Thesis for The Degree of Doctor of Technology

Blind Estimation of Sound Coloration in Rooms

Peter Mohlin

Department of Architecture and Civil Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2024

Blind Estimation of Sound Coloration in Rooms

Peter Mohlin

ISBN 978-91-8103-050-1
Doktorsavhandlingar vid Chalmers tekniska högskola, Ny serie nr 5508.
ISSN 0346-718X

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Göteborg,
Sweden
Phone: +46(0)31 772 1000

Cover:
The average distributions of estimated damping constants for different system gains using sound 1 (female speech) as source signal. Refer to fig. 9 in paper D for more details.

Printed by Chalmers Digitaltryck,
Gothenburg, Sweden 2024.

To my grandparents, Helena and Frank, whom I miss everyday, and my parents, Dan and Maria, for all the love and support.

Blind Estimation of Sound Coloration in Rooms

Peter Mohlin

Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

A common problem in sound reinforcement systems consisting of one or more microphone-amplifier-loudspeaker channels is the sound coloration caused by the repetitive amplification of strong frequency components in the loudspeaker-microphone transfer function(s). For hidden systems, such as certain reverberation enhancement systems, acoustic feedback, along with other causes for sound coloration, risk compromising the impression of a natural sounding acoustic environment. Therefore, in this thesis, a methodology for blind estimation of sound coloration is developed and evaluated. Depending on the room type and various other assumptions, the damping distribution of a room will follow a specific "reference distribution," i.e. any deviation from the distribution should indicate sound coloration. Using one microphone placed in the audience area, blind estimation of sound coloration is achieved by computing decay times of non-harmonic components in the time-frequency domain. The results show that the computed damping distributions agree well with the chi-square distributions at low system gains. As the system gain increases, the distributions are shifted toward lower damping constants, and their shapes deviate more and more from the reference distribution, thus, giving a clear indication of sound coloration. The suggested objective measures show that deviations from the reference damping distribution can be detected at substantially lower system gains compared to results of related listening tests where audible coloration is evaluated. Thus, it is safe to conclude that the proposed methodology for detecting and classifying sound coloration performs well for the studied cases. Further research is needed to optimize its robustness when using different room types, system configurations, and so on.

Keywords

sound coloration, damping constants, distributions, spectrogram, objective measures, time-frequency analysis, reassignment, audibility of tonality, pure tones, time resolution

Acknowledgment

First, I would like to thank my initial supervisor, Prof. Mendel Kleiner. Thank you for believing in me at an early stage and letting me start as a PhD student at Applied Acoustics. Lovingly, I still remember the feeling of "I did not expect that answer" which often was the case after asking you something related to acoustics, or perhaps, anything!

Second, my current supervisor, Associate Professor Patrik Höstmad, deserves a huge thanks for coping with both me and the topic of the thesis. I am truly amazed at how you have managed to give me tons of good advice. It is crystal clear that, without your support, this thesis would bury itself as a nasty singularity stuck somewhere in my sporadically functioning brain. In other words, I would have been lost and completely unmotivated and I would have been forced to give up. Fortunately though, you were there. For that, I am forever grateful.

Third, for all the colleges and friends I have met during the years as a PhD student and onwards: thank you! For some reason, I have always felt accepted and "not so weird despite doing a PhD for over 20 years". Börje Wijk and Gunilla Skog: thanks for all the great moments and all your help with technical and administrative things. The "new Gunilla" (for me at least), Susanne Pettersson: thank you for helping me finish my work and finding the final courses. Laurent Capron, I am infinitely grateful for our friendship! Wish you were here... Pontus Larsson, thanks for helping me with the measurements of all those impulse responses at Göteborgsoperan (and a lot more things). I still owe you at least one microphone stand! Prof. Wolfgang Kropp, I will not forget your support when the opinions of one reviewer was, to put it mildly, somewhat questionable. I am truly grateful. Anders Genell, thank you for interesting discussions, vibrating pant legs and the grinding. Pontus Thorsson, thank you for introducing me to a rather strange but quite wonderful loudspeaker brand (and much more!). Mikael Ögren, thank you for all electroacoustic discussions and great bass playing. Per Drougge, thank you for being part of our new journey and supporting my PhD work! I could not ask for a better colleague and friend! And to all of you who I have forgotten to mention here: I'm sorry! Remember that you matter!

Finally, Anna and all the kids, thank you for enriching my life in ways unimaginable. I am truly blessed.

The work presented in this thesis has been supported by the Swedish Council for Building Research (BFR).

Contents

Part I

Summary

Chapter 1

Introduction

1.1 Background

In many situations, it is desirable to use sound reinforcement systems (SRSs) to amplify the sound presented to audiences. The source material is often music (e.g. a symphony orchestra, rock band, a singer supported by playback, etc.) or human speech. With the technology at hand today, an SRS can be configured in virtually any way imaginable. Digital signal processors (DSPs) which operate in real-time enable advanced signal processing to add various "effects" to the source material. Examples of such effects are delays, reverbs, flangers, phasers, and so on. Also, DSPs make it possible to modify the room acoustics. For example, a room impulse response (RIR) that has insufficient energy in the very important envelopmental part ($20 - 150$ ms after the direct sound component [77]) can be improved by implementing a suitable FIR filter. Using DSPs or other equipment that affect the signals sent to the system loudspeakers could be considered to be an active method for manipulating the RIRs. Common passive alternatives are the use of absorbers, diffusors and reflecting panels, which are placed at strategic positions. Although these passive means of altering the room acoustics result in some modifications of the RIRs, there are some natural limitations. For example, if the goal is to obtain a longer reverberation time, a large number of reflective panels or diffusors must be added to minimize sound absorption. The reverberation time in a room can be estimated according to

$$T_{60} = \frac{0.16V}{S\bar{\alpha} + 4mV}\,, \tag{1.1}$$

where T_{60} is the reverberation time, V the room volume, S the total surface area, $\bar{\alpha}$ the average absorption coefficient and m the air attenuation constant [4]. Despite minimizing the average absorption, $\bar{\alpha}$, the reverberation time can only be prolonged to a certain limit, defined by the term representing the air attenuation, $4mV$. This means that if the reverberation time needs to be longer than this limit, the only remaining passive alternative is, according to eq. 1.1, to increase the room volume. However, this is not a realistic option due to the

cost of such a project. Instead, a special type of SRSs, called reverberation enhancement systems (RESs), can be used.

