Technische Universität Berlin
Fakulät I: Geistes- und Bildungswissenschaften
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation

Masterarbeit

Directivity of Real and Artificial Speakers for Room Acoustic Measurements

Jakob Struve

Matrikelnummer:	[redacted]
Studiengang:	Elektrotechnik M.Sc.
Erstgutachter:	Prof. Dr. Stefan Weinzierl
Zweitgutachter:	Prof. Dr.-Ing. Sebastian Möller
Betreuer:	David Ackermann
eingereicht am:	18. Oktober 2018

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, den 18. Oktober 2018

_____

Jakob Struve

Abstract

The DIN EN 60268-16 [1] defines the Speech Transmission Index as an objective measure for speech intelligibility and determines measurement procedures using artificial speakers and loudspeakers. In this work the influence of the directivity of these objects on the Speech Transmission Index and other room acoustic parameters is evaluated. For this purpose, as a first step, high-resolution 3D-directivity data of two artificial speakers and four small-membrane full-range loudspeakers was measured in an anechoic chamber. In a second step, a room acoustic simulation was carried out in which the measured objects were compared with existing data of real speakers and singers. The results show that in many cases the artificial speakers are in good accordance with the real speakers, whereas the loudspeakers tend to overestimate the speech intelligibility due to a more focused radiation. A second simulation shows that a vertical tilt of the loudspeakers can bring its results closer to that of the artificial speakers under certain circumstances. Due to the differing sources, the external data did not match the measurement grid used for this work. Therefore, the influence of the angular resolution of the measurement grid on the simulation results was additionally investigated.

Zusammenfassung

Die DIN EN 60268-16 [1] definiert den Sprachübertragungsindex als objektives Maß für die Sprachverständlichkeit und legt Messmethoden mit Hilfe von künstlichen Sprechern und Lautsprechern fest. In dieser Arbeit wird der Einfluss der Richtcharakteristik dieser Objekte auf den Sprachübertragungsindex und andere raumakustische Parameter untersucht. Zu diesem Zweck wurden in einem ersten Schritt hochaufgelöste 3D-Richtcharakteristiken von zwei künstlichen Sprechern und vier kleinmembranigen Breitband-Lautsprechern in einem reflexionsarmen Raum gemessen. In einem zweiten Schritt wurde eine raumakustische Simulation durchgeführt in welcher die gewonnen Daten mit existierenden Daten von realen Sprechern und Sängern verglichen wurden. Die Ergebnisse zeigen, dass die künstlichen Sprecher in vielen Punkten eine gute Übereinstimmung mit den realen Sprechern aufweisen, wohingegen die Lautsprecher aufgrund ihrer höheren Richtwirkung die Sprachverständlichkeit überschätzen. In einer zweiten Simulation kann gezeigt werden, dass in einigen Fällen die Ergebnisse der Lautsprecher durch ein vertikales Anwinkeln an die der künstlichen Sprecher angenähert werden können. Aufgrund der unterschiedlichen Quellen der externen Daten stimmen diese nicht mit dem für diese Arbeit verwendeten Messgitter überein. Aus diesem Grund wurde eine zusätzliche Untersuchung des Einflusses der Winkelauflösung des Messgitters auf die Simulationsergebnisse durchgeführt.

Contents

List of Abbreviations

C_{50}	Clarity
DI	Directivity Index
DIN	Deutsches Institut für Normung (German Institute for Standardization)
G	Gain (Sound Strength)
IR	Impulse Response
JND	Just-noticeable Difference
LF	Lateral Fraction
OpenDAFF	Open Directional Audio File Format
RAVEN	Real-Time Framework for the Auralization of Interactive Virtual Environments
RIR	Room Impulse Response
SNR	Signal-to-noise Ratio
STI	Speech Transmission Index

List of Figures

List of Tables

1. Introduction

The DIN EN 60268-16 [1] standard defines objective measures to evaluate speech intelligibility and determines procedures to measure them. In the description of the measurement procedures the fact is given, that the speech intelligibility in a room depends on the directivity of the sound source. Therefore, an artificial speaker (artificial mouth/head-mouth-simulator) is recommended for measurements in surroundings without amplification (no public address system). Alternatively a "small, high-quality" loudspeaker with a diameter of less than 100 mm should be used.
The scope of this work is to investigate the influence of the directivity of a sound source on the speech intelligibility and other room acoustic parameters and to review the recommendations of the DIN standard. For this purpose a variety of artificial mouths and small-membrane full-range loudspeakers were measured in an anechoic chamber to generate three dimensional directivity data. This data, along with directivity data of real and artificial speakers from external sources, is used in an room acoustic simulation to examine the influence of the directivity on a number of physical parameters in different rooms.
The directivity of the human voice is focused upon in a variety of research with different fields of main interest. Early studies from Dunn and Farnsworth [2] start to investigate the voice directivity, while later an artificial modelling of the voice is introduced by Flanagan [3], Flanagan [4] and Sugiyama and Irii [5]. Kob [6] and Kob and Jers [7] demonstrate methods to measure the directivity of human singers and speakers and make comparisons to an artificial singer. Although the conclusion states that "the radiation data from human singers and the data from the artificial singer are in good agreement" [6], no analytic analysis is given. Halkosaari et al. [8] compares the directivities of human and artificial speakers systematically from a telemetric point of view and places a strong focus on the mouth aperture. It was found that the artificial speakers do not always match the human speakers perfectly and that there exist a number of differences in the directivities. Due to the concentration on telemetry, only eight fixed positions are investigated on which microphones of telecommunication devices would be located. Therefore, no complete (3D) directivity data can be provided and the influence on room acoustics is not considered. In the paper of Bozzoli and Farina [9] it is stated (without further proof), that in many cases the directivity of the speaker is not relevant (particularly in room acoustics due to big room sizes). Still it is admitted that the directivity can affect the STI in certain cases and therefore the directivities of two artificial mouths and one human speaker are measured. Although an attempt was made to measure balloon data of the test objects, only the horizontal plane is evaluated in their work. The appearing differences in the directivities lead to the proposal to define a standard on the directivity of artificial mouths. The study of Katz and d'Alessandro [10] concentrates on directivity measurements of the singing voice. A professional singer is measured in a particularly detailed way using 24 microphones on a moveable arc as well as a video camera to capture the mouth aperture. This study confirms the theory that the radiation of a human singer depends mainly on the head radius and the mouth

aperture. On the other hand, no significant difference of the directivity for different vowels is found. Unfortunately it was not possible to compare the 3D directivity data of the singer from Katz and d'Alessandro [10] to the measurements in this work, because the microphones were not placed on an Gaussian grid and the number of measurement points was too small. This way the data processing described in section 2.3 could not be applied in rational way. In the study of Chu and Warnock [11] the directivity of 40 human speakers (male and female) and one artificial speaker was measured in an anechoic chamber using 16 microphones. The objective of their work was to examine differences in the directivity between normal, low and loud voice, male and female speakers, French and English speakers and average human and artificial speaker. The results show that most of these factors merely have a small impact on the directivity. The only significant difference was found between normal and low voice. The relative directivity levels are given in the appendix of Chu and Warnock [11] and are compared to the measurements in this work in section 3. Monson et al. [12] picks up the findings of Chu and Warnock [11] and provides a more detailed analysis on the horizontal plane directivity of 15 singers (male and female). Small differences of the directivity between the genders and between soft and loud speech were found in the higher frequency range. Larger differences were found between different phonemes. According to Monson et al. [12] the mode of sound production (speech or singing) has a negligible influence on the voice directivity. So far most of the previous work on the directivity of the human voice and artificial speakers concentrates on the near field and the impact on the design of telemetric devices. It seems that the influence of the speaker directivity on room acoustic measurements has not been investigated on a larger scale. As well as this, most of the available directivity data has limitations regarding the angular grid resolution and the frequency range.
The motivation for this work was to measure directivity data with a high angular resolution and a wide frequency range of artificial speakers and loudspeakers and to quantify the influence on room acoustic parameters by performing an room acoustic simulation. The simulation is done by using the software `RAVEN` and two rooms each containing two receiver positions to be considered. In the simulation, the measured objects are additionally compared to the directivity data of real speakers and singers from external sources and the influence of a vertical angle of 30° on the loudspeakers is investigated. Due to the different measurement-grid resolutions of the external data the influence of the grid on the simulation results is also examined.

2. Methods

2.1. Dataset

The influence of the speaker directivities on room acoustic measurements was examined using directivity data from various sources. Four small-membrane full-range loudspeakers and two artificial speakers were measured for this project. The detailed measuring process is explained in section 2.2. Another artificial speaker and the directivities of male and female speakers were obtained from Chu and Warnock [11]. Additionally a set of six male singers measured for Pedrero et al. [13] and a female soprano singer from Weinzierl et al. [14] were used in the simulation. Caused by the different origins of the data, there existed major differences regarding the grid resolution and the type of data (e.g. impulse responses, magnitude spectra, third-octave spectra). Therefore, it was necessary to carry out different types of signal processing for each dataset and to validate the comparability of the processed data. This process is explained in section 2.3.

Category	**Object**	**Grid Resolution**	**Origin**
Loudspeaker	6301b	5° x 5°	Measured for this project
	Mixcube	5° x 5°	
	MM201	5° x 5°	
	Talkbox	5° x 5°	
Artificial Speakers	HMS-II	5° x 5°	Measured for this project
	ITA Head	5° x 5°	
	B&K HATS	15° x 18°	Chu and Warnock [11]
Real Speakers	Male Speakers	15° x 18°	Chu and Warnock [11]
	Female Speakers	15° x 18°	Chu and Warnock [11]
	ITA Singer 1	36° x 36°	Pedrero et al. [13]
	ITA Singer 2	36° x 36°	Pedrero et al. [13]
	ITA Singer 3	36° x 36°	Pedrero et al. [13]
	ITA Singer 4	36° x 36°	Pedrero et al. [13]
	ITA Singer 5	36° x 36°	Pedrero et al. [13]
	ITA Singer 6	36° x 36°	Pedrero et al. [13]
	TU Soprano	36° x 36°	Weinzierl et al. [14]

Table 1: List of test objects

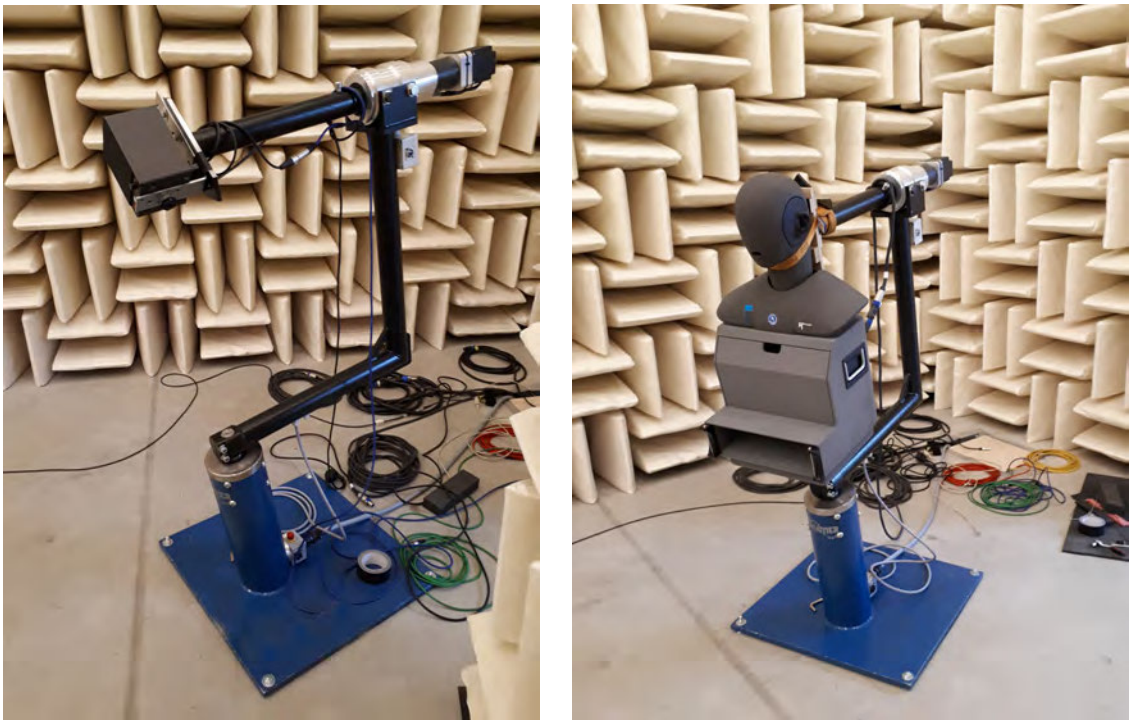
Table 1 provides an overview of all data used in this project and the corresponding grid resolution of the original measurements. The human singers in Aachen [13] and Berlin [14] were measured using a spherical array of 32 microphones [15] at fixed positions resulting in a grid resolution of 36° by 36°. The angles of this grid were chosen to match a Gaussian grid, which is helpful for the later data processing using the spherical harmonics decomposition. The data from Pedrero et al. [13] was provided as `OpenDAFF`-files containing octave band magnitude spectra and could be directly used in the simulation.

The directivity data of the female soprano from Weinzierl et al. [14] (`TU Soprano`) was published in the 4th order spherical harmonics domain using 25 coefficients. In their work, single tones (plus 9 overtones/harmonics) were recorded to generate the data. To generate a magnitude spectrum for each direction, the average of the absolute values of the single tone and the overtones was calculated in each third-octave band. In the work of Chu and Warnock [11], 14 microphones were mounted on two fixed arcs (plus two reference microphones) and a rotating chair (which moved in 15°-steps) was used to cover a half-sphere (the symmetry of the human head was assumed). This results in a resolution of 15° of the azimuth, but an irregular resolution of the elevation (the mean step size is 18°), which means that the grid is not Gaussian. For this reason, in the present work 32 grid points were selected that match the microphone array from Pollow et al. [15]. This way the further data processing could be done in the same manner as for the `TU Soprano` (see Section 2.3).

