. In all other cases the location is closer to the loudspeaker with higher level. If the level difference reaches or exceeds 18 dB, the source is located at the position of the louder loudspeaker.

Poly Microphony and Pan-Pot

If the signals are picked up with dedicated closely positioned microphones – or in the case of electronic sources directly coupled 1 – so that the sources are separated as good as possible, the preferred position of localization can be finely tuned by a so called pan-pot (**Pan**orama **Pot**entiometer).

Most rock or pop recordings use this method to place the signals on their places on the stereo base during the mix.

One important characteristic of this method is that all localization's can only be between the loudspeakers, not outside their base.

The XY Method

By using a matching assembly of two coincident (as a rule with sources located in a plane with two microphones one over another) microphone capsules with a directional characteristic that results from the set of first order characteristics, the microphone signals can be fed directly into the two stereo channels to achieve the target localization.

You have to consider: the more apart the main directions of both capsules are, the more the resulting stereo image is spread. Contrary to a very widely distributed belief this narrows down the angle in which a source must be places to get located at the same location.

Taken the other way round: If the sources are spread over a large angle, the microphones have to be directed into a *smaller* angle. You must not point them according to the geometric borders of the area to record – quite the contrary!

The Blumlein Assembly uses two microphones with figure-of-eight characteristics whose main directions are spaced at an angle of 90°. Sources located at $\pm 45^{\circ}$ are reproduced by only one channel, at $\pm 30^{\circ}$. The localization there happens already with a level difference of 18 dB, this corresponds to a source positioned at $\pm 38^{\circ}$.².

All sources that are between $\pm 45^{\circ}$ and $\pm 135^{\circ}$ appear with reverse polarity on both channels and get located diffusely out of the stereo base. This may not be problematic for a pure room sound.

You must nit forget that there is no discrimination of sources in the back. The get reproduces with full level in the front but with mirrored position.

Angled Cardioids that are coincident can also be used directly as channels of a two channel stereophony.

Again, the outer position of the stereo image is reached at the point of 18 dB level difference.

For starters here is a tabular of the recording angle against the angle between the main directions of the capsules³, arbitrary values can be computed with the application at 4 from sengpielaudio.com.

¹The typical problems that may be caused by this like hum or impedance problems can be avoided by using a DI-box, see section 6.5, page 86.

²see http://www.sengpielaudio.com/BekannteStereoMikrofonsystemeUndIhreWinkel.pdf, accessed 2016-05-02 at [24]

³from http://www.sengpielaudio.com/Stereo-Mikrofonsysteme.pdf and http://www. sengpielaudio.com/BekannteStereoMikrofonsystemeUndIhreWinkel.pdf, accessed 2016-05-02, see [24]

⁴http://www.sengpielaudio.com/HejiaD.htm, accessed 2016-05-02 see [24]

| Microphone Angle | Recording Angle |
|------------------|-----------------|
| 60° | 242° |
| 61° | 240° |
| 80° | 210° |
| 90° | 196° |
| 102° | 180° |
| 120° | 158° |
| 127° | 150° |
| 158° | 120° |

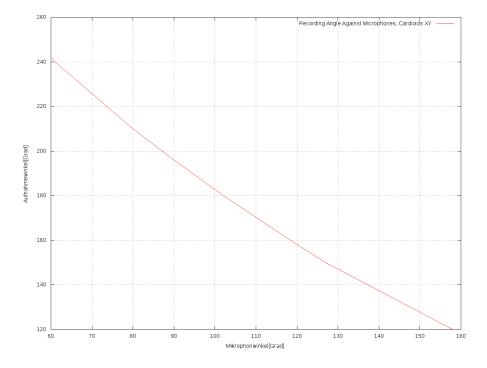


Figure 5.1: XY setup: Recording area plotted above the angle of two cardioid microphones

You see very clearly that the angles between the microphones and the recording angle move in *opposite* directions. Only if the microphones are angled at least 102° , the recording angle is less or equal to 180° , the half circle in front.

Here are sketches of the geometry for two examples of angles. You must *not* point the microphones at the outer reaches of the recording region. In the case of the angle of 60° between the microphones, these are even *behind* the microphones, not at 180° from the microphones like it looks, they are at about 190° . That it is close to the back side of the microphone is only accidentally the case here.

Other Directional Characteristics with stronger directivity shrink the recording angle for the same microphone angle. Here is a tabular for hyper cardioid capsules.

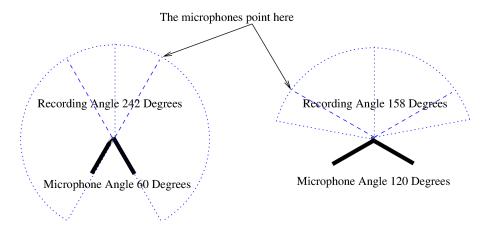


Figure 5.2: XY setup: examples of the recording area for two angles of cardioid microphones $% \left({{{\bf{F}}_{{\rm{B}}}} \right)$

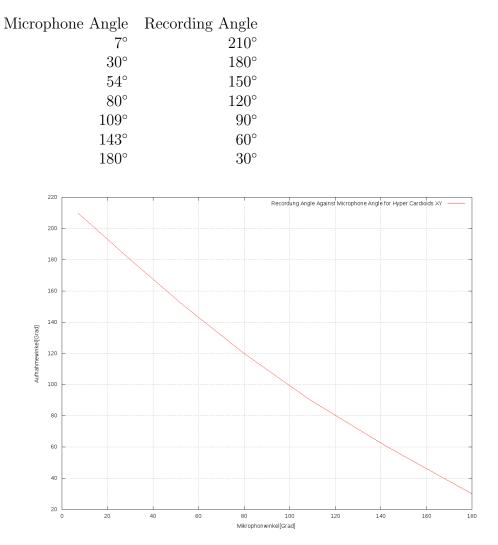


Figure 5.3: XY setup: recording areaplotted above the angle of two hyper cardioid microphones

One can see that with a microphone angle of 143° it is really possible to get source angle = localization angle, like the original idea of two channel stereophony planned it.

It is hard to use the full base with XY arrangements of cardioids except for excessively broad distribution of sources.

Apart from the theoretical aspects it is important to keep in mind that the signal in the mid, normally the most important source, does not get picked up in the main direction of any of the microphones. For most real microphones this degrades the frequency characteristics in the treble region.

One downside of all XY assemblies is the use of gradient type microphones which have problems with the correct recording of bass signals at least from a distance.

There are XY microphone assemblies readily available in one housing.

A good point is the possibility to change the stereo width during the mixing. The signals can be added as necessary because the capsules are coincident and therefore the signals have the same phase. It is trivially possible to record a distributed source with the full stereo base and later on position both extreme positions independently wherever they shall be in the stereo base by the use of the pan-pot.

The MS Assemblies

MS means mid and side signals (not to be confused with S&M...). Like the XY assembly it consists of two coincident microphones. The side signal is delivered by a figure-of-eight microphone, the mid signal by one microphone that points to the mid.

One side of the end result is built by adding both signals, the other one by subtraction, typically with a normalized sensitivity.

A source in the mid – often an important one – is picked up by the mid microphone in its main direction, in contrast to the situation with XY assembles.

The mid signal can be produced by a pressure transducer (omnidirectional). This leads to an assembly with coincident microphones that may use the good and consistent bass response of pressure transducers.

The mid signal is identical to the sum signal for monophonic reproduction, because the side signal gets canceled out in the sum. Therefore the over all frequency response is only determined by the mid microphone.

By changing the amount of side signal the width of the stereo image can be changed as necessary.

To get a discrimination of signals from the front and back, a microphone with nonomnidirectional characteristic may be used for the mid. This is often uses in movie sound production.

For all MS assemblies with idealized microphones a mathematically equivalent XY assembly can be computed. One example: The combination of omnidirectional and figure-of-eight is equivalent to two cardioids back to back.

All level stereophonic signals can be transformed by matrix computation (sum and difference again) to artificial M and S signals. These can be processed to alter the stereo image and then they get reprocessed to get the end result stereo signal.

5.2.3 Time Difference Stereophony

Additional to the level difference the localization is also dependent on the temporal correlation. For arrival differences larger that 1.5 ms, corresponding to 0.5 m path difference for 340 m/s speed of sound, the sound source is located only at the place of the first source (assuming same levels). Larger time differences lead to a delocalization and extreme differences are heard as echoes. The rule of the first wave front applies: it determines the

localization of the source, the later arriving signals are perceived as part of the surrounding. This effect is seldomly uses for recordings, but for reproduction this can help to add sources without compromising the localization.

The time difference between the channels is currently not commonly used artificially during the mix.

The use of time differences for localization is bad for the mono compatibility of the recording, because the first and retarded signal add up and lead to a comb filter effect. This colors the sound in dependence of the direction. The uncorrelated room sound can compensate this partly.

The Large AB Assemby

The large AB assembly consists of two microphones that are relatively far apart, so that all sources get received with at about the same level but time differences.

The array may be done with omnidirectional microphones. This allows for a realistic bass level without proximity effect and limits of the bandwidth to the bass. The room sound that is simultaneously present in the signal is only colored by the finite diameter of the microphones (see free field and diffuse field) and by interference.

Like the angle with the XY assemblies there is one parameter that allows to tailor the recording angle. With 0.5 m distance 180° get mapped to the stereo base, therefore a smaller distance is not suitable. This array cannot be used for a small and easily transportable solutine.

Parts of the room sound get de-correlated for large distances and are perceived as filling the reproduction room. This helps to get a convincing reproduction of a performance in a hall.

For the reproduction there are two problems:

- The mapping of recorded to perceived angles get more and more nonlinear to the large angles. Therefore the main sources should not use more than 85% of the stereo base
- The perceived angles are very sensitive against positions outside the midst. They move strongly towards the closer loudspeaker. It is important to keep in mind that 0.5 m path difference are enough to fix the localization towards the closer loudspeaker.

5.2.4 Combinations of Level and Time Difference

If level and time differences are combined it is important not to contradict one information with the other one. This happens for example if two microphones with directivity are spaced and point inwards.

ORTF, EBS & Co

The two types of localization information are combined in a way to support one another. Two directional microphones are used with a distance that is smaller that with AB assemblies and the are pointer outward. Some of the good characteristics of XY and AB are combined. For the recording the assemblies can be almost as compact as XY assemblies. The localization is more stable than that of AB assemblies, because level differences are also used. Two downsides of AB and XY assemblies are also combined:

- The signals have to be kept separated to avoid comb filtering. This means that the assembly geometry defines the localization angles without the possibility to change it in the mix.
- Omnidirectional microphones cannot be used, with all consequences for the bass frequencies.

To achieve the localization characteristics there are three degrees of freedom: distance of the microphones, the angle, the directional characteristics. Several combinations have been developed and tested and are named. The following combinations are all done with cardioids.

| Abbrev. | Name | Angle | Distance | Recording Angle | Amount Level | Amount Time Difference |
|---------|--|-------|-------------------|--------------------|-----------------|------------------------------|
| NOS | Nederlandse Omroep Stichting | 90° | 30 cm | 81° | 42% | 58% |
| EBS | Proposed by Eber- hard Sengpiel | 90° | $25~\mathrm{cm}$ | 90° | 47% | 53% |
| DIN | Proposed German norm | 90° | $20 \mathrm{~cm}$ | 101° | 53% | 47% |
| RAI | Radio Audizioni Ital- iane Radiotelevisione Italia | 100° | 21 cm | 93° | 53% | 47% |
| ORTF | Office de Radiodif- fusion- Télévision Française | 110° | 17 cm | 96° | 61% | 39% |
| 5 | - | | | | | |

The Small AB Assembly

Like Large AB this assembly is done with omnidirectional microphones. They are places so close to the sources that the distance differences produce level differences. This may be used to record large instruments (piano...) mapped to the stereo image. Like AB the recording setup determines the stereo positions if comb filtering is to be avoided.

5.2.5 Alternative Microphone Assemblies

The Decca Tree

This assembly was found by trial and error and got not deduced from a theoretical background. It is named after the record company Decca where it was developed.

It consists of two omnidirectional microphones spaces at least 2 m, too much for large AB and in the symmetry axis and about 1 m closer to the front a microphone for the mid that also has an omnidirectional characteristic. To achieve a smooth directional characteristic of the highest treble frequencies and a broad transition to a raised sensitivity of the treble region you may use microphones that are embedded in spheres with a diameter

⁵Taken from http://www.sengpielaudio.com/BekannteStereoMikrofonsystemeUndIhreWinkel. pdf, access 2016-05-06, [24]

of at about 5 cm like the originally used Neumann M50 or equivalently a combination of microphones with spheres with matching bores.

The microphones at the sides get mixed to the corresponding side channel and the mid microphone into the mid position, all with the same level.

The exact geometry has to be adapted to the situation

The result is similar to that of an AB assembly with improved stability of the localization of the mid signals. The signal sources get located at the extreme left or right position or in the mid with almost no signal located in the positions in between. The notches are less prominent than the notch in the mid for exaggeratedly spaced AB assemblies.

The Decca Tree is typically located in at least 3 m height above the head of the conductor and is often used to record an orchestra. The orchestra is places in a semi circle around the conductor. For linearly distributed sound sources it is not optimal and the assembly is used in a linear setup contrary to its name.

The setup can be complemented by two more omnidirectional microphones that are places even more extreme to the right and left. They are called outriggers. They get also places to the extreme right and left of the stereo image. Partly the original right and left microphones get a position that is closer to the mid in this case.

A very detailed paper on the history and properties of the Decca Tree can be found from Gernemann.[7]

Curtain of Microphones

The assembly consists of several microphones (typically omnidirectional) that are places equidistant in a line at a distance of the sound sources and that are places equidistant at the stereo base in the mix. Contrary to poly microphony there is no 1:1 relation between single sound sources or groups and the microphones.

5.2.6 Main Microphones and Support Microphones

There are several reasons to combine one of the microphone assemblies described above, called main microphones in this case, with several support microphones that are closer to single sound sources or groups for the recording of complex sound sources like orchestras or choirs.

The main microphones have the task of picking up the whole sound image including the surrounding, especially including the hopefully good acoustics of the room. Because of their ability to get an airy, distributed impression, time of arrival assemblies are generally preferred for this setup.

The support microphones allow the boost of signals that are too weak in the overall image, if necessary by artistic preferences to varying degrees over times. This helps to make onset of a voice clear to the listener – in contrast to the live situation the supporting optical cue for the listener is absent. The possible weakness of the treble especially for sources in the mid which can occur for XY assemblies because the microphones are used outside the optimal angle can be countered. This would often hamper the important solo sources that are typically located close to the mid.

Typically the support microphone signals do not get delayed. This means that their signals get played before the matching signal from the main microphones. Because of the law of the first wave front they define the localization. This makes the localization more stable, because an additional level localization is added.

There is a continuous path to poly microphony. If the result does not change clearly when the main microphones get muted the mixer typically made an error.

If the acoustics are extremely bad or the amount of external noise is too high, the main microphone assembly cannot be used. The only way to handle this situation is to use poly microphony and close miking. The real room sound is then better replaces by an artificial and reverberation the is tailored matching to the intended impression.

5.2.7 Ear Related Signals

Ear related signals add tailored different frequency responses of the channels to the level and time differences as means to define the localization. Like the conditions during the hearing process especially the treble frequencies are attenuated at the far side of the source. The exact characteristic is described by the HRTF (Head related transfer function).

Because during playback over loudspeakers this effect occurs a second time, this method can only be used effectively for playback over headphones, because the ear related signals get really delivered only to the matching ear that they are meant for. In the other case one ear gets the HRTF applied two times (recording + playback) and the other ear two different HRTFs. This leads to a colored sound an the cross talk between the channels is too large, because it also happens two times.

Separating Body

A coarse variant to generate ear related signals uses a separation body to create a shadow for high frequencies to a microphone on the far side of a source. Omnidirectional microphone may be used. Well known systems are

- **The sphere** at about the size of a head, reflecting or absorbing. The microphones are located at the surface or are incorporated.
- An absorbing disk between the microphones, also called Jecklin Disk 30 cm and a microphone distance of 17 cm. Practical usage showed that the resulting stereo image is too narrow, therefore newer variants are larger.

Dummy Head

The dummy head tries to copy the natural act of hearing and to reproduce the acoustical signal at the place of the dummy head in the ear of the listener. It creates a HRTF by acoustical means.

When the playback is done with a head phone, the localization outside the standard stereo position (behind, above, etc.) works surprisingly well. For the important front region it is problematic. The reason turns out to be rooted in small movements of the listeners head during the listening. The effects of the changing HRTF by these movements is taken into account during the hearing process and helps to localize very precisely in the natural listening. During the playback over headphones there is no HRTF change, because it got fixed during the recording. The resulting mismatch between the expected and real HRTF confuses the listening process.

A weakness of this setup is that the real HRTF is unique to each person and the dummy head can only provide a signal that (hopefully) matches the mean listener.

The signal cannot be modified much without destroying the dummy head effect.

5.3 Generalized Stereo: Surround

All surround settings widen the possible localization region over that of the linear front region that is typical for the normal two channel stereophony. Additionally they try to stabilize the localization in the front region against movements or less that optimal listening positions, the try to enlarge the so called sweet spot.

5.3.1 5.1

The first widely used surround format with really independent channel is called 5.1. The numbers denote the use of five full range channels and one subwoofer channel for drastic sound effects.

The playback of the bass part can differ from the original setup without changes for the end result. Useful variants are

- **without subwoofer**, the bass part is distributed to the front speakers. This is possible if they are capable of reproducing the low frequencies with the required level. This implies high power and rather large loudspeakers. This can be problematic for home theaters.
- **bass reproduction only by the subwoofer**, with five satellite speakers without large bass reproduction capabilities. All bass signals below a given frequency get split from all channels, summed up and fed into the subwoofer only (or multiple subwoofers with the same signal). The THX standard votes for a splitting frequency of 80 Hz, others below 170 Hz may be used if the subwoofer has a sufficiently low distortion, so that there are no resulting higher frequencies that allow to localize it. This is better anyway. The satellites can be built rather small in this case within limits.

In both cases the bass signal gets distributed differently from the input signal. This is called bass management.

The front row right and left channels have the same position as in the two channel stereo setup. In the mid there is an additional front loudspeaker, the so called center speaker. Its most important function when used for movies it the stable localization of the most important person in a dialogue that is typically shown in the center at the matching point without the instability of a phantom source against asymmetrical hearing positions. This is necessary to maintain the quality of the room reproduction at the different seating positions in a cinema.

There is no definitive method how the three front speakers shall be treated during the mixing process. It is possible to create phantom sources between a side channel and the center. The downside is a comb filter effect in the treble region that is more prominent than in the two channel stereo case, because there is no such strong shadow of the far side ear by the head. Both channels can reach the ear in a direct path. Sometimes a standard two channel stereo signal is given to the front right and front left speakers and a complete distinct signal – typically dialog – is fed to the center channel. This can only work as intended if the center signal and the two channel stereo signal are not correlated to each other. In all other cases the localization gets very inconsistent when the stereo image gets asymmetrical for non ideal listening positions while the center signal remains stable.

Depending on the kind of reproduction there are two setups of the reproducing loud-speakers.

The setup is complemented by two surround channels that are ideally located at 110°. They have the task of providing the whole area of 360° for localization. It has been shown that phantom sources at higher angles than the front speakers can be achieved, but they are extremely unstable. At least it is possible to get the effect of a complete surrounding by ambient noise of room reverberation. If the room reverberation in the listening room is not too high, the perception of being in the simulated location can be improved a great deal which helps to feel present in the room that is shown. Therefore it is useful to damp a room for listening to surround recordings more than is optimal for two channel stereo production.

