Ulrich Schütze

The Head-Vocal-Booth as Acoustic Treatment for Voice Recordings

bachelor's thesis

submitted to

Graz University of Technology

Supervisor

DI Julian Koch BSc.

Signal Processing and Speech Communication Laboratory

Graz, February, 2025

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used.

date	(signature)

Abstract (English)

The head-vocal-booth (HVB) is a rather small box made of sound absorbing material, in which a microphone is meant to be placed. It is marketed as an alternative to conventional acoustic treatment for voice recordings. The aim of this work was to describe the HVB's effect on such recordings with a focus on reverberation, flutter echoes, and colouration due to room modes. To achieve this, the HVB was analysed theoretically and reverberation chamber measurements and a simulated free-field measurement were taken. Although conventional reverberation time parameters do not apply to the HVB, it was shown that the HVB noticeably reduces reverberation. Flutter echos are presumably well attenuated while the effect on colouration due to room modes presumably is mitigated moderately. The HVB also introduces considerable coloration to the recording, especially in the mid and low frequencies.

Abstract (German)

Die Head-Vocal-Booth (HVB) ist eine recht kleine Box aus Absorbermaterial, in der ein Mikrofon platziert wird. Sie wird als Alternative zu herkömmlicher akustischer Raumbehandlung bei Sprach- und Gesangsaufnahmen vermarktet. Ziel dieser Arbeit war es, die Auswirkungen der HVB auf solche Aufnahmen zu untersuchen; mit Fokus auf Nachhall, Flatterechos und Klangfärbung durch Raummoden. Dafür wurde die HVB theoretisch analysiert und es wurden eine Hallraummessung und eine simulierte Freifeldmessung durchgeführt. Obwohl herkömmliche Nachhallzeit-Parameter auf die HVB nicht direkt anwendbar sind, konnte gezeigt werden, dass sie den Nachhall merklich verringert. Flatterechos werden voraussichtlich deutlich abgesenkt, während der Effekt auf die Klangfärbung durch Raummoden als moderat einzustufen ist. Gleichzeitig führt die HVB zu einer deutlichen Klangfärbung, besonders im mitten- und tieffrequenten Bereich.

Contents

1 Introduction

The world of acoustic media production has changed greatly over the past few decades, which can be attributed in great parts to advancements in personal computing, digital signal processing and the internet. Until the 21st century, most professional work in the music industry and the audio section of the film and entertainment industry happened in music studios. However, through the shift towards computer based sound production and the availability of affordable sound production gear, semi-professionals and professionals alike have moved at least parts of their work into their homes where many have constructed so-called "home studios" or "bedroom studios" [1]. These home studios are seldom purpose-built to be used for audio applications, which often poses a particular challenge when it comes to acoustic treatment.
Acoustic treatment is important for monitoring as well as recording. One can avoid the need for acoustic treatment in monitoring by using headphones and when recording electronic or electroacoustic instruments directly into a recording interface, room acoustics do not have any effect. However, there is one important element that has to be recorded using a microphone and is, therefore, always affected by room acoustics: Vocals. Whether for singing or for spoken word, a dry vocal recording is preferred in most cases. Meaning, the recording should not contain effects such as reverberation, flutter-echoes, and colouration due to room modes; effects that are often prevalent in living spaces, due to a lack of appropriate acoustic treatment.
Since conventional acoustic treatment using absorbers and diffusors along the walls and ceiling can be expensive, alternatives aimed specifically at the case of recording voices have emerged. One of these alternatives is the head-vocal-booth (HVB), as it will be called in this work. A HVB is a small box made of absorbing material. A performer can insert their head through an opening in the box and the performance, usually singing or spoken word, is recorded with a microphone inside the box (see Figure 2.2). This device is supposed to isolate the performance from the room. Whilst eliminating room-reflections and their resulting effects in the recording, the HVB is also meant to reduce background noise and improve signal-to-noise ratio. Furthermore, it is supposed to limit noise emission into neighboring areas, which can be a problem when recording loud performances in particular.
It is the aim of this work to investigate the effectiveness of the HVB as a form of acoustic treatment in regard to reverberation, flutter-echoes, and colouration due to room modes. Meanwhile, possible side-effects are considered. In order to evaluate the HVB in this manner, it is analysed theoretically as well as practically in the form of measurements.

2
Theoretical Analysis

In order to predict how the HVB behaves when used in vocal recordings, properties of the human voice as the main sound source have to be considered. Then, three acoustical problems that are often encountered in untreated rooms are presented, after which the HVB itself is analysed.

2.1 Properties of the Human Voice

The human voice is a complex system that produces speech or singing in a two stage process. The sound source (glottis) produces a sound that is then filtered by the vocal tract. The resulting sounds and noises form three main groups: voiced sounds, unvoiced sounds, and plosives. These sounds mainly occur in the frequency range between 100 Hz and 4 kHz with little energy outside of these bounds. The region around 2- to 3 kHz is particularly important because voice sounds resonate in this region [2]. Apart from frequency response, source directionality is an important factor in recordings. For higher frequencies, the directionality of the human voice is similar to a cardioid pattern due to the acoustic shadow of the head. For lower frequencies, diffraction around the head becomes more pronounced and sound radiation is almost spherical [3]. As Figure 2.1 shows, the mentioned pattern is true for the horizontal plane (2.1b) as well as the vertical plane (2.1a) with the exception of the 270° region due to the acoustic shadow of the body. In all further considerations, the voice will be approximated as a spherical sound source for the entirety of the relevant frequency spectrum.

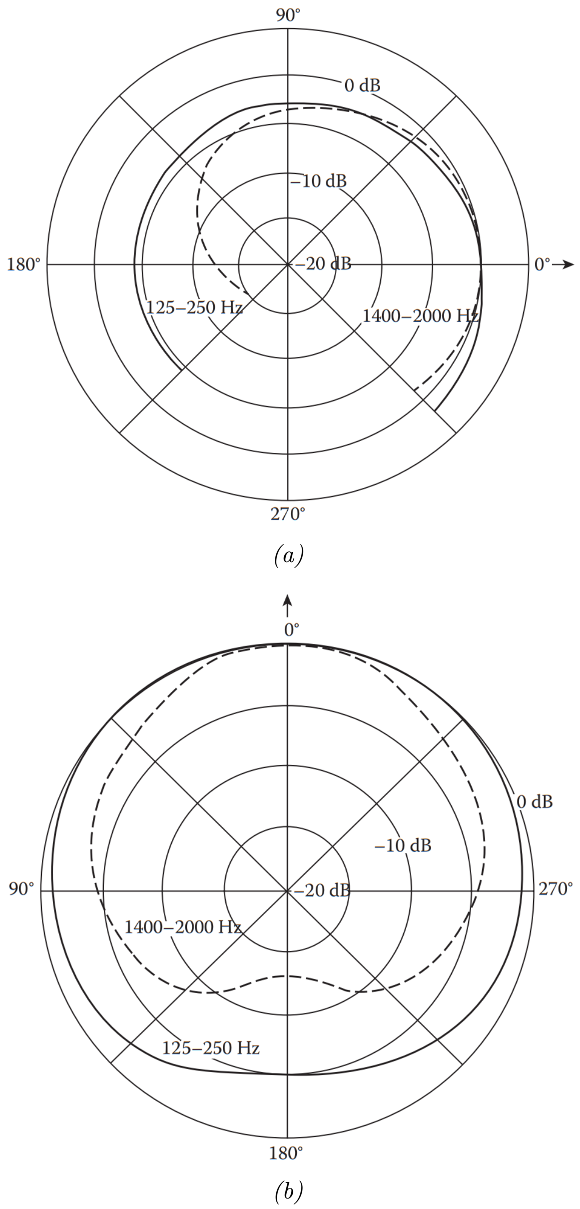


Figure 2.1: Directionality of the human voice in the vertical plane (a) and in the horizontal plane(b) [3].