In essence, there are two types of electroacoustic enhancement systems, which operate on the principles of either enhancing the early reflections or the reverberation time [87]. The first type is called in-line or non-regenerative systems, since they are intended to pick up the direct sound using directional microphones positioned where the reverberant energy is low, apply signal processing followed by amplification and finally feed the resulting sound into the room using numerous loudspeakers. Acoustic feedback needs to be minimized to avoid audible coloration artifacts. Thus, an in-line system should enable a direct link (or "line") between the source and listener(s) without any acoustic feedback, hence its name. The second type of system focusing on enhancing the reverberation time is called non-in-line or regenerative. Here, the acoustic feedback is an important part of the system design and the system microphones are placed where the reverberant energy is high so that the system reacts to *any sound* in the room. In contrast to the in-line systems, which are similar to a standard SRSs and therefore lack the "natural" control of the room reverberation, the non-in-line systems are suitable as RESs. Ideally, a RES could control both the early reflections and the reverberation, but optimizing both aspects is not trivial. In practice, however, there is some overlap between the two system types. For example, a non-in-line system might affect the early reflections and an in-line system might alter the reverberation time.

Feedback control in RESs can be considered to be of special importance, since many systems are "hidden", i.e. an audience is supposed to be unaware of any electroacoustic enhancement. The feedback can cause audible coloration or, in the extreme case, sustained howlback. The coloration is often identified when the sound becomes hard and metallic. Systems closer to instability will have additional ringing artifacts. Finally, when the system is unstable, the ringing is constant and is only limited by system non-linearities.

To minimize the acoustic feedback, time-variance can be introduced, which modulates the microphone signals in some way before they are fed to the system loudspeakers. Common modulation types are delay and phase modulation which are implemented using DSPs. A second type of time-varying systems controls the gain of the system channels so that the gain increases only when the sound level in the room decreases (i.e. a form of amplitude modulation). This also reduces the risk of sound coloration due to acoustic feedback. The third and final example operates in the frequency domain, where potential "ringing frequencies" are tracked. These frequencies vary over time and if a certain frequency peak becomes too strong, a very narrow notch filter is placed over it. A DSP designed for this purpose is often called "feedback destroyer" and such DSPs are commercially available from numerous manufacturers [64–69]. Although suitable for maintaining system stability, these methods do not guarantee that the audible sound coloration is minimized. For this to be true, they need to be more sensitive, since sound coloration caused by acoustic feedback can be detected by listeners even at low system gain settings [19].

Sound coloration measures are often encountered in research related to sound reinforcement and reverberation enhancement systems. In general, sound

coloration is a subjective attribute, related to a "reference", which could be defined as the listener's idea of how the source sound should sound in a given context, and any audible differences from the reference, usually classified as having detrimental effects on the source sound (i.e. sound coloration is normally not considered to improve the source sound in the given context).

For the majority of the more refined sound coloration measures mentioned in this thesis, there is a need to either measure the impulse response or (perfectly) capture the source signal(s). In practice, it is difficult to obtain the anechoic source signals, especially for larger bands or orchestras. Also, measuring the impulse responses without alerting the audience and interfering with the performance (or speech) is cumbersome. Thus, these two facts severely limit the use of the above measures during live performances. A third limitation is related to the sensitivity of the measures. Arguably, all proposed methods intended for feedback control which require feedback or howling detection can be modified into objective measures of sound coloration. However, for successful detection of acoustic feedback, the system will typically be close to instability, i.e. sound coloration and ringing artifacts are clearly audible. Hence, it is not possible to accurately estimate sound coloration at lower system gain settings. More fundamental measures could be based on various pitch detection methods [28] - [47], signal modeling (e.g. all-pole and ARMA models) or the algorithms implemented in feedback destroyers mentioned above. A more detailed discussion concerning these measures is found in chapter 2.

1.2 Objectives

The main goal of this thesis is to develop a method for estimation of sound coloration caused by acoustic feedback in rooms. In view of the limitations of the current sound coloration measures briefly mentioned above, a different strategy is presented in this thesis. The most important criterion is to be able to compute an accurate measure without a measured reference, i.e. "blindly". Therefore, using distributions of damping constants is a promising way forward, since theoretical "reference distributions" exist. If it is possible to develop a robust coloration measure, the system will be easier to optimize and/or operate.

The damping constants are computed from spectrograms, which require various computational parameters. This motivates a study of the audibility of the tonality in short decaying pure tones, which could be seen as unmasked ringing artifacts due to acoustic feedback. In other words, what are the lowest tone durations at which tonality is audible? The answer should be a suitable "time constant" for the coloration detector, which translates to a time constant for the spectrogram computations.

Finally, if listening tests can reveal how the sound coloration is perceived, it might be possible to link the subjective and objective data.

1.3 Limitations

All aspects concerning the robustness of the coloration measure, including parameter selection, the use of different source signals, rooms, absorbers and/or diffusers, sound reinforcement or reverberation enhancement system configurations, etc, are suitable suggestions for future work and will not be discussed in great detail in the following text.

Using the iterative method described by Svensson [1] for system simulations usually implies that the system is time-variant. However, in this thesis, the SRS is time-invariant and the entire audible frequency range is used. The iterative method was selected because an initial goal was to introduce time variance to study various feedback control methods. However, this idea was eventually abandoned due to the complexity of the main objective, i.e. the development of the coloration detector.

In order to quantitatively verify the implementation of the coloration detection method, an attempt was made to mimic or emulate the behaviour of a rehearsal room sound reinforcement system by computing impulse responses based on the summation of a large number of damped sinusoids. To emulate the effect of different system gains, a set of damping distributions with decreasing median values and variances was applied to the summation. Thus, considering that simulated and emulated systems are used in this thesis, measurements in real rooms equipped with a multi-channel SRS would be a very interesting extension to this work. An attempt was made to perform such measurements, but the results were not as expected due to poor SNR and too few system channels.