Cinema The three front loudspeakers are located at the sides of the screen and one in the mid, optimally behind an acoustically transparent screen. The surround channels get reproduced by two arrays of identical loudspeakers on each side of the room, so that the acoustical reproduction is as similar as possible for all listening locations in the room. The front speakers are positioned in one row to minimize space requirements. The main goal is a good enough compromise of the reproduction quality over the whole seating area.

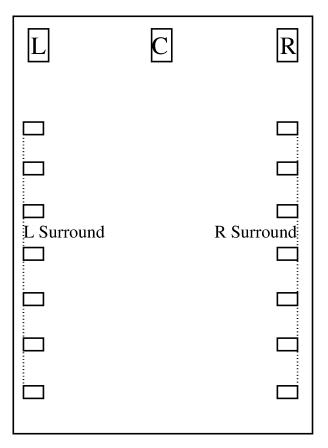


Figure 5.4: 5.1 loudspeaker setup in a cinema

Studio/At home All speakers are located at the same distance from the main listening position. The reproduction at this place is optimized. The delay and the relative listening level of all channels are consistent. At an angle of 110° there are the two surround channel speakers.

The circle in the image does not show a desired wall, it just represents the equal distance to the speakers.

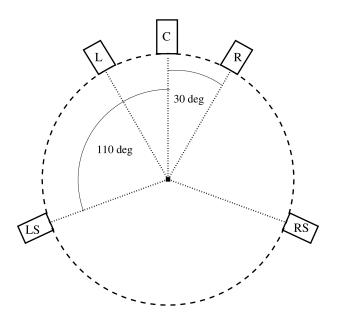


Figure 5.5: 5.1 loudspeaker setup in studio or home cinema

Poly Microphony

By poly microphony and generalized pan pot technology different signals can be picked up separately and placed at will in the surround panorama, either at a single channel or between two of them. Care should be taken that the phantom images outside the front speakers are not stable.

Microphone Assemblies

⁶ In analogy to the situation for two channel stereophony there are several microphone assemblies for level and time difference localization as well as combinations. Partly two channel assemblies get generalized. The principles stay the same, the orders of magnitude differ sometimes, as shown by Gernemann.[8]

Level Localization (coincident) The coincident assemblies for surround sound all face the same problem: the spherical functions of first order (and thus all common directional characteristics without support by interference) can't provide a sufficient separation of five channels, so a strong cross talk is inevitable. The situation is most problematic for the front channels with an angular distance of only 20°. The narrow characteristics have no full cancellation at 180°. The theory shows that in the plane there are only three independent signals. They can be provided by three cardioid microphones of – theoretically equivalent – a double MS system. It can be obtained by the combination of a Blumlein assembly with a coincident omnidirectional microphone. From these raw signals the surround channels can be computed by matrix calculations.

Several attempts at microphones with higher order spherical functions often lead to level losses by the double derivative with and high noise levels.

A commercially available solution is the Eigenmike® from MH-Acoustic⁷.

⁶See the chapter "Allgemeine Betrachtungen zur Mehrkanal-Stereofonie", Mikrophonaufsätze[36] and Wittek, "Mikrofontechniken für Atmoaufnahme in 2.0 und 5.1 und deren Eigenschaften" [34]

⁷http://mhacoustics.com/products#eigenmike1, accessed 2017-02-15

The em32 Eigenmike® uses a sphere with a diameter of 8.4 cm with 32 omnidirectional microphone capsules. The signal can be converted as needed by matrix calculations.

In the documentation 8 the lower frequency limit for the different orders is given as

- $\begin{array}{rrrr} \text{Order} & f_0 \\ 0, 1 & 30 \text{ Hz} \\ 2 & 400 \text{ Hz} \\ 3 & 1000 \text{ Hz} \\ 4 & 1800 \text{ Hz} \end{array}$
- **Time Difference Localization** For the front channels this may be achieved with a surround mixing of the Decca tree. The signals get sent to the right, center and left channel. Two additional microphones in the back may complete this for a full five channel signal. The result is similar to that of the regular Decca tree, with added stability of the mid signal localization. It gets almost independent from the listener's position.[7]
- **Combination Assemblies** Like in the two channel stereo case they are used to combine the airy sound image of the time difference localization with the lowered cross talk of level localization.

With INA3 and INA5 three or five unidirectional microphones are used in a triangle or pentagon.

If this is generalized for flexible directional characteristics and distances in the similar setup called ASM5 (Adjustable Surround Microphone).

Perfect isolation at least for the angle of 180° is achieved with the IRT cross by M. Williams. It uses four unidirectional microphones each in an angle of 90° an a distance of 25-30 cm. At the optimal point for reproduction it delivers a good surrounding sound field. The signals are fed to the left, right, left surround and right surround channel. The imaging is not 1:1, because the front direction have too small angles in the localization and the back have exaggerated levels.

Helut Wittek [35] proposes the use of two varied ORTF arrangments, one to the front and one to the back. The pairs are angled 100° apart and separated by 10 cm, both pairs are separated by 20 cm. To improve the separation, super cardioid capsules are used. The separation of the channes is not complete like in the case of the IRT cross, but the assembly is so compact that is can be easily transported and it can even be purchased prebuilt.

This setup has been enhanced to add height information m see chapter 5.3.2, page $80\,$

Arrays Better results can be achieved by phase controlled microphone arrays. They can discriminate the number of directions that is equal to the number of microphone capsules. By using DSP technology it is possible to get uncolored recordings of highly resolved different directions. They may be transferred numerically for a given playback setting.

⁸https://mhacoustics.com/sites/default/files/Eigenbeam%20Datasheet_R01A.pdf, accessed 2017-02-15

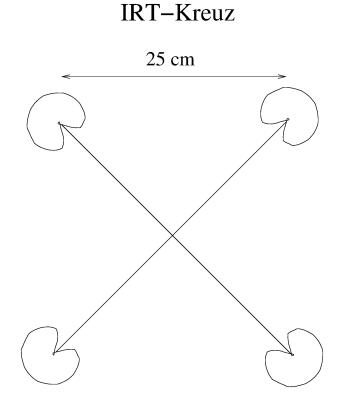


Figure 5.6: The IRT cross

The mix may put everything evenly to the five channels. For movies the mix often puts the dialog into the center loudspeaker, the music into the right and left channel with a standard two channel mix and the room reverberation to the left and right surround loudspeaker and ambient noise (the "atmosphere") to the right, left, right surround and left surround channel for the acoustical immersion. It is necessary to achieve the right amount of coherence to achieve the immersion and prevent a split between a front and back sound field.

5.3.2 More Channels

The 5.1 setup may be complemented by additional channels. The first attempts tried to get more channels into the plane to achieve a more even and precise localization for all angles there with up to 9.1 configurations.

Later it was found that for a natural impression the most important addition it the inclusion of sources outside of the plane. Sources from below are seldom needed and hard to achieve for playback, so that the additional channels and speakers are typically meant for angles from above. They may reproduce real sources (aircrafts, birds,...) and reflections from the roof of a room. Because the localization for these directions is limited in precision this is typically accomplished with two loudspeakers above the plane in the front region. Sometimes a loudspeaker in the nadir is added ("voice of god").

The hassle to create such a setup may be eased by using directional speakers in the front that are pointed to the roof, so that the reflection comes from above.

The assembly shown in chapter 5.3.1, page 79 can get added microphones for height information[35]. Four additional super cardioid capsules are used. Their signals are separated from the first four ones by level only by putting them at an angle of 90° relative

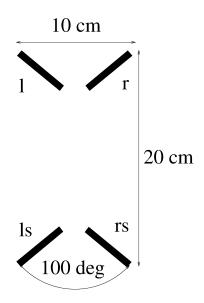


Figure 5.7: The ORTF surround setup

to them. The documentation does not mention the absolute angle of the capsules, but the drawings seem to depict that the first four capsules are facing 30° downwards

5.3.3 Virtual Sound Sources with Direction Information

A strong generalization of the fixed surround configurations and and formats can be achieved by using distinct channels that carry direction information. They can be matched to the optimal distribution of replay channels for an arbitrary given loudspeaker setup.

5.3.4 Ambisonics

This technology records the complete set of spherical functions at least of first order. They build a base of four dimensions. Two mathematically equivalent descriptions are

- **B-Format** consisting of the components 0 (pressure part, omnidirectional), X, Y, Z (pressure gradients along the relevant coordinate axis). It can be directly recorded with an omnidirectional and three perpendicular figure-of-eight microphones, all ideally coincident. This is a three times MS setup.
- A-Format consisting of four components directed at the corners of a tetrahedron. This may be approximated by four unidirectional capsules that are very close together and pointed into the right directions. A practical solution is the Soundfield microphone. [30]

These formats are not only transformable to each other, they stem from a general mathematical theory and can be transformed to the optimal distribution of signals for arbitrary loudspeaker setups, starting with the 0 signal alone for monophonic reproduction.

The other way round any collection of signals of arbitrarily oriented microphones with one of the classical directional characteristics that work coincidentally can be transferred with matrix computation into one of the bases, provided there are enough independent channels.

5.3.5 Wavefield Synthesis

Like the name implies this technology tries to reproduce the whole wave field during playback. It uses an array of loudspeakers that get fed with signals that create the desired wave fronts outside the near field of the loudspeakers by using the Huygens principle. It should match the original (or emulated) sound field in direction and amplitude. To get close to this goal it takes a lot of channels for recording, mixing and playback.

6 Electrical Transmission

6.1 Voltage Adaption

In audio technology voltage adaption is used almost exclusively. This means that outputs are built with the lowest possible impedance and inputs with the highest impedance. This ensures that small deviations in impedance over the frequency do not show in the resulting transmission, because practically the whole output voltage is transmitted to the input.

Another advantage of low output impedance is that typically loads by even long cables by there capacitance do not lead to an upper frequency limit inside the bandwidth of audio signals. Otherwise the treble region would be impaired.

It also puts a high load to external noise sources and diminishes their level.

The general rule is that the input impedance must be at least ten times that of the output.

Not often used but possible is that there may be several inputs connected to one output. This impairs the quotient of impedances. The split must be done close to the output, so that both cables are directly connected to the low output impedance (Y-cable).

6.2 Asymmetrical transmission

For short distances an asymmetrical connection can be sufficient. It directly connects the ground and signal wires. This simple kind of connection can lead to problems.

- Input and output are forced to have the same ground potential or at least one of the components must be completely floating (decoupled from ground).
- The whole connected setup must not be grounded at more than one place, except for a complete cross connection with very low impedance that forces an equal ground potential. In all other cases a hum loop may occur. It may result in a loud hum noise that completely suppresses the signal.
- Every external noise at the ground of a cable or at a device cannot be separated from the real signal.

If these problems arise, it may help to separate the ground potentials by means of a (good) audio transformer.

The typical connectors for unsymmetrical connections are jack and cinch. The pin assignment is

| Jack | Cinch | Signal |
|--------|--------|--------|
| Sleeve | Screen | Gnd |
| Tip | Center | Hot |

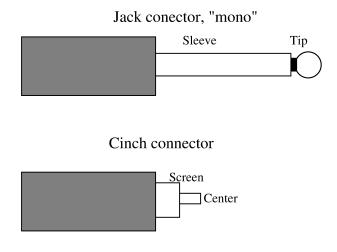


Figure 6.1: The jack and cinch connector

6.3 Symmetrical Transmission

To further reduce stray noise, a symmetrical transmission is used. Often this is done by sending the original signal over one wire and the inverted one over another wire. The input uses the difference of both voltages, either active or by means of a transformer.

For the noise cancellation is is only necessary to have both wires connected with the same impedance.¹

As usual the cable is screened and the ground level is also available at both ends. To keep the stray noise on both wires as close as possible, they are twisted. This also removes magnetic stray signals, because every round of the twisting delivers an area for induction that inverts the phase and therefore cancels of the last one.

These measures ensure that the remaining stray noise on both wires is equal. The difference that is produced on the input therefor cancels it.

Active difference amplifiers can work down to arbitrarily low frequencies and their maximum voltage swing is close to frequency independent. They have difficulties with large potential differences, because at least one of the inputs may be taken outside its linear region or even into the region of destruction when there are no protection circuits.

An input transformer is mostly immune against potential differences. It does however have problems with high levels for low frequencies. Its momentary voltage is proportional to the time derivative of the magnetic flux density. Therefore this must grow proportional to the period time and there is a limit for the maximal magnetization of the core material. This is very problematic for high level of extremely low frequencies, which may occur for pop noises or wind eddies. It is important to avoid any form of direct current at the input. It leads to a high current which may permanently magnetize the core (remanence). The transformer subsequently does not work symmetrically which limits the maximum input level and leads to high asymmetrical distortion.

Normally the symmetrical connection is done via three polar XLR connectors. They have a notch. There are male and female connectors. The output us always male.

The connector transmits the following signals

- 1 Gnd
- 2 Hot (+)
- 3 Cold (-)

¹Same as in: as closely matched as possible, deviations maximally in the low single digit percent region.

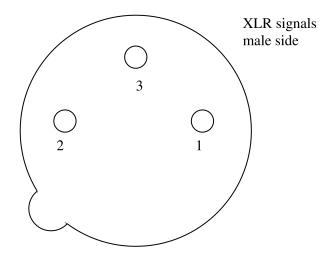


Figure 6.2: The XLR connector

A three pole ("stereo") jack connector can also connect symmetrical signals, its poles are

| Sleeve | Gnd |
|--------|------------|
| Ring | Cold $(-)$ |
| Tip | Hot $(+)$ |

Jack connector, "stereo"



Figure 6.3: Stereo jack

The downside of the jack connector is the unavoidable short circuit during the plugging or unplugging. This forbids the use for phantom powered connections.².

6.4 Cables, Connectors

The main requirements for cables in audio technology are

- **mechanical stability**, it must not break or change its characteristics temporarily or even permanently under mechanical stress (bending, twisting, tension),
- **reasonably low resistance,** so that also for long cable lengths the voltage is not changed relevantly,
- **low capacitance**, to avoid treble loss on long cables by capacitive loading of the output impedance or cable resistance,
- **good, HF proof shielding,** to prevent the stray noise on the signal wires. It should be done with interwoven wires, possibly combined with a conductive foul.

 $^{^{2}}$ See section 6.6, page 86

The same requirements have to be met by the connectors. They have to provide an interface and mechanical characteristics that guarantee secure and low resistance connections even after many plugging/unplugging cycles.

The wave impedance of the cable is only relevant for HF or digital signals. For LF the wave length is far larger that the usual cable lengths. This means that wave effects will not apply on LF cables.

Dubious products with extremely high prices cables that provide additional unnecessary or only alleged characteristics belong to audio technology voodoo and have no right to exist.

6.5 The DI-Box

A DI-Box converts unsymmetrical signals, even those with high level for direct loudspeaker connection, in a way that the output is suitable for the direct connection to symmetrical inputs (**D**irect Injection).

Typically they provide a switchable voltage divider at the input and parallel output connector to forward the unsymmetrical signal typically to an amplifier.

After that there is a circuit that creates a symmetrical low impedance signal, possibly with a lowered level.

6.5.1 Passive DI-Box

A passive variant uses an LF transformer. It gets transformed downwards to get a low impedance. The voltage gets transformed with the quotient of windings, the impedance with its square.³

The circuit does not need any power source and can handle large potential differences. It is not suitable for high impedance input signals like those from passive electric guitars, because they get excessively loaded which leads to massive sound changes.

6.5.2 Active DI-Box

The active variant consists of a buffer amplifier with a high input impedance, a phase reversal stage and two low impedance buffer amplifiers for the output. The symmetry is built electronically. The circuit can be powered by a battery of via phantom power (see below). The input impedance may be extremely high, and a large range of frequencies and amplitudes may be converted linearly. The missing transformer limits the possible potential difference between input and output.

6.6 Phantom Power

Phantom power allows to power a device (typically a microphone) over the signal cable without changing the (voltage difference) signal of the output. Hence the name. Therefore the same power voltage is fed over resistors to both differential wires.

It is necessary to keep the symmetry as close as possible. The resistors that distribute the power should differ by less that 0.4%(!). Otherwise there is a net direct current

³An example circuit can be seen from Sengpiel[24], URL http://www.sengpielaudio.com/ PassiveDI-Box.pdf, accessed 2015-05-18

between the signal paths which can permanently magnetize an input transformers of the microphone preamplifier. In all other cases the resistor difference weakens the impedance symmetry.

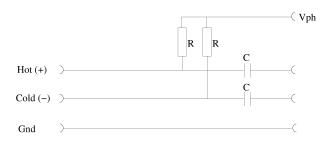


Figure 6.4: Schema of phantom powering

The standardized parameters of phantom powering are

| Name | Voltage | R | Maximum Current |
|------|-----------------|------------------------|--------------------|
| P48 | $48 \mathrm{V}$ | $6.8~{ m K}\Omega$ | 10 mA |
| P24 | $24 \mathrm{V}$ | $1.2~\mathrm{K}\Omega$ | 12 mA (presumed) |
| P12 | $12 \mathrm{V}$ | $680~\Omega$ | 15 mA |
| | _ | | |

There are many devices that do not provide the necessary current (it has to be provided on all inputs simultaneously). Often the resistors are not matched closely enough. P48 is most prominently used.

6.7 Analog to Digital Conversion and Back

When a conversion to the digital domain is done, the original signal is put into discrete values in two domains.

The amplitude is expressed by an integer. This replaces the continuous analog value by a number that stands for an interval of possible input values.

Without further measures theoretically this leads to a distortion of the wave form, the transfer function consists of stair steps. This loss of information is not real as soon as the stair steps are smaller than the level of the noise in the signal. The information about the signal proper remains complete.

If the digital signal resolves less than the precision of the analog signal, there are really artifacts by discretization. They consist of stair step distortions or (soft) clicking noises for low frequency input signals with very low level. This kind of noise is correlated heavily and in an extremely annoying way to the input signal and must be avoided at all cost. This can be achieved by adding an artificial noise to the analogue input signal that has an amplitude that is at least as high as on step in the digital signal. Its frequency characteristic can be tailored to keep its level low for the frequencies where our ears are most sensitive. As result the converted signal has a ground noise. Small band signals whose levels are far below the resolution of the digital signal are retained with this method. Far too few people are aware of that fact.

The time domain gets sampled at equidistant places. This appears to be a drastic information loss. According to Shannon's information theory it is necessary that

the input signal has a limited bandwidth. In that case two samples per period are sufficient to retain the full information that was present in the input signal.

Signals in the input with frequencies above that maximum appear at frequencies that are mirrored at the maximum frequencies. This effect is called aliasing. It has to be avoided at all costs, because these frequencies are inharmonic to the input signal and thus these components of the output signal are extremely annoying.

In the beginning of digitization in audio technology the deep pass filtering was done by extremely steep analog filters. These had their problems. They were massively complex with components that had to ensure tight tolerances and in the region below the frequency limit the phase was strongly changed.

In the mean time there are almost universally so called 1-bit converters. I will sketch their principle and try to show the major aspects of their operation.

They create a difference signal between their momentarily approximation and the input signal. This is done with a frequency that is a rather high integer factor of the conversion frequency. The analog filter must only ensure that the input signal has no components that are higher than the half of this internal frequency. It does not have to be steep or very exact. The approximation process is built in dedicated (and integrated) hardware. In principle it creates a digital (computed) filter on the bit stream which is phase linear and removes all frequencies that are too high with a steep slope. It is compared to the instantaneous input signal. The inherent high error of the one bit signal is reduced by averaging and filtering until the resolution of the output signal is met. This is an extreme case of dithering. The internal noise is shaped in its frequency response in such a way that the error noise is moved above the bandwidth of the input signal and removed by the lowpass filtering.

Because the effective lowpass filter is computed digitally there are no complex analog filtering circuits with their problems any more.⁴

The remaining critical parameters with this method are only the constant frequency and the stable reference voltage.

The conversion from digital to analog works with the same principle: the internally generated one-bit data stream is filtered. Again, the result is almost ideal without large dependency on tolerances of electronic components.