2.2 Common Acoustical Problems

The following acoustical problems are commonly found in untreated living spaces.

2.2.1 Reverberation

Reverberation is the decay of sound inside a room after excitation stops. It is defined as the time it takes the sound pressure level to decrease by 60 dB after the excitation source in a stationary sound field has been turned off. This time is called reverberation time (RT) (see Section 3.2.2). Reverberation is caused by sound reflections at the boundaries of a room. RT can also become longer because of resonances or room modes [4]. For conventional vocal-booths, RTs of 70 ms to 0.3 s are aimed for. These very low RTs are especially sought after for voice recordings in motion picture productions like voice-overs [5].

2.2.2 Flutter-Echoes

Flutter-echoes occur if sound bounces between two parallel reflective boundaries. The unique sound of this effect is unwanted in nearly every situation. Spectrally, it is located mainly in the 1-2 kHz region, which makes it highly relevant for vocal recordings [6].

2.2.3 Colouration due to Room-Modes

This effect is present at low frequencies, and also depends on the position of the microphone inside the room. Depending on the position relative to the wavelength of the resonating mode-frequency and the type of the mode (axial, tangential, oblique), either peaks or dips in the frequency response can occur. Below the Schroeder frequency f_s mode density is low and individual cancellation effects or resonance peaks can be measured or perceived. Above f_s, mode frequencies are too close together and do not cause the same narrow-band effects. The Schroeder frequency f_s is defined as

$$f_s = 2000 \cdot \sqrt{\frac{RT}{V}}$$

where RT is the room's reverberation time and V is its volume. In large rooms or halls, f_s is typically below the relevant frequency range for voices [7]. However, in living spaces such as bedrooms and living rooms, f_s may sit well within the range of speech and singing, especially if the space is not acoustically treated. For example, in a room with an area of $20\,\mathrm{m}^2$ and a height of 2.6 m with RT of 0.6 s, f_s would be 215 Hz.

2.3 Head-Vocal-Booths

As mentioned in the introduction, a HVB is a box made of absorber material that is supposed to improve the quality of voice recordings by eliminating unwanted acoustic phenomena. Four different products from two different manufacturers are currently available on the market. For this work, the t.akustik Vocal Head Booth (Figure 2.2a) serves as the base of all considerations. The following analysis including the measurements, results, and discussion refer to the t.akustik Vocal Head Booth only. It is, however, presumable that much of the discussed properties would be similar for other HVBs due to similar dimensions and similar used materials.

This specific HVB has a simple cuboid-like shape and is made from only one type of homogeneous absorber material of a constant thickness, which makes it well suited for a general analysis of HVBs as a concept.

(a) t.akustik Vocal Head Booth [8] *(b) Isovox Mobile Vocal Booth V2 [9]*

Figure 2.2: Two Head Vocal Booths from two competing manufacturers. t.akustik Vocal Head Booth (left) with the back-piece removed, Isovox Mobile Vocal Booth V2 (right) with one side-piece removed and the back-piece flipped up.

2.3.1 Physical Properties

The t.akustik Vocal Head Booth's base shape is that of a cuboid with a cutout for the performer's head and torso. All walls are made from 40 mm thick PET (Polyethylene terephthalate) absorber material and the booth is set on a base-plate made from MDF (Medium-Density Fibreboard) wood.
Figure 2.3 shows the inner dimensions of the HVB. Because of its dimensions, the HVB only allows for a certain range of distance from the microphone with the maximum being around 25 cm, depending on the mounting and shape of the microphone. This almost entirely disqualifies the HVB as an alternative to conventional acoustic treatment for situations where greater distances between performer and microphone are preferred. By removing the back-piece, the performer could theoretically reach a greater distance from the microphone. Since this is an edge-case and it is not the intended use of the HVB, this case is not part of any considerations in this thesis. Furthermore, the booth somewhat restricts freedom of movement of the performer, which could cause discomfort or hinder expressiveness in artistic performances.

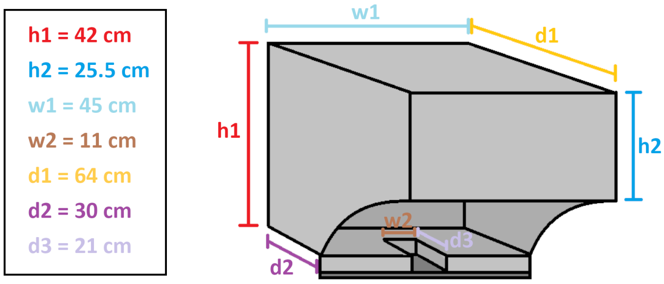


Figure 2.3: Inner dimensions of the t.akustik Vocal Head Booth

2.3.2 Acoustic Properties

The only information on the absorbing properties of the PET absorber material used in the HVB that is available online is the statistical absorption coefficient, also known as random-incidence absorption coefficient, of the same material of a different thickness (50 mm as opposed to 40 mm in the HVB) [10].

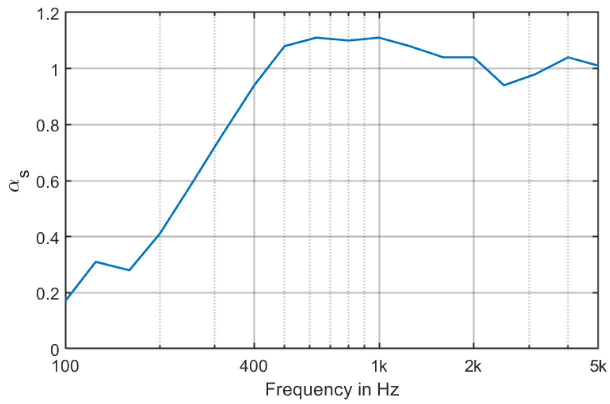


Figure 2.4: Statistical absorption coefficient α_s over frequency for the same PET absorber material found in the HVB, but at a larger thickness of 50 mm [10].