2.2. Measurements

As part of this project, the 3D directivities of four full-range loudspeakers with small membrane diameters and two artificial speakers (artificial mouths) were measured. The `Fostex 6301b` and the `Klein & Hummel MM201` are active full-range studio monitor loudspeakers with 4 inch (100 mm) drivers. The `NTI Talkbox` is an active loudspeaker that is specially designed for testing purposes and room acoustic measurements and contains a 4.5 inch (115 mm) driver. The `Avantone MixCube` is the only passive loudspeaker in the test field and is also designed as a full-range monitor. It has a driver with a diameter of 5.25 inch (135 mm).
The `Head Acoustics HMS-II` is a commercially available artificial head with an integrated artificial mouth. As a second artificial speaker a head built at ITA Aachen [7] was measured (later called `ITA-Head` in this paper). This head contains a two-way system built from one midrange loudspeaker at the mouth and two low-range loudspeakers at the torso with a bass reflex opening in the neck. The transition frequency between the systems is at 150 Hz. Each way of the system was measured separately and later combined during the signal processing.
The measurements were conducted in an anechoic chamber at FourAudio in Aachen, Germany. This chamber is an anechoic half-space, which means that its floor is reverberant and the walls and the ceiling are sound absorbing. It can be considered as anechoic down to 100 Hz due to the depth of the absorbing wedges of 85 cm. A calibrated 1/4 inch condenser microphone (`Brüel & Kjær 4939`) was used for the measurements. The microphone was placed directly on the floor to ensure that it was as close as possible to the reverberant surface. The data was captured using the `FourAudio WinMF` measurement system along with its `ROBO3` frontend. The frontend also contains a power amp to drive the passive loudspeakers. To measure the 3D directivity (balloon-data), the speakers were mounted on the `FourAudio ELF` system. This stand allows the automatized movement of the speaker around two axes. Figure 1 exemplary shows the measurement setup for a loudspeaker and an artificial speaker. The acoustical center of

each object was placed on the intersection of the two rotary axes.



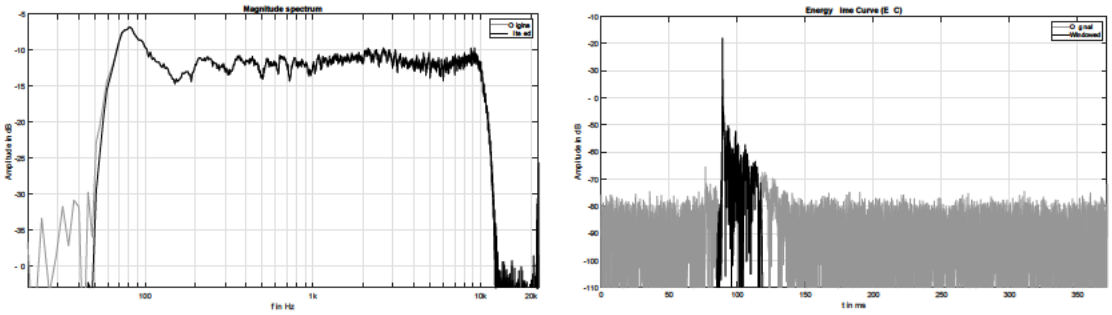
(a) Loudspeaker (b) Artificial Speaker

Figure 1: Exemplary measurement setup

In the first step the frontal frequency response was measured for each test object. In order to minimize the influence of reflections coming from the floor, the speaker was placed on the ground of the anechoic half-space. In this arrangement the frequency response was measured at different distances (varying from 8 to 0.5 m plus an additional near-field measurement directly in front of the membrane). By doing this, the influence of floor-reflections can be reduced and an impulse response that is close to free-field conditions can be calculated using `WinMF`.
Afterwards the speakers were mounted on the rotating stand to measure the 3D directivity with a distance of 8 m to the microphone. Depending on the geometry of the test object, axes of symmetry could be used to minimize the duration of the measurements. Most of the test objects are vertically symmetric, therefore a half-sphere was measured. The `NTI Talkbox` and the `Avantone MixCube` are also symmetric regarding the horizontal axis, therefore it was sufficient to measure a quarter-sphere. The speakers were automatically moved in 5° steps. At each position an impulse response was measured using a sine-sweep of 14th order (2^{14} samples, $f_s = 48\,\mathrm{kHz}$) with slower sweep speed at low frequencies to improve the SNR (the according impulse response was automatically calculated by the `WinMF` software).

2.3. Data Processing

In order to use the measured data in the room acoustic simulation it was necessary to further process the raw data and convert it into a file format that is suitable for the simulation software. In this case the Open Directional Audio File Format (`OpenDAFF`) [16] was chosen as it can be used in the simulation software `RAVEN` (described in Section 2.4). All further processing in this section was done in `MATLAB` using `AKtools` [17] and the `ITA Toolbox` [18].
The results of the measurements described in section 2.2 already possessed the appropriate grid resolution for the use in the simulation. Nevertheless some steps of data processing had to be carried out. The impulse response of each direction was filtered using a Butterworth high pass filter of 8th order at 60 Hz. This was carried out in order to avoid artefacts coming from reflections caused by the room, which is not perfectly anechoic at low frequencies. Additionally each impulse response was windowed to cut out negligible parts of the signal (before and after the impulse response) and to reduce the overall data size. These two steps are shown in figure 2 (exemplary for the `NTI Talkbox`).



(a) Filtering (b) Windowing

Figure 2: Processing of the measurement data

To use the data for the simulation in `RAVEN` it was necessary to split the signals into third-octave bands and calculate the energy of each band. This was done using the function *ita_fractional_octavebands* from the `ITA Toolbox`, which splits the frequency range of the human hearing into 31 bands according to DIN EN 61260-1 [19]. The third-band data was normalized to the spectrum of the frontal direction, to allow the comparison to the data from external sources which is also given in relative levels to the frontal direction. In the final step, the quarter- and half-spheres were mirrored to receive a full sphere for each object and the data was assigned to the `OpenDAFF`-grid. For each object one `OpenDAFF`-file containing third-band energies and one containing impulse responses was exported for further usage.
The two measurements of the ways of the `ITA-Head` were complexly added in an additional step. This was done before the third-band filtering using the `AKfilter`-function for high- (for the mid-range way) and low-pass (for the low-range way) filters at 150 Hz

(crossover filtering).
In addition to the objects which were measured especially for this project, data from various sources should be taken into account. One common problem regarding all external data is the resolution of the measurement grids which differ to the grid used in this work. To guarantee that the comparability of results from the room acoustic simulation between the data from different sources is given, it should be examined if the different grids have a relevant influence on the simulation results.
For this purpose, a low-resolution version of each test object from section 2.2 was created. This was done by selecting 32 points out of the 2522 data-points of the high-resolution grid that correspond with the 32 points of the microphone array from Pollow et al. [15] (this can be described as spatial downsampling). Afterwards a spherical harmonics decomposition was carried out to again create high-resolution `OpenDAFF`-files from these down-sampled objects (described below). Figure 3 shows the high-resolution measurement grid (blue) and the points corresponding to the low-resolution microphone-array (red). By comparing the simulation results of a high- and a low-resolution version of the same object, the influence of the measurement grid can be quantified. The comparison of the results between the spatially downsampled data and the original data is given in section 3.5.

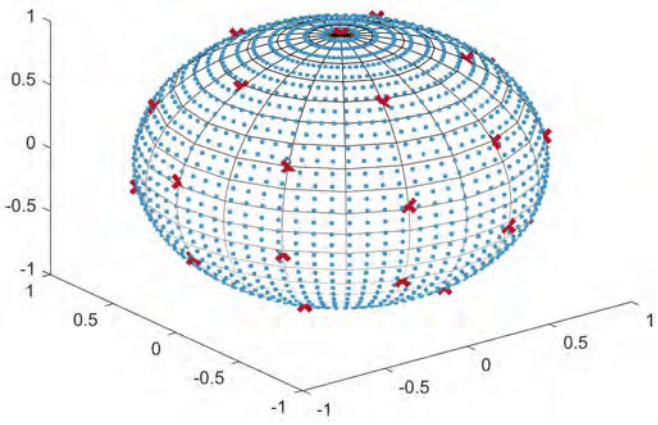
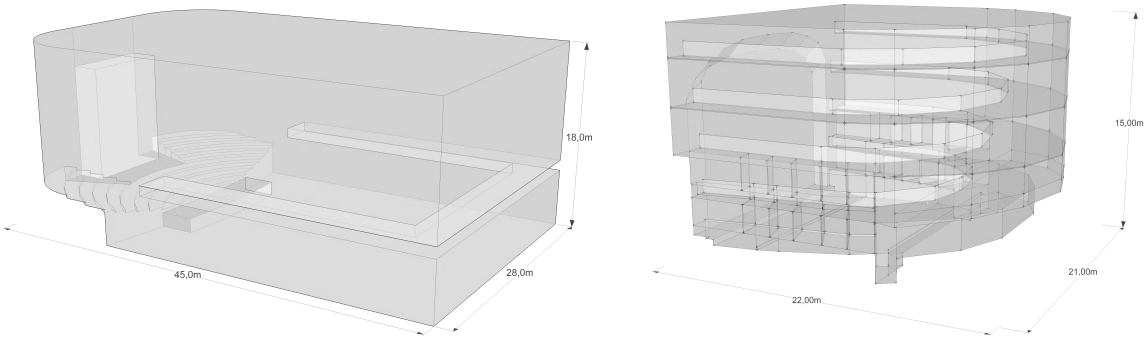


Figure 3: Measurement grid

In order to create high-resolution `OpenDAFF`-files from the down-sampled data as well as from the data from from Chu and Warnock [11] and Weinzierl et al. [14] the spherical harmonics decomposition [20] should be used. In the first step, a discrete spherical harmonics transformation was calculated using the low-resolution grid and the third-octave spectrum at each data point. For this purpose the *AKsht*-function from the `AKtools`-toolbox was used. The order of the spherical harmonics transformation was set to 4. Subsequently the spherical harmonics coefficients of the high-resolution grid were calculated using the *AKsh*-function and the 5°x5°-angles determined by the `OpenDAFF`-file. Finally, an inverse transformation using the results of the steps before was performed to receive third-band energies at each data point of the high-resolution grid.

2.4. Room Acoustic Simulation

All room acoustic simulations for this project were carried out with the software `RAVEN` (Room Acoustics for Virtual ENvironments). `RAVEN` is a hybrid room acoustic simulation software, which means that a deterministic image source-method and a stochastic ray tracing-algorithm are combined to calculate a sound field that is as realistic as possible in a reasonable processing time. The simulation of the Room Impulse Response (RIR) is split in three parts: The direct sound and the early reflections must be very precise regarding timing and spectral information to guarantee an exact localisation [21, p. 48]. For the later reflections, a high temporal resolution is not as critical as the reverberation energy is integrated over certain time slots and angle fields by the human hearing [21, p. 48]. In `RAVEN`, the early reflections are calculated using the image-source method. This method is a good approximation of reflections of first and second order, but omits surface or obstacle scattering which is an effect that is dominant for higher order reflections. Therefore, stochastic ray tracing is used to calculate the later reflections. This algorithm allows it to build an energy histogram which represents the envelope of the human hearing. The temporal fine structure of later reflections is synthesized by using a Poisson-distributed noise process [21, p. 70]. The complete RIR is generated by a superposition of the results from the described methods.
To take the influence of different room geometries and acoustically relevant properties into account, the simulation was performed with two largely different rooms. The models were simplified for the acoustic simulation, meaning that acoustically invisible structures (objects that are smaller than 0.5 m) were replaced by flat surfaces with corresponding absorption and scattering parameters. The room models were built in the CAD-software `SketchUp` and later were exported for the use in `RAVEN`. In this process one *.ac*- (containing the names and coordinates of the surfaces in the model) and one *.rpf*-file (containing the simulation parameters) was created for each room.
The *Concertgebouw* (Fig. 4a) is a fairly large shoebox shaped concert hall with a volume of 20 786 m^3 and a $RT60$ of 2.28 s. The room model was taken from the GRAP-database [22]. As a second room the *Theater an der Wien* (Fig. 4b) was chosen, which is a historical theatre room with a volume of 5312 m^3 and a $RT60$ of 1.31 s. The room model was taken from Ackermann et al. [23]. In comparison to the *Concertgebouw* it is smaller and contains a broader variety of materials which makes the sound field more inhomogeneous. For each room the source position and two different receiver positions are taken from the sources of the room models. The positions are in accordance to DIN EN ISO 3382-1 [24] and DIN EN ISO 3382-2 [25], which means that the sender is placed in a position similar to the source in a typical use case of the room, but not directly on an axis of symmetry and the receivers are placed on typical listeners positions with a distance more than two times that of the critical distance (Concertgebouw: Receiver 1 $d = 11.15$ m Receiver 2 $d = 15.34$ m, Theater an der Wien: Receiver 1 $d = 4.93$ m Receiver 2 $d = 9.94$ m).