About common misunderstanding in the domain of DA and AD conversion I strongly recommend the video and accompanying wiki by xiph.org "Digital Show and Tell" [40].

For typical analogue audio signals 24 Bit are sufficient to represent them lossless.

6.8 Digital Transmission

An important characteristic of digital audio signals is a much higher bandwidth than the analog signal. For the transmission the typically HF aspects are important and the wave impedance of cables must be taken into account. Therefore other kinds of cables must be used unless the length is very short.

6.8.1 S/PDIF

S/PDIF delivers a serial signal which may contain several channels. The signal itself is coded in the same way as with AES/EBU below. The metadata are encoded in a binary

 $^{^{4}}$ See Swen Müller, [12], pages 27 ff

format, not ASCII text.

With the right converters AES/EBU and S/PDIF interfaces can get connected to each other.

The audio data get transmitted with 2, 2.8 or 3.1 MHz. The format may transfer 24 bit long data or a data stream in one of the formats MP2, AC3 or DTS.

The electrical connection may be done asymmetrically (75 Ω , coaxial, max. 10 m, TTL for very short distances) or optically (TOSLINK, polymer cable). The optical connection avoids problems with differing electrical potentials. The protocol is meant to be used by end users and is often used for the audio transmission in audio-video setups.

6.8.2 AES/EBU

AES/EBU is a format that transmits an audio stream serially. For 24 bit data the meaning of the trailing four bits is not specified. Metadata are transmitted as ASCII strings. There are two different electrical variants.

AES3

The electrical specification is that of RS422. The cable is specified as 110 Ω shielded twisted pair.

AES3i

The transmission is done asymmetrically via a coaxial cable.

6.8.3 AES42

This norm describes the digital connection and protocol for professional microphones to the matching inputs over a symmetrical transmission. It gets powered over 10 V DPP DPP (Digital Phantom Power) and includes a back channel with an amplitude of 2 V which may change the parameters of the microphone.

The clock of several microphones may be synchronized centrally. The audio signal is encoded in AES3 format.

If the microphone uses two A/D converters in parallel with different input levels which are calibrated against each other it is possible to convert such a high level interval that it is possible to work with a fixed sensitivity.[10]

6.8.4 Integration into Video Signals, HDMI, DisplayPort

For digital transmission of video signals the audio signal is often included. This is the case for HDMI and Display Port.

6.8.5 USB

Some devices use the USB connection to transfer audio signals as USB Audio. This includes microphones that have a built in AD converter and a phase locked DA converter for headphones. For standard microphones there are also adapters that can be plugged into the XLR connector.

6.9 Wireless Transmission

There are several methods to do a wireless transmission. The range depends on the transmitter power, the sensitivity of the receiver, the environment and the wavelength used. Short wavelengths (and thus high frequencies) of radio waves limit the range. Some materials like steel enforced concrete that contain conductive material can limit it severely.

Standing waves lead to cancellation at the knot areas. To prevent interruptions in the transmission it is therefore absolutely necessary to use receivers with two distinct antennas (diversity). A synchronous cancellation at two places is much more seldom than at one place.

Modern systems use additional communication paths like IR transmission to exchange parameters between transmitter and receiver.

6.9.1 Analog, FM

The classical form of radio wave transmission uses frequency modulation (FM). To improve the signal to noise ratio it often uses a compression of the dynamics in the transmitter and a reverse expander in the receiver. This may lead to artifacts that may disturb in extremely demanding applications.

Most radio frequencies can only be used if the necessary permissions have been obtained. The rules are differ between nations and got changed rather often in history.

It is important to take care of interferences and simultaneous use of the same or too close frequencies.

6.9.2 Digital

Newer gear uses digital transmissions. The quality is not really limited by the wireless technology.

The standard audio format of BlueTooth is not suitable for professional live use, because most codecs use strong digital compression to use a low bandwidth. Most of them lead to a long delay which prevents them to be used for multi channel recordings or live events.

6.10 Digital Storage - Formats

Digital audio formats contain a header which describes the internal structure, followed by the audio signal and possibly additional information.

Typical formats are .wav (Microsoft), .aiff, .au. There are several combinations of number of channels, bit depths, bit formats and sampling rates.

Often used are

Channels 1 (mono) bis 6

Bit depth 16, 20, 24, all linear; 32 Bit float

Conversion rate 44.1 KHz, 48 KHz, 96 KHz, 192 KHz

During recording it is a good idea to work with plenty of dynamic reserve to prevent overdrive and still keep enough level above the noise floor.

32 bit floating point data deliver a practically unlimited dynamic. This format is very useful for the computation of intermediate steps because it limits the stacking of rounding errors by long algorithms (mostly when there are lots of very low values). This is important for the computation of filters when there are extreme values (notch filters...).

To deliver the end result the 44.1 KHz and 16 bin – the CD standard – are still sufficient. Higher rates allow to shift the residual noise completely out of the hearing range and allow for a small increase of signal to noise ratio. The added storage requirements are however much larger than in the case of a greater bit depth with less technical overhead.

6.11 Digital Compression - Lossless

For digital audio data you may spare a lot of space.

Lossless compression can be inverted later. The best known formats are flac and shorten. For everything that is not currently in use there is no reason not to use these formats for storage instead of the uncompressed variants. The typical applications can directly play back the flac format. Its definition is completely open and there is free software and free libraries to work with them.

The size is halved as a rule.

6.12 Digital Compression - With Losses

If you take advantage of the masking effect in the hearing process and leave away or simplify the masked regions of the audio signal, you may spare a lot of space.

Because this removes information, the process cannot be reversed.

Well known formats that work this way are mp3, off, aac. With data rates above 256 Kbit/s listeners were not able to discern a difference to the original signal in a blind test by Heise Verlag for the computer magazine T. However, if the underlying base for masking goes missing, artifacts can be spotted. Because all is about notches in the frequency response, it sounds a bit like a phasing/flanging effect. Examples are processing by EQ or a hearing loss that prevents the hearing of the masking frequencies.

If used with reasonable data rates these formats are suitable for the transport on mobile storage and to deliver the final product. They are not usable if something shall be changed on the signal later on, e.g. single tracks before the mix.

7 Electrical and Digital Processing of Signals

7.1 Amplifiers, Level

Amplifiers have the task of increasing the level of the input signal with respect to the voltage or current. The idea is that they should not degrade it in the process. They should be as neutral as possible and not add noticeable noise of any kind.

7.1.1 Microphone Preamplifier

Microphone preamplifiers raise the level of microphone signals that typically is very low into the usable level for further processing. The typical line signal levels in the studio are -10 dBu for consumer grade equipment and +4 dBu for professional equipment. Some devices can be switched to any of both working levels.

Because of the usually low level of the input signal the input of the preamplifier must have a very low level of self noise. This must be maintained in the case of a lowered input sensitivity for higher levels. It is not trivial to ensure this because this leads to high impedances for the internal negative feedback resistors.

Bad preamplifiers often have a minimum sensitivity that is too high. This is problematic if even a dynamic microphone with its inherent low sensitivity overloads the input. This may be countered by using a resistor circuit in form of a T. It is important to put this circuit directly in front of the amplifier input. Otherwise the signal in the cable gets more sensitive against stray noise.

The lowest possible noise level may be achieved by using a transformer that is tailored to the circuit used. Another advantage is that this leads to a high degree of symmetry and immunity to potential differences to the source. The downside is the danger of overloading the magnetic core when the input includes high levels of very low frequencies (e.g. wind noise).

It is important to have a good suppression of HF stray signals (cell phones!).

Most microphone preamplifiers include the option of providing phantom power.

Sometimes a high pass filter may be added at an early stage of the signal path. It can help against sub-bass signals that might otherwise overload the amplifier.

7.1.2 Power Amplifier

The power amplifiers are used to raise the input signal level and feed it with very low impedance into loudspeakers with an impedance of typically between 4 and 8 Ω .

The strongest power amplifiers are used for PA. They have to be extremely robust (electrically – against short circuits, against extreme ambient temperatures and mechanically) and precise. In the studio they either work with digital signal processing or analog frequency dividers integrated into the loudspeakers (active loudspeaker) or stand alone and silent – without fan. This limits compact designs.

Each and every reasonable modern amplifier is neutral towards the sound. For precisely same levels and without overload they cannot be discriminated by ear in blind tests.¹

7.2 Automatic Level Control: Limiter, Expander, Gate

The variation in level of a raw recording is often too large for playback for technical or aesthetic reasons or in view of the realities during playback. Aside from manual correction there are circuits or algorithms that can change the levels according to parameters.

7.2.1 Compressor

Just to clarify: the compressor has nothing to with the digital compression of audio formats, although the same word is used!

The compressor is an automatic volume control that lowers the amplification of an audio signal dependent on the momentary level.

In analog technology different circuit concepts were used to achieve this.

- **Optoelectronic compressor.** A lamp that can react quickly (electrolumeniscence, LED) lights up an light dependent resistor (LDR) that is the lower part of a voltage divider. Reaction and release times are dependent on the property of the LDR.
- **FET as voltage controlled resistor.** The control voltage is connected to the gate of a field effect transistor that again works as the lower part of a voltage divider. The time constants depend on the circuit that provides it. A problem is the high variation of parameters between specimens and the limited linear voltage range between source and drain.
- VCA. An amplifier with voltage controlled amplification allows the most neutral and most freely parametrizable control of the amplification. The circuit is built around tranconductance amplifiers.

All of those circuit designs have a level at which the control begins, called threshold. Otherwise the lowest levels would be emphasized too much, increasing the noise floor.

Another parameter is the remaining slope of the output level related to the input level over the threshold. It is normally called ratio.

These static parameters are complemented by two time parameters, the attack time until the control sets in and the time until the neutral state is reached after the end of the excitation, the release time.

A long attack time keeps short impulses, a short one can help against overload, but not completely prevent it.

A short release time restores the normal amplification fast. This may restrict rhythmic variations in level, which is not always useful. The strong and fast change of amplification for background signals can irritate ("pumping").

¹See http://sound.whsites.net/amp-sound.htm and about a test setuphttp://sound.whsites.net/ absw.htm at [4]

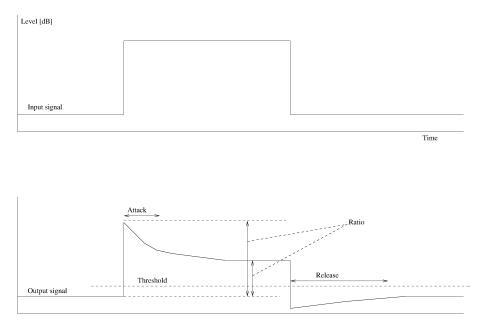


Figure 7.1: The parameters of a compressor

The time constants should be chosen to be suitable for the typical timing of the input signal.

The compressor leads to less difference in level. Softer moments of a signal may be perceived better. The signal gets thickened.

To find the useful combination of parameters for a given signal can be very hard and time consuming. Ethan Winer once noted that this time can be used to achieve an even better result by controlling and automating the volume setting. The result is often better.

For short time constants the control may change noticeably during the period of low frequencies and thus change the wave form. This creates harmonics, the compression gets closer to that of a distortion. Indeed, a compressor with time constants both set to zero is a distortion circuit.

A compressor that only works for high frequencies via crossover is called de-esser. It can limit exaggerated hiss signals from speech.

Multiband compressors use crossovers to compress frequency ranges independently. This prevents cross modulation of different frequency ranges.

The control signal can sometimes be processed externally $(EQ, \ldots, this is called side chain$ ing) or obtained form an external signal. This is useful to give the external signal priority. This is called ducking.

Changing the Decay Time Constant of Percussive Signals

For percussive signals with an amplitude falling over time, the time constant can be changed by a compressor – in both ways.

Prolonging can rather trivially be achieved by a compressor. If it divides the dynamic by a denominator and the time constants of the compressor are very short compared to that governing the original signal, the slope (in a logarithmic depiction over time) gets divided by the same number at least in the initial section of the delay. This results in a prolonging of the decay.

As side result, the dynamics of the whole signal is diminished, including the attack. This may be desired or not.

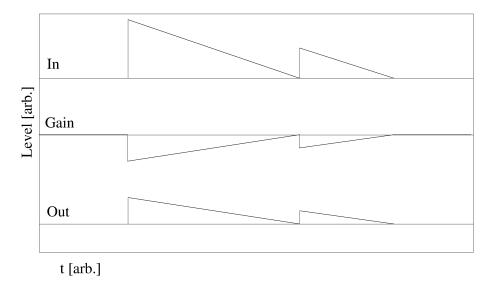
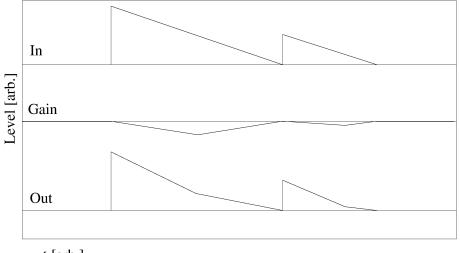


Figure 7.2: Compressor: prolonging a delay

Shortening the decay can be done by taking a short release time constant (to start fresh without gain reduction on an fresh pulse) and using an attack time constant that reduces the gain with its time constant. The slope of this gain reduction over time (in a logarithmic view) adds to that of the original signal.

The starting phase of the signal is not altered at all in the ideal case, so the dynamic of the onset of each pulse is unchanged.



t [arb.]

Figure 7.3: Compressor: shortening a delay

7.2.2 Limiter

If the slope is set to the extreme case of zero, this sets a fixed highest output level. If the controlled signal is delayed before the control takes place, this can be used to prevent any

higher output level, because the limiter "can see into the future". This is called a brick wall limiter.

7.2.3 Expander

If the direction of the control signal is inverted, the dynamic gets increased instead of being lowered.

One possible use of that principle is to use a compressor before a signal gets transmitted or stored in a noisy channel and to reverse that change after the transmission or recall. This was used for tape recordings (Dolby A/B/C, HighCom) and for wireless analog transmission of microphone signals.

Expanding the treble region below a threshold can be used as one sided noise reduction method, e.g. during restoration of old recordings.

7.2.4 Gate

The extreme case of an expander mutes the signal below a threshold. This removes low level noise in signal pauses completely. If this cutting off is perceivable it may be more disturbing than the original noisy signal.

7.3 Mixing

The most useful analog variant to mix signals uses a potentiometer to weigh the level of each signal and generates a current source by use of an output resistor. These current signals get combined at the virtual mass of a current to voltage amplifier that is built in an op-amp circuit.

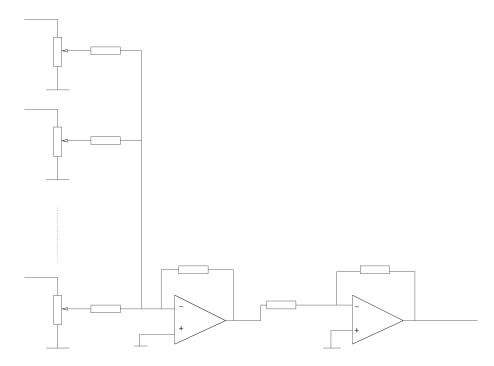


Figure 7.4: Basic analog mixer circuit

In the digital domain it is only necessary to compute a weighted sum of the signals. If this is done with integer computation it can lead to overflow or it this is prevented to a loss of resolution if the digital levels get too low.

In the case of floating point computation there are no real problems. One could counter possible cancellation effects by sorting the summands according to their absolute values and starting the addition beginning with the lowest absolute values. I do not know whether this is done somewhere in real life.

7.4 Filter

Filters are used to limit the frequency domain. For acoustic recordings of signals that are known to have a limited frequency region you may remove all frequencies that cannot be provided by the source. This is often true for bass frequencies that are not reached by an instrument (low bass for a piccolo flute, a triangle, ...), but are present in the signal (noise from outside in the case of non-ideal acoustical separation, vibration on the microphone stand, wind noise).

It is very useful to use a high pass filter for all channels that do not need low frequencies. This avoids most unwanted masking effects, spurious triggering of compressors and more.

7.5 Equalizer

Equalizers change the frequency response of a signal by controlled attenuation or amplification of some frequency regions. As the name implies this war first meant to remove deviations of an unequal frequency response.

7.5.1 Shelving Filter

The simplest useful circuit to change the overall frequency response allows to attenuate of amplify the bass and treble region independently.

It can be built with relatively few components in analog form. With logarithmic potentiometers there are pure passive variants which lower the overall level strongly. In a feedback loop it is possible to work without level loss in the neutral setting. It also allows the use of linear potentiometers that have much better precision and the frequency response of attenuation and amplification are symmetrical. This is called the Baxandall circuit.².

More generalized variants add one or two mid channels with fixes center frequencies and bandwidths.

7.5.2 Graphical Equalizer

If you generalize the principle of the single sound control circuits by adding more and more controls for different frequencies with bell shaped response the result is a device that aims to reproduce the frequency characteristic that is depicted graphically by the setting of the different typically slide controls.

There are many variants to achieve this with analog circuits. The simple circuits that use LRC circuits or most of the time a gyrator circuit to replace the inductance have a Q

²See http://sound.whsites.net/articles/eq.htm auf [4], search for "Basic Tone Controls"

factor for each frequency that is dependent on its amplitude setting. More complicated variants provide active decoupled filters for each frequency and provide a constant Q.

Even those provide a frequency characteristic that deviates from the setting on the front panel, because the neighboring frequency settings have a residual effect. An extreme example is that the maximum of a frequency response is higher if all controls are put to the maximum than if they are alternating on maximum and minimum.

By means of digital signal processing it is possible to reproduce the control settings in the resulting frequency response because the interaction can be corrected.

The graphical equalizer allows to find a matching setting for a given deviation of the frequency response quickly and trivially by mirroring the levels.

The usual versions give between seven and 31 frequency channels. For 10, 15 and 32 channels the frequencies have be standardized to allow direct comparison between analyzers and the correcting equalizers.

Because the frequencies are quantized and fixed it is often not possible to match a needed frequency characteristic exactly, for example if a given arbitrary resonance frequency has to be countered.

7.5.3 Parametric Equalizer

The concept of the parametric equalizer was introduced by George Massenburg. With relatively few bands whose parameters (hence the name) can be modified completely, they allow a precise an targeted change of the frequency response.

For each band there are

- **Bell/Shelf:** only for the highest and lowes band: either the band provides a resonance (bell formed characteristic) or shelf like in the case of the simple shelving equalizer, the maximum or minimum is reached at the relevant end of the spectrum.
- **Frequency:** the mid frequency or for the edge frequency in the case of the shelf of the band.

Level: the amount of level change at the mid frequency (bell) or extreme frequency (shelf).

Q: the Q-factor, sometimes replaced by the bandwidth which is inversely proportional.

With these parameters it is possible to notch out problematic frequencies or to amplify or diminish arbitrary frequency regions.

The correct usage of this effect is everything but trivial because it is hard to find the matching parameter combination for a given sound change. A lot of knowledge and experience is necessary. Like many extremely efficient tools it is dangerous in the hand of inexperienced people.

7.6 Time Manipulation Effects

These effects work with delayed copies of the input signal.

7.6.1 Delay, Echo

Analog circuits work wit bucket chain filters or mechanically with tape loops to delay a signal. If the original signal is mixed this results in a one time echo (slap back, often used in rock'n'roll). If the delayed signal is fed back to the input of the delay unit, a succession of equidistant echoes is created. This either gets lower in level exponentially with each repetition (flutter echo) or they build up in volume until the point of saturation is reached (UFO-effect).

Digital technology can achieve delays in integers of the clock time trivially, the stored value is read out. Intermediate delay times are also possible, but this requires reading out several values to compute a band limited interpolation.

The delayed signals often get attenuated in the treble region (similar to the effect of a long distance through air).

7.6.2 Reverb, Room, Early Reflection, Delayed Onset, Plate, Spring

If very many delayed copies of a signal with an overall exponential decrease in level and a density that rises over times and best still coming from different angles are combined, you get a reverberation tail, the last part of a real room reverberation.