When testing the sound coloration measures, four source signals, two speech and two music samples, are used. Similar to using only one room type for developing the coloration measure, four source signals is arguably a low number. However, this number was chosen considering the computational times and more practical aspects related to the presentation of results.

1.4 List of publications

Paper A

P. Mohlin, "The just audible tonality of short exponential and Gaussian pure tone bursts," J. Acoust. Soc. Am. 129(6), pp. 3827–3836 (2011).

Paper B

P. Mohlin, "Improving the readability of noisy reassigned spectrograms by all-pole modeling," *to be submitted*.

Paper C

P. Mohlin, P. Höstmad, "Objective measures for time-frequency distributions based on well-known signal quantifiers," *to be submitted*.

Paper D

P. Mohlin, P. Höstmad, "Blind estimation of sound coloration in rooms using chi-square distributions of damping constants," J. Acoust. Soc. Am. 152, 456 (2022).

1.5 Thesis outline

The organization of this thesis is as follows. In chapter 2, a literature review focusing on possible methods for detecting sound coloration due to acoustic feedback is found. The basics of reverberation enhancement systems, which typically are so called non-in-line systems, are discussed in chapter 3. The fundamental expressions concerning system stability are derived. In chapter 4, the time constant of the coloration detector, based on the audibility of tonality, is discussed in the context of spectrogram computations. In addition, the "gray area" related to the loudness of short pure tones exceeding one critical bandwidth is mentioned. The proposed sound coloration detector is using time-frequency data as input. The spectrogram and its reassignment are presented in more detail in chapter 5, including the limitations of reassignment when analysing signal decay times. Ending the chapter, a number of objective measures for time-frequency analysis are introduced. In chapter 6, it is shown how the simulated sound reinforcement system used for developing the coloration detector is created. Both the measurements of the room impulse responses and the iterative computational method to simulate the system and its acoustic feedback are discussed. Some early, and perhaps less successful, attempts at coloration detection are brought up in chapter 7 followed by the much more promising and final version of the detector in chapter 8. The last two chapters are the general discussion and conclusions. In the general discussion, the most important aspects of the results presented in the appended papers are highlighted as well as how the papers relate to each other. Analogously, in the final conclusions chapter, the most important contributions of all papers are summarized. A psychoacoustic study is presented in Paper A where the audibility of tonality in short-duration decaying pure tones is investigated. The results are used to motivate the required time resolution of the "coloration detector" presented in paper D. In paper B, numerous methods for improving reassigned spectrograms of noisy signals are explored and new expressions are derived for the reassignment operators. A set of objective measures for time-frequency analysis are suggested. The set of measures is refined and expanded to a total of eight measures, which can be found in paper C. In addition, a more systematic evaluation of the measures is presented, including varying key computational parameters. Paper D sums up the main focus of this thesis. A coloration detector for sound reinforcement and reverberation enhancement systems in rooms is proposed.

Chapter 2

Literature review

This review will focus on different ways of detecting sound coloration caused by acoustic feedback. Both potential and actual methods will be discussed, starting with the former.

There are numerous ways to obtain an objective measure of sound coloration and howlback. Essentially all methods currently used for pitch detection [28] - [47] can be modified in one way or another to work for this purpose. However, the robustness of some of these methods for potential coloration detection is questionable since most of them are developed for speech signals.

In essence, there are three main groups of pitch detection algorithms: the ones that mainly operate in the time domain (see e.g. [30], [36], [38], [43], [40], [29], [39]), frequency domain (e.g. [37], [46], [41]) and both time and frequency domain (e.g. [44], [33], [34], [47], [45], [31], [32], [42]). A good overview of several algorithms and how they perform is given in [28]. Among these methods, various correlation methods (especially autocorrelation) are common, as well as the short-time Fourier transform, cepstrum analysis and the computation of useful time-frequency distributions and/or instantaneous phase. More exotic approaches are, for example, the use of MUSIC (multiple signal classification) and neural networks.

Another approach is to compute various models of the signal and use the models for pitch analysis (see e.g. [48] - [63]). Autoregressive models, also called all-pole models, and ARMA (AutoRegressive Moving Average) models are often used and the corresponding time varying modeling techniques (TVAR and time varying ARMA) are studied by several authors. The benefits of modeling the signal is that the model can give direct information about e.g. signal frequencies. TVAR assumes time varying all-pole coefficients and enables extremely high time-frequency resolution. However, the model order determination is critical and a low signal to noise ratio (SNR) might be problematic.

More potential coloration detectors are found in several patents related to feedback destroyers (e.g. [64–69]). In [64], the signal is analyzed in the frequency domain. The strongest peak located at frequency f_0 is monitored along with two of its subharmonics ($\frac{1}{n} \cdot f_0, n > 0, n \in \mathbb{Z}$) and/or harmonics ($n \cdot f_0, n > 0, n \in \mathbb{Z}$). If the (sub)harmonics are strong enough compared to the

maximum peak value, a decision is made that no "resonating frequencies" are present. However, if the (sub)harmonics are lower in level, more specifically 33 dB lower, than the strongest peak, there are problems due to acoustic feedback and a notch filter is placed over the frequency in question. Also the time domain is utilized, since spectra are stored and compared over time. However, this approach is not that common. Instead, the signal is often studied in the time domain using phase locked loops [70] - [72], [66], [67], [69]. A similar method is used in [75] to efficiently filter audio signals which have an additional interfering periodic signal, for example typical 60 Hz hum (which have strong harmonics). In [71], an LMS adaptive notch filter is used in conjunction with a phase locked loop. The FFTs of the signal elements directly following spoken syllables are analyzed in [72] and peak indices are calculated for feedback detection. Other examples are the use of up down counters (to check for periodicity) [65] and synchronous signal analysis [68]. The unwanted frequency components are removed using notch filters or frequency dependent gain control.