Figure 2.4 shows these absorption values plotted over frequency. For the material used in the HVB even less absorption in lower frequency bands is to be expected. In theory, however, the statistical absorption coefficient, derived from reverberation chamber-measurements, does not hold much value in this case because the absorber is very near to the source. The fact that the absorber is in the near-field of the source for a large section of the relevant frequency range (100 Hz - 4 kHz) poses a potential problem of the HVB. For spherical waves, the Helmholtz number kr indicates whether a near-field or a far-field consideration is appropriate. k is the angular wavenumber given by $k = \frac{\omega}{c}$, where ω is the angular frequency, c is the speed of propagation, the speed of sound in this case, and r is the distance from the source.

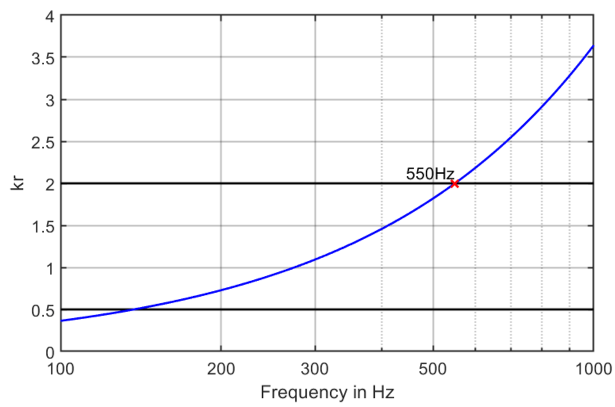


Figure 2.5: kr over frequency f at a distance from the source r = 20 cm. The two thick black lines indicate the borders between near-field, transition region, and far-field (from bottom to top)

Figure 2.5 shows kr for r = 20 cm, which is the worst case for distance between mouth and absorber material inside the HVB for a correct setup (mouth is centered between the walls and pointing at the microphone). The start for the far-field region is usually seen as kr = 2, which corresponds to a frequency of 550 Hz. This means that for most fundamental frequencies in speech and a large portion of fundamental frequencies in singing, especially for male voices, the absorber material lies in the near-field, or in the transition region of the sound source. As a consequence, the often used approximation of plane wave fronts is not applicable. When analyzing the reflection and absorption of spherical waves above an absorbing plane, the angle of incidence along the absorbing plane varies depending on the distance from the source. This is relevant because the absorber material of the HVB is homogenous and non-locally-reacting. Such materials possess reflection coefficients which depend on the angle of incidence [11]. Extensive analytical analysis of this topic has been conducted and summarized by Mechel in his German book "Schallabsorber, Band 1" (see [11]). Dragonetti and Romano compared different approximations for the same setup in their work using FEM simulation [12]. A more vivid expression of the difference between plane wave propagation and spherical wave propagation becomes obvious when looking at source impedance. The expression for spherical wave impedance

$$\underline{Z}_{sph} = Z_0 \frac{k^2 r^2 + jkr}{1 + k^2 r^2}$$

where Z_0 is the specific impedance of air, k is the angular wavenumber and r is the distance from the center of the source, shows how $\underline{Z}_{sph}$ becomes smaller for lower values of kr and, therefore, at a constant distance, for lower frequencies. This results in greater impedance mismatch at the surface of the absorber and a higher complex reflection coefficient, compared to plane wave incidence [12]. This should, in theory, lead to a build-up of energy at frequencies of about 550 Hz and below inside the HVB.

Another property that presumably holds significant acoustic importance is the hole in the cuboid box that is the HVB. Analysis of the effects of this hole would be very complex because the effects are probably very frequency dependent and the acoustic properties of the performer’s torso partly blocking the hole would have to be taken into account.

To examine the mentioned aspects and hypotheses, simulation using the finite element method (FEM) could be helpful and interesting.

3 Measurements

This chapter describes the measurements conducted in order to check the validity of the theoretical analysis. First, the unique concept of the measurement is presented and then results are shown and described.

3.1 Methodology

Since no measurements of a HVB can be found on the internet or elsewhere, the goal was to gather as much data as possible from one measurement setup. Therefore, both a reverberation chamber (RC) measurement and a free-field measurement were taken at the same time. As the location, the RC at Graz University of Technology was chosen. The free-field measurement was achieved by truncating the impulse response (IR) before the first reflection. According to the formula

$$\Delta f = \frac{c}{\Delta d} = \frac{344\,\mathrm{m/s}}{4.7\,\mathrm{m}} = 73.19\,\mathrm{Hz}$$

where c is the speed of sound and Δd is the difference in distance travelled between the direct sound and the reflected sound, maximum frequency resolution Δf would amount to around 73 Hz with the height of the RC (4.9 m) being the limiting factor. Parameters like the EDC (energy decay curve) or T_c (center time) could be computed by using the full IR.
The primary function of the HVB is to eliminate room reflections and reverberation. However, RT as a parameter is not a good fit for describing the HVB because it is defined for one single room and the situation with the HVB present can be approximately interpreted as two coupled rooms (see Sections 3.3.2, 3.3.3). T_c as a parameter for the reduction of the influence of the room and speech-transmission-index (STI) as a measure of speech intelligibility were also considered.

Measurements were conducted on October 22, 2024, with the help of supervisor DI Julian Koch BSc. Inside the RC, the temperature was 21.6°C and the humidity was 56.3% (measured with permanently installed equipment). The following equipment was used in the measurements:

Device	Exact Name
DUT	t.akustik Vocal Head Booth
Source 1	Genelec 8020c
Source 2	HEAD acoustics HMS II.3
Sink 1	NTi Audio M2230
Sink 2	Austrian Audio OC818
Audio Interface	RME Fireface UCX
Measurement System	Windows Laptop with ITA-toolbox in MATLAB
Accessories	Microphone Stands, Speaker Stands, Cables

Table 3.1: Equipment used in measurements of the HVB

As apparent in Table 3.1, two different sinks and sources were used respectively. The small loudspeaker Genelec 8020c (hereafter called "loudspeaker") was chosen for its rather flat frequency response in the relevant spectrum, whereas the HEAD acoustics HMS II.3 artificial head with artificial mouth (hereafter called "artificial head") should approximate the radiation characteristics of a human mouth and the resulting impedance. As sinks, the NTi Audio M2230 omnidirectional measurement microphone (hereafter called "measurement microphone") was chosen for its neutral frequency response, while the Austrian Audio OC818 studio condenser microphone (hereafter called "studio microphone") should represent a more realistic use case of the HVB. Additionally, the OC818 allows for retroactive changes to the polar-pattern. All combinations of the sinks and sources above were conducted with the HVB and without the HVB in order to allow for calculation of the differences between these two states since this, in theory, eliminates all other influences and extracts only the effect of the HVB.
In all measurements, the microphone sat at a height of 2 m. For all setups that include the loudspeaker, there was a distance of 25 cm between the loudspeaker and the microphone membrane. The artificial head was placed 20 cm away from the microphones. Figure 3.1 shows photos from two of the eight measurements. All measurements used an exponential sweep from 50 Hz to 10 kHz.

(a) Measurement microphone, artificial head with HVB

(b) Studio microphone, loudspeaker without HVB

Figure 3.1: Photos taken during measurements in the reverberation chamber. The back-plate of the HVB was only lifted up for the photo; during measurements it stayed down in its original position.