(a) Concertgebouw (b) Theater an der Wien

Figure 4: Room models

In `RAVEN` the test objects (which are represented by the `OpenDAFF`-files) were used as source directivities and one simulation was carried out for each object in each room. The front pole of the `OpenDAFF`-files was vertically aligned to 0°, which means no vertical angle was applied. By taking a first look at the directivity of the artificial speakers and the real speakers it could be seen that the maximum of the radiation is slanted slightly downwards, which is in accordance with the findings of Dunn and Farnsworth [2], Flanagan [3] and Chu and Warnock [11]. Hence it was of interest to investigate whether the results of the loudspeakers will become more similar to the real and artificial speakers by tilting them downwards vertically. The results of an additional simulation with tilted loudspeakers are provided in section 3.4.
As an output, the simulation generates the RIR for each receiver, as well as an energy histogram. A number of room acoustic parameters can be calculated using these results. The room acoustic parameters are physical quantities and therefore objective criteria to describe acoustical qualities. The DIN EN ISO 3382-1 [24] shows a number of parameters which should be considered in describing acoustical settings. In addition, the *Speech Transmission Index* (STI) (defined in DIN EN 60268-16 [1]) is discussed as it is the most relevant parameter for this project.
The STI is an objective measure of speech intelligibility and is calculated using the reduction of signal modulation between the sender and receiver in the testing environment. The basic procedure is the calculation of the modulation transmission function $m(F)$ at 98 discrete data points. These data points are determined by 14 modulation frequencies in third-octave bands between 0.63 Hz and 12.5 Hz and seven octave bands between 125 Hz and 8 kHz (this corresponds to the third-octave bands from 100 to 10 000 Hz) in which each modulation frequency generates one value. The modulation factors are basically signal to noise ratios (SNR) and are calculated using equation 2.1 [26, p. 194].

$$m(F) = \frac{1}{\sqrt{1 + \left(2\pi F \cdot \frac{T}{13.8}\right)^2}} \cdot \frac{1}{1 + 10^{-\left(\frac{S/N}{10\,\mathrm{dB}}\right)}} \tag{2.1}$$

In equation 2.1 F depicts the modulation frequency in Hz, T the reverberation time in s and S/N the signal to noise ratio in dB. The modulation factors are weighted to achieve a higher correlation to the actual speech intelligibility and are used for the calculation of the effective SNR X_i (for each modulation factor m_i). Equation 2.2 shows the calculation of X_i.

$$X_i = 10 \cdot \lg \frac{m_i}{1 - m_i} \mathrm{dB} \tag{2.2}$$

Through averaging these values in the 7 octave bands, the modulation transfer indices are obtained. The STI value can then be calculated by using one more weighting of the octave band values. The resulting STI is a value in a range from 0 to 1, which can be divided into five categories of speech intelligibility: 0...0.3 "bad", 3...0.45 "poor", 0.45...0.6 "fair", 0.6...0.75 "good", 0.75...1 "excellent" (see [26, p. 196]).
The next important parameter is the *clarity* C_{50}. It is defined by the ratio between the early and the late part of the sound energy in decibel. For speech signals the sound energy within the first 50 ms is considered as the early part. This parameter describes the temporal clarity at the listeners position. In a range from -3 to 2 dB good speech intelligibility is assumed.

$$C_{50} = 10 \cdot \lg \frac{\int_0^{0.050} p^2(t)\,\mathrm{d}t}{\int_{0.050}^{8} p^2(t)\,\mathrm{d}t} \tag{2.3}$$

The *sound strength* G is used to measure the subjective sound pressure at the listener position. It gives an impression of the proportion of the room in the overall sound field. The sound strength can be measured by comparing the sound pressure of the test object at the listeners position to the sound pressure of the same source in the free field in the distance of 10 m ($p_{10}(t)$) in dB [24].

$$G = 10 \cdot \lg \frac{\int_0^{8} p^2(t)\,\mathrm{d}t}{\int_0^{8} p_{10}^2(t)\,\mathrm{d}t} \tag{2.4}$$

To generate the free field impulse response $p_{10}(t)$, a separate simulation in `RAVEN` was carried out. For this purpose a room with an absorption factor $\alpha = 1$ for all surfaces was designed and the distance between the sender and the receivers was set to 10 m (while keeping the directions of the two receivers towards the source).
The values of C_{50} and G are calculated seperately for each third-octave band and the resulting value is calculated by the arithmetic mean in the octave from 500 to 1000 hertz (third-octave bands from 400 to 1250 Hz) [24].
The *lateral fraction* LF can provide information about the apparent source width. It is

determined through the relation of the early reflections (between 5 ms and 80 ms) from side directions and the overall sound energy in dB and can be measured using a gradient-microphone to capture the side directions ($p_L(t)$, 0°-direction of the microphone pointing to the source) and an omnidirectional microphone at the same position to capture the overall sound energy ($p(t)$). The gradient-microphone used for the LF mathematically corresponds to a weighting with the square of a cosine of the angle of the lateral sound.

$$LF = \frac{\int_{0.005}^{0.080} p_L^2(t)\,\mathrm{d}t}{\int_{0}^{0.080} p^2(t)\,\mathrm{d}t} \tag{2.5}$$

Like the other parameters, the LF is calculated individually for each third-octave band in `RAVEN`. The mean value of LF is calculated by the energetic mean in the octave bands from 125 to 1000 Hz (third-octave bands from 100 to 1250 Hz).
Independent from the room acoustic simulation, the *Directivity Index* DI is calculated for all objects in the 31 third-octave bands. The DI is defined in DIN EN 60268-5 [27] as the logarithmic ratio of the sound pressure level of the test object on the reference axis and the sound-pressure of an omnidirectional source emitting the same acoustic power. As a reference axis, the frontal direction of the test objects was chosen and the according level could be directly extracted from the `OpenDAFF`-file. To calculate the emitted acoustic power, the levels of all directions were integrated over the complete sphere.

3. Results

3.1. Magnitude Spectrum

Although the spectrum of the measured objects is not considered in the room acoustic simulation, it is relevant to classify the overall characteristic of a loudspeaker. Figure 5 shows the third-octave band magnitude spectrum (normalized to 1 kHz) of the four loudspeakers and the two artificial speakers which were measured for this work.

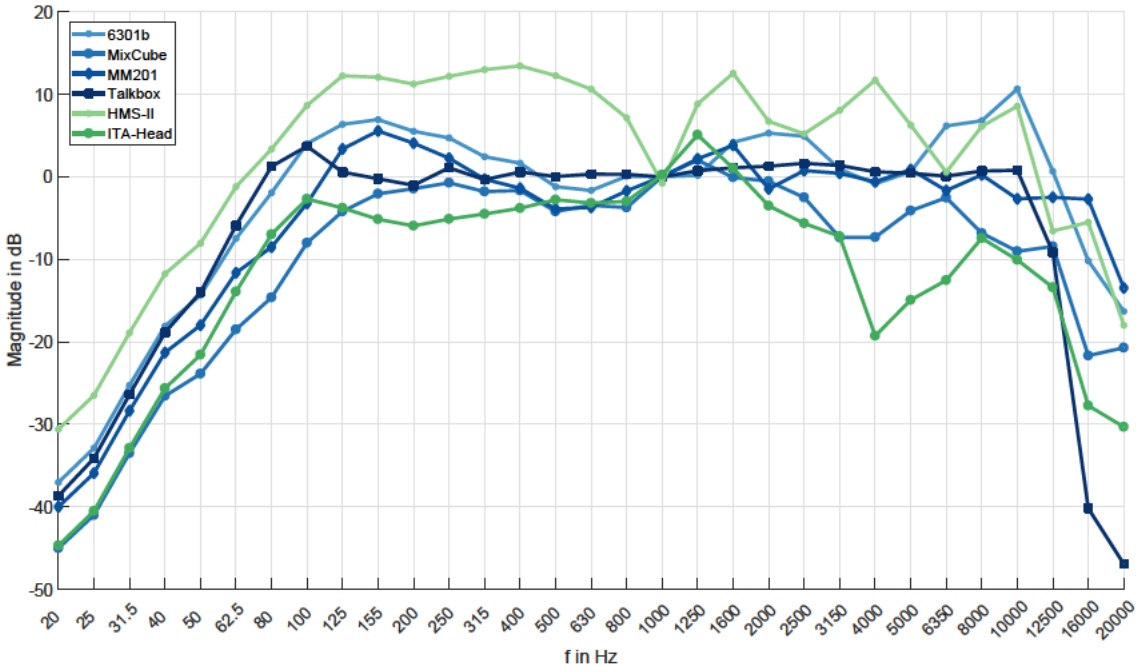


Figure 5: Normalized third-octave magnitude spectrum (frontal direction)

It can be seen that all objects show a drop of the magnitude below 125 Hz and above 10 kHz. The `NTI Talkbox` has the most flat spectrum of all objects in the range from 125 to 10 000 Hz, whereas the `Fostex 6301b` and the `Klein & Hummel MM201` have a small boost 155 Hz and the `Avantone MixCube` a depression at 4 to 5 kHz. The `Fostex 6301b` also shows a peak at 10 kHz and has the most uneven spectrum of the four loudspeakers. The unevenness of the loudspeakers is small in comparison to the spectrum of the artificial speakers. The `Head Acoustics HMS-II` shows a fairly large notch at 1 kHz and smaller notches at between 2 kHz and 3.15 kHz as well as at 6.35 kHz. The `ITA-Head` as a small peak at 1.25 kHz and a large recess at 3.15 kHz.

3.2. Directivity Index

The first important parameter to classify the directivity of the test objects is the *Directivity Index DI*. Figure 6 gives an overview over all objects in third-octave bands between

20 Hz and 20 kHz. The following figures show the DI sorted by the three different categories: loudspeakers (Fig. 7), artificial speakers (Fig. 8) and real speakers/singers (Fig. 9). The DI in this figures is calculated in relation to the frontal axis of the objects.

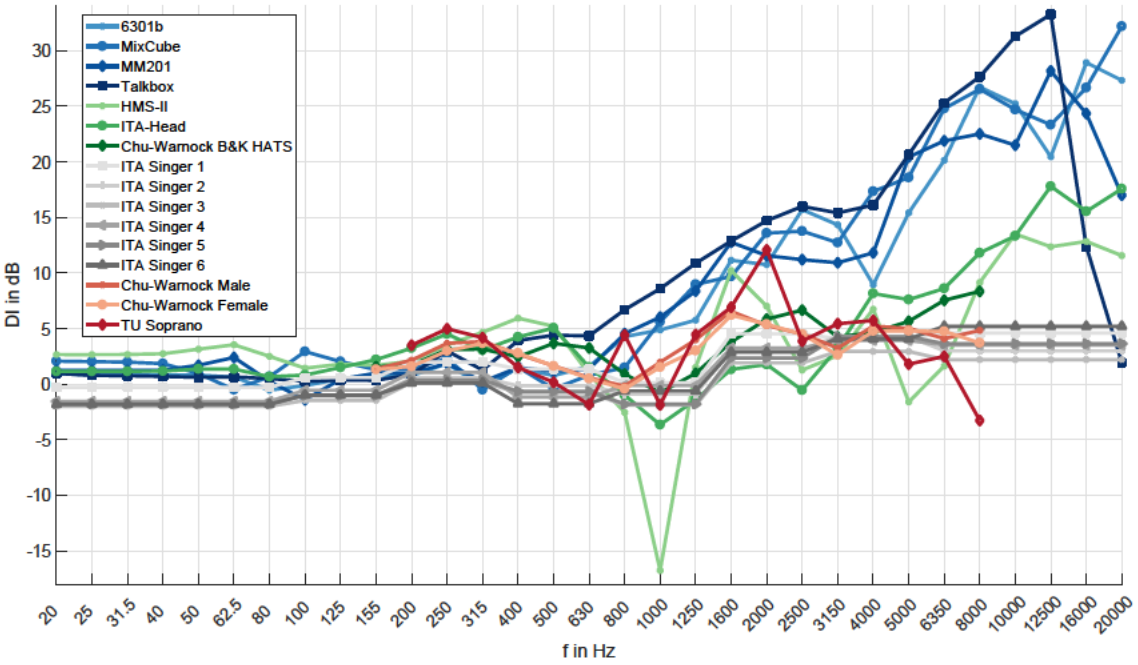


Figure 6: Directivity Index: Overview

Figure 7 shows that all four measured loudspeakers have a similar curve in a wide frequency range. In the range from 20 to 315 Hz the mean DI for all four loudspeakers is 0.95 dB with a maximum of 2.98 dB and a minimum of −1.41 dB. Thus the loudspeakers can be considered as omnidirectional in that range. At 400 Hz the DI of the `NTI Talkbox` starts to increase, whereas the other three loudspeakers start to focus above 800 Hz. In the whole range from 400 Hz to 20 kHz the `NTI Talkbox` has the highest DI of the four loudspeakers. Above 12.5 kHz the DI of the `NTI Talkbox` and `Klein & Hummel MM201` begins to decrease. This characteristic can be neglected as this frequency range is not relevant for the calculation of the room acoustic parameters that are evaluated in this work and the energy in this bands is low (see Figure 5).

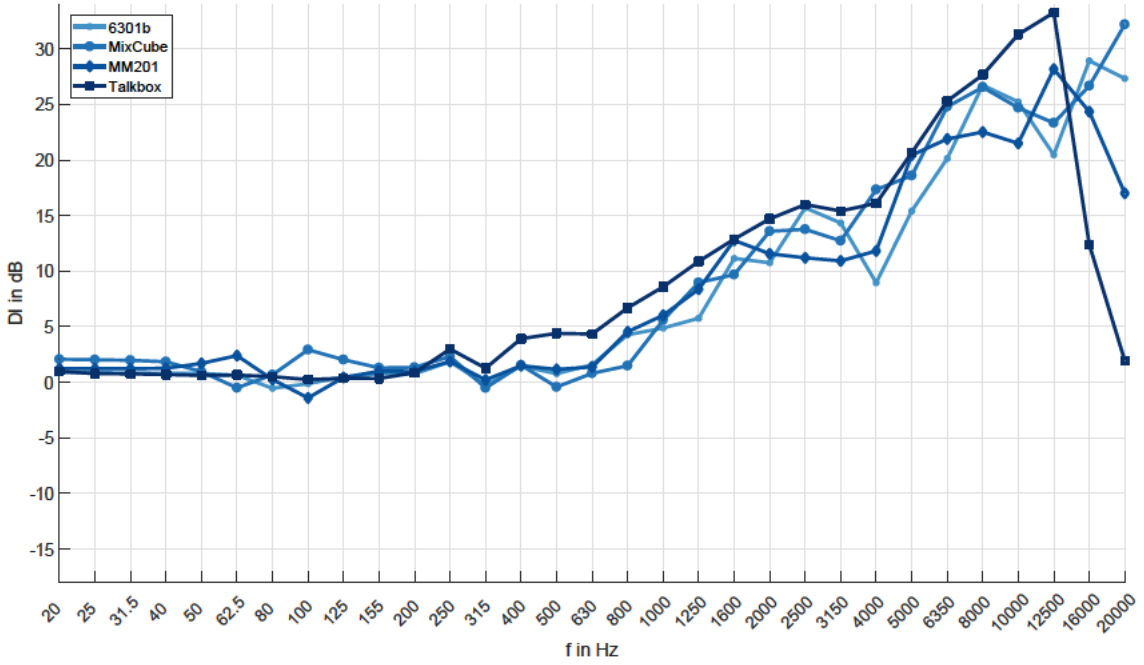


Figure 7: Directivity Index: Loudspeakers

The three artificial speakers show noticeably differing trends. Most remarkably, none of them reach the high DI values of the loudspeakers at higher frequencies. All three artificial speakers have the lowest directivity in the 1 kHz-band. This is most drastic for the `Head Acoustic HMS-II` which reaches −16.78 dB at this frequency. It should be noted that the `B&K HATS` was measured in the range from 155 Hz to 8 kHz by Chu and Warnock [11], therefore its curve is only displayed in this range.