For actual coloration measures proposed in the literature, an impulse response and/or a perfect source signal are usually needed. Some examples are mentioned in the following text. Meynial and Vuichard [20] use Rayleigh distributions and histograms of frequency response magnitudes of measured room impulse responses to detect sound coloration caused by one or more electroacoustic channels and/or by poor room acoustics. A method to apply suitable time domain windowing in order to target and amplify the late part of the impulse response is described. The resulting frequency responses are "equalized" by the low-Q variations (imagine a smoothed response) of the responses and windowed (only frequencies above the Schroeder frequency are considered) before their magnitude histograms are compared to the Rayleigh distribution using various statistical measures. The measure which gives high correlation to listening test results and also shows promising numerical properties such as low uncertainty, turned out to be the standard deviation of the high-Q, or more irregular, variations of the room transfer function (RTF). By dividing the original RTF by a smoothed version of it, the high-Q variations of the of the RTF is obtained. Since human hearing performs a type of "autogain" operation when listening to long decaying sounds, the decay of the RIRs is compensated for. Additionally, a number of time and frequency windows are used to treat the data in a statistically correct way (mixing time [5], Schroeder frequency [4], etc). These windows are also used to avoid corrupted results due to background noise, which occur for a RIR signal to noise ratio lower than about 10 dB. An obvious limitation of the measures is that they are based on measured impulse responses, i.e. using program material for coloration analyses would require custom deconvolution algorithms (which rely on perfectly captured source signals).

A similar objective measure is proposed by Watanabe and Ikeda [21], i.e. the standard deviation of the equalized and rapidly varying frequency response is computed and compared to the Rayleigh distribution. Again, this requires measured impulse responses.

Poletti [22] also presents a method based on the Rayleigh distribution, but only distributions obtained from stochastic simulations of reverberation

enhancement systems are analysed. No recorded systems or measured impulse responses are used for sound coloration analyses.

Several measures of sound coloration are computed by Korany [23]: the temporal diffusion index, based on the autocorrelation function, the power cepstrum and the spectral coloration index. For all measures, simulated impulse responses are used.

Room impulse responses are also required for the measures discussed in a paper by Rubak [24]. To estimate the (timbre related) spectral coloration, a measure is proposed based on the modulation depth of the spectrum obtained after applying auditory filters to the impulse response. The temporal coloration is estimated using the autocorrelation functions of an octave band filtered (from 125 to 4000 Hz) impulse response. Finally, a time-frequency distribution is computed for which it is proposed to use an auditory filter bank instead of the short-time Fourier transform, with a time window length of 30 ms. The modulation depth of each resulting spectrum is then computed, leading to a time dependent spectral coloration measure.

Another method to estimate sound coloration is to use spectrograms of the source signal with and without added reverberation, denoted as input and output signals respectively [25]. The proposed measure is defined as the difference in spectrogram magnitudes of the input and output signals at detected input signal onsets. As for the method proposed by Nielsen [19], the source signal is required for estimating the coloration.

One of the more advanced methods, called adaptive feedback cancellation (AFC), models the impulse response of the acoustic feedback path using an adaptive filter [26]. Considering that the modeled impulse response is intended to cancel out the acoustic feedback path, ideally resulting in a completely uncolored sound, the impulse response of the full system is not used to estimate sound coloration. Since the coloration due to acoustic feedback is minimized, the audible sound coloration in AFC systems is often a result of the signal processing needed to decorrelate the input signal of the adaptive filter and the disturbance signal in order to avoid slow convergence speed or convergence towards a biased solution. A better approach is to apply the decorrelation in the adaptive filtering circuit instead of the closed signal loop. An objective measure, first proposed by A. Spriet et al. [27], is used to assess the sound quality of the resulting feedback compensated signal. Again, the source signal is required to compute the coloration measure.

A method using program material for the detection of sound coloration in reverberation enhancement systems is proposed by Nielsen [19], [73]. The coloration detection is based on modulation transfer functions (MTFs), which are defined as the absolute value of the normalized Fourier transforms of the squared room impulse responses. An estimation of the MTF is obtained from the cross power and power spectra of the source and receiver envelope functions. It is shown how MTFs corresponding to increasing system gains are clearly separated, especially if frequency domain windowing is applied. This, according to the author, shows the potential of using MTFs for coloration detection. However, the problem of (perfectly) capturing the source signal is not solved. It is suggested to use highly directional microphones close to the source(s), but

the difficulty using this approach for larger orchestras is apparent. Additionally, the author does not propose a robust coloration measure based on the MTFs as the coloration is estimated based merely on visual inspection of the MTF plots.

The complex modulation transfer function (CMTF) was suggested by Schroeder [18] as

$$M(\omega_m) = \big|M(\omega_m)\big|e^{j\phi(\omega_m)} \tag{2.1}$$

by showing that if the envelope of a white noise signal is amplitude modulated by a sinusoidal modulation function, $s(t) = s_0(1 + \cos(\omega_m t))$ and sent through a noise-free linear system, the output envelope will be

$$r(t) = s_0\Big[1 + \big|M(\omega_m)\big|\cos(\omega_m t + \phi(\omega_m))\Big]. \tag{2.2}$$

Here, ω_m is the modulation frequency, ϕ the phase of the complex number when expressed in polar form and s_0 serves as an amplitude scaling factor for the modulated envelope.

If it is assumed that the squared and smoothed envelope of an impulse response can be modeled as

$$h^2(t) = e^{-2\delta t}, \tag{2.3}$$

where δ is the (average) damping constant of the room and is linked to the reverberation time, T_{60}, according to

$$\delta = \frac{6.91}{T_{60}}, \tag{2.4}$$

one can show that the -3 dB low-pass cutoff frequency, f_c, of $20\log\big|M(\omega)\big|$ is

$$f_c = \frac{2.2}{T}. \tag{2.5}$$

Thus, an increase in the reverberation time of a room can directly be seen in the CMTF.

For a regenerative system consisting of one channel with amplifier gain $G_(\omega)$ and loudspeaker (L) - to - microphone (M) transfer function $H_{\mathrm{LM}}(\omega)$, it can be shown that

$$T_{\mathrm{on}} = T_{\mathrm{off}}\frac{1}{1 - |G(\omega)H_{\mathrm{LM}}(\omega)|}, \tag{2.6}$$

where T_{on} and T_{off} are the reverberation times with the system turned on and off [73]. Inserting eq. 2.6 into eq. 2.5 leads to

$$f_c = \frac{2.2(1 - |G(\omega)H_{\mathrm{LM}}(\omega)|)}{T_{\mathrm{off}}}, \tag{2.7}$$

i.e. f_c decreases with increasing gain for this simple case.