3.2 Acoustic Parameters

A number of different parameters were chosen to evaluate the measured IRs. These parameters are meant for use in rooms or halls and are typically measured by averaging over at least two source-positions and six microphone-positions spread out across the room.

3.2.1 Energy Decay Curve EDC

The energy decay curve describes the decrease of energy inside a room after its sound field has reached a steady state and the excitation source has been turned off. It is calculated by Schroeder-backwards-integration:

$$R(t) = \int_t^\infty p^2(\tau)d\tau$$

where $R(t)$ is the EDC and $p(t)$ is the pressure amplitude response, in this case, the IR. The EDC is essential for further calculating reverberation time parameters.

3.2.2 Reverberation Time $RT60,\ RT30,\ RT20,\ EDT$

Reverberation time is a measure for the duration of the decay of sound inside a room with a diffuse sound field. It is commonly calculated from the EDC and typically describes a decrease of 60 dB, measured from -5 dB to -65 dB ($RT60$). In practice, $RT30$ and $RT20$ are more common, because they require less SNR (Signal-to-noise-ratio) with the relevant ranges being -5 dB to -35 dB ($RT30$) and -5 dB to -25 dB ($RT20$). The resulting times are then doubled or tripled respectively, to make them comparable. EDT (Early Decay Time) is evaluated from 0 dB to -10 dB and then multiplied by 6. Due to these characteristics, none of these RT parameters could be used for HVB measurements. The short distance between source and sink and the resulting EDC called for a different form of reverberation analysis as explained in Sections 3.3.2 and 3.3.3.

3.2.3 Center Time T_c

T_c marks the x-coordinate of the geometric center of the squared IR. It is calculated using the formula

$$T_c = \frac{1}{W_{total}} \int_0^\infty t \cdot p^2(t)dt$$

where T_c is the center time, W_{total} is the total energy of the signal and $p(t)$ is the pressure amplitude response; the IR in this case.

3.2.4 Speech Transmission Index STI

STI is a common measure used to quantify the intelligibility of speech. It describes the reduction of modulation-depth in the speech envelope caused by reverberation and background noise. Originally it was measured using a noise signal with a frequency spectrum similar to speech and sinusoidal modulation of the intensity envelope [13]. In this application, STI is estimated from the IR using the so-called indirect method. In the ITA-Toolbox, this is done in accordance with ISO 60268-16 [14].

3.3 Results

Measurement analysis and visualisation took place in MATLAB, using the ITA-toolbox [14]. The results can be split into two groups: Results from the free-field analysis are presented first, with frequency response being the main focus. Afterwards, results from the RC analysis, including reverberation, center-time T_c, and speech transmission index STI, are shown.

3.3.1 Free-Field Frequency Response

First, the IRs were truncated so that they only include direct sound. As Figure 3.2 shows, direct sound is recorded 49 ms after the start of the measurement and the first reflection appears at the 60 ms mark. Therefore, the start point was set to 45 ms in order to include the full startup-transient and IR was truncated at 59.7 ms, which gives a time-delta of 14.7 ms, which in turn results in a frequency resolution $\Delta f = \frac{1}{\Delta t} \approx 68$ Hz. This value is very close to the approximation in Section 3.1.

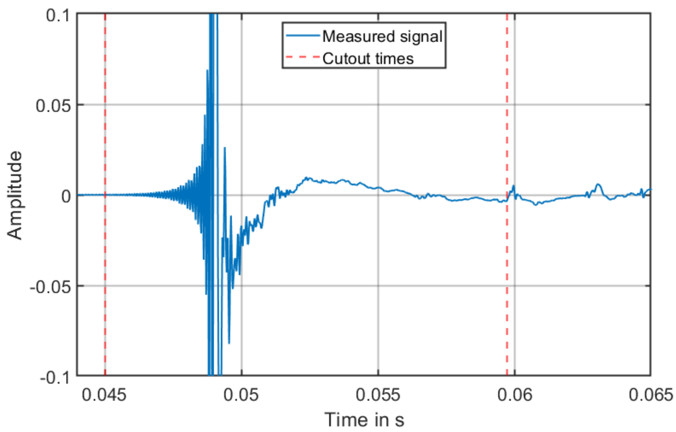


Figure 3.2: Start of impulse response measured in reverberation chamber with Measurement microphone and loudspeaker, but without HVB. Cutout times are marked.

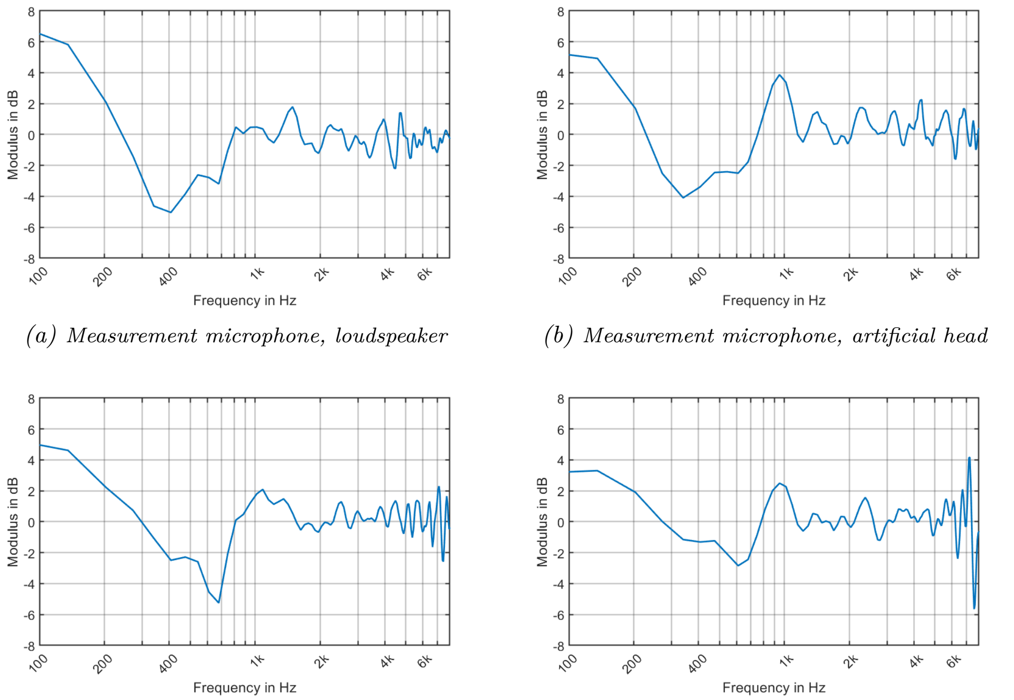
These truncated IRs simulate a free-field environment and allow for an analysis of the effect the HVB imposes on a recording in the frequency domain, without the effect of room-reflections. To further eliminate influences from the used microphones and loudspeakers, differences between the truncated IR with the HVB and the truncated IR without the HVB were calculated for every source-sink combination in the form of spectral division:

$$H(f) = \frac{B(f)}{E(f)}$$

where $H(f)$ is the resulting free-field frequency response of the HVB, $B(f)$ is the frequency response of the whole measurement system with the HVB, and $E(f)$ is the frequency response of the measurement system without the HVB. This calculation returns a frequency response that describes the HVB in the given setup. Figure 3.3 shows all four resulting frequency responses for all four combinations of sources and sinks. Although a frequency range from 100 Hz to 4 kHz would be sufficient as discussed in Section 2.1, the plots of Figure 3.3 show values for up to 8 kHz because the data was available.