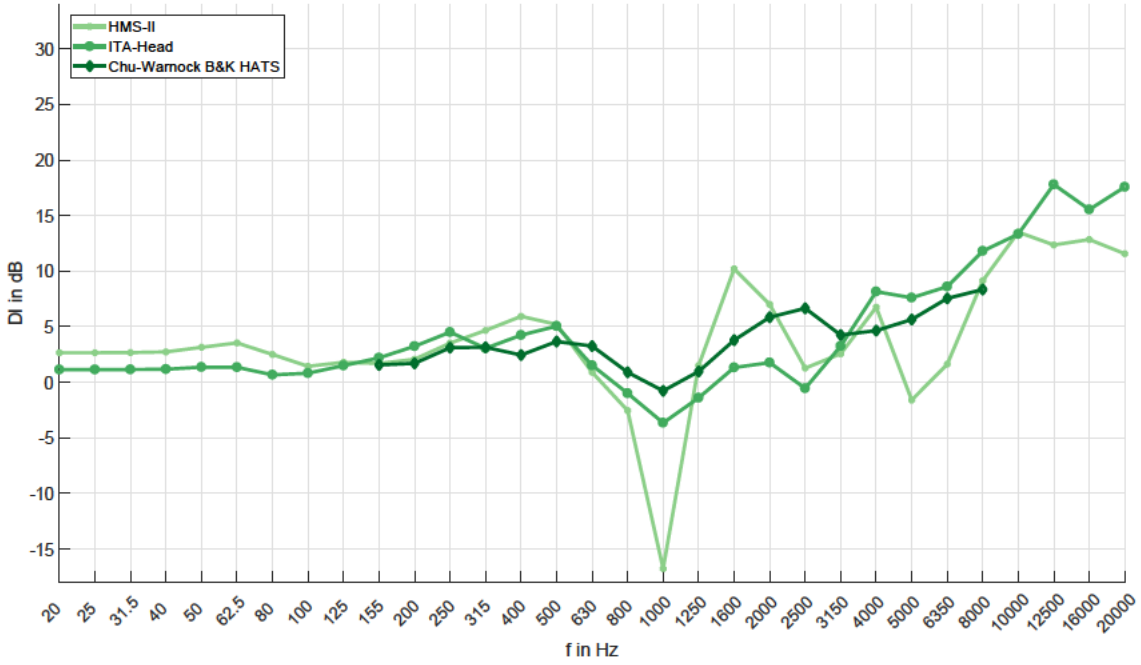


Figure 8: Directivity Index: Artificial speakers

Figure 9 shows the directivity indices of the real speakers and singers from the differing sources (see Table 1). Again it should be noted that not all objects were measured in the whole frequency range: The data from Chu and Warnock [11] ranges from 155 Hz to 8 kHz, the singer from Weinzierl et al. [14] (`TU Soprano`) ranges from 200 Hz to 8 kHz. The six `ITA-Singers` show similar DI-curves over the whole frequency-range. The male and female speakers from Chu and Warnock [11] also differ in less than 1 dB between each other. This is in accordance with the results presented in their paper. The `TU Soprano` shows a different and more irregular DI-trend with peak of 12.07 dB at 2 kHz and decreasing directivity in the range of 2 to 10 kHz.

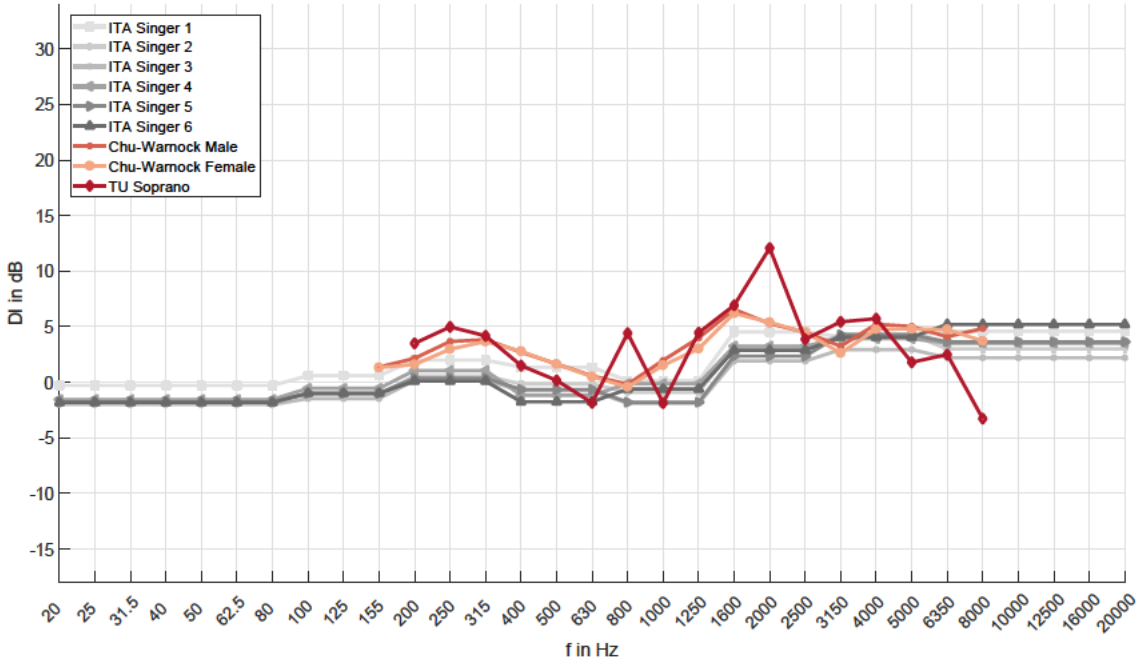
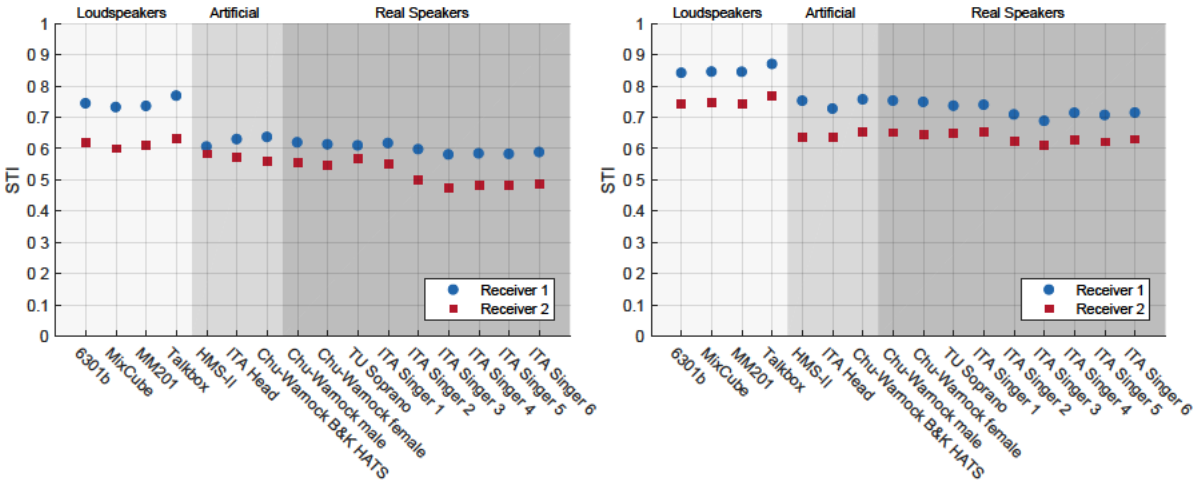


Figure 9: Directivity Index: Real speakers

3.3. Room Acoustic Parameters

A selection of the most important room acoustic parameters which are associated with speech intelligibility is given in this section. Each figure shows one parameter for one room sorted by the tested objects. From left to right the first four values belong to the loudspeakers followed by the three artificial speakers and the nine real speakers. The values for two different receiver positions are shown for each room. Additionally a table for each parameter shows the mean, maximum and minimum values for each of the three categories of test objects. For the objects that were not measured in the complete frequency range, the values of the lowest and highest measured band were extrapolated to calculate the room acoustic parameters.
Figure 10 depicts the *Speech Transmission Index*. The mean STI (over all objects and both receiver positions) of the Concertgebouw is 0.60 whereas the mean STI of the Theater an der Wien is 0.71. There are significant differences between the categories of test objects: For the Concertgebouw the mean STI of the loudspeakers at Receiver 1 is 0.74, but only 0.62 of the artificial speakers and 0.60 of the real speakers (Receiver 2: Loudspeakers 0.62, Artificial speakers 0.57, Real Speakers 0.52). The values of the Theater an der Wien show a similar trend: Loudspeakers 0.85, Artificial Speakers 0.74, Real Speakers 0.72 (Receiver 1) and Loudspeakers 0.75, Artificial Speakers 0.64, Real Speakers 0.63 (Receiver 2).



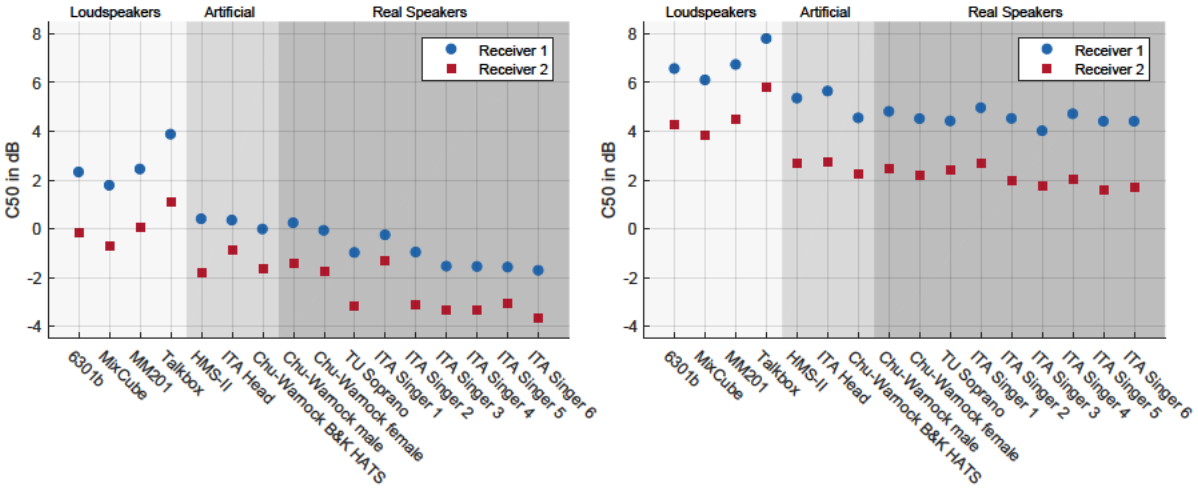
(a) Concertgebouw (b) Theater an der Wien

Figure 10: Speech Transmission Index (STI)

		Loudspeakers			Artificial Speakers			Real Speakers		
Room	Pos	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Concert-	1	0.73	0.74	0.77	0.61	0.62	0.64	0.58	0.60	0.62
gebouw	2	0.60	0.62	0.63	0.56	0.57	0.58	0.48	0.52	0.57
Theater	1	0.84	0.85	0.87	0.73	0.74	0.76	0.69	0.72	0.75
an der Wien	2	0.74	0.75	0.77	0.63	0.64	0.65	0.61	0.63	0.65

Table 2: Mean, maximum and minimum values of STI

In Figure 11 the *Clarity* is given. Again the loudspeakers reach a higher value than the other tested objects and as well as this the differences within the categories show a wider spread. The overall trend of the values is similar for both rooms and for both receiver positions. The mean value over all objects and positions, again, is significantly higher for the Theater an der Wien (4.01 dB) in comparison to the Concertgebouw (−0.80 dB). The mean values of the loudspeakers (Concertgebouw Receiver 1: 2.60 dB Receiver 2: 0.08 dB, Theater an der Wien Receiver 1: 6.79 dB Receiver 2: 4.61 dB) are higher than the values of the artificial speakers (Concertgebouw Receiver 1: 0.24 dB Receiver 2: −1.43 dB, Theater an der Wien Receiver 1: 5.17 dB Receiver 2: 2.56 dB) and real speakers (Concertgebouw Receiver 1: −0.94 dB Receiver 2: −2.68 dB, Theater an der Wien Receiver 1: 4.52 dB Receiver 2: 2.09 dB).