The first few echoes of a real reverberation signal all have an influence on the perceived room characteristics and position of the source in the room. Therefore these echoes can often be parametrized independently from the reverberation tail (early reflections).

The closer the source is located to the listener, the longer is the delay of the first echoes. For the ceiling, the bottom and the side wall this happens non-linearly after Pythagoras' law, for the back wall even two times linear (forward and back). This effect can be utilized in the simulation of reverberation to achieve a controlled perceived distance. For equal levels or reverberation the source appears closer when the reverberation signal gets delayed as a whole (pre-delay). The effect gets stronger when the relative level of the reverberation signal gets lowered.

This effect was first used for reverberation rooms. In a strongly live but diffracting room there was a loudspeaker and several spaced microphones. The signal was delayed by tape loops to achieve the requested impression.

You have to be aware that very early reflections lead to sound changes by interference (comb filter). After that there is a time region that matches typical time constants of speech and thereby degrades the legibility of speech, because the reverberation of one phoneme leaks into the next one. This can also degrade the perception of music, because our ear is trained to perceive effects in this time range very detailed. After this time the reverberation tail leads most of all to the perception of room size and surrounding by the room.

To sound naturally the reverberation should decrease the treble over time similar to the situation in the air when sound travels long distances (and times), depending on the dampness. Many porous materials and thin absorbers also work mostly in the treble region.

If this treble change is omitted, you get the characteristic of one of the first devices to emulate a reverberation with less space requirements than a real room. It is a metal plate which reflects bending waves at its ends. The programs that emulate this behavior often are named plate reverb. Another emulation of reverberation used the delay of torsion waves on spiral spring (spring reverb). Single springs lead to problematic flutter echoes. For high quality spring reverb, several different springs were used in parallel and sometimes there were additional reflection points in form of bends or weights on the springs to get an even denser signal. One noticeable effect of all spring reverberation units that is missing on other reverberation units is the strong dispersion. The bending wave with short wavelength have a stronger force that they work against so they travel faster. For each echo you can hear the time difference depending on the frequency (it sounds a bit like 'tshuuuiiii'). This is not prominent for source signals that are relatively constant over time like e.g. organ sounds, but it is very strong for wide band short signals like drum sounds.

7.7 Modulation

Very noticeable are all effects that lead to a cyclical modulation of the signal, because such modulations transport emotions in speech to a high degree.

7.7.1 Tremolo

Tremolo means the cyclic modulation of the level of a signal. If you do this on both sides of a stereo setup with inverted phase, it results in a panorama modulation.

If the modulation frequency is raised into the hearing region, the sound gets very rough because of the resulting sum and difference frequencies. If the frequency gets even higher the mirror frequencies stand out as inharmonic frequencies which either stand at one frequencies of even wander in opposing directions (sum, difference) when one component changes its frequency. If the modulation is so strong that the polarity gets inverted, the effect is called ring modulation.

7.7.2 Vibrato

Vibrato is the name for a cyclical sinus modulation of the frequency of a signal. Afterwards this can be obtained by the modulation of the delay of a signal.

7.7.3 Phasing

A phasing effect consists of the mix between the original signal and the result of a modulated all pass filter. This creates several cancellation frequencies that change cyclically. This effect is very effective on broadband and spectrally dense signals with noise components, because there are no original notches in the signal that mask the effect.

7.7.4 Flanging

Flanging originally consisted of two tape drives that played the same signal, and the fine control of the speed was changed by breaking one of them manually at its flange (hence the name). For few notches (low time difference) this creates an effect that is similar to phasing. It is a modulated comb filter effect. These days the effect is done by adding the original signal with a delayed copy and modulating the delay time in a range that leads to strong comb filtering. There may be many more canceled frequencies than for typical phasing, and the frequencies are equidistant. Sometimes the delayed signal is fed back to the delay circuit input, which can lead to a obvious frequency change in the resulting signals after several repeated changing delays.

7.7.5 Chorus

Chorus combines the original signal with one or several modulated delayed copies like flanging does. This time the delay is so long that there is no noticeable comb filter effect, the cancellation frequencies are too many. The effect is that the ear hears each of the signals as single uncorrelated source, for example a choir (sic!) instead of one singer.

7.7.6 Rotary

The rotary effect is a very complex modulation. The original is created acoustically and mechanically. The first one to deliver this effect was the brand Leslie, and often the effect is called after them. The main use is the electro mechanical Hammond organ. There are several variants, the most elaborate one uses a frequency divider. This treble region gets fed from a horn driver to a curved and rotating horn. The Doppler effect leads to a vibrato because of the changing path and subsequent acoustical delay. Depending on the details of the horn mouth and its directivity (originally equipped with a diffusor) and the distance to the listener/microphone this also results in level changes, an additional vibrato effect. If a reflecting wall is located behind the construction it creates not only a simple echo, but its frequency modulation is phase reversed. When the distance to the listener/microphone increases, the distance to the mirrored source decreases and vice versa. If the echo is not too weak, this leads to a chorus effect.

The bass part gets delivered by a bass speaker the points to the ceiling an feeds a curved one sided drum that also rotates. This affects the directivity of the mid frequencies so that the listener gets a tremolo in the mid range, while the bass is barely affected because of diffraction.

The construction can be switched to run in two fixed speeds. In the slow setting the main effect is a complex chorus effect. By the rotation and variable dispersion of the sound the stereo image is also strongly modulated. In the fast setting the main effect is a strong and fast vibrato.

The most prominent effect is the change between both speeds. The bass channel has a much larger moment of inertia so that its change of speed is much slower. This speed change effect can be used as means of expression by skilled musicians.

7.8 Frequency Change

7.8.1 Harmonizer

Like the chorus effect (section 7.7.5, page 102) this creates a second (or more) voice from the signal. This time the frequency gets changed by a constant factor. This is done by using a delay that reads out the copy with a different rate than the sample rate. Basically it is a delay whose delay time is modulated with a saw tooth signal. If the delay time increases, the signal gets slowed down, its frequencies are lowered. For a convincing signal there are some additional measures to be taken to prevent harsh artifacts.

• The jump to the begin of the modulation cycle has to be done without a jump in amplitude, for example from zero crossing to zero crossing or by blending to a signal

with a different phase of the modulation signal. This requires at least a second delay line.

• To prevent a noticeable echo the delay time must not get too long. On the other hand the lowest frequency of the input signal must fit into the modulation cycle to prevent phase jumps and a missing transformation of the lowest frequencies. To find a compromise, a frequency divider may be used, so that different frequency ranges are treated with their matching maximum delay times, dictated by their lowest signal frequencies.

Factor close to one create a convincing choir effect, without the typical cyclical modulation of the original chorus effect described above.

The effect can also be used to create signals that are shifted a given musical interval, e.g. an octave above or below with factors of two of 0.5 respectively. The linearly transformed formants change the sound of the newly created signal against the original, what may be desired, especially when the original signal is omitted. The result can be describes as Micky Mouse or monster effect.

7.8.2 Autotune

If the frequency of the input signal is analyzed and the harmonizer converts the fundamental frequency to match a given musical key, the signal gets normalized to these frequencies ex post. If the target frequency jumps because the input frequency slides the resulting jump in formants is very noticeable ("Cher effect"). The Autotune effect is often used inflationary alongside time quantization to automatically adjust shortcomings of the input signal.

7.8.3 Double Tracking

If the frequency of the input signal is analyzed and the harmonizer creates a second voice in a useful musical frequency, you may create a backing voice automatically. This works best if the formants are analyzed and steered by automated filtering in a way to make the sound as realistic as possible.

7.9 Nonlinear Characteristic: Saturation, Distortion

It is possible to send signals through a nonlinear characteristic, that means to distort them non-linearly. This may be the emulation of some real electronic or electrical components that have this characteristic. This should be seen as an effect and not as the standard optimal method of improvement.

The overdrive of an AD converter, of typical transistor or other circuits with strong negative feedback generate a hard clipping. Everything above a certain threshold is limited to this maximum. It is a drastic effect with sudden onset that immediately generates lots of harmonics even of high order. Normally this should be avoided, there are not many useful applications of this effect in music.

The overdrive of tube circuits with a weak negative feedback or none at all, similar FET circuits of anti-parallel diodes (especially germanium type) behind a resistor or magnet tapes generate a controllable overdrive, because the characteristic does not have a sudden

sharp edge but a rounded transition. The strength of the effect can be finely tuned by the input level. This is called overdrive or saturation. For pure tones this effect is not obvious until it is very strong, because the generated harmonics match those of the input signal. For several tones the effect creates sum and difference frequencies that are inharmonic and easily detectable. In the extreme case of complex and broadband signals the additional frequencies can be masked by other signal components (noise, cymbal).

Transformers do not overdrive at a fixed level but when the input voltage's time integral exceeds a limit. Depending on the core they can remain magnetized without signal.

The characteristic of a transformer may be simulated if needed. Therefore the signal has first to be integrated over time (-6 dB/octave), clipped and then differentiated again (+6dB/octave). It is better to leave the remanence out of the simulation, because the hysteresis makes the characteristic dependent on the whole history and thus irreproducible. The material remembers the last strong signal.

Many prefer the clipping of transformers as necessary part of old well known studio gear (Compressors,...) and want to have them reproduced by plugins. Personally I prefer to invoke effects one at a time in a controlled way, that means a compressor should be as neutral as possible, an overdrive circuit can always be added if needed.

7.10 Analog Processing

The first realization of all the described effects wad done by analog circuits. Frequency dependent control was done with the frequency dependent impedances of condensers or – more seldomly – inductors. After operational amplifiers were readily available, inductors were often replaced by gyrator circuits.

Delay was achieved by tape loops of for short time constants (flanger, sometimes echo) by bucket-brigade circuits. Complex reverb was achieved with real rooms or mechanical analogies in form of plates or springs.

7.11 Digital Processing

Digital processing has the advantage of being exactly reproducible and not necessarily adding noise to the signal.

The linear processing of signals is done by modeling the relevant impulse response (IR). The signal gets folded with this response.

As a rule all you need for a signal that is available in a computer is the necessary program (typically called plugin, because they get applied inside a host program) and can readily be applied in the computer. This avoids the need for space, cabling and sources of noise or errors.

7.11.1 Finite Impulse Response, FIR

The finite impulse response is a predefined impulse response of finite length. It is generally created by Fourier transformation of a requested frequency response and a fade out at the end instead of simple truncation (windowing). Apart from the restrictions by the finite length it is possible to create arbitrary frequency responses.

The two downsides of this approach are

- The finite length limits the absolute resolution in the frequency domain (uncertainty relation), which bothers most in the bass region.
- You cannot simply integrate variables for a parameter change in real time. For each change, the answer function has to be recomputed and loaded into the algorithm

For ling impulse responses the tail of the impulse response may be handled with less burden on the computation by creating its Fourier transform and multiplying that with the Fourier transform of the relevant part of the signal and back-transforming the product.

An extreme case of a long impulse response is a pre-computed reverberation signal taken from a real or virtual room.

7.11.2 Infinite Impulse Response, IIR

IIR algorithms work with feedback. For bad choices of parameters this can lead to a resulting oscillation. The requested differential equations get directly modeled as difference equations. Because the width of the differences is pre-set and fixed, the characteristic of the limes of the differential equations for infinite frequencies is now achieved at the frequency limit of the signal. The whole characteristics for high frequencies are a compressed version of the original characteristics.

Very complex frequency responses down to the lowest frequencies can trivially be achieved without much burden to the processors. The parameters can be changed transparently in real time.

For high frequencies the plugin should correct the mid or edge frequencies and the equally compressed bandwidths (or reciprocally the Q factors) to get close to the desired value. For bell shaped courves, the upper half is still more compressed than the lower one, but this does normally not concern much and is normally ignored.

Müller[12]³ gives the following equation for this transformation, f_s being the Nyquist frequency of the signal, f_{ale} the frequency used in the IIR algorithm, f_{res} the resulting frequency in the changed signal.

$$f_{\rm res} = \frac{f_s}{\pi} \arctan\left(\pi \frac{f_{\rm alg}}{f_s}\right) \tag{7.1}$$

The effective frequency must be multiplied in the algorithm by a factor of

$$C = \frac{\tan\left(\pi \frac{f_{\text{res}}}{f_s}\right)}{\pi \frac{f_{\text{res}}}{f_s}} \tag{7.2}$$

At least the lower flank of a parametric EQ band may be corrected if the effective factor Q_{eff} is transformed for the algorithm to

$$Q_{\text{alg}} = Q_{\text{eff}} \cos\left(\pi \frac{f_{\text{res}}}{f_s}\right) \tag{7.3}$$

7.12 Programs for Audio Processing

State of the art today requires that the programs at least internally work with 32 bit floating point data. During the audio processing the intermediate values can vary extremely in magnitude and a cumulative data loss would otherwise be unavoidable.

 $^{3}[12]$, page 57

7.12.1 Wave Editors

Wave editors are the simplest programs to process audio data. As a rule they allow the recording, storage and replay of audio data and the application of plugins. The processing is typically done directly on the audio data without keeping the original data. These programs are not meant to add tracks while simultaneously listening to the relevant existing tracks or to apply complex routing or to provide the original data and the steps of processing in a transparent way.

They are very well suited to cut tracks or provide measurements and apply finishing touches on a signal. The user interface can stay simple to use.

On free example of this kind of program is Audacity.

7.12.2 Digital Audio Workstation, DAW

The digital audio workstation is the Swiss army knife of audio technology. It features at least

- **Recording of additional tracks while monitoring the existing ones.** This is necessary to create overdubs, new tracks that shall be played back in parallel to the previously recorded ones.
- **Import of existing tracks.** Aside from recording, existing audio files can be imported as tracks and integrated into the project.
- Mixing of all tracks. From all tracks, a mix including panorama settings can be created.
- **Sound processing and analyzing with plugins**. Each track separately can be processed or analyzed with plugins or internally by the host program.
- Routing, intermediate mixing, subgroups. The signal flow can be determined freely. This allows to include external hardware effects with in- and outputs or to combine tracks to process them in the same way. A virtual track that is created by an internal submix is sometimes called a bus.
- Storage of all processing steps as history (total recall). The program stores the original data (unless these are cut and deleted by the user) and not their intermediate processing results. The parameters of processing get also stored and can be edited. Thus every step is reproducible and can be altered again.

Sometime there are added features like

- **Integration and processing of Midi tracks.** Midi tracks are edited and processed directly in the program. Thus there is no need for specialized external programs whose results have to be imported. This is much more flexible.
- **Integration of video tracks.** This is necessary if the result will be part of a video and the time of video and audio has to be synchronous.
 - A free program with these abilities is Ardour (GPL license).

8 Reproduction

8.1 Loudspeakers

8.1.1 Quality of Reproduction

The loudspeaker has the job to reproduce the sound signal optimally in a given region. The quality of reproduction is ruled by a set of important parameters.

Frequency response. This is by far the most important aspect. Optimal is a large region in which the deviation from an ideal constant line for the whole bandwidth is to be minimal. Normally it is only documented in the main direction.

Because of the way our hearing process works resonances increasing some frequencies are much worse for the sound quality than dips. One reason can be that resonances help identify the characteristics of the original sound sources, but dips by comb filtering is present in all setting including reflecting planes. Therefore they get more or less ignored during the hearing process. For low listening levels low frequencies (below 100 Hz) cannot be heard and do not have to be reproduced at all.

Most kinds of music use only frequencies down to 40 Hz, with the main exception of large organs. Even most drums reach only down to at about 60 Hz. There days there is no real excuse not reproducing frequencies up to 20 KHz.

Directivity characteristic. This aspect has been underestimated for a long time. Work by e.g. Anselm Görtz made clear how important it is. The characteristic should be as constant as possible at least from the mid frequencies to the upper limit, and the frequency response should be as constant as possible in the main directivity region.

Görtz uses a color coded display of the level (relative to the response at 0°) in dependency of frequency and angle (plotted horizontally and vertically). Sean Olive and his group use the mean frequency response in the whole room for three distinct areas as addition frequency response curves instead.

There are two important aspects of the angular characteristic. One is that the sound is not only defined in one narrow region, but the listener is free to move without big downsides.

The other is important even for a resting listener, and this is what Olive puts in perspective: The sum frequency response of the excitation of the room matches what the listener hears as direct sound. Linkwitz notes that our ear is trained to suppress the room sound part during the hearing process to recognize the sound source. If the components differ strongly, this is harder to achieve and the room sound interferes much more with the perception.

All monopole constructions reach a 4π radiation (omnidirectional) in the bass region. For an even frequency response in the free field this leads to an increase in level for the bass reproduction. This can be even stronger depending on the location of

8 Reproduction

the loudspeaker in the room. Active loudspeakers therefore often have switches to correct the bass response depending on the location of the speaker.

One way to keep the directivity characteristic constant even including the bass is the use of omnidirectional loudspeakers (at least horizontally). This is what Linkwitz favors.

The complementary method is the use of directed bass radiation by using dipoles or combinations like cardioid loudspeakers. Cardioid loudspeakers omit the dependency on the back wall, which makes it easier to find a suitable location. It has to be noted that all dipole constructions have diminished sensitivity in the low bass and less maximal sound pressure than monopoles.

Maximum sound pressure level with acceptable distortion factor. One information first that does not seem to be widely known: at least in the bass region as rule loudspeakers reach distortion factors that are orders of magnitude higher that what is common for power amplifiers these days.

Görtz depicts frequency curves of the maximal sound pressure level at a distance of 1 m with a limiting distortion factor of 3%, in the bass additionally the level for 10%. They define the upper useful dynamic range. The high limiting distortion values show clearly which compromises loudspeakers enforce in this respect at the current state of the art.

8.1.2 Types of Transducers

Dynamic Loudspeakers

A voice coil is positioned into the gap of a pot magnet and creates the driving force for the connected diaphragm as soon as a current runs through it.

The response of the system for the lower end of its bandwidth is ruled by parameters that were defined by Thiele[29] and Small[26] [27] [28]. They are now normally provided by the vendor.

The Thiele-Small parameters are at least

 $R_{\mathbf{e}}$ the electrical resistance of the voice coil.

 $F_{\mathbf{S}}\,$ the free air resonance frequency of the chassis.

 $Q_{\mathbf{ms}}$ the mechanical Q factor of the resonance without electrical load on the connectors.

 $Q_{\mbox{es}}$ the electrical influence of a short circuited coil to the Q factor.

 $Q_{\mathbf{ts}}$ the combined Q factor for the chassis in the case of a short circuit or a low impedance amplifier output. It can be computed by the other two Q factors with the equation

$$\frac{1}{Q_{\rm ts}} = \frac{1}{Q_{\rm ms}} + \frac{1}{Q_{\rm es}} \tag{8.1}$$

- $V_{\rm as}\,$ the theoretical volume of a closed box that provides the same back driving force as the diaphragm mounting.
- X_{max} the maximal displacement of the diaphragm with a distortion factor less than 10%.

Richard Pierce provides a free spreadsheet [21] and corresponding paper [22], which calculates matching closed back and bass reflex enclosures for given Thiele-Small parameters. It allows to enter a requested volume and gives the resulting characteristics including frequency response and maximum power input limited by X_{max} .

Cone is the form of diaphragm that is used most. To prevent a bending as much as possible, the coil is located at the end of a cone or a modified cone that has a curved cross section (NAWI Nicht AbWIckelbar). The NAWI form has the advantage that it prevents the onset of a bending vibration with half the frequency of the lowest signal frequency (alternating to the inside and outside), that result in the generation of pseudo bass sounds. In the mid of the cone there is a dome or a phase plug. Normally the chassis is open in the rear and gets fitted into a housing to get the optimal performance.