All four IRs have a few things in common: A rather large boost below around 230 Hz,

a dip from around 230 Hz to about 800 Hz, and all except 3.3a feature a peak around 1 kHz. This response is easily explained when taking into account the reflective properties of the HVB's walls and the HVB's dimensions. When considering the sidewalls, for example, one can calculate as follows: The difference in path length between the direct path and the reflected path is around 25-30 cm (cf. Figure 2.3), depending on the setup, which leads to maximum cancellation at 573-688 Hz. At low frequencies, reflections arrive almost in-phase which leads to constructive interference.
In measurements taken with the studio microphone, the low-frequency boost is less pronounced than in measurements with the measurement microphone. Another difference between the two microphones is the location of the biggest dip, which is 330 Hz for the measurement microphone and 630 Hz for the studio microphone. These differences and general variability across the four setups can be attributed to different directivity characteristics of the sinks and sources and to slightly different sink-source-distances (20 cm for the loudspeaker, 25 cm for the artificial head).



(a) Measurement microphone, loudspeaker

(b) Measurement microphone, artificial head

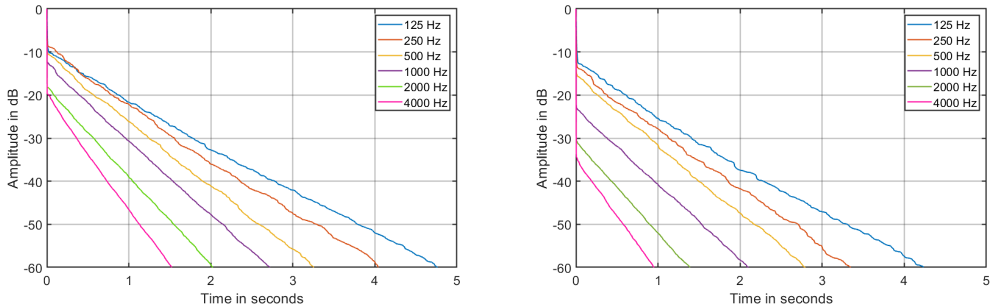
(c) Studio microphone (Cardioid), loudspeaker

(d) Studio microphone (Cardioid), artificial head

Figure 3.3: Effect of HVB on frequency response for all four combinations of sources and sinks. The graphs depict the magnitude of a frequency response acquired by spectral division: $\frac{B(f)}{E(f)}$, where $B(f)$ is the FFT of the truncated impulse response measured with the HVB, and $E(f)$ is the FFT of the truncated impulse response measured without the HVB. The studio microphone was set to a cardioid polar-pattern.

3.3.2 Energy Decay Curve *EDC*

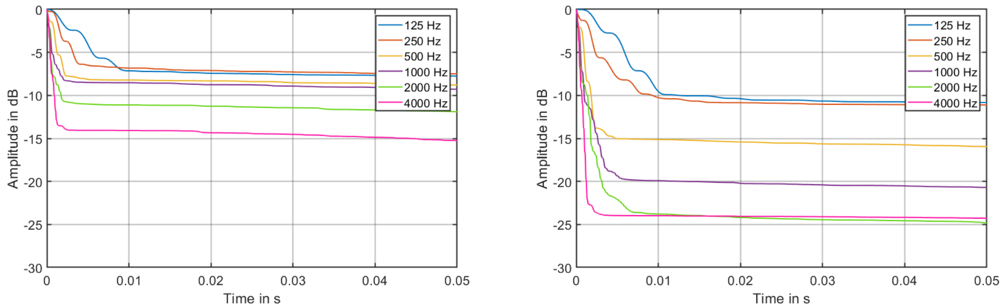
Figure 3.4 shows the *EDC* for the case of the studio microphone and loudspeaker, both with and without the HVB. The *EDC* in Figure 3.4a is as expected from a RC-measurement, except for the immediate decrease of 10 dB (and more for higher frequencies) in the first milliseconds. This is probably a result of the small distance between source and sink.
A decrease in level ranging from 3 dB for the lowest band to 15 dB for the highest



(a) EDC without HVB *(b) EDC with HVB*

Figure 3.4: Energy Decay Curve EDC using the studio microphone (cardioid) and the loudspeaker without the HVB (left) and with the HVB (right). The EDC is evaluated in octave-bands from 125 Hz to 4 kHz.

band due to the HVB is apparent when comparing to the *EDC* in Figure 3.4b. While the HVB seems to attenuate reverberation over all considered frequencies, it does so a lot more effectively at frequencies over 500 Hz. Unexpectedly, there seems to initially be more energy in the 250 Hz band than in the lower 125 Hz band at the very start of the decay curve. This is also true for the other setups. The big initial drop in level makes it obvious that consideration of conventional RT parameters is pointless (cf. Section 3.2.2).



(a) EDC without HVB *(b) EDC with HVB*

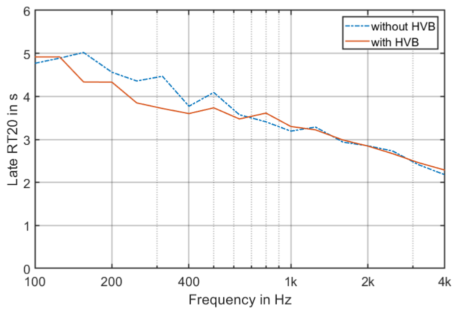
Figure 3.5: First 50 ms of the Energy Decay Curve EDC using the studio microphone (cardioid) and the loudspeaker without the HVB (left) and with the HVB (right). The EDC is evaluated in octave-bands from 125 Hz to 4 kHz.

Figure 3.5 suggests that the HVB inside the RC can be viewed as two coupled spaces. The 250 Hz octave shows a more gradual transition into the long decay of the RC when the HVB is used (Figure 3.5b), which can be interpreted as multi-exponential decay; a superposition of the decay inside the HVB and the decay of the RC. This behaviour is another reason, why conventional RT parameters are not applicable for the HVB. A thorough analysis of this concept and its implications can be found in Balint's work [15].