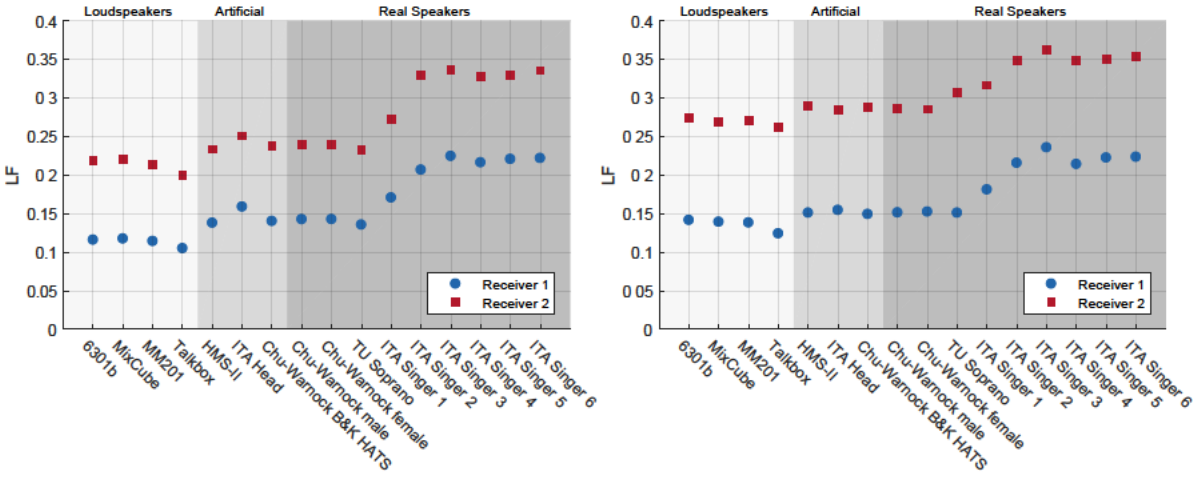
(a) Concertgebouw (b) Theater an der Wien

Figure 11: Clarity (C_{50})

		Loudspeakers			Artificial Speakers			Real Speakers		
Room	Pos	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Concert-	1	1.77	2.60	3.87	−0.02	0.24	0.40	−1.72	−0.94	0.23
gebouw	2	−0.68	0.08	1.09	−1.82	−1.43	−0.85	−3.65	−2.68	−1.31
Theater	1	6.09	6.79	7.79	4.54	5.17	5.64	4.01	4.52	4.95
an der Wien	2	3.83	4.61	5.83	2.27	2.56	2.72	1.60	2.09	2.69

Table 3: Mean, maximum and minimum values of C_{50} in dB

Figure 12 shows the *Lateral Fraction*. It behaves inversely to the previous parameters: The `ITA-Singers` show the highest values, whereas the loudspeakers tend to have lower values. This applies particularly to the Concertgebouw (Fig. 12a). It can be seen that for this parameter the differences between the different real speakers are fairly high. The difference between the maximum and the minimum at one position is up to 0.11, whereas for the loudspeakers and artificial speakers it is not greater than 0.02.

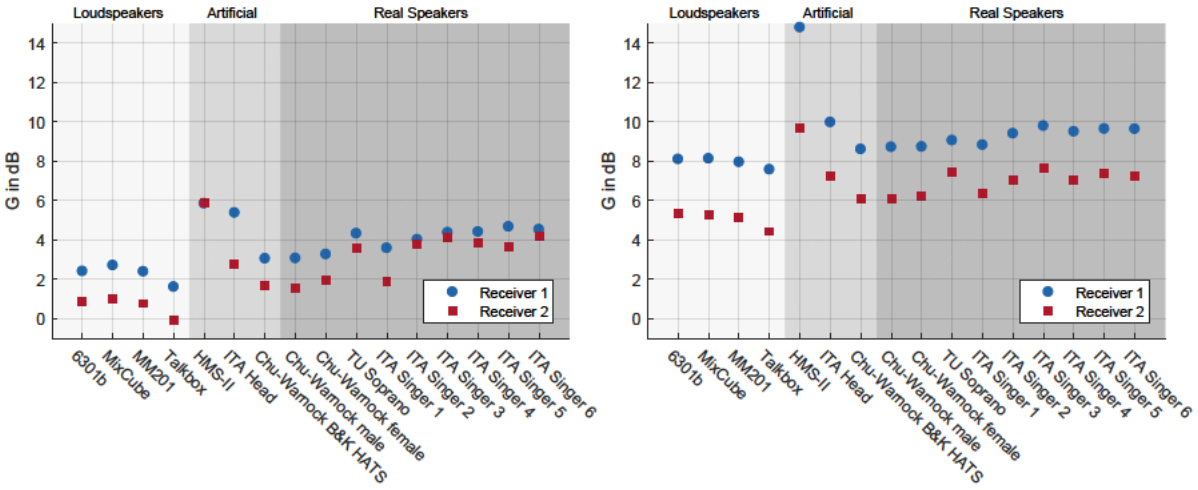


(a) Concertgebouw (b) Theater an der Wien

Figure 12: Lateral Fraction (LF)

Room	Pos	Loudspeakers Min	Loudspeakers Mean	Loudspeakers Max	Artificial Speakers Min	Artificial Speakers Mean	Artificial Speakers Max	Real Speakers Min	Real Speakers Mean	Real Speakers Max
Concert-	1	0.10	0.11	0.12	0.14	0.15	0.16	0.13	0.19	0.22
gebouw	2	0.20	0.21	0.22	0.23	0.24	0.25	0.23	0.29	0.34
Theater	1	0.12	0.14	0.14	0.15	0.15	0.15	0.15	0.19	0.24
an der Wien	2	0.26	0.27	0.27	0.29	0.29	0.29	0.29	0.33	0.36

Table 4: Mean, maximum and minimum values of LF



(a) Concertgebouw (b) Theater an der Wien

Figure 13: Sound Strength (G)

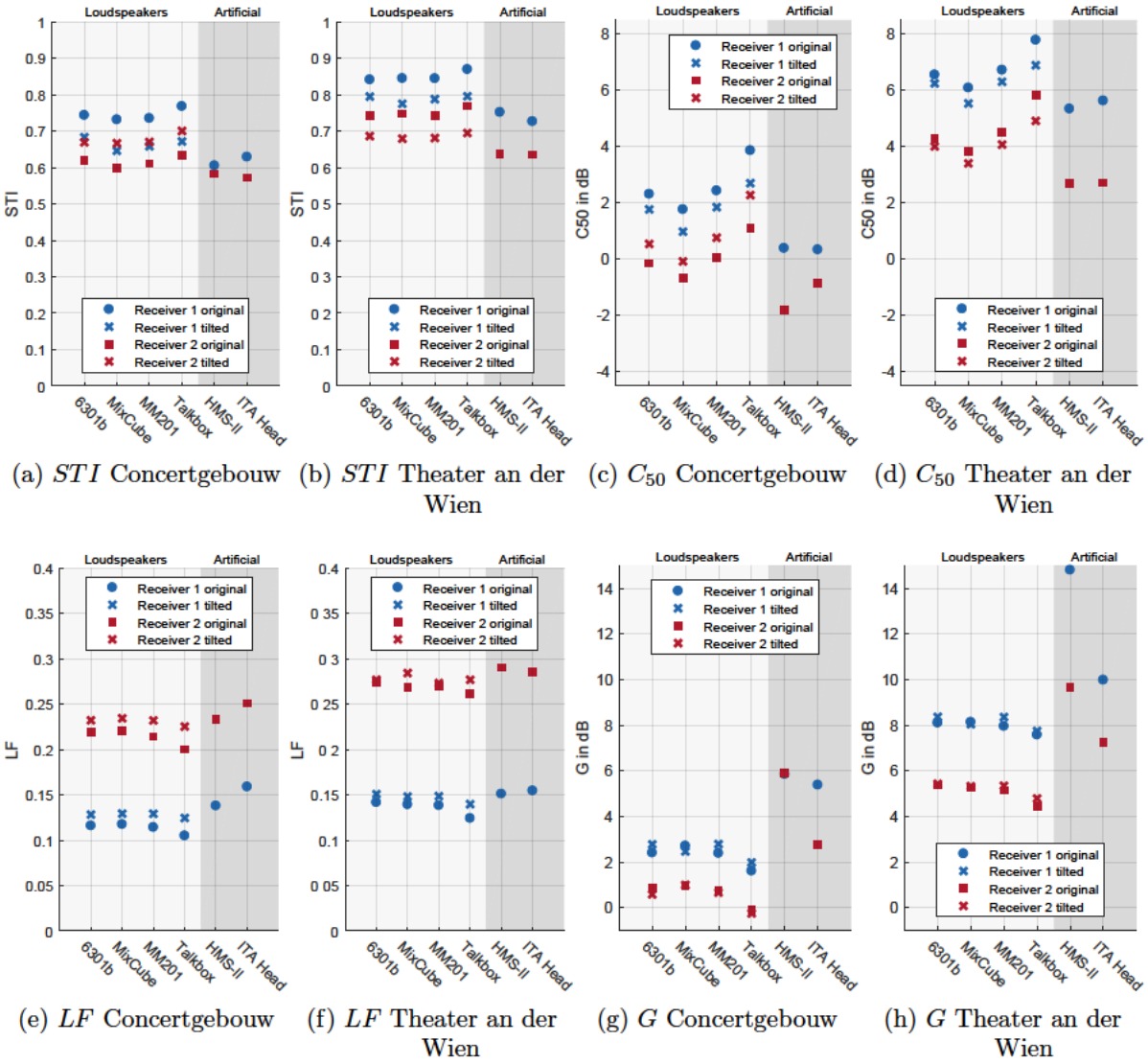
Figure 13 depicts the *Sound Strength*. The highest values of this parameter can be found for the artificial speakers with the `Head Acoustics HMS-II` peaking at 5.85 dB (Concertgebow) and 14.8 dB (Theater an der Wien). As for the LF, the loudspeakers show lower mean values than the other categories, for example 2.48 dB lower than the artificial speakers and 1.75 dB lower than the real speakers at receiver 1 in the Concertgebouw.

		Loudspeakers			Artificial Speakers			Real Speakers		
Room	Pos	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Concert-	1	1.62	2.28	2.71	3.06	4.76	5.85	3.07	4.03	4.68
gebouw	2	−0.09	0.63	0.98	1.70	3.45	5.89	1.58	3.19	4.23
Theater	1	7.58	7.94	8.14	8.61	11.13	14.80	8.72	9.26	9.80
an der Wien	2	4.43	5.05	5.37	6.09	7.68	9.68	6.09	6.94	7.66

Table 5: Mean, maximum and minimum values of G in dB

3.4. Tilting of the Loudspeakers

In a second simulation it was of interested whether the results of the room acoustic parameters from the loudspeakers could come closer to the real and artificial speakers if the loudspeakers were tilted vertically. To determine the tilting angle, the directivities of the real and artificial speakers are weighted in third-octave bands with the weighting-factors of the STI which are given in DIN EN 60268-16 [1] and the direction of the maximum of the sum of this weighted directivities is calculated. Using this method, a mean angle of this direction of 30° was found for the real and artificial speakers. This value corresponds to the findings of Dunn and Farnsworth [2], Flanagan [3] and Chu and Warnock [11]. Figure 14 shows the results of the loudspeakers which are tilted by 30° compared to the previous results of the loudspeakers and artificial speakers without a tilt.



(a) STI Concertgebouw (b) STI Theater an der Wien (c) C_{50} Concertgebouw (d) C_{50} Theater an der Wien

(e) LF Concertgebouw (f) LF Theater an der Wien (g) G Concertgebouw (h) G Theater an der Wien

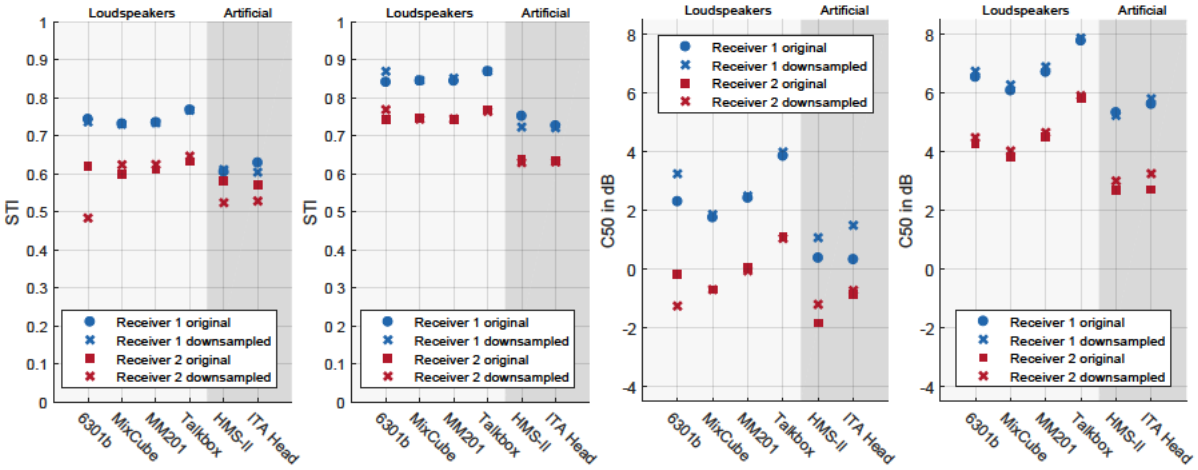
Figure 14: Influence of the tilting of the loudspeakers on room acoustic parameters

Figure 14a shows that the mean STI of the loudspeakers is lowered by 0.08 at Receiver 1, but increased by 0.06 at Receiver 2. In the Theater an der Wien (Fig. 14b) the mean STI is reduced by 0.06 at both receivers. The differences of C_{50} follow the trend of the STI with a maximum of difference of 1.19 dB. The LF increases in both rooms at both receivers by maximum 0.025, while G shows only very small deviations.