Many of the different feedback detection algorithms discussed above will work only when the (potential) howlback frequency components are quite strong. Therefore, most of these algorithms will work fine if the desired

objective measure would describe the existence of sustained howlback (true or false). However, for coloration detection, just a couple of measures are interesting: the CMTF and the analysis of the statistics of the RTF according to [20].

Several coloration measures mentioned above including [20] are based on RIRs, which means that the RIRs must be measured or modeled in order to calculate the proposed objective measures. This is very hard to do during e.g. a concert and since the involved RIRs are constantly changing, the RIRs should be analyzed over time. Thus, a robust model of the RIR(s) needs to be computed in real-time, which is a complex task requiring perfectly captured source signals.

The CMTF is limited in similar ways. However, the CMTF can be estimated using program signals, but this requires calculations of the cross and auto power spectra of the source and receiver envelopes. This method is limited by the fact that picking up only the source signal with a microphone is impossible. There are always additional signals from the room and SRS.

Chapter 3

Basic RES layouts

For a non-in-line system, there will typically be numerous feedback paths between the microphones and loudspeakers used in the system. Each of these paths can be said to generate coloration of the sound in the room. In general, by increasing the number of paths, the coloration will be less audible. If the assumptions in [4] are fulfilled, e.g. averaging the magnitudes of all transfer functions over frequency (and channels) and using a large number, n_L, independent microphone-amplifier-loudspeaker channels placed inside a room, the reverberation time will increase according to [1]

$$T_{\text{on}} = T_{\text{off}} \frac{1}{1 - n_L \cdot \text{MLG}}, \tag{3.1}$$

where T_{on} and T_{off} are the reverberation time of the room with and without the added RES respectively and the MLG is defined according to

$$\text{MLG} = \overline{|G(\omega) H_{\text{LM}}(\omega)|^2}. \tag{3.2}$$

For multichannel systems, the MLG is computed by averaging over both frequency and all channels ("spatial averaging"). Additionally, this assumes that the loudspeakers and the natural source are uncorrelated. As mentioned above, this very basic RES solely relies on acoustic feedback in order to prolong the reverberation time. An example of a commercial system which uses this approach is the somewhat dated Assisted Resonance System [78] - [86]. Unfortunately, for some systems, sound coloration due to acoustic feedback turned out to be problematic and, as a result, system tuning was difficult and time consuming.

If each loudspeaker receives the signal from all n_M microphones, eq. 3.1 is modified according to [1]

$$T_{\text{on}} = T_{\text{off}} \frac{1}{1 - n_L n_M \cdot \text{MLG}}. \tag{3.3}$$

Here, the same assumptions are made as for the previous expression. It is important to note that if the microphones receive a significant direct sound

component, the source and microphones are correlated and the assumption is invalid.

The most basic SRS is shown in figure 3.1. $H_{\mathrm{SM}}(\omega)$, $H_{\mathrm{LM}}(\omega)$, $H_{\mathrm{LR}}(\omega)$

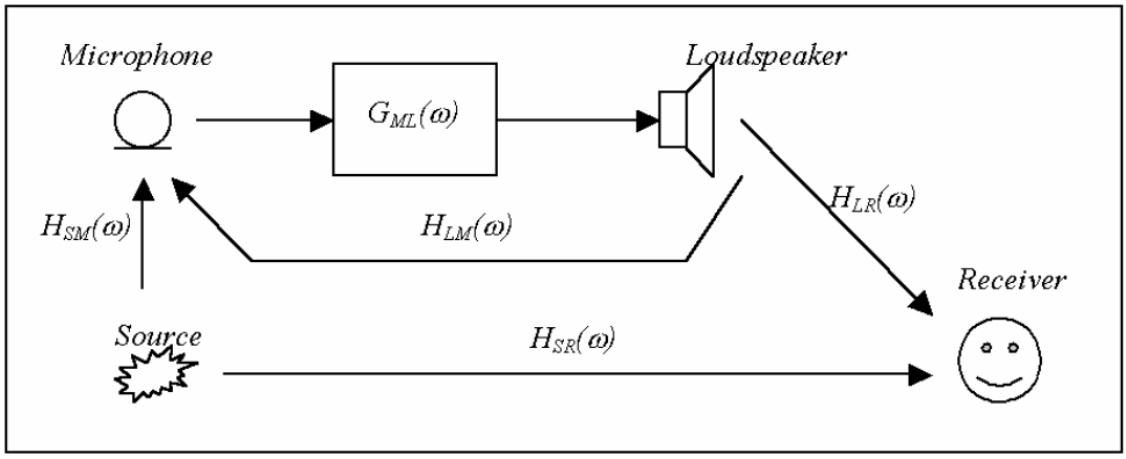

Figure 3.1: A single electroacoustic channel inside a room.

and $H_{\mathrm{SR}}(\omega)$ denote different room transfer functions (RTFs), which are the frequency domain equivalent of a RIR. SM denotes source-microphone (from-to), LM loudspeaker-microphone, LR loudspeaker-receiver and SR source-receiver. $G_{\mathrm{ML}}(\omega)$ represents the frequency response of the gain, i.e. an amplifier (and possibly an equalizer). As shown in the figure, $H_{\mathrm{LM}}(\omega)$ characterizes sound transmission between the system loudspeaker and microphone. This means that sounds radiated by the loudspeaker will be picked up by the microphone, amplified again, radiated, picked up, amplified and so on. This is a so called feedback loop.