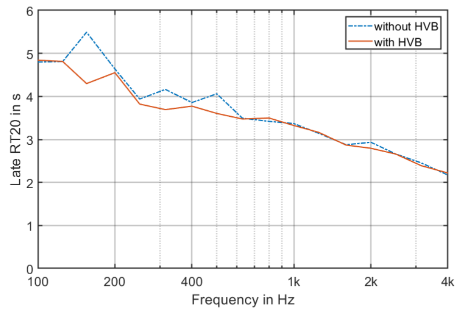
3.3.3 Reverberation

As mentioned, consideration of conventional *RT* parameters is not appropriate for the case of the HVB. Figure 3.4 shows that rather than shortening the length of the decay, the HVB reduces the overall level of the decay. The decay curves are

shifted towards lower dB values while their slope almost stays the same, especially for higher frequencies (see Figure 3.6).
Figure 3.7 shows values at which the regression line used for late $RT20$, which describes a drop of 20 dB starting from the EDC value at 40 ms, would intersect the y-axis for each third-octave band. 40 ms was chose as a starting point because it is the point where the EDC has reached the region of decay attributed to the RC in all third-octave frequency bands. This measure was named reverberation level and it was calculated for the four combinations of sources and sinks, with two additional plots showing the results for the studio microphone with an omnidirectional polar-pattern. This was, however, acquired through the same measurements with the polar-pattern changed retroactively.
When the HVB is used, a clear level-decrease in reverberation is apparent over all frequencies and source-sink combinations. While the achieved values vary substantially over all four setups, attenuation of reverberation generally becomes higher as frequency rises. This is probably due to better absorption of the HVB at higher frequencies. Overall differences can be attributed to the different dispersion characteristics of the used sources and varying directivity of the sinks.

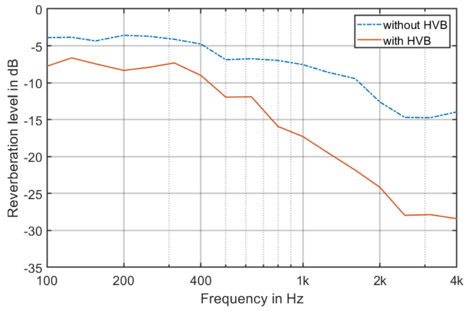


(a) Measurement microphone, loudspeaker

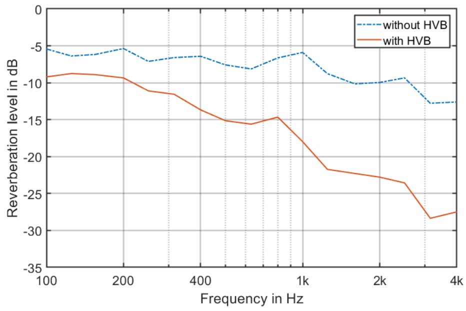


(b) Studio microphone (cardioid), loudspeaker

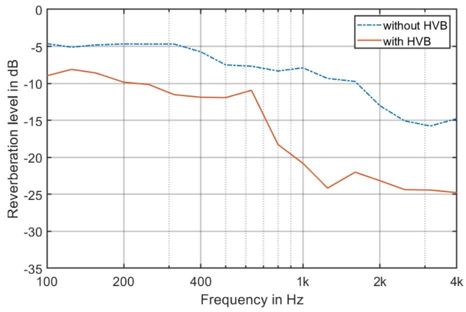
Figure 3.6: Late reverberation time $RT20$, taking into account a level-drop of 20 dB in the respective third-octave band of the EDC starting from 40 ms after the start of the EDC. Late $RT20$ was calculated without the HVB (blue) and with the HVB (orange). Used equipment were the loudspeaker, the measurement microphone (left), and the studio microphone (right).



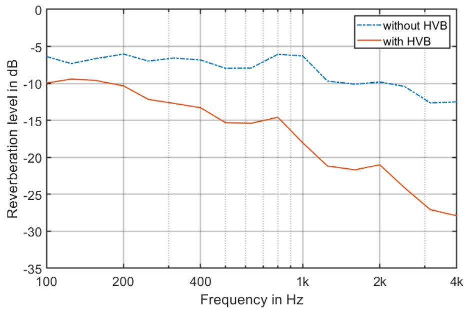
(a) measurement microphone, loudspeaker



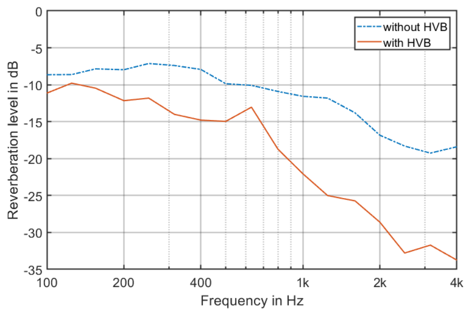
(b) measurement microphone, artificial head



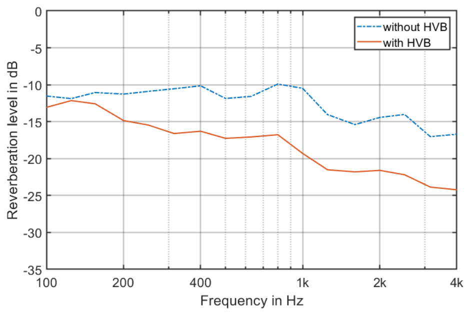
(c) studio microphone (omni), loudspeaker



(d) studio microphone (omni), artificial head



(e) studio microphone, loudspeaker



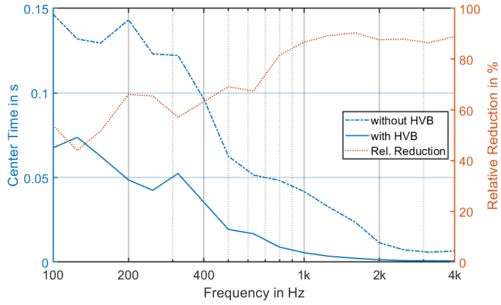
(f) studio microphone, artificial head

Figure 3.7: Reverberation level, which is the theoretical level of the EDC at 0 s, extrapolated from the regression lines used to calculate late RT20 (cf. Figure 3.6) without the HVB (blue) and with the HVB (orange) for all four setups. EDCs were calculated in third-octave bands.

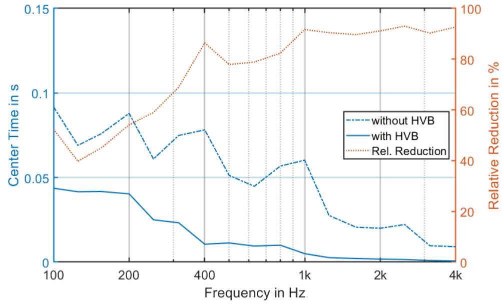
3.3.4 Center Time T_c

Figure 3.8 shows results for T_c with, and without the HVB. The values, especially those without the HVB, are a lot smaller than one would expect from RC-measurements according to EN ISO 354 [16]. This is again a result of the small distance between the sound source and the microphone, which is not in accordance with EN ISO 354. Unlike conventional RT parameters, however, T_c is more robust concerning the shape of the EDC. Therefore, consideration of T_c still holds value in the case of the HVB.
Figure 3.8 consists of 6 plots showing results for the four combinations of sources and sinks, again with two additional plots showing the results for the studio microphone with an omnidirectional polar-pattern. The change in polar-pattern brings the results for the studio microphone closer to the results of the measurement microphone, which is omnidirectional as well. From this, it can be presumed that