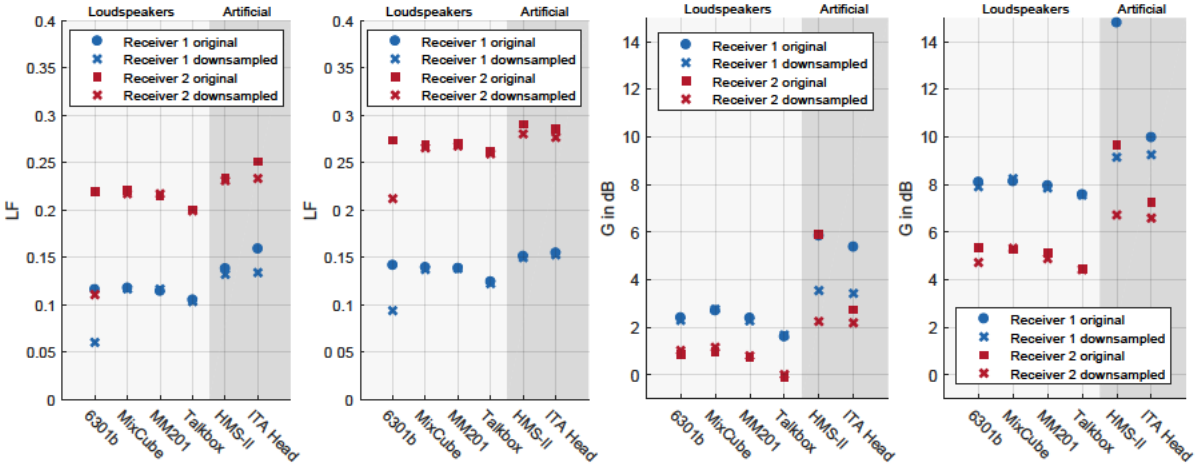
3.5. Comparison between Original and Down-Sampled Objects

Additionally the simulation is carried out with the spatially down-sampled objects as described in section 2.3. Figure 15 shows the resulting room acoustic parameters at the two receiver positions for the two rooms. It can be stated that on a large scale the

parameters resulting from the down-sampled objects show values that are very close to the values from the original objects. Nevertheless some bigger deviations exist for single objects and receiver positions. This particularly applies to the `Fostex 6301b` which in the Concertgebouw shows a larger difference of 0.14 of the STI at receiver 1 as well as of C_{50} at both receivers (up to 1.1 dB) and fairly large differences of the LF (up to 0.11) at both receivers in both rooms. The `Head Acoustics HMS-II` shows larger deviations of G (up to 5.65 dB) at both receivers in both rooms.



(a) STI Concertgebouw (b) STI Theater an der Wien (c) C_{50} Concertgebouw (d) C_{50} Theater an der Wien



(e) LF Concertgebouw (f) LF Theater an der Wien (g) G Concertgebouw (h) G Theater an der Wien

Figure 15: Comparison of room acoustic parameters from original and downsampled objects

4. Discussion

The results of the room acoustic simulation show a number of similarities within the categories of the test objects. Comparing the loudspeakers against each other, the room acoustic parameters show only small differences. A connection between the DI and the room acoustic parameters can be found, most evident for the `NTI Talkbox` which has the the highest DI of all loudspeakers and also has the highest values for the STI and C_{50}, whilst having the lowest values for G and LF. The `Fostex 6301b`, `Klein & Hummel MM201` and `Avantone MixCube` show very similar results of the room acoustic parameters and also of the DI. It should be mentioned that all differences between the loudspeakers are very small compared to the differences between the other categories.
Looking at the artificial speakers, a number of similarities can be found between the three different models. For the STI, for which a wide frequency range from 100 to 10 000 Hz is considered, the differences between the artificial speakers are fairly small. This is in accordance with the trend of the DI which is within a similar range overall. Still there are some weighty differences in singular third-octave bands, most significantly at 1 kHz, where the `Head Acoustics HMS-II` shows a very low DI of -16.78 dB. This can be explained by a noticeable notch on the front axis of the directivity at 1 kHz (see Fig. 16) possibly caused by interferences due to the geometry of head.

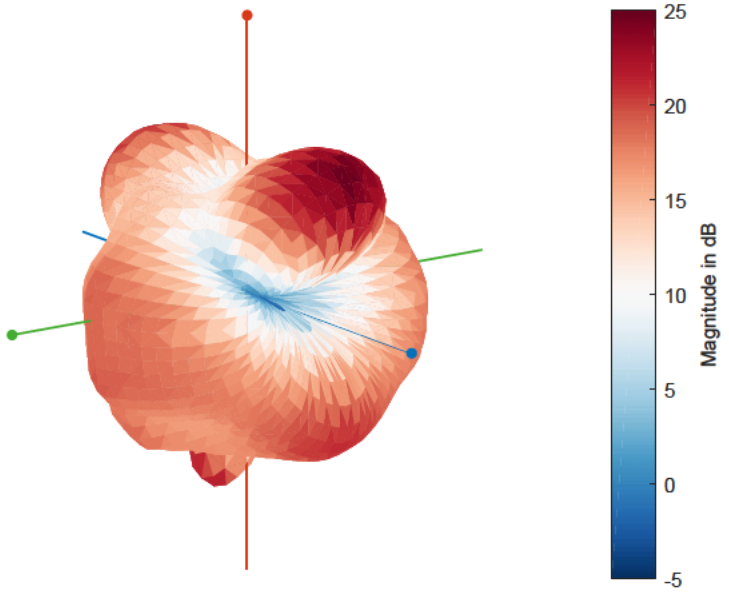


Figure 16: 3D directivity of the `Head Acoustics HMS-II` at 1 kHz

This remarkable characteristic makes a particular difference at the calculation of G as it is calculated in the frequency range from 400 to 1250 Hz, leading to unusually high values of this parameter for this object.
In reviewing the room acoustic parameters in section 3 in detail the most evident finding

is identified to be the distinctive differences between the categories of test objects. For the majority of room acoustic parameters, the loudspeakers show a different behaviour than that of the artificial speakers and the real speakers. It can be stated that the loudspeakers overestimate the STI in most cases (see Fig. 10). This can be explained by the higher DI of the loudspeakers in the frequency range of 630 to 10 000 Hz (see Fig. 6). According to Bradley et al. [28] the just noticeable difference in the STI is 0.03. Table 2 shows that the mean difference in the STI between the loudspeakers and the real speakers independently of the room and the listener position is greater than 0.1. This means that the difference should be clearly noticeable in listening tests. In Ahnert and Tennhardt [26, p. 196] a table is shown, that defines different ranges for the STI to classify the speech intelligibility. It can be found that the difference between loudspeakers and real speakers can change the rating of the speech intelligibility from "good" to "reasonable". Contrary to this, the artificial speakers show very similar results to the real speakers. The difference between the mean values of the STI ranges from 0.01 to 0.04. This can lead to the thesis that artificial speakers are an adequate replacement for a real speaker and therefore suitable to be used in measurements according to DIN EN 60268-16 [1].
The results of C_{50} support these findings. For this parameter a just noticeable difference of 1.1 dB was found by Bradley et al. [28]. The difference of mean C_{50} between loudspeakers and real speakers ranges from 2.27 to 3.54 dB, whereas the difference between artificial speakers and real speakers is only 0.47 to 1.25 dB. Therefore again, the artificial speakers are in good accordance to the real speakers, while the loudspeakers differ more drastically.
A just noticeable difference of 0.05 for LF is given in DIN EN ISO 3382-1 [24]. This parameter shows an inverse trend compared to the parameters discussed before: The real speakers reach higher values compared to the loudspeakers. This can be explained when the relevant range of the DI from 100 to 1250 Hz in Figure 6 is considered. A lower DI in this range means that the radiation of the source is tending to be more omnidirectional. This means that the lateral sound from a source with a lower DI has more energy than one from a more focused source with a higher DI. It should be noted that strong differences between the different real speakers can be found for this parameter. These differences, to a certain point, can be explained by the different origins of the data. The `ITA Singers` show higher values than the `TU Soprano` and the speakers from `Chu-Warnock`.
The results for *Sound Strength* support the findings that artificial speakers are more suitable for classifying the actual room acoustic conditions caused by a human speaker in a room, whereas the loudspeakers tend to underestimate G. In a more general conclusion, the results of the simulations show that each room acoustic parameter is highly dependent on the room geometries, the structure of the materials and the receiver position. It can be seen that in general, the speech intelligibility is higher in the Theater an der Wien, which is strongly related to the shorter reverberation time of this room. This is supported by the room acoustic parameters, which show a higher STI and a higher C_{50} for this room. As for the influence of the receiver position, it can be said that a greater distance between the source and the receiver leads to a lower STI, a lower C_{50}

and a lower G, but a higher LF.
The results in section 3.4 demonstrate that applying a vertical angle of 30° to the loudspeakers so they are tilted slightly downwards can have a certain influence on the resulting room acoustic parameters. The STI is altered quite drastically by this: Depending on the receiver position and the room geometry it can be reduced or even increased. For Receiver 1 in the Concertgebouw the mean STI of the four loudspeakers is reduced by 0.08 which brings it closer to the STI of the artificial speakers. At Receiver 2, the mean STI is increased by 0.06, whilst in the Theater an der Wien is reduced by 0.06 at both receivers. This means that tilting the loudspeakers can help to obtain results that are closer to the results of an artificial speaker using a loudspeaker. However, this is not given in every room, at every position. Figure 14c and 14d show that C_{50} behaves in a similar manner, but the differences between the original results and the tilted loudspeakers are relatively small. The same applies to LF and G (Fig. 14e - 14h).
One problem that this work reveals is the availability of high quality data from real speakers and singers, as the used data has limitations regarding the grid resolution, frequency range and frequency resolution. Due to this issue it is difficult to determine an exact range of values of the room acoustic parameters (particularly of G and LF) for real speakers. High-resolution directivity measurements of the full sphere would be of interest for further studies and a more detailed comparison to the objects which were measured for this work.
As described in section 2.3, additional `OpenDAFF`-files with low spatial resolution have been created to verify the comparability of data with different measurement grids. After running the simulation with both high- and low-resolution data of the test objects, the resulting room acoustic parameters show only small differences (see section 3.5). The mean and standard deviation of the difference between the high- and low-resolution files are calculated over six objects in the two simulated rooms and at two receiver positions. Table 6 shows these values for each room acoustic parameter.

	STI	$C50$	LF	G
Mean	0.019	0.316	0.016	0.862
Standard Deviation	0.029	0.340	0.027	1.432

Table 6: Mean and standard deviation of the difference between high- and low-resolution data

According to these findings it is possible to compare the simulation results of data from different sources which were measured with different grids, as long as the grid has at least the 36°x36°-(Gaussian-)resolution used for the down-sampling.

5. Conclusion

The objective of this work was to examine the influence of the directivity on room acoustic measurements. For this purpose, the directivity of four small-membrane full-range loudspeakers and two artificial speakers were measured in an anechoic chamber. The resulting data features a high angular grid-resolution and covers the complete frequency range of the human hearing. The measurements were used in a room acoustic simulation, in which they were compared to the directivity data of one more artificial speaker and a total of nine real speakers and singers which were obtained from external sources. In the simulation, two receiver positions for two significantly different rooms were considered. Additionally the data was compared regarding the Directivity Index.
The results show that there are distinctive differences between the loudspeakers on one side and the artificial speakers and real speakers on the other side. The most unambiguous finding is, that the loudspeakers generate higher values of the STI in the simulation than the other objects. This is supported by the results of C_{50}, LF and G which also show similarities between the real and artificial speakers, but fairly large differences for the loudspeakers. Furthermore, a relation to the DI can be found: The loudspeakers show a higher DI and therefore a more focused radiation in a wide frequency range. It is assumed that this is the reason for the differences in the room acoustic parameters. In an additional simulation it was intended to tilt the loudspeakers by 30° to achieve results that are closer to the the artificial speakers (this could be achieved relatively easily during measurements in reality). The results show that this can lead to a STI that is closer to the artificial speakers, but not at every receiver position. The other parameters are only slightly influenced by this procedure.
One problem which became apparent during the study was the data quality of the real speakers. No data was available which matches the high resolution measurement grid that was used for the loudspeakers and that captures the complete frequency range up to 20 kHz. Therefore, an attempt was made to determine the influence of the measurement grid in an additional simulation with spatially down-sampled objects of the high-resolution data which was measured for this study. It could be shown that the influence of the grid resolution in many cases is below the JND, but still some differences are existing. For future research and a more detailed comparison, a better data quality of the human speakers would be helpful.

References

[1] DIN EN 60268-16 (2004): *Elektroakustische Geräte – Teil 16: Objektive Bewertung Der Sprachverständlichkeit Durch Den Sprachübertragungsindex (IEC 60268-16:2003); Deutsche Fassung EN 6026816:2003*. DIN Deutsches Institut für Normung e.V.

[2] Dunn, H. K. and D. W. Farnsworth (1939): "Exploration of Pressure Field Around the Human Head During Speech." In: *The Journal of the Acoustical Society of America*, **10**, pp. 184–199. doi:10.1121/1.1915975.

[3] Flanagan, James L. (1960): "Analog Measurements of Sound Radiation from the Mouth." In: *The Journal of the Acoustical Society of America*, **32**, pp. 913–921. doi:10.1121/1.1936423.

[4] Flanagan, James L. (1972): *Speech Analysis; Synthesis and Perception*. Berlin, Heidelberg, New York: Springer Verlag. doi:10.1007/978-3-662-01562-9. OCLC: 568936170.

[5] Sugiyama, K. and H. Irii (1991): "Comparison of the Sound Pressure Radiation from a Prolate Spheroid and the Human Mouth." In: *Acustica*, **73**, pp. 271–276.

[6] Kob, Malte (2002): *Physical modeling of the singing voice*. Berlin: Logos Verlag. OCLC: 51699861.

[7] Kob, Malte and Harald Jers (1999): "Directivity Measurement of a Singer." In: *The Journal of the Acoustical Society of America*, **105**(2), pp. 1003–1003. doi: 10.1121/1.425813.

[8] Halkosaari, Teemu; Markus Vaalgamaa; and Matti Karjalainen (2005): "Directivity of Artificial and Human Speech." In: *Journal of the Audio Engineering Society*, **53**(7/8), pp. 620–631.

[9] Bozzoli, Fabio and Angelo Farina (2003): "Directivity Balloons of Real and Artificial Mouth Simulators for Measurement of the Speech Transmission Index." In: *Audio Engineering Society Convention 115*. New York, NY, USA.

[10] Katz, Brian and Christophe d'Alessandro (2007): "Directivity Measurements of the Singing Voice." In: *International Congress on Acoustics*. Madrid, Spain.

[11] Chu, W. T. and A. C. C. Warnock (2002): *Detailed Directivity of Sound Fields Around Human Talkers*. Tech. Rep. RR-104, National Research Council Canada. doi:10.4224/20378930.

[12] Monson, Brian B.; Eric J. Hunter; and Brad H. Story (2012): "Horizontal Directivity of Low- and High-Frequency Energy in Speech and Singing." In: *The Journal of the Acoustical Society of America*, **132**(1), pp. 433–441. doi:10.1121/1.4725963.

[13] Pedrero, Antonio; et al. (2012): "Mozarabic Chant Anechoic Recordings for Auralization Purposes." In: *ACÚSTICA 2012 : VIII Congresso Ibero-Americano de Acústica*. Évora, Portugal.