The upper frequency limit is determined by the inductance of the voice coil, phase cancellation by different path lengths of the sound from the diaphragm (minimized by the phase plug if present) and (irregularly) by a break up into partial oscillations of the diaphragm.

Dome is the name for a diaphragm that is formed as partial sphere. It creates an approximate spherical sound wave field with a high angular dispersion. It gets used as direct radiator or as source in a wave guide (see below).

This form is mostly used for tweeters, very seldom in the mid range. The chassis are closed to the back.

- **Compression drivers** are not meant to be used as direct radiators but as sources for horns or wave guides. An inverted cone creates sound waves that are focused to the output. To prevent cancellation by differing path lengths there is a component with bores of equal length, the phase plug. It allows the driver to work up to the highest necessary frequencies. Without the strong acoustical resistance of the horn or the displacement can easily exceed the maximum values, the driver can be destroyed when used this way, even for a short test.[42]
- **Double cone** combines two cones with a elastic mounting between them for a high bandwidth. The back cone is a normal bass speaker cone with the exception of a decoupled flexible inner part that damps its reproduction of high frequencies. In front of that is a second smaller cone that is directly connected to the voice coil for the reproduction of the high frequencies. The exact sizes and forms and the distance between the cones are extremely critical for the overall frequency response. As a rule the frequency bands of both cones overlap. This easily leads to a uneven response in the intersecting frequencies. The construction works as mechanical frequency divider. This principle is used for most simple and cheap broadband loudspeakers.
- **Coaxial loudspeakers** consist of two separate loudspeakers that are combined into one unit. The bass speaker has a cone shaped hole which is fed by the tweeter that is mounted behind it, typically a compression driver. It uses the diaphragm of the woofer as horn.

Both drivers get driven separately by means of a frequency divider. If done correctly this allows for a rather good broadband reproduction, but the highest frequencies are seldom provided. A very good effect of this arrangement is that the bass and treble have a common acoustical center which lessens the angular changes of the directional characteristics because of interferences, and the characteristics can be axially symmetrical.

Magnetostatic

The magnetostat replaces the voice coil by a field of current driven conductors in a magnet field. This results in a driving force that is distributed over the whole area of the diaphragm. The principle is used for midrange and treble speakers.

Air Motion Transformer

In a region of the foil that is formed zig-zag each fold contains a conductor that inverts the current direction with each bend. This construction is located inside a magnetic field which causes under current to open or close all folds in one direction. This presses air out or sucks it in. As consequence there is a homogeneous sound radiation with high air speed because a complete fold gets opened or closed. This principle is used for tweeters.

Bending wave transducers

These loudspeakers do not try to prevent partial oscillations, they use it as principle of sound radiation. At one end of the diaphragm the wave is excited and it runs as mechanical wave towards the other end which is damped. Loudspeakers that work by this principle are able to achieve a broadband radiation with high phase precision. One example is that they are able produce square wave sound realistically and they are therefore some times used as reference source for microphone calibrations. This principle was developed by Manger.

Piezo

Piezo is an effect that translates an electrical field inside a suitable crystal into a change of thickness. If two alternatively poled crystals are mechanically combined, this can cause a bending (the same way as bi-metal for temperatures), that may excite a diaphragm and possibly a subsequential horn.

Piezo transducers provide a high electrical impedance (dominated by a capacitance) and a high efficiency. They are not limited mechanically or by a maximum power input but by a maximum voltage, limited by electrical breakdown. With each breakdown the sensitivity decreases, in the end the device can completely break.

A problem is the typical resonance at the lowest transmitted frequency. They can be used without frequency divider because of the capacitance characteristic. It is possible to shunt them with a resistor (limiting the overall sensitivity) and provide a normal high pass filter which may be tailored to suppress the resonance.

Piezos often work up to the ultrasonic frequencies. They can be used as tweeters, some models work down to $1.8~\mathrm{KHz}.$

Plasma

The use of plasma is the most direct way to excite air movement. A very pointy tip gets strongly charged positively and leads to a corona discharge that drives the generated positive ions away from the tip (electrostatic repulsion). The voltage gets modulated with

the signal and thus the speed of the plasma stream. In principle this can be highly linear, because nothing but the plasma moves.

A huge downside it the creation of ozone which is at about ten times more poisonous that chlorine. In open air the maximum sound pressure level is not really high enough. This effect may be used for tweeters.

Electrostatic

Like the magnetostat the electrostatic drives the whole diaphragm, this case by electrostatic forces. The conductive diaphragm is fed with the upwards transformed alternating voltage of the signal, with an added constant also high bias voltage. Like for any large area transducer the directional characteristic is very frequency dependent, except the case of an area that is bent in two dimensions. A cone transducer in a matching enclosure should add the bass region.

Modulated Fan

The latest variant to provide extreme bass (sound pressure level and lower frequency limit) is a fan with modulated angle of its blades. Contrary to the Cone transducer it is not limited by the extreme displacement and subsequently volume, but the maximum air speed. The maximum air speed is proportional to the sound pressure level, so this level can be reproduced down to arbitrary low frequencies. The transducer is active, the signal steers the angle and the energy comes from the motor of the fan.

A cone speaker can create such low frequencies only in a small tightly closed room, without exciting waves.

8.1.3 Enclosures

Without being mounted somewhere dynamic loudspeakers at least for bass reproduction have a diameter that is extremely lower than the wave length of the lowest frequencies to be reproduced, therefore there is an acoustical short circuits already at relatively high frequencies. The waves that are radiated to the back get diffracted to the front and with ever lower path difference relative to the wave length and the 180° phase difference this leads to a fall off with a slope of 6 dB/octave in sensitivity. For high frequencies this does not happen because the radiation there is very directive, in between there is a region where the sensitivity is even increased above that of one side of the speaker alone. Here the sound of the backside gets diffracted to the front (and vice versa) and the path difference and the corresponding phase shift leads to an increased sound pressure level.

Baffle

By mounting the driver into a baffle the lower frequency limit may be lowered by introducing a detour. The transition is more regular if the loudspeaker is not put into the mid. Each edge may provide a different transition frequency and its location may be chosen to optimize the irregularities of each edge.

Open Cabinet

If an enclosure with completely open backside is used, the detour for the backward sound is still longer. A downside is that in this case the propagation delay and the acoustical impedance mismatch on the back end reflects back a part of the sound pressure (half open pipe). This leads to resonances if odd multiples of four times the depth of the enclosure match the wave length.

Transmission Line

If the opening at the back side is put in a distance that puts the lowest resonance under the resonance of the driver, the low cutoff frequency may be shifted to low values. This length is rather large for low frequencies, therefore the path is often folded to keep the construction compact. This concept is called transmission line. It is important to suppress higher resonances, e.g. by damping the transmission line.

Dipole

The baffle, like the naked driver, works as dipole. In the bass region there is the same sound towards the back side as to the front, but with inverted polarity. If drivers for higher frequency get added as pairs with inverted polarity, this can be extended over the whole hearing range. The resulting construction is then a dipole. This is a possible way to maintain a very consistent directivity over the whole operational frequency range.

Because an ideal dipole does not radiate any sound at an angle of 90° this can be used to radiate only indirect sound to the listener when the construction is turned perpendicular to him. The sound gets perceived to be very distant.

Closed Cabinet

The closed cabinet seals the backward part of the diaphragm to the outside. The loud-speaker works as monopole.

The backside chamber contains compressible air and works to the diaphragm as additional back-propagating spring. To easily compare the characteristics of this effective spring and the mechanical spring of the chassis, the latter one in combination with the diaphragm area is transformed into an air volume that provides the same spring strength, the equivalent volume V_{as} .

It the volume is filled with distributed fiber material, this damps standing waves and internal resonances. Additionally it increases the heat capacity, so that compression does not happen adiabatically as usual, but with the heating of the fiber material the combination moves more to isothermal behavior. This increases the acoustically effective volume, up to a factor equal to the adiabatic exponent of air which is 1.4. Let's call this volume $V_{\rm ak}$ for now. The effective volume $V_{\rm eff}$ can be computed from $V_{\rm as}$ and $V_{\rm ak}$ by

$$V_{\text{eff}} = \frac{V_{\text{as}}V_{\text{ak}}}{V_{\text{as}} + V_{\text{ak}}}$$
(8.2)

Because the resonance frequency is inverse to the square root of the effective volume, the closed enclosure raises it by

$$f_{\text{box}} = f_0 \cdot \sqrt{\frac{V_{\text{as}}}{V_{\text{eff}}}}$$
(8.3)

The Q factor changes proportional to the resonance frequency.

$$Q_{\text{box}} = Q_{\text{ts}} \frac{f_{box}}{f_0} \tag{8.4}$$

A resulting Q factor of $\frac{1}{\sqrt{2}}$ (at about 0.7) delivers the maximal smooth transition without intermediate peak at the lower frequency limit.

It is possible to go for a different Q factor and provide an arbitrary requested Q factor and resonance frequency by means of the Linkwitz transform. If the frequency gets lowered this needs a large increase of level. If the mechanical characteristics of the loudspeaker, especially a large maximum excursion X_{max} , its maximal input power and the power of the amplifier allow this, this it is possible to construct very compact solutions especially for small rooms. Because this solution has to closely match the existing chassis and enclosure this should best be provided by an optimized active solution.

Far below the resonance limit the sensitivity falls of by 12 dB/octave.

Bass reflex Cabinet

A well defined leak in a loudspeaker cabinet is one possible way to widen the frequency response in the bass. The cabinet is no longer a closed volume which puts back pressure to the loudspeaker, now it has an additional own resonance. The air in the cabinet is still a compressible volume that works as spring at the area of the opening. The mass of the air in the opening and partly the air before and behind it that also moves according to pressure changes provides mass. Together this acts as Helmholz resonator. Sometimes an additional diaphragm is used instead.

Simply put the resonance peak shall be placed below the resonance of the closed cabinet and widen the frequency response, if possible with a transition that is as flat as possible.

Below the resonance of the Helmholz resonator the opening is just a mass damped leak. This creates an acoustical short circuit which adds another slope of 12 dB/octave at the cutoff which adds up to 24 dB/octave. Therefore it is impractical to widen the frequency range electronically. Because there is no back pressure from the air in the cabinet below the resonance, the excursion for very low frequencies is increased considerably compared to the closed cabinet. For high power applications it is very important to cut off the frequencies below the resonance electronically. Because of both effective resonances the phase response is more steep than that of the closed cabinet.

As a rule for the same lower cutoff frequency the bass reflex cabinet can have half as much volume as the comparable closed cabinet. For this reason this principle is used frequently.

For a long time bass reflex constructions were dimensioned by trial and (often) error. Lots of examples were created with exaggerated bass, which lead to a distrust of this principle. Some rules of thumb (try to get the two resonance peaks in the electrical impedance plot to the same height) brought more reproducibility.

The papers by Thiele and Small were a breakthrough. They standardized the parameters that describe the characteristics of the driver and computed the results of the construction mathematically. The whole system is a high pass of fourth order.

The ideal bass reflex cabinet can only be built with a driver that has a total Q factor $Q_{\rm ts}$ of at about 0.36. For a given $V_{\rm as}$ and Resonance frequency $F_{\rm s}$ the volume and resonance of the cabinet are also fixed.

The equation also allow to find the best compromise for other values of the Q factor. Lower values lead to a gradual transition at the lower frequency limit with a bass roll off in the region. Higher Q factors provide a peak at the resonance frequencies that extends to a oscillating frequency characteristic above (Tschebycheff filter).

At the frequency of the Helmholz resonator the driver hardly moves at all, the radiation occurs almost exclusively through the port.

The same resonance frequency can be obtained with a short port and a low diameter or the other way round. An additional restriction results from the fact that the small opening can lead to extreme air velocities for high levels (typically it is advisable to keep it below 10% of c, still a rather large value). High velocities lead to non linear damping, which is proportional to the square of the velocity and not directly proportional. This distorts the wave form. Additionally there may be eddies which create broadband noise modulated by the fundamental frequency. Both can be minimized by using large ports with long tunnels and streamlined form at the ends of the tunnel.

8.1.4 Sound Guides

The typical ridiculously low effectively of a loudspeaker (for HiFi loudspeakers sometimes less than 0.25%!) is mostly a result of the extreme mechanical impedance mismatch between the mechanical characteristics of the driver (moved mass, effective area) and those of the very compliant and light air, especially if it distributes the movement to spherical angles of up to 4π (free placement, omnidirectional characteristics).

For this extreme impedance mismatch, with each directly adjacent plane (bottom, wall, edge) this angle gets halved and the sound pressure level in the deep bass doubles by coherent addition, which also doubles the efficiency. To repeat it again: 3 dB more would be the doubled power density by halving the spherical angle of the radiation, the coherence shifts this towards 6 dB, therefore the power density is raised by a factor of four, which results in the doubled efficiency. The same effect is used for public address reproduction by stacking of several phase coherent subwoofers.

It the loudspeaker array is smaller than the quarter wave length, this increases the efficiency. For higher frequencies the directivity is narrowed instead, also resulting in more sound pressure level.

Another way to improve the matching of impedances is the usage of sound guides. A pipe with the area equal to that of the diaphragm creates a much better matching because only one angle is filled with sound instead of the omnidirectional space. The bad news is that now the mismatch happens at the end of the pipe, where now a unmatched part of the wave gets reflected back, resulting in the typical pipe resonances. These resonances get used with added damping for the transmission line speakers.

Horn

The pipe resonance can be avoided it the transition between loudspeaker area and the free air happens gradually. This distributes the reflection in a way that ideally the integral over all partially reflections cancel out by different phases. This construction is called a horn.

The usual theory of horns uses an one dimensional model to compute the effects of a given transition of the cross section relative to the distance from the driver.

This model is only correct if the wall of the horn is always perpendicular to the wave front. This can only be true for a torus or conical horn exactly.

All other horn functions are computed with more or less drastic simplifications. Because the underlying assumption of one dimension is not valid, the wave front does not propagate as imagined and splits into several modes. The result is a frequency dependent irregular directional characteristic of the horn. On the other hand you can gain a lot of efficiency (up to 90%) and sound pressure level.

The most extreme simplification assumes a plane wave and sets the cross section at a given main coordinate equal to the effective area of the wave front at this distance from the source. This does not take into account that the waves far from the mid axis have to follow a curved path so that the cross section is to far away from the driver to be correct. The simplification is less severe when the wave front is distorted and mixed anyway because of a curved horn, reflectors or simply folds. This holds rather well for bass horns that often get folded to keep the dimensions at manageable sizes.

A more detailed approximation that at least partly addresses the curvature of the wave front is the spherical wave horn. For its computation the wave form is assumed to be spherical, with a radius equal to the wave length of the lowest transmitted frequency. One positive aspect is that the resulting horns are shorter than those computed with the assumption of plane wave fronts.

The reflection of the sound is minimized when the length of the horn is as long as the wavelength and if this is also true for the circumference at the mouth. Shorter horns approach a transmission line more and more, with a horn function setting in for higher frequencies.

Real horns for low bass are extremely large, even when folded. Smaller than the famous Schmacks horns in the corner of a room (the effective output area if eight times that of the construction because of three reflections) they cannot be effective, unless stacked in nontrivial large numbers to achieve the necessary area together.

In the bass region the most prominent versions of horns are the

- **Front Loaded Horn.** The Horn is mounted at the front side of the chassis and the whole (limited) frequency range is radiated by the horn. The back side is typically put to a relatively small closed cabinet.
- **Back Loaded Horn**. The horn construction is mounted to the rear of the chassis, often folded multiple times to keep the setup compact. A pressure chamber in combination with the horn throat works as acoustical low pass filter to suppress the higher frequencies that are problematic for the folded horn. In the frequency area where the horn is effective the diaphragm barely moves and the radiation happens almost exclusively through the horn via the back side. Above the transition frequency the loudspeaker (often a broad band chassis) radiates directly. To keep the frequency response as consistent as possible, the chassis should have a weak bass without the horn (this requires a relatively low value for $Q_{\rm ts}$). Otherwise the sound pressure level is higher below the transition frequency because of the higher efficiency of the horn compared to direct radiation. An extremely low value for the Q factor is also problematic, because it often correlates with a very low X_{max} , which limits the maximum sound pressure in the bass, also for bass horns. Because the transition between the frequency ranges is rather slow (6 dB/octave), there is some overlap between the front and back radiation. The sound that comes from the rear is inverted in polarity and delayed by the longer path length. This leads easily to cancellations in the overlapping frequency range, at least the sound pressure level is almost never smooth there. Most problematic are the shortened horns that also add the transmission line pipe resonances.

The typical compromises for bass horns are unnecessary for mid and treble frequencies. There is no need for folding or backloaded horns. These kinds of horns are often used in PA to produce as much sound pressure level as possible with each driver without getting the interference problems that are typical for multiple sources in the mid and treble frequencies (lobing). For compact two way PA loudspeakers the treble horns have to be reduced in the input level to get an even frequency response. The added advantage is an increased input power ability of the whole loudspeaker.

Typical horn cross section functions dependent on their axis are

The Exponential Horn changes the cross section exponentially over the sound patch. In the one dimensional approximation it changes as

$$A(x) = A_0 \cdot e^{kx} \tag{8.5}$$

with

$$k = 4\pi \frac{f_0}{c} = \frac{4\pi}{\lambda_0} \tag{8.6}$$

The frequency response over the lowest frequency (determined by λ_0) is linear until the directivity at the horn throat begins. An ideal infinite horn cuts of every frequency below.

- The Tractrix Horn uses the tractrix as contour. This is the curve that results when a damped mass is dragged by movable sticks along a line that does not cross its position.
- **The Hyperbolic Horn** has a hyperbolic contour. The throat opens slower that the exponential curve, which rises the efficiency for the lowest frequencies.
- **The Conical Horn** actually maintains the perpendicular angle between wave front and horn surface. It can be described in one dimension. It is the base for wave guides and constant directivity horns.

Wave Guide

Horn constructions that do not break up the sound propagation into several modes because their walls remain perpendicular to the wave front are called wave guides.

They have a conical form. If the wave front is formed spherically and the propagates at the right angle through the horn, the requisite is met.

The theory of horns shows however that the horn function (high efficiency etc.) begins very smoothly and for very high frequencies compared to a similarly sized exponential horn. The mouth causes back reflections.

Both can be optimized. The throat of the horn can be formed as exponential horn until the deviation from the plane or spherical wave gets so bad that modes begin to form. This increases the efficiency because the impedance match at the throat gets better for lower frequencies. The reflection at the mouth of the horn can be minimized by rounding the transition to the front plane of the loud speaker. The remaining smooth change of efficiency at the onset of the optimal impedance match can be filtered out rather easily.

Wave guides define very clearly the angle up to which all transmitted frequencies can be radiated with constant sound pressure level. Most excellent PA and studio monitor loudspeakers therefore use this concept.

An introduction into the construction of wave guides can be found at[4], http://sound. whsites.net/articles/waveguides1.htm.

8.1.5 Passive Loudspeakers and Crossover Networks

If several chassis shall be used to reproduce different frequency regions, the signal frequencies should be split and delivered to the matching driver.

The most primitive variant cuts off the unsuitable bass frequencies with simple capacitors before the mid drivers and/or tweeters. This prevents their overheating and unbearable mechanical stress. The matching low pass filter has to be provided by the chassis for the lower frequencies by their acoustical frequency characteristics.

An improvement is to really split the frequencies with coils and capacitors and efficiency matching by voltage dividers consisting of resistor networks – if needed.

The crossover networks can be of different order.

1. Order uses only a capacitor or coil for high and low pass filtering. The cutoff happens with 6 dB/octave. The chassis are required to remain linear far into the region of the adjacent driver and be able to handle the associated power because of the large overlap.

2. Order uses an Coil in series and a capacitor in parallel as low pass filter, reverse for high pass filtering. The cutoff happens with 12 dB/octave. Aside from the cutoff frequency it is possible to choose between a smooth transition with Bessel filters, the maximum linear transition with Butterworth filters and the steep cutoff with Tschebycheff filters that overshoot in the pass band.