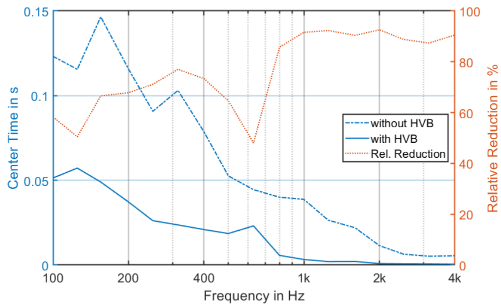
the difference between plots 3.8e and 3.8f and the other four plots stems from the difference in polar-pattern: A cardioid polar-pattern seems to result in a shorter T_c because it attenuates sound from all directions, except for on-axis incidence. The cardioid polar-pattern further seems to result in a lower reduction of T_c.
In general, the reduction in T_c produced by the HVB is greater for higher frequencies. Also, substantial differences between the loudspeaker and the artificial head are apparent. This might mainly stem from their different directivity characteristics



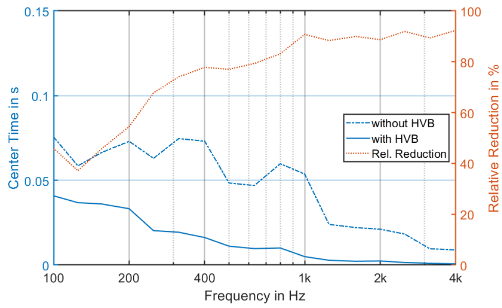
(a) Measurement microphone, loudspeaker



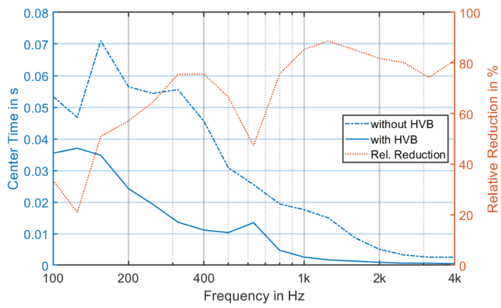
(b) Measurement microphone, artificial head



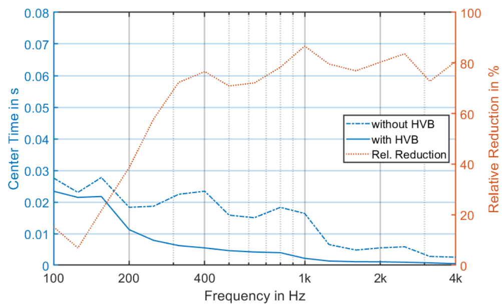
(c) Studio microphone (omni), loudspeaker



(d) Studio microphone (omni), artificial head



(e) Studio microphone (cardioid), loudspeaker



(f) Studio microphone (cardioid), artificial head

Figure 3.8: Center Time T_c in seconds with the HVB (blue) and without it (blue, $- \cdot -$), evaluated in 1/3-octave bands. Relative Reduction due to the HVB is also given (orange, $\cdots$).

3.3.5 Speech Transmission Index *STI*

Table 3.2 shows results for *STI* for all setups including the studio microphone changed to an omnidirectional polar-pattern. All values, even those without the HVB could be classified as "excellent" according to ISO 60268-16, but this is again a consequence of the short distance between sound source and microphone. Nevertheless, all setups show improvements in STI when the HVB is used, bringing the values very close to the theoretical maximum of 1. While having a high STI even in the empty RC, a cardioid-polar pattern seems to result in less improvement of STI.

STI Results

Setup	Without HVB	With HVB	Rel. Improvement
Measurement mic, LS	0.89	0.977	9.8%
Studio mic (omni), LS	0.908	0.986	8.6%
Studio mic (cardioid), LS	0.96	0.996	3.7%
Measurement mic, AH	0.89	0.991	11.3%
Studio mic (omni), AH	0.899	0.993	10.5%
Studio mic (cardioid), AH	0.977	0.999	2.3%

Table 3.2: STI (Speech Transmission Index) results calculated from impulse responses. LS designates the loudspeaker; AH designates the artificial head

4 Discussion

This thesis shows an analysis of a HVB in the form of theoretical considerations as well as measurements. The goal of this analysis was to examine how effective a HVB is in its use as acoustic treatment for vocal recordings in regard to three common problems: Reverberation, flutter-echoes, and colouration due to room modes.

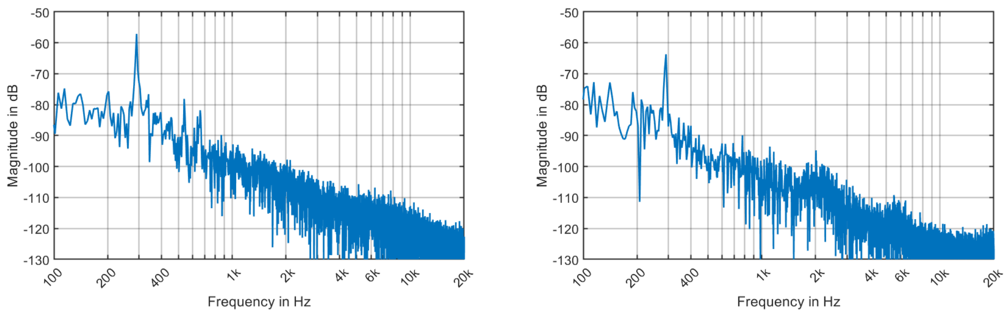
Because insufficient data on the absorption properties of the material the HVB is made of is available, theoretical considerations concerning the influence of the HVB on frequency response and on reverberation are hardly possible. Moreover, conventional RT parameters could not be used in a meaningful way. Therefore, reverberation was quantified as reverberation level, which describes the EDC at the intersection between the regression line used to calculate late RT and the y-axis. Results indicate a clear reduction in reverberation level over the entire relevant frequency range of 100 Hz to 4 kHz when the HVB is used (see Section 3.3.3). While late RT itself is affected marginally, reverberation level is noticeably attenuated across all relevant third-octave bands. Center time T_c is also reduced considerably when using the HVB. To put these effects into perspective, a comparison to reverberation level values and T_c values for well-treated recording rooms or vocal booths would be interesting.

For the problem of flutter-echoes, a notable improvement is suggested not only by measurement results, but also when considering the absorption characteristics of the HVBs material. Even if flutter-echoes might still occur in the room when the HVB is used, the high statistical absorption coefficient in the 1-2 kHz range indicates that the recording would remain largely unaffected. T_c values of only a few milliseconds in this frequency band suggest the same. To test this assumption, measurements in a room plagued by noticeable flutter-echoes could be conducted.

From the statistical absorption coefficient (Figure 2.4) and the increased impedance mismatch due to the short distance between absorber and source, it is obvious that the HVB absorbs very little sound energy at low frequencies. Therefore, it would likely not prevent room modes from forming. Assumably, it would, however, still reduce the sound colouration that those room modes create because the microphone is shielded from almost all directions by surfaces that are at least partly reflective below the Schroeder frequency of typical living spaces. Additionally, the HVB introduces a boost of frequencies below 230 Hz. When flattening the frequency response by equalisation, the sound level of potential room modes is automatically reduced.