[14] Weinzierl, Stefan; et al. (2017): *A Database of Anechoic Microphone Array Measurements of Musical Instruments*. Tech. rep., Technische Universität Berlin. doi: 10.14279/depositonce-5861.2.

[15] Pollow, Martin; Gottfried Behler; and Bruno Masiero (2009): "Measuring Directivities Of Natural Sound Sources With A Spherical Microphone Array." In: *Ambisonics Symposium*. Graz, Austria.

[16] Wefers, F (2010): "OpenDAFF - A Free, Open-Source Software Package for Directional Audio Data." In: *Jahrestagung Für Akustik DAGA*. Berlin, Germany.

[17] Brinkmann, Fabian and Stefan Weinzierl (2017): "AKtools—An Open Software Toolbox for Signal Acquisition, Processing, and Inspection in Acoustics." In: *Audio Engineering Society Convention 142*. Berlin, Germany.

[18] Berzborn, Marco; Ramona Bomhardt; Johannes Klein; Jan-Gerrit Richter; and Michael Vorländer (2017): "The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing." In: *43th Annual German Congress on Acoustics*. Kiel, Germany, pp. 222–225.

[19] DIN EN 61260-1 (2014): *Elektroakustik – Bandfilter Für Oktaven Und Bruchteile von Oktaven – Teil 1: Anforderungen (IEC 61260-1:2014); Deutsche Fassung EN 61260-1:2014*. DIN Deutsches Institut für Normung e.V.

[20] Rafaely, Boaz (2015): *Fundamentals of Spherical Array Processing*, vol. 8 of *Springer Topics in Signal Processing*. Berlin, Heidelberg: Springer Verlag. doi: 10.1007/978-3-662-45664-4.

[21] Schröder, Dirk and Michael Vorländer (2011): "RAVEN: A Real-Time Framework for the Auralization of Interactive Virtual Environments." In: *Forum Acusticum*. Aalborg, Denmark: European Acoustics Association, pp. 1541–1546.

[22] Ackermann, David; et al. (2018): *A Ground Truth on Room Acoustical Analysis and Perception (GRAP)*. Tech. rep., Technische Universität Berlin. doi:10.14279/depositonce-7003.

[23] Ackermann, David; Christoph Böhm; and Stefan Weinzierl (2018): *A Database on Musicians' Movements During Musical Performances*. Tech. rep., Technische Universität Berlin. doi:10.14279/depositonce-7469.

[24] DIN EN ISO 3382-1 (2009): *Akustik – Messung von Parametern Der Raumakustik – Teil 1: Aufführungsräume (ISO 3382-1:2009); Deutsche Fassung EN ISO 3382-1:2009*. DIN Deutsches Institut für Normung e.V.

[25] DIN EN ISO 3382-2 (2008): *Akustik – Messung von Parametern Der Raumakustik – Teil 2: Nachhallzeit in Gewöhnlichen Räumen (ISO 3382-2:2008); Deutsche Fassung EN ISO 3382-2:2008*. DIN Deutsches Institut für Normung e.V.

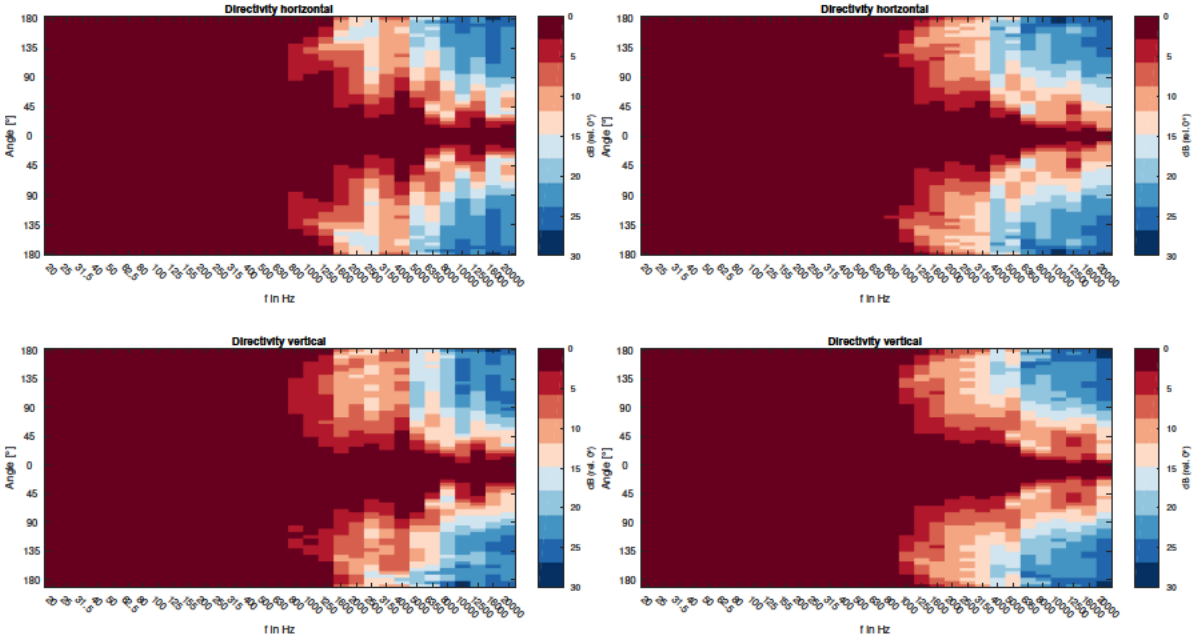
[26] Ahnert, Wolfgang and Hans-Peter Tennhardt (2008): "Raumakustik." In: Stefan Weinzierl (Ed.) *Handbuch der Audiotechnik*. Berlin, Heidelberg: Springer Verlag, pp. 181–266. doi:10.1007/978-3-540-34301-1.

[27] DIN EN 60268-5 (2010): *Elektroakustische Geräte – Teil 5: Lautsprecher (IEC 60268-5:2003 + A1:2007); Deutsche Fassung EN 60268-5:2003 + A1:2009*. DIN Deutsches Institut für Normung e.V.

[28] Bradley, J. S.; R. Reich; and S G Norcross (1999): "A Just Noticeable Difference in C50 for Speech." In: *Applied Acoustics*, **58**(2), pp. 99–108. doi: 10.1016/S0003-682X(98)00075-9.