In principle the order can be arbitrarily high, but higher orders that three are not used often for passive designs because the effort is considerable.

The typical computations rely on real valued impedances at their output. Therefore the chassis should have a parallel circuit which equalizes their impedance characteristics. This can be done by a resistor / capacitor combination against the impedance rise through the inductance of the voice coil and by a RLC network against the impedance peak caused by the fundamental resonance of the chassis.

Large inductors with low inner resistance are big and expensive. They are necessary for low splitting frequencies and high power levels.

Aside from the usage of possible expensive components this technology has the disadvantage that it relies on the parameters of the chassis remaining constant. For high power the voice coil gets heated up and increases its resistance, changing most other parameters. One result is a decrease in sensitivity, which is often called power compression. The frequency limit is changed as well as the Q factor of the chassis, which increases. If the chassis is behind a passive high pass filter this lowers its cutoff frequency, so that the chassis gets even more power and more mechanically demanding frequencies. This can have a massive impact on the stability of the device.

A detailed look at the whole situation can again be found at [4], http://sound. whsites.net/lr-passive.htm

8.1.6 Active Loudspeakers and Crossovers

Active crossovers split the signal with analog circuits before the power amplifiers for each chassis. This makes the crossover stable against changes of the parameters of the chassis. It is simple to create these circuits with operational amplifiers and gyrator circuits using only R and C components for arbitrary frequency characteristics, including corrections for the frequency characteristics of the drivers. There is no more problem with the inherent resistance of the voice coil of the drivers. It is now only dependent on the chassis itself and the cabling between the power amplifier and the chassis. The crossover network may

also include limiters to prevent an overloading of the chassis. If the crossover and power amplifiers are integrated to the loudspeaker, the setup is called active loudspeaker.

8.1.7 DSP Based Loudspeakers

If you use a DSP before the power amplifiers you may reach the limits of what is technically feasible. It is possible to create crossover networks with arbitrarily steep splitting, errors in the frequency characteristics can be countered and the power limitation does not have to be made dependent the level. It is possible to make the limiter react to a detailed model of the mechanical and thermodynamic situation in the driver and react precisely to its values. If the cabinet and drivers deliver the desired directivity and can deliver the desired sound pressure level, most remaining deviations from the ideal can be corrected by computation. It is necessary that individual deviations of the driver parameters or those over time are not so strong that they undermine the correction attempts.

8.2 Room

Except for the case of open air the surrounding room has a massive impact on what gets heard.

If the loudspeakers are placed at a problematic position, comb filtering may completely cancel several frequencies. The mode structure of the room may completely alter the frequency response most of all in the bass. The ringing of modes and the reverberation prolong the duration of parts of the signal. Repeating echoes disturb the recognition of the time structure of the signal, they are very noticeable.

However, this does not mean, that open air or a completely anechoic room is optimal for listening. For example, the immersion into room sound (also the part which is present in the recording) is completely missing for the back in the case of two channel stereo reproduction without room.

Lets first have a look at the potentially negative influences of a room.

The modes should not be too few (that as consequence stand out) and not too strong. Otherwise some bass notes are too load and others too soft. Long ringing resonances of some modes destroy the time signature of bass pulses by prolonging them. This disturbs also because masking effects transfer this disturbance to other frequencies that are not affected themselves. The modes can be diminished best by large volume bass traps. These are porous damping substances like open sponges, mineral, glass or other fiber material. For bass traps they should be packed into some interface that prevents a too large damping of higher frequencies. They are often and easily damped in rooms by small or thin absorbing items (carpet, armchairs, etc.).

If this is not feasible or not possible in the required scale the peaks can be countered by inverse parametric equalization, which diminishes the large sound pressure and the occlusion of higher frequencies by large amplitudes in the bass. Floyd Toole[31] [32] demonstrated that, other than often said, this also improves the response in the time domain. The equalizer cannot however change anything about the extreme dependency of the frequency response with respect to the hearing position. This requires a damping or the usage of several subwoofers on strategic places (Toole) or on arbitrary and diverse places and globally optimized individual parametric equalization (Earl Geddes).

For a consistent reproduction of low frequencies in a room you should either use at least three independently and optimally filtered subwoofers as proposed by Geddes[6],

two subwoofers places centrally at the front and back or four of them in the room corners. The last two alternatives were found by a large series of brute fore modeling computations by Welti[33]. The original presentation seems to have vanished from the web.

The reverberation should not be too long and not be change too much over the frequency range. The sound of the reverberation itself should be similar to the original signal with the exception of a steady roll off in the treble region. This helps the ear to find the correlation. The overuse of thin absorbers is only effective in the treble region and can lead to a very muffled reverberation.

Strong early reflections should be avoided to prevent comb filtering effects by them. Other than subwoofers all loudspeakers should therefore be located at least several tens of centimeters away from the back or side walls. In studios, loudspeakers are sometimes integrated seamlessly into the wall, which completely prevents comb filtering from that wall.

For two channel stereophony the area around the loudspeakers and the areas causing the first reflection (as a rule the side walls) should not reflect too strong. The backwards region however should reflect to achieve a immersion into sound, because the recording can only be heard from the forward directions.

Detailed recommendations can be found at [2].

Surround reproduction asks for low reflectivity also for the back, because sound that shall be heard from behind is part of the recording. In real live the same setup is often also used for the reproduction of two channel stereo recordings, therefore a compromise should be reached.

Detailed recommendations can be found at [3].

If there are still standing waves and slap back echoes although the room is damped enough, you may and should also work with diffraction and diffusors.

8.3 Headphones

Headphones completely eliminate the influence of the room. Closed headphones isolate the external sound, open constrictions are acoustically transparent – in both directions.

To judge a sound in a loud surrounding and to prevent unwanted leakage of the headphone signal into microphones in the case of overdub recordings closed headphones should be used.

Headphones are optimal for the reproduction of head related recordings, especially for dummy head recordings.

Every other recording is not reproduced optimally, because the natural cross talk between both ears is missing. The sound is located at a line between the ears and not as coming from some angle and distance from the front. Without massive training this makes it impossible to judge the location of the perceived sources, and even with training is remains hard and lacks precision. The missing mechanical impression for high bass levels prevents to realize that the hearing is done at rather high level ¹ and whether the bass level is correct. Most of the time the headphone is used in the studio as a kind of "acoustical looking glass" to hear details. Otherwise it is better that no control at all and the natural cross talk should be modeled with a matching frequency dependent cross feed.

¹You easily and unknowingly choose a much too high level. This makes prolonged hearing with headphones dangerous.

8 Reproduction

Another big problem of headphones at the time being is that it has not be clearly defined which frequency response should be targeted. Neither the reproduction of the impression of a free field not of a reverberation field is preferred by test listeners. Right now Sean Olive does ground breaking work in this realm. The reproduction of the impression of a stereo hearing in an optimal reproduction room seems to be optimal, but the exact parameters have yet to be finalized.[14] [15] [16] [17]

The papers by Sean Olive et. al. also prove that the exact position on the ears has a large impact at the frequency response. The position over the hearing canal is critical for correct treble reproduction and leaks affect the bass level for closed headphones.

9 Practical Recording

9.1 Requirements

9.1.1 Artists, Music/Text, Instruments

The most important aspect to achieve a good result are the quality of the performers, the performers and their mental condition, and the material to be performed (lyrics/text, composition, arrangement).

During the recording and (most of the time later) and mixdown it is important to develop a concept that can be approached step for step. It starts with a useful sound that is depending on useful microphone positions and the acoustical situation. The concept is important in order to systematically approach the correct balance and not to try blindly. Technical knowledge helps to find the way, artistic understanding defines the goal.

It is hard to match both worlds but absolutely necessary for optimal results.

9.1.2 Together, One by One?

This is an important question, if several tracks shall be combined to achieve the result.

From a technical viewpoint, the best isolation between tracks can be achieved if they are recorded one by one. All parameters can be optimized individually for each track. Each track can be re-recorded as often as necessary.

On the other hand it must not be forgotten that the situation of the performers is completely differing from that of live performance. The performance may shift into a technical region towards perfection and in the worst case away from the artistic expression or the inherent meaning. Also missing is the spontaneous interaction of performers that listen and react to each other. This is replaces by advanced planning, which may lead in extreme cases to static results that might get boring.

Especially big ensembles should be recorded simultaneously. This does not rule out recording as many tracks as needed. This is closer to the situation that live performers are used to.

There is some middle ground. It is possible to record the group at once and later repair errors or/and add overdubs later (background vocals, strings,...).

9.1.3 Psychology of the Recording

Like always when dealing with real persons the psychological aspect must not be forgotten. It has an essential influence on the quality of the performance.

The surrounding during the recording should be as comfortable and free from distractions as possible. Frequent interruptions because of technical problems are a disaster.

There should be no unnecessary pressure. If the limits are reached – psychologically or possibly physically (voice, cramps), the next step should be delayed until the performer is back in shape.

If a recording is to be repeated, the message to the performer must never be perceived as degradation. Otherwise you will not get better performances later.

9.1.4 Perfection - Genie - "Magic"

In the studio you have everything needed to work on a recording until everything fits together and you get very close to perfection. You may add lots of sessions which allow to pick the best part to create the optimal track, edit small time inconsistencies or even tonal problems.

You must keep in mind, however that you are dealing with art and the you must never only work towards the technical and precision aspect of the performance alone. It happens again and again that over-productions removes every trace of spontaneity an feeling from a recording.

If you have decide between a genial magical moment and perfection, choose wisely. Some remarkable recordings contain errors, e.g. in an improvised solo, because it is part of the best take or they even turned an error into some interesting idea.

On the other hand there is no excuse to leave trash in a studio recording.

9.1.5 Live - Studio

For live acts there are lots of constraints that stem from the nature of real time, there is exactly one try for everything. The often random acoustic and the wish for a good presentation put additional strains on the technical possibilities for a recording.

You have to master the acoustical situation as it is, the "contributions" of the audience should be in a reasonable level and the PA and monitor system cause additional noise. Often the high level of noise prevents using an optimal microphone placement and they have to be put extremely close to the sources. The choice of microphones is limited, because specialized vocal microphones are used for the sound reinforcement, and it may be necessary to use dynamic microphones because they are less sensitive to weather conditions (open air) and mechanically more robust.

In contrast a studio is made to prevent extra noise and to provide everything that is necessary for an optimal recording. The acoustical situation is optimized. Sometimes there is not enough reverberation to allow a large group of musicians to hear each other the way they are used to. In those cases it may be necessary to provide a rough mix including reverberation to their headphones or to use monitor loudspeakers to add artificial reverberation to the room. This causes a trade-off between technical optimal situation and better surrounding for the artists. It might be worth it.

Multitrack recording of live events are sometimes enhanced in the studio by adding or replacing tracks.

9.2 Microphone Technique

9.2.1 Close Miking

For classical music the room is typically added directly in the microphone signals to get a recording that matches the expectation of the hearer. This adds depth, groups the sounds together and delivers the feeling of a large room.

For almost all other cases the microphones are placed close to each sound source. If the room sound is to be included, this is done by additional microphones and tracks. This allows to decide during the mix how high its level should be. It can even by modified.

Only close miking allows to get the maximum clarity, because there is a high resolution in the time domain and the full treble content is on the track. If several channels are recorded at once, this technique allows the best possible isolation between tracks. Sometimes this is used in combination with increased distance or acoustical sound obstacles between sources, absorbing or transparent.

Directional microphones can be used to cancel other sources. For example a guitarist may be singing and play his guitar, and the guitar amplifier can be put at a canceled angle of the vocal microphone. This allows the artist to work as he is used to. Directional microphones also minimize the room influence.

You always must remember that each microphone that is not omnidirectional shows the proximity effect and is not able to automatically record the bass part as it is. This can be circumvented if an omnidirectional microphone may be used, possibly located more closely to the source to get the same signal to noise/reverberation ratio.

The directional and geometrical characteristic of the source must also be taken into account. Otherwise it may happen that characteristic frequencies are missing, too weak or strong. Some aspects of the sound may be muffled or exaggerated. This is obvious for all string instruments. The treble part is mostly radiated from the strings themselves, everything else comes from the resonance body - and this part is absolutely essential. Typical fret, picking or handling noise must be incorporated correctly. They must not be exaggerated and a good musician will minimize them. If the are missing it may be hard to recognize the instrument and they are often essential to clearly depict rhythm details.

Either there is a position that provides the desired sound or the parts of the sound are recorded at different positions and combined in the mix – sometimes even at different locations in the stereo image. The phase between the different channels must match, otherwise some frequencies cancel in the mix, at leas if played mono.

9.2.2 Multiple Microphones

If several sources shall be recorded simultaneously or the room shall be included or a wide source shall be put into a stereo image, multiple microphones are used synchronously. For digital recordings it is absolutely necessary that all A/D converters used run with the same clock, otherwise the phases of the tracks get inconsistent. Professional converters allow to be fed with an external clock, otherwise a converter with enough channels has to be used.

If several sources shall be distributed in the stereo image, one of the well known stereo combinations can be used. Either the stereo position is fixed at the time of the recording (large AB, small AB, ORTF, EBS,...), or a coincident combination (XY, MS) is used that can be later placed as needed at positions in the stereo image. For all combinations that include time of arrival for localization at least partially, this would lead to the summation in one channel and result in comb filtering.

For a good room acoustic that matches what is needed in the recording, it may be a good idea to put some room sound microphones far away from the source – large AB might be a good combination. The downside is that the mix inside the reverberation signal cannot be altered later.

If the maximal flexibility is maintained by using multiple microphones with close miking

and low cross talk, it is consequent to generate the matching room sound electronically an with the desired optimized parameters.

9.2.3 Some Concrete Sound Sources

Below are some musings about the recording of some important sound sources with examples of configurations. Those are not meant as definitive recipes but as suggestions and starting points for the own search for the best suited method in a given situations. Never forget: the context is an important integral part of the setup.

- **Voice** The voice is by far the most important sound source of all. Our hearing has been prepared by evolution and daily training to extract as much rational and emotional information from its sound. As soon as one voice (or more) are in the equation, it is almost every time the most important component. This requires a careful handling during recording and mixing. The following microphone techniques can be useful:
 - **Directly before the mouth** . A word of warning: in this region the full impact of plosive sounds and the breath is present, therefore wind shields are a necessity. This is extremely important for all directional capsules. For loud surroundings and live reproduction this is normally the optimal position for a microphone because of the high sound pressure level of the signal. It is useful to use microphones that were specifically optimized for this case, so called (live) vocal microphones. Other microphones should be used with an extra external windshield, because none is integrated. Directive capsules should have a rather high lower transition frequency, because the proximity effect raises the level of the low frequencies in this application. Except for omnidirectional microphones you always have to be prepared to lower the bass level with an equalizer. For the remaining popp sounds and noise by handling and mechanical vibration a high pass filter of at least 12 dB/octave is suitable.

It is important to record mouth and nose. Without the nose signal it sounds nasal. If the microphone is used from the side the proximity effect vanishes, but this is very dependent on the exact angle, for example if the artist holds the microphone in his/her hand.

The handheld usage increases the danger of feedback in live situation, because it can not be ruled out that monitor or other loudspeakers get into the angle of high sensitivity or close to the microphone by chance.

A trained singer increases the distance to the microphone for loud parts to control the overall volume. This also changes the sound for directive microphones, because it weakens the proximity effect.

A microphone stand keeps the position and angle of the microphone consistent to the stage (as long as nobody moves it), a head mount keeps it consistent relative to the head. This keeps the sound consistent but the artist can not simply walk away to prevent e.g. messages to the sound engineer from being picked up.

At the side of the mouth keeps the microphone away from the direct region of the breath. If a omnidirectional microphone of only a few mm size and mounted at the neck is used, you get a very usable signal without proximity effect, even suitable for live sound reinforcement.

- At the lapel (Lavalier microphone) omits problems by breath and the microphone can easily be installed and it is not too prominent optically. On the other hand the sound pressure level is rather low, which increases the risk of feedback in live situations. Additionally the sound at this location is strongly colored. The treble region is extremely weak because the mouth is directional in this region and the chin in the sound path. The chest with its acoustic-mechanical resonances increases the level in the region around 600-700 Hz massively, which has to be countered by a channel of a parametric equalizer or by the construction of a specialized microphone. Otherwise the sound is extremely boomy and this does not only sound displeasing but it also decreases the speech intelligibility. For sound reinforcement it also puts a rather large need on the power for reprodution which is far higher than what a clear signal requires.
- At the crest the sound is remarkably similar to what can be heard in the front at some distance. It can be used without large equalizer changes. For live applications of theater or musical performances sometimes a small microphone is integrated into the front of the hairdo without attracting attention.
- At some distance before the mouth speech and vocals sound the way that we expect from daily routine. For distances exceeding 50 cm you should always use directional microphones to achieve enough discrimination against the surrounding sound. Smaller distances with directional microphones always require the use of pop filters with foam elements or in the studio in the form of frames that hold a fabric or dense metal grids. These frames should be spaced at least 10 cm away from the microphone capsule to allow residual eddies and pressure peaks by the breath behind the screen to be eliminated by the distance, so that their effect is not increased by the proximity effect. The frame should not be directed normal to the line connecting the microphone and the mouth, they should be angled. This prevents the reflection by the capsule to be redirected to it, causing a weak but avoidable comb filter effect. For distances lower that 20 cm it is possible to use omnidirectional microphones in the studio and avoid all problems inherent to directional microphones. Even with these precautions it is advisable to prevent using a position that is directly in the line of the breath. It should be put below or above the mouth but directed to it.

In classical recordings in halls the typical large reverberation radius comes to help and makes it possible to work with distances that are larger than typical for rock or pop music, and you should take advantage of this. For live recordings the microphone may be placed rather low to avoid the line of view.

Boundary layer microphone at the speaker's stand or desk. Contrary to the case of a microphone on a stand in some distance to the surface of a desk or speaker's stand the boundary microphone avoids comb filtering by reflections on the relevant surface in which it is integrated. This is often used for video conference setups.

A downside is that this position is suited even better to transmit every rustling of paper or handling noises of coffee cups or similar. The worst case is the placement of folders or other large items directly over the microphones without noticing it. This does not improve the sound of the voice and can be mistaken with a technical problem in the connection.

- **Strongly directional microphones from the ceiling** can prevent the mentioned problems of boundary layer microphones. The large distance is countered by its directivity (e.g. super cardioid). If the angle is chosen to bring the reflection by the ceiling (which should be weakened by acoustical measures anyway) into the cancellation angle of the microphone, broadband alteration of the sound by comb filtering are avoided. The comb filtering by the reflection of the desk remains, however. At least the reflected sound has a longer path and is therefore weaker than the original signal and the directivity of the mouth weakens the treble part in the important high mid and treble region.
- Half super cardioid boundary layer microphones at the bottom. Again, this position avoids the negative effects by the reflections from the bottom. If the microphone is directed towards the sound source, the strong directivity counters the level decrease by larger distances over a wide range because the source gets more and more into the angle of maximal sensitivity. This is a useful setup to record theater performances.[36]
- **Microphone arrays** in one dimension have the normal wide directivity in one direction and perpendicular to it they only pick up a very narrow angle. This excludes a lot of surrounding sound.

Strong vertical directivity picks up a horizontal region of e.g. a speaker's stand. The speaker may move to the side without getting out of the active region.

On a conference table the strong directivity may be used horizontally. The neighbor is suppressed, the speaker may sit of stand and remain in the active region.

Upright piano, grand piano belong to the very few instruments that have the size and spatial distribution of sound generation that can render a stereophonic recording useful (not strictly necessary). Therefore all the common microphone assemblies for stereophonic recording may be utilized with the exception of large AB, because the instrument is no big enough to make this useful.

The full frequency range of the piano sound can only be caught by omnidirectional microphones.