While partly preventing colouration due to room modes, the HVB introduces notable colouration itself. The boost in low frequencies is likely caused by the high reflectivity of the porous materials - such as the one used in the HVB - at these frequencies. This effect stems from an impedance mismatch at the absorber's surface, which is likely further exaggerated by the proximity between absorber and source inherent in a HVB setup. Such colouration is generally unwanted, as true-to-life recording of a voice performance is often the goal. A parametric equaliser could effectively flatten the response.

In addition to the measurements in the RC, an example of male speech and singing inside the HVB was recorded a very difficult room that showed lots of reverberation, noticeable room modes in the tested vocal range (c3-c4 ≙ 130-261 Hz fundamental frequency), and background noise/humming. Upon listening to the recorded vocal takes, it is clear that the HVB reduces the perceived level of reverberation, especially in speech. The perception of the background hum at 290 Hz is not affected by the HVB, although an attenuation of 6 dB is visible in Figure 4.1; background noise sounds slightly darker with the HVB, which Figure 4.1 confirms. Inside the booth, lower frequencies sound very exaggerated; equalisation is necessary, but done quite easily. Higher frequencies, on the other hand, sound a bit muffled. The dip in mid frequencies from 350-700 Hz is not as noticeable in this case and does not necessarily need to be equalised.



(a) without HVB *(b) with HVB*

Figure 4.1: Frequency response of background noise and hum calculated from the recorded example.

In further research, more measurements, tailored towards the different effects of the vocal booth would be interesting. An analysis of the sound isolating properties of the HVB with microphones or sources outside the HVB would deliver more insights since this aspect was mostly left out in this work. Additionally, measurements for determining relevant specifications of the HVBs material could lead to a more accurate understanding concerning the mentioned impedance mismatch. Moreover, measurements taken with a different HVB, for example the Isovox 2 mentioned in Section 2.3, would make comparisons possible. Considering other HVBs would also allow for better analysis of the HVB as a concept, as opposed to analysis of a single product. Alternatively, an analysis of the HVB using FEM simulation could help in understanding the physics of the HVB.

Furthermore, it is obvious in Sections 3.3.2-3.3.5 that conventional parameters in room acoustics are not a perfect fit for quantifying the effects of the HVB. The main reason is the small distance between source and sink that the HVB is meant for. Even STI, the main application of which is speech intelligibility, is hardly of use when analysing the quality of voice recordings since high values achieved even without the HVB suggest good recording quality although voice recordings taken in an RC would be unusable in almost all situations. Therefore, development of a new parameter aimed at close-mic situations would be a useful next step. This parameter should take into account reverberation or "dryness" of the recording and colouration or "distortion of natural timbre". With a parameter such as this, entire recording setups could theoretically be evaluated and compared.

References

[1] Adam Patrick Bell. *Dawn of the DAW: The Studio as Musical Instrument*. Oxford University Press, 2018. Chap. A History of DIY Recording: Striving for Self-Sufficiency.

[2] F. Alton Everest and Ken C. Pohlmann. *Master Handbook of Acoustics*. 7th Edition. New York: McGraw-Hill Education, 2022. Chap. 5 Signals, Speech, Music, and Noise.

[3] Heinrich Kuttruff. *Room Acoustics*. 6th Edition. CRC Press, 2017. Chap. 1.7 Sound sources, pp. 20–21.

[4] Heinrich Kuttruff. *Room Acoustics*. 6th Edition. CRC Press, 2017. Chap. 5 Reverberation and steady-state energy density.

[5] F. Alton Everest and Ken C. Pohlmann. *Master Handbook of Acoustics*. 7th Edition. New York: McGraw-Hill Education, 2022. Chap. 25 Acoustics of Isolation Booths.

[6] Tor Halmrast. 'Why do flutter echoes 'Always' end up around 1-2 kHz?' In: *Proceedings of the Institute of Acoustics Vol. 37 Pt.3*. 2015.

[7] Heinrich Kuttruff. *Room Acoustics*. 6th Edition. CRC Press, 2017. Chap. 3.6 Steady-state sound field, p. 67.

[8] *t.akustik Vocal Head Booth*. URL: `https://www.thomann.de/de/t.akustik_vocal_head_booth.htm` (visited on 12/02/2025).

[9] *Isovox Mobile Vocal Booth V2*. URL: `https://www.thomann.de/at/isovox_mobile_vocal_booth_v2.htm` (visited on 12/02/2025).

[10] *Data Sheet PET Absorber*. URL: `https://images.static-thomann.de/pics/atg/atgdata/document/specs/522257_etc_v4_datasheet_en_online.pdf?_gl=1*3jojqs*_gcl_aw*R0NMLjE3MzYxNzQ3OTMuQ2p3S0NBaUFtLTY3QmhCbEVpd0FFVmZ0TmxtX1phMXB6Qk03Wkxqbkl2dnlxOVYwbGtQYkxhWVNRYTJwRmhkUzBYVVlvODJYdEZtb0h4b0NnVEFRQXZEX0J3RQ..*_gcl_au*MzYyOTI1NTY0LjE3MzA1NDExMzU.*FPAU*MzYyOTI1NTY0LjE3MzA1NDExMzU.*_ga*MTAzMjc2NTEwMy4xNTc0NzAwNDQ4*_ga_QNTG1E3BFT*MTczNjE3ODQxOS4xNzQuMS4xNzM2MTc4NDk1LjAuMC4xMDQ1Nzg5MTE5*_fplc*a0JVZkc2SUhQWmdTbVpYMCUyRkFPJTJGcCUyQklrMnU3NExxN0s0dTZwWTBpU0xmUGtOWVFCTFcwbHF3T0U3aUw4MEVxSWhrTTRDclRYYjQ0TnppV3dTbVFTQ3BsakE2elR4THVIYzZLUzZBV0FqT01yN25wc2dGQjlGUTlXdVRKZnhRJTNEJTNE` (visited on 12/02/2025).

[11] Fridolin P. Mechel. *Schallabsorber Bd. 1 Äussere Schallfelder*. S. Hirzel Verlag, 1989. Chap. Kugelwellen über Absorber-Ebene.

[12] Raffaele Dragonetti and Rosario A. Romano. 'Condsiderations on the sound absorption of non locally reacting porous layers'. In: *Applied Acoustics 87*. Elsevier Ltd, 2014, pp. 46–56.

[13] Sander J. Van Wijngaarden, ed. *Past, present and future of the Speech Transmission Index*. TNO Human Factors, 2002. Chap. 1 and 2.

[14] *ITA-Toolbox*. URL: `https://www.ita-toolbox.org/` (visited on 12/02/2025).

[15] Jamilla Balint. 'Evaluating the decay of sound'. PhD thesis. Graz University of Technology, 2020. Chap. 2 Multi-exponential decay model & Bayesian decay time estimation.

[16] ISO. *EN ISO 354: Acoustics — Measurement of sound absorption in a reverberation room*. 2003.