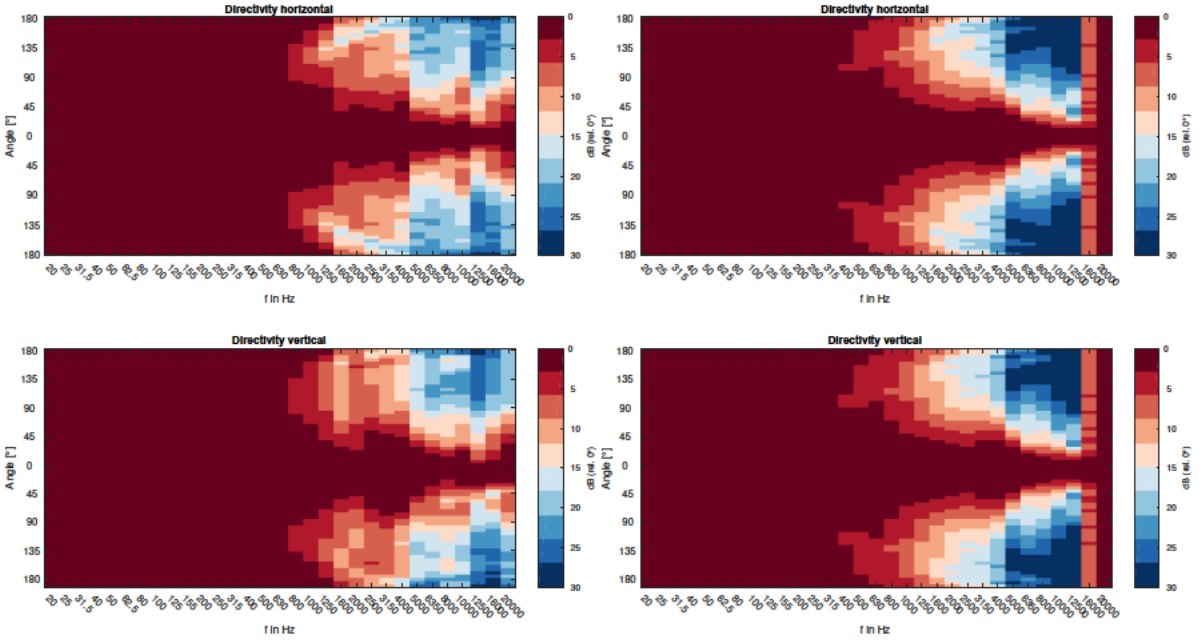
A. Appendix

A.1. Isobar Plots



(a) 6301b

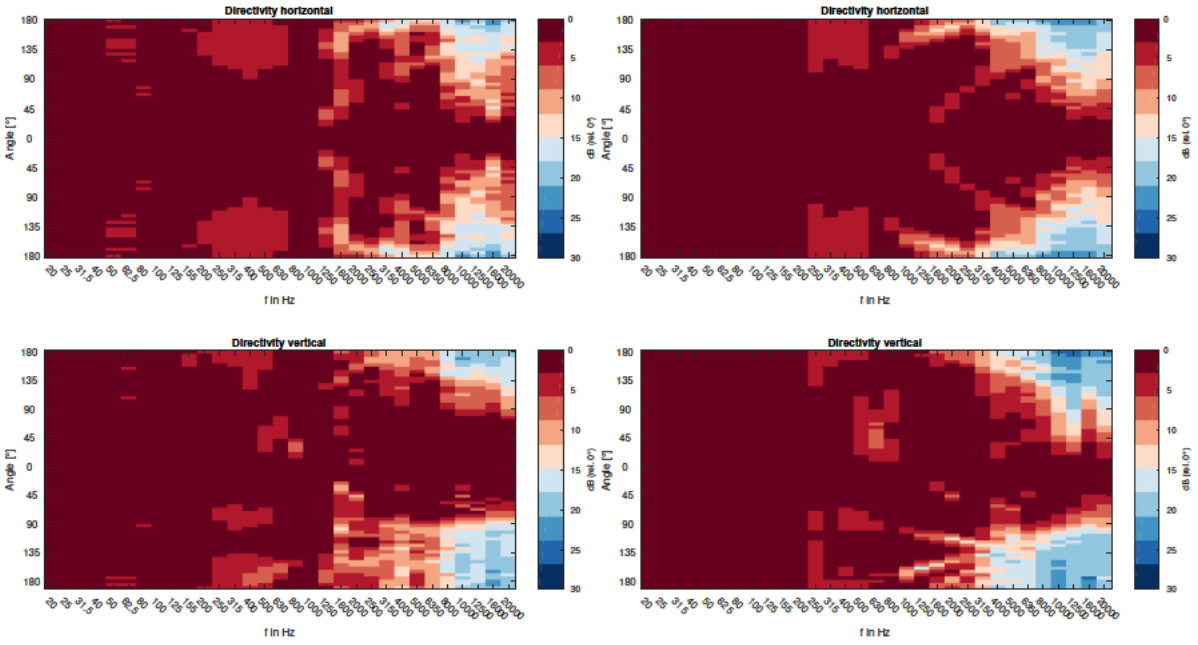
(b) MixCube



(c) MM201

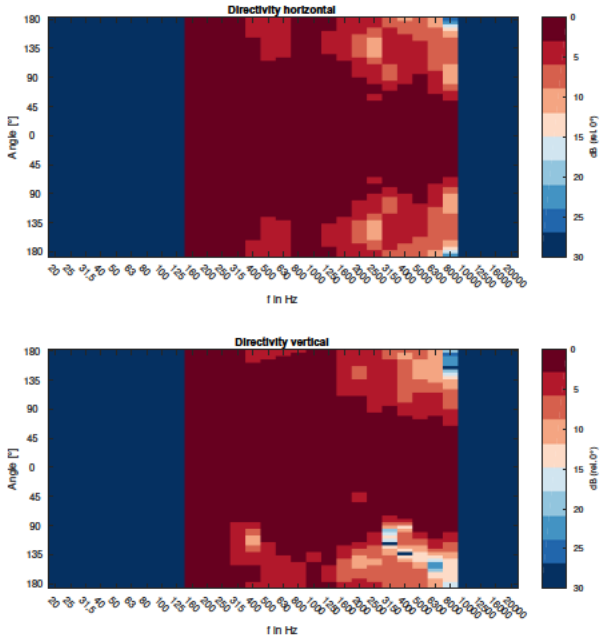
(d) Talkbox

Figure 17: Isobar plots - loudspeakers



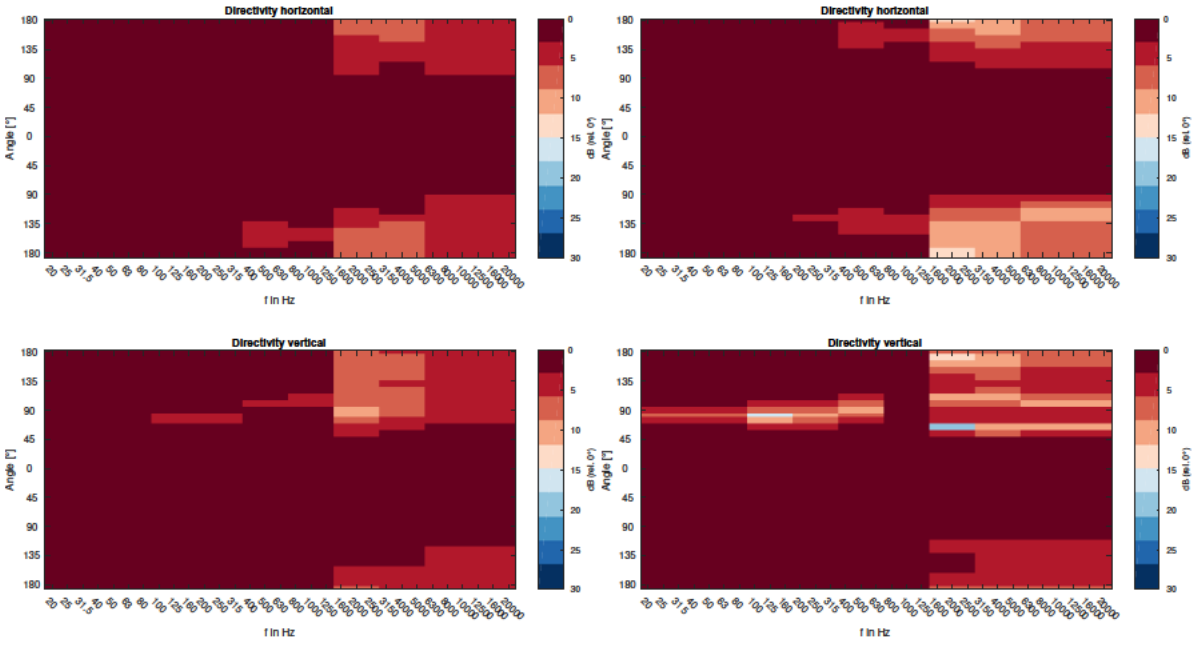
(a) HMS-II

(b) ITA-Head



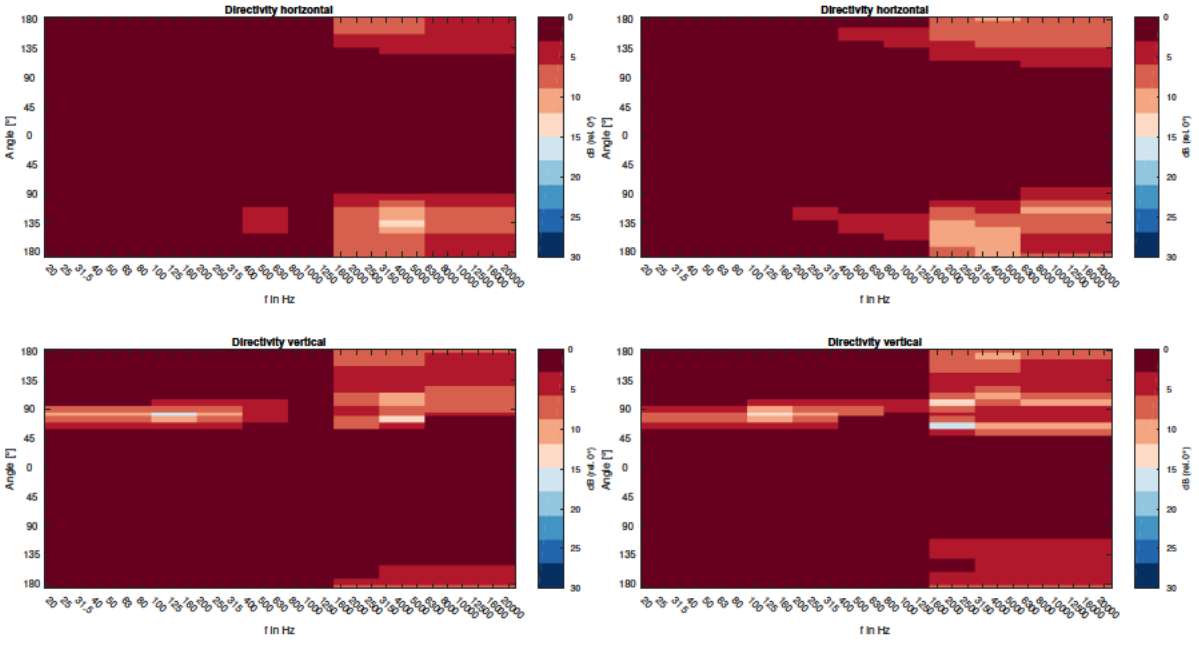
(c) B&K HATS

Figure 18: Isobar plots - artificial speakers



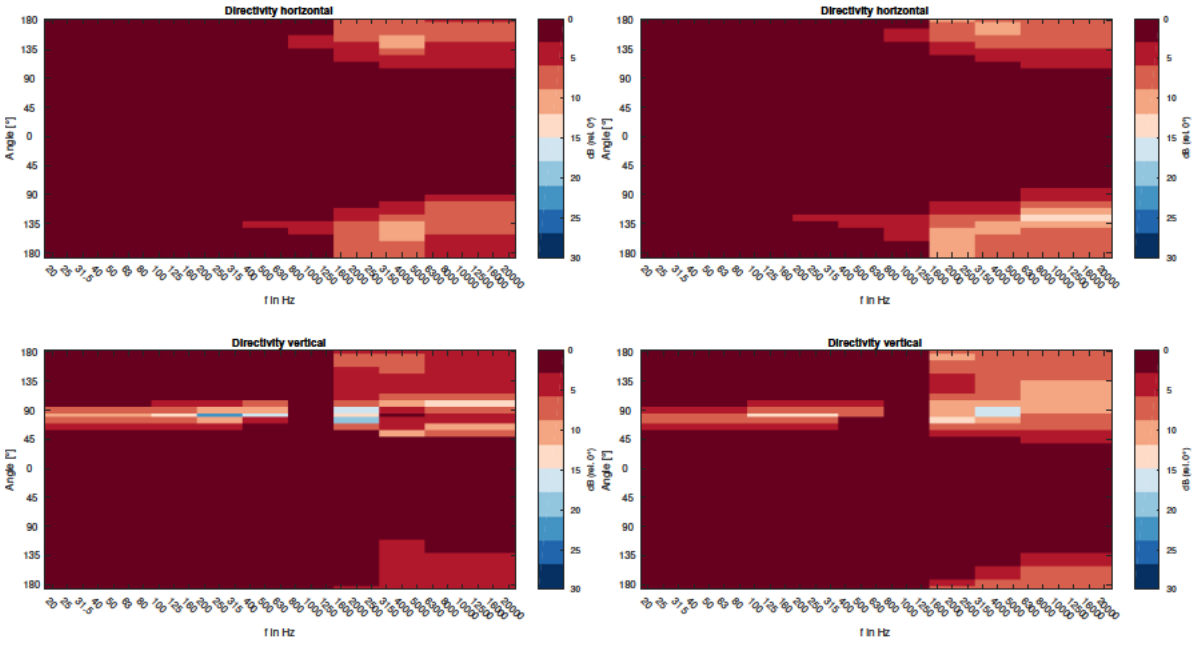
(a) ITA-Singer 1

(b) ITA-Singer 2



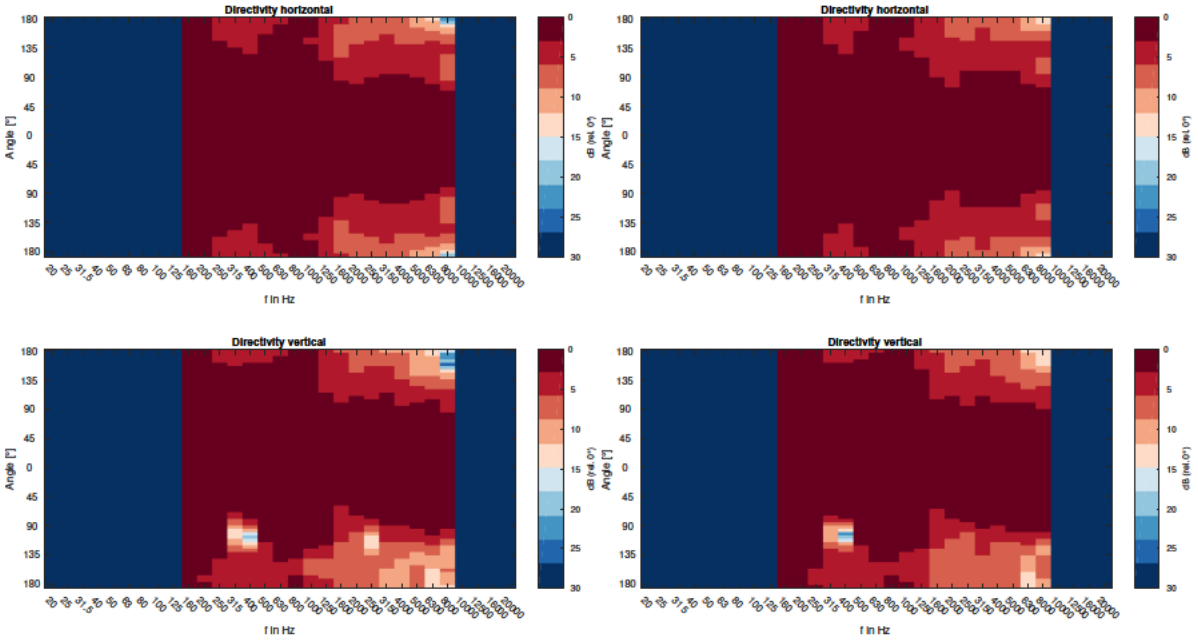
(c) ITA-Singer 3

(d) ITA-Singer 4



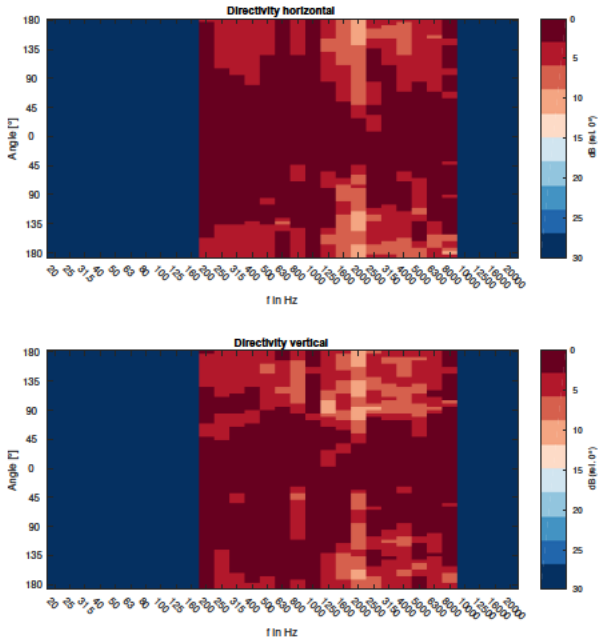
(e) ITA-Singer 5

(f) ITA-Singer 6



(g) Chu-Warnock Male

(h) Chu-Warnock Female



(i) TU Soprano

Figure 19: Isobar plots - real speakers

Figure 17a - 19i shows the horizontal and vertical isobar-plots of the test objects used in this work. The levels are normalized to the 0°-axis.

A.2. Documentation of the Measurement Data

The directivity data measured for this project is prepared for a data publication in the future. The measured impulse responses of the full sphere are presented along with the angles of the measurement grid as comma-separated values (CSV) and **MATLAB** data-files on the attached data-DVD. Figure 20 shows the coordinate convention of the sampling grid used for the measurements. This front pole convention determines that Φ gives the orientation in the y/z-plane (frontal plane), while Θ gives the orientation in the x/z-plane (median plane). The angular resolution of Φ and Θ is 5° which results in 2522 sampling points in total. The on-axis impulse response ($\Phi = 0$ and $\Theta = 0$) is given in the first row of each *.csv- and *.mat-file.

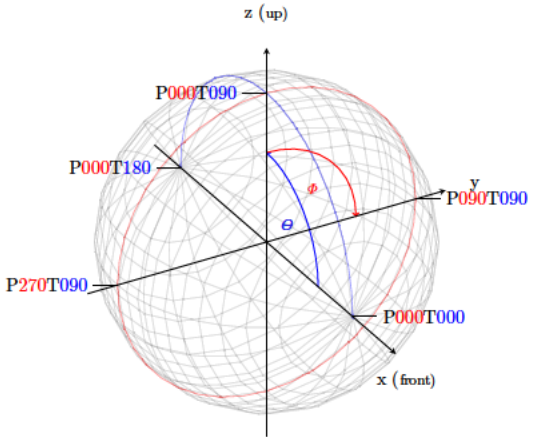


Figure 20: Front pole coordinate convention

The data structure of the attached *.csv-files is given in Figure 21. Each line contains the grid-angles as the first value followed by the 3744 sample values of the impulse response (with a sampling rate of 48 kHz). Figure 22 shows the structure of the corresponding *.mat-files.

```
   1    P000T000,0.0000000,0.0000000,0.0000006,0.0000050,0.0000147,0.0000041,-0.0000091,...
   2    P000T005,0.0000000,0.0000000,-0.0000006,0.0000043,-0.0000017,-0.0000124,-0.0000785,...
   ⋮
2521    P355T175,0.0000000,0.0000000,-0.0000004,-0.0000017,-0.0000031,-0.0000066,-0.0000155,...
2522    P000T180,0.0000000,0.0000000,-0.0000005,-0.0000025,-0.0000073,-0.0000124,-0.0000154,...
```

Figure 21: Data format *.csv-files

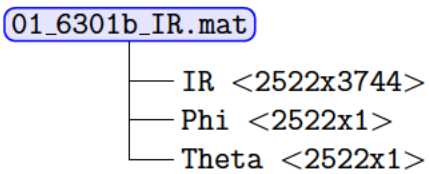


Figure 22: Data format *.mat-files

A.3. DVD Content

The attached data-DVD contains the following folders:

01_Measurement Data and Processing
The raw data is included as *.spk-files as generated at FourAudio. A `MATLAB` script for each measured object is given that is used for the conversion to `OpenDAFF`. Each script can generate one `OpenDAFF`-file containing impulse responses and one containing third-octave band energies (*.mat-files containing the measurement grid and the data are generated additionally).

02_Data_Publication
Contains the data for the publication as described in Section A.2.

03_External Data
Contains the data from the external sources and `MATLAB` scripts to convert it to `OpenDAFF`-files.

04_Simulation
Contains `MATLAB` scripts for the simulation using `RAVEN` and the used data (Room models and `OpenDAFF`-files).

05_Plotting_and_Tools
Contains the `MATLAB` script to generate the plots that are used in this work and additional `MATLAB` scripts for the data processing (e.g. downsampling, calculation of room acoustic parameters).

06_Literature
This folder contains all literature used in this work as full text `PDF` files and the bibliography as a `BIBTEX` file.