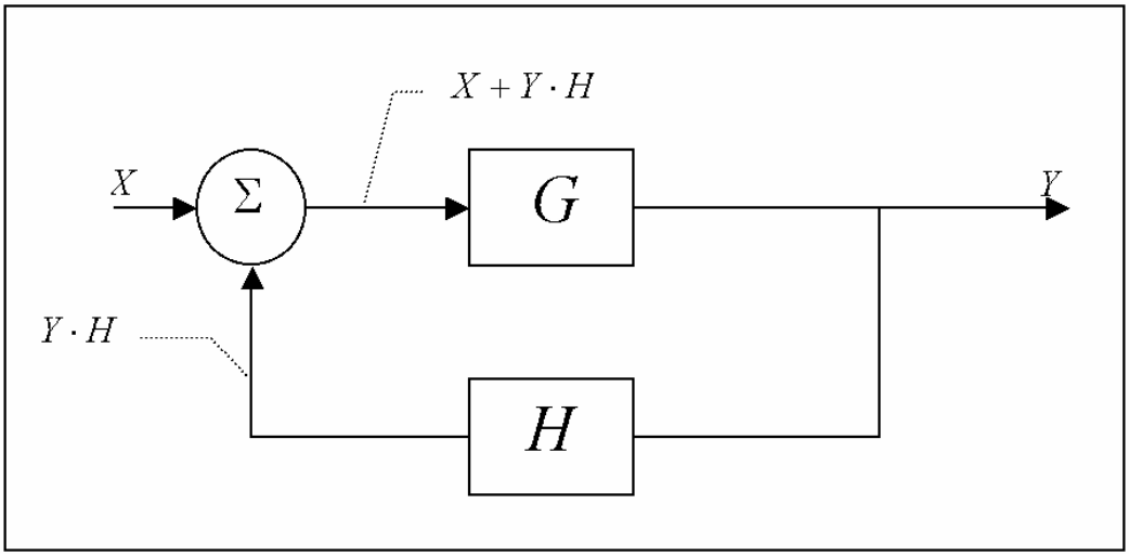

Figure 3.2: Block representation of a feedback loop.

The "total" transfer function, from the source to the listener, is

$$H_{\mathrm{TOT}}(\omega) = H_{\mathrm{SR}}(\omega) + H_{\mathrm{SYS}}(\omega), \tag{3.4}$$

i.e. "direct" + "system" sound. $H_{\mathrm{SYS}}(\omega)$ can be derived with the help of figure 3.2, which only shows the feedback loop. From the figure, it is clear that

$$Y = G(X + Y \cdot H) \tag{3.5}$$

and

$$G = \frac{Y}{X + Y \cdot H}. \tag{3.6}$$

Now, let G_{EQ} denote an equivalent block which represents the entire feedback loop, i.e.

$$G_{\mathrm{EQ}} = \frac{Y}{X}. \tag{3.7}$$

Some basic rearrangements follow:

$$\begin{aligned} Y &= G(X + Y \cdot H) = G \cdot X + G \cdot Y \cdot H \\ G \cdot X &= Y - G \cdot Y \cdot H = Y(1 - G \cdot H) \end{aligned}$$

$$\frac{Y}{X} = \frac{G}{1 - G \cdot H} = G_{\mathrm{EQ}}. \tag{3.8}$$

According to figure 3.1, it is clear that $H_{\mathrm{SYS}}(\omega)$ and $H_{\mathrm{TOT}}(\omega)$ should have the following forms:

$$\begin{aligned} H_{\mathrm{SYS}}(\omega) &= H_{\mathrm{SM}}(\omega) \cdot G_{\mathrm{EQ}}(\omega) \cdot H_{\mathrm{LR}}(\omega) \\ &= H_{\mathrm{SM}}(\omega) \frac{G_{\mathrm{ML}}(\omega)}{1 - G_{\mathrm{ML}}(\omega) H_{\mathrm{LM}}(\omega)} H_{\mathrm{LR}}(\omega) \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} H_{\mathrm{TOT}}(\omega) &= H_{\mathrm{SR}}(\omega) + H_{\mathrm{SYS}}(\omega) \\ &= H_{\mathrm{SR}}(\omega) + H_{\mathrm{SM}}(\omega) \frac{G_{\mathrm{ML}}(\omega)}{1 - G_{\mathrm{ML}}(\omega) H_{\mathrm{LM}}(\omega)} H_{\mathrm{LR}}(\omega). \end{aligned} \tag{3.10}$$

Eq. 3.9 tells us that there is a problem with system stability if the denominator becomes zero. Nyquist showed that for a feedback loop, the stability criterion is [76]

$$\Re\big\{G_{\mathrm{ML}}(\omega) H_{\mathrm{LM}}(\omega)\big\} < 1. \tag{3.11}$$

Seen in the complex plane, one can write

$$G_{\mathrm{ML}}(\omega) H_{\mathrm{LM}}(\omega) = r e^{i\theta}, \tag{3.12}$$

where

$$r = \big|G_{\mathrm{ML}}(\omega) H_{\mathrm{LM}}(\omega)\big|. \tag{3.13}$$

Thus, it is clear that the magnitude of $G_{\mathrm{ML}}(\omega) H_{\mathrm{LM}}(\omega)$ or r is allowed to exceed one as long as the phase rotates the complex vector away from the real axis according to

$$\cos\theta < \frac{1}{r}, \;\; r > 1. \tag{3.14}$$

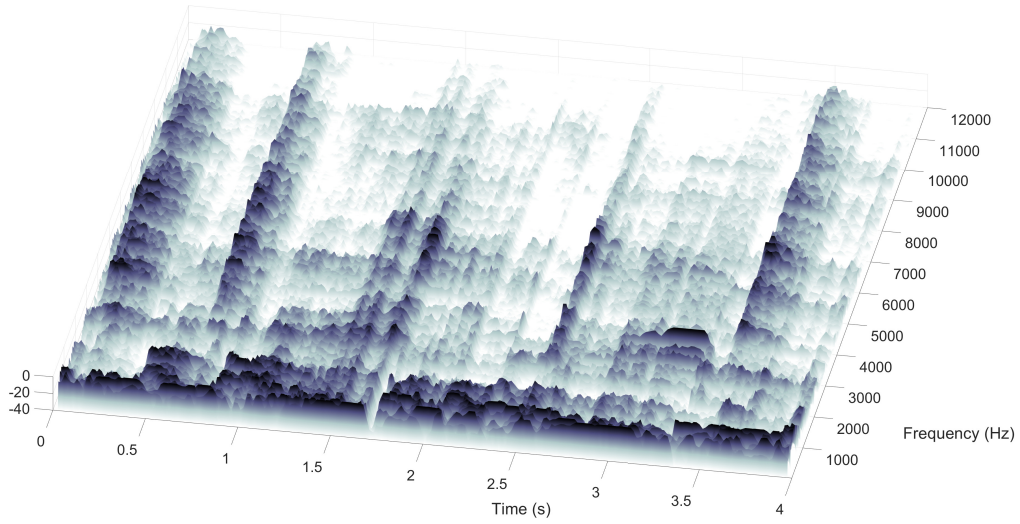

Figure 3.3: Spectrogram of a male speech signal. The SRS is close to instability.