For coincident assemblies (MS with omnidirectional mid channel, in an ensemble possibly XY, if the extremely low frequencies shall be reserved for other instruments) the width and position of the stereophonic image can be changed after the recording. This is often important if the instrument is not used solo.

All assemblies that include time of flight localization require the assembly to be placed and directed in a way to deliver the stereo image the way it should be on the final mix, because a later change of the stereo image leads to the addition of two very similar and highly correlated signals with different delays on the same channel with all the resulting comb filtering.

The characteristics of the sound are extremely dependent on the location of the microphones, because the parts of the total sound are weighted differently on different places.

A very voluminous sound that is mainly suitable for accompanying in the background can be achieved by small AB at the end opposite to the player. The noise by the hammers and the mechanics are almost completely missing, but the strings and the resonance by the sound board are picket up very strong. The sound is perceived as being very indirect and usable mainly for some classical music styles.

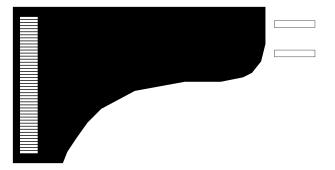


Figure 9.1: Piano recording: end position

A more direct sound is achieved by a small AB assembly in the mid of the curve of the open side of the grand piano. Here all components of the sound (strings, sound board, mechanics, hammers) are present in the recording. This improves the rhythmical characteristics. For a solo piano mainly in the classical domain this can be a good starting point to search for the optimal setup.

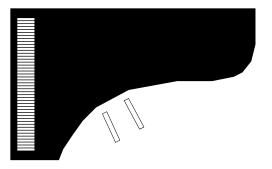


Figure 9.2: Piano recording: side position

For music that depends on the rhythmical aspect, especially rock jazz or pop, the noise part of the mechanics and the hammers should be emphasized, because they carry the main part of the time signature of the rhythm, combined with the treble part of the beginning of a tone. If the short strings of high notes get also recorded with sufficient high level the resulting sound can cut through dense and loud mixes. Usually this is achieved by small AB and MS/XY assemblies above the hammers. For live applications with loud stages the relevant distances are often too high to suppress the ambient sound.

In this case you may use several microphones that are very close to the strings. In extreme cases it may be useful to use specialized pickups or even electronic stage pianos. Often the sound is manipulated extremely, mostly by largely boosting the mid and treble frequencies.

Brass and saxophone are very directional for mid and treble frequencies. If the whole spectrum shall be recorded the microphone has to be places in the main direction of sound projection. In live circumstances it is a good idea to use compact microphones that are mounted at a fixed position directly at the output opening.

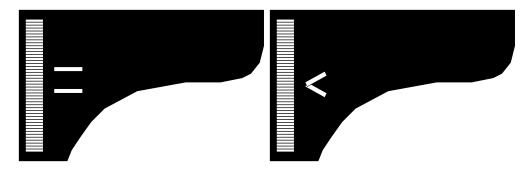


Figure 9.3: Piano recording: hammer position, AB and XY setup

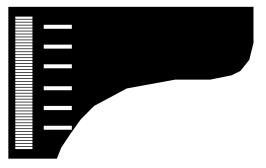


Figure 9.4: Piano recording: hammer position, microphone array

Horns are a special case. Their opening is traditionally pointed away from the listener. Therefore they sound very different from e.g. a trumpet. Behind the instruments there has to be a reflective wall, so the onset of the sound is recorded indirectly with an imminent delay.

Aside from condenser microphones, typical live vocal microphone for close miking can deliver very good results. The acoustical requirements are similar to these for close miking of vocals.

- **Other woodwind** should be recorded from up front or slightly from above, because this gives the optimal mix of the pure tone and the auxiliary handling noses. Those should be minimized. If they are completely absent, the characteristic of the instrument is often missing.
- Acoustical guitar. It is necessary to find the correct balance between the picking noise, the sound of the strings and the guitar body and handling noise. They depend heavily on the style and playing technique. Close to the resonance opening the bass is very strong because of its Helmholtz resonance, therefore a microphone should not be put too close to it.

For monophonic reproduction a starting point is to place one microphone in the mid of the full string length at some distance. It is possible to reproduce the instrument stereophonic. The full width of the stereophonic base should be avoided, however. It does not match the expectations about the size of this instrument.

If a good instrument is played by an able artist, the sound is very well balanced. Therefore it is important to use a microphone that is as neutral as possible. Either pick a small diaphragm condenser microphone or a dynamic microphone that has the typical mid peak compensated, either fixed (Beyerdynamic M201) or switchable (Sennheier MD441, Shure SM7B).

Built in piezo pickups are a good thing for live reproduction. As a rule, they are not able to deliver a balanced sound for high quality recording. A lot of tweaking by EQ is typically necessary.

- **Electric Guitar.** The classical amplifier for electric guitar is not only a pure amplifier, it is an inherent part of the instrument, because it forges the resulting sound. The relevant influences are:
 - the preamplifier and power amplifier with their typically nonlinear and often completely saturated valve stages and output transformer
 - the mostly simple three channel equalizer classically built with a simple potentiometer/resistor/condenser circuit that only allows to emphasize the bass and treble region¹
 - a high pass in the preamplifier that cuts of the basic frequencies massively
 - the loudspeaker that deliberately omits any tweeter to suppress the sharp parts of the distorted sound but raises the level of the upper midrange, especially when driven by a tube power amplifier which has a high output impedance, which is even stronger when in overdrive mode the negative feedback can't diminish it.

As long as the saturation of the output transformer and power stage is not prominent in the sound, it is typically possible to cut back the volume before the last stage without loosing the saturation (Master Volume).

The sound pressure level close to the loudspeaker is often extremely high. Because there are not treble frequencies to be reproduced, it is no problem to work with dynamic microphones. For close miking the position of the microphone is extremely critical for the sound characteristic. In the mid directly before the voice coil the level of the high mid frequencies is maximal. Because the loudspeaker chassis types that are often used for this task often have a very distinct compliance of the diaphragm, the mid frequency levels taper off towards the edge. Often several microphones are used, sometimes to catch an additional room sound component.

There are more and more emulations of the whole combination of amplifier and loudspeaker available and they often even include the typical effect pedals. This separates the sound creation and amplification. The good part about this is the independence of the exact position of a microphones and the isolation from other sounds, both really important in live situations.

If it is not clear whether the ideal setup of the amplification has been found for a recording, it is possible to record an additional track the raw signal of the instrument. This has to be done with an active and high impedance DI box. Later this signal can be fed to arbitrary amplifiers with arbitrary settings until the sound is convincing (re-amping). Other than often assumed, the output fed into the amplifier does not have to be artificially of high impedance like the passive electric guitar itself is.

 $^{^1 \}mathrm{See}$ [4], Article about equalizers, paragraph 8: Guitar Amp Tone Stacks

9 Practical Recording

- **Electric bass guitar.** The combination of amplifier and loudspeaker can also be part of the sound creation for a bass guitar. If picked up by microphone it is necessary to have a suitably low bass frequency limit, because the instrument is able to produce tones down to 40 Hz. If in doubt, there should always be a track with the raw signal to have the option of mixing it with the microphone signal. This signal is sure to contain even the most extreme frequencies in case they are needed. Re-amping is also an option here.
- **Electronic keyboards** are typically recorded directly with DI-boxes. For some of them the amplifiers and loudspeaker are also part of the sound forging.

Classical electronic pianos are often played over suitcase amplifiers for electric guitars. The recipes for electric guitar apply here, too.

Classical electronic or electro-mechanical organs and their emulations sometimes use rotating loudspeaker setups as mechanical effect units to get their typical sound, described before². In this case there rotors should be picked up by microphones, typically separate for bass and treble rotor. For the bass pay attention to the lower frequency limit of this instrument that is located at 30 Hz. The treble is limited to $\tilde{6}$ KHz. The microphones can be placed at different distances for different sounds. For close microphones the strong relative distance changes add a prominent volume modulation, for larger distances and a rear wall their reflection of a sound with inverted Doppler effect there appears a prominent chorus effect for slow speed of the rotor. The rotation can be brought into the recording by using a suitable stereo setup.

9.2.4 The Large Distributed Sound Source - Orchestra, Choir, Organ

For large distributed sound source first of all one of the microphone assemblies described above 3 is used to get the over all sound consisting of the room component and the sum and distribution of all single sound sources.

Orchestra

At least for orchestras this is complemented by auxiliary microphones that pick up single sound sources of groups of sources (in this case at some distance to let them fuse to a coherent section).

The signal of these microphones can be used to raise the volume of parts that are not clearly reproduced. Additionally the inherent level stereophony of these signals improves the precision of the localization and they sound more direct (with less reverberation) than in the over all sound. This effect is even stronger when the time delay of the over all microphone signal is not compensated by delaying the single signal (Haas effect).

Choir

A choir should be recorded from relatively far above so that the back rows are not too far relatively to the front row and their sound is not shadowed by the heads of the front row,

 $^{^2 \}mathrm{see}$ subsection 7.7.6, Page 102

³level (section 5.2.2, page 67), time difference (section 5.2.3, page 71) and combination stereophony (section 5.2.4, page 72)

reducing treble. For homophonic compositions, the voices should be placed one behind the other to fuse the overall sound. For polyphonic compositions the voices should be separated on the stereo base to ease the distinction of the distinct voices.

Organ

An pipe organ is best recorded including its room with a large AB assembly. The stereo image does not have to separate the single registers. Organ builders even try to keep the sound symmetrical by interleaving the position of large pipes on each side. A no too sharp stereo image that mainly shows the size of the instrument and the room best matches the character of the instrument. For the room sound it should be tried to show the different depths of back positive (if present), main part and echo part of the instrument.

9.2.5 The Drum Set

The drum set – provided it should sound more or less natural – is also recorded by a main microphone assembly and auxiliary microphones. The main microphones are located over the head of the artist (hence the name overhead microphones). The approximately record the sound that the artist experiences while playing. For AB assemblies it is important that the distance of both microphones to the snare drum is equal, because otherwise its frequency response is poisoned by comb filtering in case of monophonic reproduction.

The overhead signal should already contain everything important, except the bass drum that comes to weak, especially when using an XY (e.g. to narrow down the stereo base of the set during mixing) with the weak bass of directional microphones. The frequency response must contain the whole treble region because the cymbals deliver such frequencies. If there are auxiliary microphones for everything but the cymbals, it is not necessary to have the full bass spectrum in this signal.

If the roof is not very high, it is important to have very little signal of their reflections, either by damping them or by using directional microphones angled to suppress the angle of these reflections.

Alternatively the extreme treble can be damped by thin absorbers and the microphones can be so closely put to the ceiling that the remaining frequencies have little phase shift when they reach back to the microphones.

The auxiliary microphones are in descending order of importance:

- **Bass Drum(s).** The frequency of the fundamental tone is typically in the region between 50 to 60 Hz. The sound pressure level close to the instrument is extreme. Either use omnidirectional microphones with high maximum SPL or a directional microphones with their proximity effect. Even then it is necessary to look for a sufficient low limit of the transferred frequencies. The microphone can be located behind the drum head opposed to the player or, if it is open, in the drum body itself to pick up the noise of the beater hitting the drum head with the added rhythmical clarity that can provide.
- **Snare.** The microphone should be directional and be placed at the rim and pointed towards the mid of the drum head. If the snare strings shall be picked up strongly, an additional microphone may be placed at the lower drum head. To prevent cancellation of low frequencies, it has to be inverted in polarity.

- **Hi-Hat.** For some styles its rhythm is very prominent (swing), in other cases it marks the base rhythm and, in combination with the cymbals it can mark and highlight specific moments in a composition. It should be recorded from an angle above in a way to shadow the snare drum acoustically. The exception is a setup with one microphone shared by the snare drum and the hi-hat. The region at the side is to be avoided because of the air flow when the instrument is opened or closed.
- **Toms.** With the exception of the bass heavy standing tom these microphones do not have to be very broadband in the frequency response this might help diminish bleeding of the other instruments. They should be located close to the rim, for the hanging toms one microphone might be sufficient for two of them if placed in the mid.

9.3 The Mix of the Recording

9.3.1 The Goal: A Simulated Perfect Event, Not a Documentation

Let me start with the bad news: even with the best available technology and the best possible methods it is not possible to create a perfect match of the real event.

The good news: This is not necessary or even a good idea. If all technology and methods are used with competence, it is possible to create a convincing simulation of a perfect event that never was and that by far surpasses what is possible in reality. Errors can be fixed, the balance optimized, it is possible to even out things, to achieve maximal clarity and get very close to an ideal sound.

9.3.2 Cleaning Up

Every part of a track that (at least in the case of poly microphony) has no part at the result at a given time range should be muted. For signals that have a slow attack or release component this has to be done carefully by listening to the soloed track. If not it is too easy to remove important parts of the signal start or end.

Care has to be taken if the channel contains a room signal. The muting can cause a sudden change in the image of the background or "atmosphere", for example in its panorama.

9.3.3 Balance of the Sound Sources

The initial meaning of mixing and still its most important aspect is the control if the balance of the volume of the different sound sources. Most of the times this has to be controlled and changed continuously during the mix dependent on the needs (technically, artistically) at a given time. These days this can be done easily by mix automation that can record reproduced and edited as needed. Alan Parsons said in an interview with Google, that the old name "balance" of the mixing is still the main task.⁴

Effects that control the volume of a channel (mostly compressors) can do a part of this control. It is far from trivial to find the optimal parameter combination, however. Ethan Winer says that the effort is typically so large that it can be better used to create the optimal fader setting automation for the time range of the recording by hand in all detail.

 $^{^{4}}$ Video at [19]

9.3.4 The Sound of the Individual Tracks

The sound that the microphones deliver from the chosen position can be changed as needed in the total mix by equalizers. It is extremely important to keep in mind all the time that in doubt the total sound is more important than that of a single track. Therefore each change has to be counter-checked in the resulting mix.

There mus not be too many tracks that add to the bass part, otherwise the bass part may mask other parts and weaken the transparency. The proximity effect has to be compensated everywhere unless it matches the specific needs, and every unnecessary or artificial (wind or handling noise, vibrations) bass part is to be cut off by high pass filtering.

A rule of thumb is to raise frequencies with broad bands and cut out with narrow bands. A narrow band boost changes the character of the source because our ear is trained to recognize sound sources partly by observing characteristic resonances. To find the exact frequency of annoying parts of the signal, it is a good practice to use a narrow band boost and change the frequency until the annoyance is strongest. This is then the suitable frequency to cut out. This is described by Falconer[5].

9.3.5 Localization

By use of pan pot and the assignment of the main microphones to the output channels the sources can be distributed long the stereo base or (for surround mixes) in the possible angles. If there is a common distribution as is the case for orchestras, the mix should match the expectation.

Single solo signals should appear in the center of attention, in the mid.

The distribution as such eases the distinction of the different sound sources and is one of the means to create transparency in the recording aside from balance and tailored sound of the individual tracks.

The overall left to right balance must be kept and one side should not be prominent.

An extremely full sound can be achieved by adding unison instruments with different sounds at both extreme locations of the stereo image (Wall of Sound). A game of question and answer can be emphasized by a left - right localization of two sound sources. You should not go as far as the very early stereo recordings and create a ping-pong stereo.

The reverb should fill the stereo image (out of phase parts can widen it even more), for surround mixes this should include the surround loudspeakers to enclose the listener and sound natural.

9.3.6 Room Information

Room information that is part of the recording should be added into the mix in a way to achieve a convincing and coherent unity with the matching direct signal(s). The signals must not appear to be separated from the room.

9.3.7 Combination of Effects: Creating a Depth Scaling

Aside from the angle in the ground plane (or possibly the elevation) there is one more dimension in which a source is located: the perceived distance. Additional to the recording at a suitable microphone position there are some means during the mixing that can create a depth scaling:

- **Volume.** This has already been covered in the section about balance. Soft signals are perceived more distant.
- Amount of extreme treble. The air attenuates the treble at large distances (extreme case: thunder). With the treble control it is possible to simulate distance.
- **Presence frequencies.** By the works of Blauert we know frequencies that can cause a sound source to be perceived more or less present.

Presence is generated by attenuation at 1 KHz and a boost at 300 Hz and 2-4 KHz, everything rather broadband, Sengpiel recommends 2/3 of an octave. Distance is created with the opposite measures. Up to 6 dB are useful for this effect.

Delay of reverb. The sooner the reverb (especially the early reflections) appears, the more distant the source appears. In real life this is caused by diminishing detours of the grazing reflections and even more so of reflections by a rear wall with growing distance.

You may artificially create different distances by using a reverb signal that gets fed by different mix buses that contain different delays. Sources that shall sound distant get fed directly into the reverb effect, the closer they shall appear, the longer the pre-delay by the corresponding effect bus.

Reverb volume. The more distant a source is the higher is the level of the reverb compared to the direct signal.

By combination of these effects with stereo localization even two channel stereophony allows to distribute the sound sources on a two dimensional plane. This improves the distinction of the different sound sources and transparence without the need to keep the mix unnecessary dry (with little reverb).

Her is an illustrative example for a pop ballad, it is not necessary to use all of these parameters:

9.3.8 Focus

The means described above also allow to steer the attention of the listener. Pure audio recordings without visual clues can replace them by emphasizing some parts of the recording by volume or sound. This is another way a recording may get superior to the real event.

9.3.9 Dynamics

Voices and musical instruments often allow dynamics that cannot be usefully reproduced in realistic listening environments. Therefore it is often necessary to change the dynamics of the whole recording, as a rule, to compress it.

For a long time that should hopefully be ended soon there was a catastrophic race towards the loudest possible recording. It lead to many records in the domain or rock and pop music, in which each deviation from full scale level or – even worse: controlled distortion – was suppressed by means of brick wall limiters, compressors and multi band compressors. The effect cannot be reverted, because the information about the original dynamics is destroyed in the process. Rhythmical accents and dynamics as means of

| Strings — Background Vocals | | > Background Vocals | Back Less treble Less 3–4 KHz, 300 Hz, more 1Khz Softer More reverb |
|-----------------------------|---|----------------------------------|--|
| < | —— Drum Set | > | Reverb not delayed |
| Keyboard | Bass Solo Vocals | Guitar | More treble More 3–4Khz, 300Hz, less 1Khz Louder Less Reverb Reverb delayed Front |
| Left earlier Left louder | Left and right at the same time same volume | Right earlier e, Right louder | |

Figure 9.5: Example: creating a two dimensional sound stage

expression got lost. Luckily the onset of online streaming that often forces a constant lower mean level for all recordings can invalidate this approach.

A useful limitation of the extreme dynamics especially for some classical compositions can be done manually by changing the volume settings gradually before the original dynamic changes. This keeps the perceived change of dynamics intact. An example: before a fortissimo passage begins, the volume gets lowered slowly and barely noticeably. The dynamic change that follows remains intact although the overall dynamic change has been decreased to a value that can be used practically.

10 Practical Sound Reinforcement

10.1 Real Time

After the typically short sound check everything else happens in real time. The sound engineer must remain present and react to every upcoming change if and when needed, as fast as possible.

To achieve this, it is absolutely necessary to understand the used technology and acoustical situation in depth. A long and deep thinking about the situation may later on help to prevent a problem next time, however.

A good preparation can avoid many possible pitfalls from the beginning. This includes a good technical equipment that is optimally configured and measurement tools like a cable tester (most electrical problems are related to contacts and cables).

Cables must be laid out in a way that prevents tripping over them, everything should be fixed if possible (often by means of the famous Gaffa tape). The time table should be fixed in advance, everything should be arranged up front (effects, playbacks).

10.2 The Feedback

Because the reinforced sound is released in the same room or even in open air situations in close proximity, part of that sound will always end back at the microphones that are used.