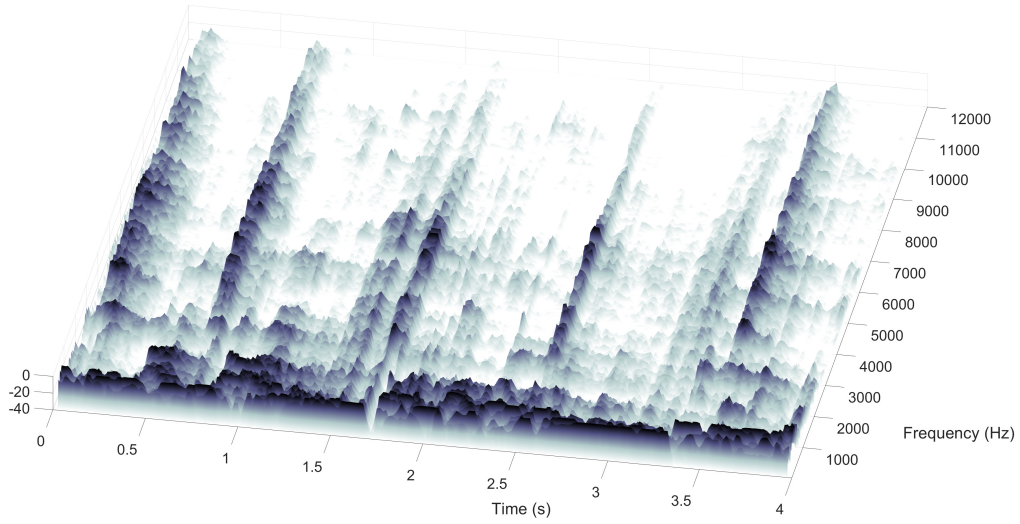

Figure 3.4: Spectrogram of a male speech signal. The SRS is far from instability.

An unconditional, and thus, much "safer" stability criterion is simply to limit r according to

$$\left|G_{\mathrm{ML}}(\omega)H_{\mathrm{LM}}(\omega)\right| < 1. \tag{3.15}$$

Now the phase is allowed to vary in any way possible, without causing system instability.

It is interesting to study the effect of an SRS close to instability by looking at the spectrogram of the signal presented to a listener. The signal is calculated using measured transfer functions and the procedure described in chapter 6. The source is a male speech signal and the simulated SRS has 6 channels. Figure 3.3 and 3.4 show the result for two different gain settings, -18.2 and -60 dB respectively. At a high gain setting, when the system is approaching

instability, one can see that the decay times increase substantially for numerous frequency components. Comparing both figures, the changes in decay times in figure 3.3 at about $0 < t < 1$ s and $3.5 < t < 4$ s are especially clear, as well as the clear ringing at around 3 kHz ($t = 3$ s).

Chapter 4

The audibility of ringing

In paper A, listening tests are performed to investigate the audibility of tonality in short decaying pure tones. The idea is that the results would give an indication of the required time constant for the time-frequency computations performed in the coloration detector. If unmasked and loudness compensated pure tone bursts, as implemented in paper A, correspond to single feedback components in an SRS close to instability, the "best case scenario" for detecting sound coloration in real SRSs is arguably reached. Thus, the resulting time limits for the audibility of tonality should correspond to a "minimum" or "best case" spectrogram time constant. It was found that tonality could be perceived in pure tones with a total duration (i.e the sum of the -60 dB attack and decay times of the signal envelope, where the attack and decay functions are Gaussian) of around 3 ms and frequencies above 3.4 kHz. For lower frequencies, the required tone durations rapidly increased, from around 5 ms at 3.4 kHz to 20 ms at 150 Hz.

Since harmonics analysis is a vital part of the coloration detector, the lower frequency limit needs to be around 20 Hz to incorporate both synthesized and natural sounds. This corresponds to a window length of at least $1/20 = 50$ ms. Thus, the 3 ms time constant suggested above, is not applicable to the time window length. Instead, $h(t)$ in eq. 5.3 (which is a 65 ms Blackman window), is shifted 3 ms for each new STFT computation. If the window length is fixed, a large window overlap greatly reduces the lower limit of the estimated decay times. This is illustrated by the examples in figure 4.1 - 4.3. The Blackman window was chosen due to its low spectral leakage, which should be an advantage when analyzing speech and music signals with their complex harmonics. Any additional coloration component could be located close in frequency to a signal component, which makes it important to minimize side lobes.

As can be observed, the higher window overlap with the 3 ms time step produces decay curves which are nearly parallel to the reference envelope for levels below approximately -20 dB. This is true for all three -60 dB decay times. For the lower overlap, corresponding to 50 % overlap since the time step is half of the window length, i.e. 32.5 ms, correct decay time estimation will clearly