In the best case this degrades the sound at the microphone position to a degree as delayed and colored noise signal. In the worst case it builds up to a undamped feedback howl or (often when the loop is closed by vibrations of the floor and microphone stands) as drone sound up to the limits of the available sound power. In extreme cases this can lead to destroyed equipment or even hearing loss.

Everything useful to prevent this from happening should be used. Some measures against this feedback are:

10.2.1 Limitation of Volume or Amplification

The feedback begins the moment that the volume of a reinforced component or the overall mix attains the original level at one microphone at the matching phase even for one single frequency. This means that the overall amplification reaches the factor one or 0 dB. Therefore it is important to never use more amplification than necessary for the application, not even a little bit.

You have to be aware that the negative consequences of feedback set in before the undamped oscillation and can be catastrophic to the sound. Before the undamped oscillation there is already a resonance boost at the critical frequency that can massively color the sound, also there is a comb filter effect with repetitions and a growing ringing at this frequency. This are all very good recipes against each form of useful mixing result. The volume of the sound sources on stage, especially instrument amplifiers, should be as low as possible. These amplifiers are not intended to reinforce the sound for the audience if a PA system is used for this. If the high volume is not needed because for example the sound including loudspeaker distortion is needed, such amplifiers should be located as close as possible to the artist's ear and pointed towards him, possibly on a loudspeaker stand or angled upside if used on the floor. This way the artist can hear enough without flooding the rest of the stage and additional microphones there with this sound, complicating everything.

If a signal is hard to be heard, first check whether a different signal masks it and try to lower its volume. This is always possible without danger in contrast to the boosting of some signal.

10.2.2 Useful Position and Angle of Directional Loudspeakers, Isolation

The reinforcing loudspeakers, most of all the PA, should only cover the regions they are meant to deliver sound to. Every bleeding at the region of the stage where microphones are located complicates the situation. In extreme cases they should be isolated by use of acrylic glass sheets or curtains.

10.2.3 Close Miking

If a strong amplification is to be used, it is absolutely necessary to place microphones for reinforcement as close as possible to their sources. Contrary to the situation in a studio compromises in the resulting sound are less of a concern. A theoretically better sound is no use at all if it is not attainable because of the onset of feedback howling. EQs may be used to optimize the resulting sound as much as possible in the live situation.

10.2.4 Directional Microphones, at Suitable Places and Angles

The stage is one domain where good directional microphones can shine. The rear opening of their capsules must not be covered by hand to maintain the directivity, although this is sometimes done.

Auxiliary sound sources, for example the monitor loudspeakers or unrelated loud instruments (e.g. drums) should be located at the minimum of the sensitivity characteristics if possible.

It is important that the microphone has a consistent directivity without strong frequency dependence. Often diffraction weakens the cancellation just for the important upper mid frequencies that are needed for good intelligibility and should otherwise be boosted. Again, one single frequency whose sensitivity in the critical angle is too high raises the danger of feedback.

10.2.5 Take Care when Boosting Frequency Regions

Every boost of frequency regions raises the danger of feedback. Cutting out complementary frequencies is inherently safer.

10.2.6 DI Boxes, Wherever Suitable

If possible at all, electronic signals should be fed into the equipment by means of DI boxes. The signal is not in danger of leading to feedback, and the isolation against other signals is perfect, allowing for a more transparent mix.

If amplifiers or loudspeakers are part of the formation of the sound (guitar amplifier, rotational loudspeakers), an emulation should be considered. Not every artist will consent, but sometimes the emulation may be used and the emulated sound delivered over a neutral monitor box towards the artist.

10.2.7 Transducers Without Strong Resonances

Every resonance, no matter how narrow or weak, adds up to the danger of feedback. This is true for loudspeakers (most of all monitors) and all microphones. Sadly, most frequency response curves provided by manufacturers are smoothed, downplaying the imminent danger quite a bit. For some modern dynamic stage microphones this aspect is taken into account lately.

10.2.8 Little Reverb in the Monitor

The reverb is always bound to deliver an extremely nonuniform frequency response, because each contained echo adds its own comb filtering.

Therefore reverb should be used only carefully in the monitor signal. It should not be ruled out, however. It can improve the detection of frequency changes as for example in a vibrato (the reverb signal works as reference) and the feeling of the artist. Never forget: the performance of the artist is the most important aspect.

10.2.9 Careful Usage of Compressors & Co

Compressors and limiters lower the amplification for high levels. If the mix is changed during a loud moment, it can be that the amplifications gets raised into the region of feedback when the level falls off.

Expanders and noise gates lower the amplification or switch off the signal altogether at soft moments. An onset of the signal can now trigger a sudden feedback.

10.2.10 Equalizers

With an EQ you may specifically lower the amplification of frequencies that are in danger of feedback. Often a very narrow cutting down is sufficient, which minimizes the impact on the sound. In the mean time a real time frequency analyzer is available for free for each smartphone. It helps detecting the frequencies of the feedback signal and can thus help to find the right EQ setting.

10.2.11 Feedback Killers

Feedback killers watch the signal and automatically place a matching notch filter at the frequency where a feedback is detected. The good part is that this happens in real time without supervision.

The problem is that this can happen with signals that are mistaken for a feedback like long lasting high level organ sounds or with deliberate feedback sounds form a electric guitar.

10.3 Vibrations, Wind and Pop Sounds

All directional microphones are extremely sensitive about vibrations, wind and pop sound.

It no extremely deep frequencies can be part of the desired sounds, typical for voice signals, it is useful to use specialized vocal microphones for on stage use with their built in proximity effect compensation. They provide also measures against wind and pop noises. These measures are often not working perfectly, often against pop noises when used extremely close to the mouth in direct breath. An additional pop and wind protection by foam material can be useful. It is also easy to clean and exchange. The protective grilles with their built in foam protection can typically simply been screwed off and should be cleaned regularly. Normal dish detergents are suitable for the task.

For amplified instruments with high volume the decoupling against vibrations by a microphone stand can be sufficient.

In open air you have to take care whether wind blows lead to noise with ultra low frequencies. If this is the case, it is necessary to utilize foam protections of that are rather large. Compact protectors either do not work efficiently or they have to be so dense that they suppress the treble too much[36].

It is necessary to fix microphones for instruments in a way that ensures that they keep the optimal place once it has been found. Microphone stands, cable and the microphones themselves must be secured against tripping over them.

10.4 Weather Conditions

Dampness and water (rain) are natural enemies of everything that has even remotely something to do with electricity. In open air, care has to be taken that the whole equipment is not endangered.

Vocal microphones close to the mouth are endangered by spit drops and condensing water from the breath. Dynamic capsules have the advantage that there is no discharge of the polarization voltage or in the impedance matching buffer. Again, an additional foam protection may help.

10.5 PA - Sound for the Audience

The PA (Public Access) is meant to produce the sound for the audience. It has to provide a high acoustic power with great bandwidth with good linearity and a directional characteristic that is as independent of the frequency as possible and matches the need of the event. If more than one row of audience is present – in other words, always – at least the treble region must be radiated from over the height of the heads, because otherwise the heads of the first row provide an acoustical shadow for the rows that are behind it, and so on.

The needed power raises dramatically if the necessary sound pressure level or the needed distance grow, especially in open air without the support of additional reverberation.

Please recall the calculation with the dB level: doubling the pressure level means four times the power, a felt doubling of loudness ten times the power, doubled distance in open air four times the power.

PA loudspeakers therefore are always built to have a high power handling and a high sensitivity (given in dB at one meter distance for 1 watt of power to the nominal impedance in open air).

10.5.1 Minimal Solution, no Bass

A very basic setup for signals with little bass (speech, small acoustic instruments) can be done with two small loudspeakers on stands on each side of the stage. One single central loudspeaker would either stand in the way between artist(s) and audience or behind the stage at the worst place for the danger of feedback.

10.5.2 Full Spectrum, Subwoofer

The bass components from a loudspeaker on a stand get attenuated by comb filtering by the first reflection on the floor in typical setups. If the whole spectrum is to be delivered, it is necessary to provide at least one subwoofer. It should be placed directly on the floor, so that the reflection in the bass region is in phase. This improves the sound pressure level by 6 dB and the efficiency by 3 dB (addition of two signals with the same sound pressure level, both in phase).

10.5.3 Classical Stack

The classical variant to create a large acoustical power uses a stack of loudspeakers on both sides of the stage. On the floor there are large subwoofers, most of the times as bass reflex bins or (more and more seldom) bass horns. They work in phase to enlarge the sound pressure and improve the efficiency. With growing need you may add as many loudspeakers as needed. If in the upper bass region a growing directivity sets in because of a large base that is more narrow than optimal, they can be arranged along a part of a circular curve. Before the directivity matches the angle of the curve there is a intermediate frequency region in which the phase difference by the curvature is not sufficient so that there is a more narrow directivity and increased sound pressure level there.

The method of stacking several loudspeakers is more problematic in the region of upper midrange and treble. With classical horn loudspeakers (they are almost always used because of their efficiency) it is not possible to match the phases of the loudspeakers at the rims and without holes in the region that generates the sound. When lots of speakers get combined this way this creates an effective grid with a resulting reciprocal and frequency dependent grid in the directivity (lobing). Therefore, several equal loudspeakers should be mounted above each other to keep the sound consistent in the lateral dimension, or one should try to achieve the result with one strong mid and treble speaker with enough output power. Because the needed raw power in the bass is much higher, it is possible to obtain a rather high acoustical output this way.

10.5.4 Vertical Array

For a vertical array the lobing is suppressed by using loudspeaker components that are built to create an even or smoothly curved (5°) wave front over the whole height. These can

be mounted directly touching each other and by their angle it is possible to form the total wave front that gets radiated piece by piece. To compensate the loss of sound pressure level for large distances the upper components get mounted at an angle perpendicular to the direction of the most distant listeners and the lower components get mounted at an angle to each other so that a curved and more distributed and therefore softer wave gets sent to the closer listeners. Less components beam towards them.

As a rule, the bass is again created by subwoofer stacks at the floor.

If there is enough power, it is possible to generate a directivity towards the bass by using subwoofers with partially opened backs that can generate a cardioid characteristic analogous to the cardioid microphones.

This class of equipment can only be utilized when the preparation and planning includes simulation of the setup.

10.6 Monitor

10.6.1 Monitor Loudspeakers

To help the artists to hear themselves and each other clearly, you may use monitor loudspeakers that get fed a dedicated mix. This may be one global mix from the front of house mixing console or several mixes for each artists, created on a specialized mixing console, distinct from the PA signal. Good communication with the stage is important. It may be a good idea to provide a monitor loudspeaker for the monitor mixer that can be switched to each monitor signal for control.

The main types of monitor loudspeakers are the floor monitor in wedge form, which should be positioned in the direction of the cancellation by the microphone directivity and the side fill monitors at the sides of the stage. They are more or less a small PA system to flood the stage with a monitor signal.

Because the monitor loudspeakers are directed towards the artist(s) it is extremely important to do everything possible against the danger of feedback caused by them.

The measures include (again) planned positioning of monitor loudspeakers and the stage microphones, usage of loudspeakers with as little resonances as possible, no or minimal use of reverb with their inherent uneven frequency response, equalizers to cut down dangerous frequencies and automatic feedback suppressors.

A bad monitor mix may present a group worse that a bad PA mix. The timing may suffer and the artists may appear unable to do a good performance.

Monitor loudspeakers do not always require a good bass response, because often the PA signal leaks large amounts of bass towards the stage, depending on the setup, room acoustics and the PA system type. Problems may arise by the inherent delay of the PA signal by its path length towards the stage. This can endanger the timing of the performance.

10.6.2 In Ear

The in ear monitoring utilizes soundproof ear buds that are fed with a monitor signal for each artist, often wireless. This prevents any danger of feedback by the monitor system. Some artists find it uncomfortable to be completely isolated form the real acoustical surrounding. It may be helpful to add a signal of the audience reaction into the mix.

Bibliography

- Overview by DLR about microphone arrays http://messtec.dlr.de/de/technologie/dlr-aerodynamik-und-stroemungstechnik/ mikrofon-array-mikrofon-array/, accessed 2016-05-30
- [2] EBU Tech. 3276 2nd ed. Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic, Mai 1998, accessed 2016-11-09 URL: https://tech.ebu.ch/docs/tech/tech3276.pdf
- [3] EBU Tech 3276-E Listening conditions for the assessment of sound programme material Supplement 1 Revised May 2004 Multichannel sound, accessed 2016-11-09 URL: https://tech.ebu.ch/docs/tech/tech3276s1.pdf
- [4] Rod Elliot: Web page about electronics, especially audio electronics. URL: http://sound.whsites.net, accessed 2018-03-27
- Joel Falconer: How to Use a Parametric Equalizer, http://music.tutsplus.com/ tutorials/how-to-use-a-parametric-equalizer--audio-2301, accessed 2016-06-07
- [6] Earl Geddes: Multiple Subwoofers in Small Rooms, Presentation http://www.gedlee.com/downloads/OptimalBassPlaybackinSmallRooms.pptx, Video https: //www.youtube.com/watch?v=SCWL-zusyqw, accessed 2016-06-01
- [7] Andreas Gernemann: "Decca Tree" gestern und heute, http://www.uni-koeln. de/phil-fak/muwi/ag/tec/deccatree.pdf, accessed 2016-06-02
- [8] Andreas Gernemann: Psychoakustische Werte für die Lokalisation mit Wei's und direr Frontlautsprechern, http://www.uni-koeln.de/phil-fak/muwi/ag/tec/ lokalisation.pdf, accessed 2016-06-02
- [9] Jürg Jecklin: Musikaufnahmen: Grundlagen, Technik, Praxis München: Franzis Verlag, 1980
- [10] Stephan Leschka: Digitale Mikrophontechnik, http://www.neumann.com/download.php?download=lect0052.PDF, accessed 2016-05-18
- [11] Volker Meitz: Die historische Entwicklung von Kondensatormikrophonen unter Berücksichtigung von Schaltungstechnik und Klangverhalten https://www2.ak.tu-berlin.de/~akgroup/ak_pub/abschlussarbeiten/2005/ MeitzVolker_MagA.pdf, accessed 2016-06-10
- [12] Swen Müller: Digitale Signalverarbeitung für Lautsprecher http://publications. rwth-aachen.de/record/56081/files/Mueller_Swen.pdf, accessed 2016-07-20

- [13] Nyal: Hard proof that equalization kills room modes http://www. acousticfrontiers.com/201163hard-proof-that-equalization-kills-room-modes-html/
- [14] Sean Olive: A Virtual Headphone Listening Test Method, http://seanolive. blogspot.de/2016/04/a-virtual-headphone-listening-test.html, accessed 2016-06-01.
- [15] Sean Olive: Perceiving and Measuring Headphone Sound Quality: Do Listeners Agree on What Makes a Headphone Sound Good? http://seanolive.blogspot. de/2014/06/the-2014-loudspeaker-industry.html, accessed 2016-06-01.
- [16] Sean Olive: Harman Researchers Make Important Headway in Understanding Headphone Response http://seanolive.blogspot.de/2013/07/ harman-researchers-make-important.html, accessed 2016-06-01.
- [17] Sean Olive: The Relationship between Perception and Measurement of Headphone Sound Quality http://seanolive.blogspot.de/2013/04/ the-relationship-between-perception-and.html, accessed 2016-06-01.
- [18] Harry F. Olson, Acoustical Engineering, 2nd ed. 1948
- [19] Alan Parsons & Julian Colbeck: Art and Science of Sound Recording, Talks at Google https://www.youtube.com/watch?v=9pDUwak1GA4&list= PL9Jp8w1LJcfIrwpBwFCyrpq7QhBLSLogV&index=6ş accessed 2016-06-06
- [20] Stefan Peus: Modern Acoustic and Electronic Design of Studio Condensor Microphones, AES Convention Paper 6131, Berlin, 2004
- [21] Richard Pierce: A Novel Approach to Rapid Loudspeaker Design and Prototyping, 1993, http://www.rdrop.com/~billmc/dick_pierce/spreadsheet.ps, accessed 2016-05-30
- [22] Dick Pierce, Loudspeaker.sc Spreadsheet für Unix-Tabellenkalkulation sc, http:// www.rdrop.com/~billmc/dick_pierce/, accessed 2016-05-30
- [23] Martin Schneider: Kondensatormikrophone im Vergleich Comparing Condenser Microphones https://www.neumann.com/download.php?download=lect0057_ Schneider_TMT_2010_Kondensatormikrophone_im_Vergleich.pdf, Zugriff 2017-06-12
- [24] Eberhard Sengpiel URL: http://www.sengpielaudio.com, accessed 2016-05-02
- [25] Sivian, O'Neil, Journal Acous. Coc. Amer., Vol. 3, No4, p 483, 1932
- [26] R. H. Small: Direct Radiator Loudspeaker System Analysys, J. Audio Eng. Soc, vol. 20, pp383-395, (1972)
- [27] R. H. Small: Closed-Box Loudspeaker Systems, Pts I and II, J.Audio Eng. Soc., vol 20, pp 798 -808, vol. 21 (1972), pp. 11-18 (1973)
- [28] R. H. Small: Vented-Box Loudspeaker Systems, Pts. I-IV, J.Audio Eng. Soc., vol 21, pp 363-372, pp. 438-444, pp. 549-554, pp. 635-639 (1973)

- [29] A.N. Thiele: Loudspeakers in Vented Boxes, J. Audio Eng. Soc, vol 19, pp. 282-392, pp471-4831971
- [30] Wikipedia article about the Soundfield microphone https://en.wikipedia.org/ wiki/Soundfield_microphone, accessed 2016-07-12
- [31] Floyd E. Toole: The Acoustical Design of Home Theaters http: //eu.infinityspeakers.com/tl_files/themes/shop/skins/infinityshop/ stylesheet/02/series/technologies/acoustical_design.pdf, accessed 2016-06-01
- [32] Floyd E. Toole:Getting The Bass Right http://www.harman.com/ sites/default/files/white-paper/12/11/2015%20-%2006%3A25/files/ LoudspeakersandRoomsPt3.pdf, accessed 2016-06-01
- [33] Sumit Chawla, Todt Welti, Kevin Voecks: Multiple Subwoofers Benefits Setup http://www.rapalloav.co.nz/blog/Multiple-Subwoofers-Benefits-Setup, accessed 2016-06-01
- [34] Wittek: Mikrofontechniken für Atmoaufnahme in 2.0 und 5.1 und deren Eigenschaften http://www.hauptmikrofon.de/HW/Wittek_Atmo_TMT2012.pdf, accessed 2016-07-12
- [35] Wittek: "ORTF-3D": A microphone setup for 3D Audio and VR ambience recording http://www.hauptmikrofon.de/3d/ortf-3d, Zugriff 2019-02-30
- [36] Jörg Wuttke: Mikrophonaufsätze URL: http://www.schoeps.de/documents/Mikrofonbuch_komplett.pdf accessed 2016-05-11
- [37] Jörg Wuttke: personal web site URL: http://www.ingwu.de/, accessed 2016-05-11
- [38] Jörg Wuttke: Wissenswertes rund ums Mikrophon, http://www.ingwu.de/mikrofontechnik/mikrofonaufsaetze/ 34-20-wissenswertes-rund-ums-mikrofon.html, accessed 2017-08-04
- [39] Christopher (Monty) Montgomery and the Xiph.Org Community: A Digital Media Primer for Geeks, URL: https://wiki.xiph.org/Videos/A_Digital_Media_Primer_For_Geeks, accessed 2016-05-19
- [40] Christopher (Monty) Montgomery and the Xiph.Org Community: Digital Show and Tell, URL: https://wiki.xiph.org/Videos/Digital_Show_and_Tell, accessed 2016-05-19
- [41] Wolfram Wagner, video list to demonstrate audio themes, URL:https://www. youtube.com/playlist?list=PL9Jp8w1LJcfJF60H4xynTjqXfzjYzNeoT
- [42] Manfred Zöllner und Eberhard Zwicker: Elektroakustik, Springer, 2003