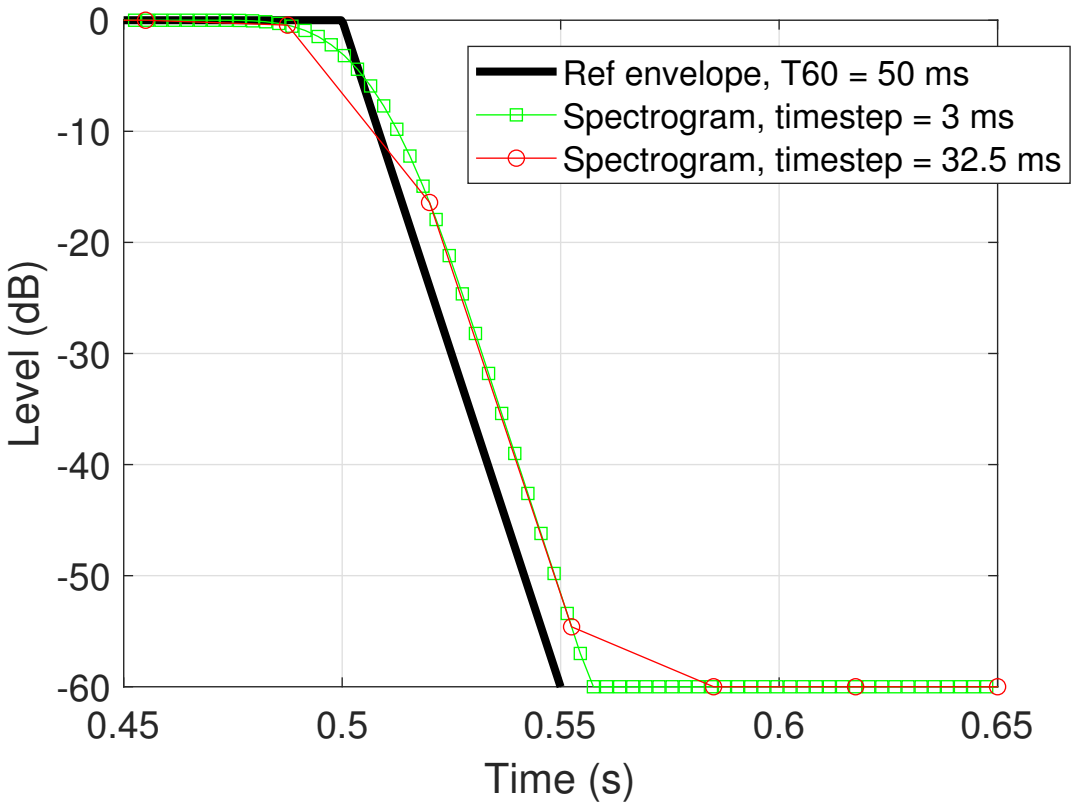

Figure 4.1: Comparison between a reference envelope with a -60 dB decay time of 50 ms applied on a 1 kHz pure tone and spectrogram output at 1 kHz using two different window overlap settings.

be problematic for the two shorter -60 dB decay times.

4.1 Pitch selectivity or pitch discrimination

It is important to note that the research presented in Paper A differs from pitch discrimination studies. The main reason for this is that the listeners are free to define tonality themselves (based on instructions and examples), i.e. no reference related to tonality or pitch is presented. Here, it is useful to introduce the term frequency *selectivity* [9], i.e. the ability to resolve frequency components in complex sounds. In contrast, frequency *discrimination* research focuses on the audibility of frequency changes over time. Typically, experiments are designed using either two successive steady tones of slightly different frequencies or two tones where one is frequency modulated (using low modulation frequencies). In the former case, the listeners reports if they can detect any pitch changes between the two tones and the resulting measure is called the Difference Limen for Frequency (DLF). In the latter case, the listeners try to identify the frequency modulated tone, which results in the Frequency Modulation Detection Limen (FMDL).

If one assumes that the pitch perception mechanism on the basilar membrane is defined by a "place theory", i.e. signals with different frequencies excite

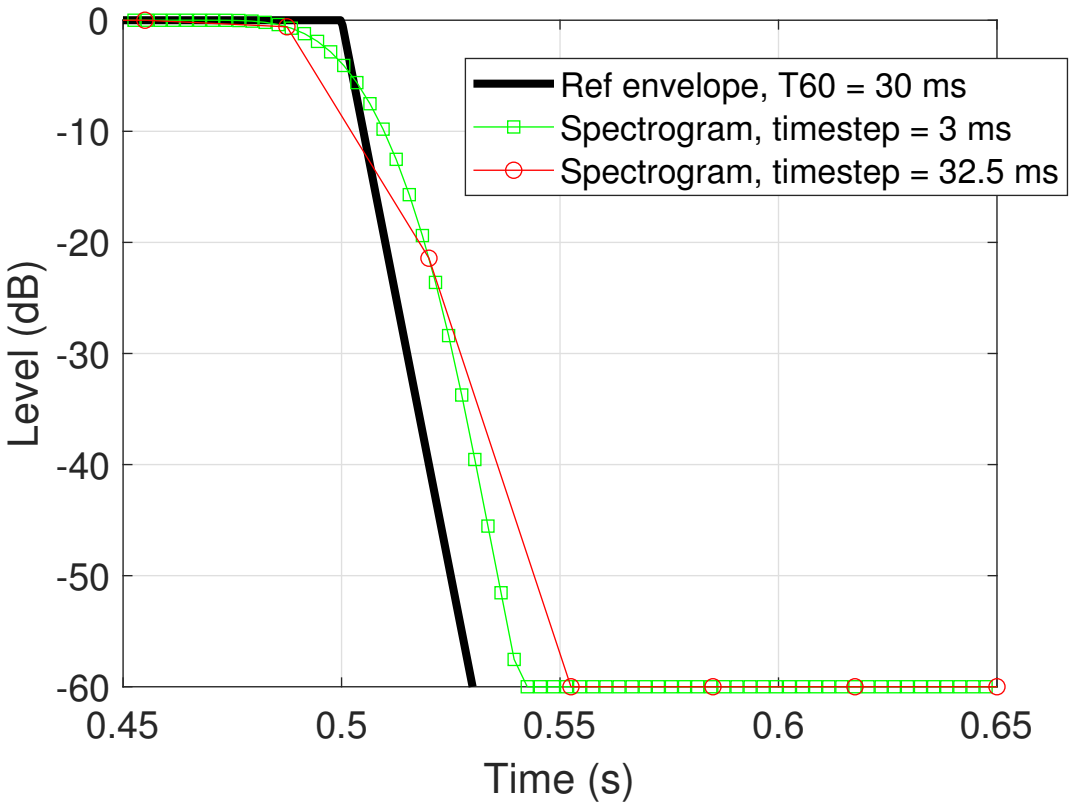

Figure 4.2: Comparison between a reference envelope with a -60 dB decay time of 30 ms applied on a 1 kHz pure tone and spectrogram output at 1 kHz using two different window overlap settings.

different parts of the membrane (leading to different neurons being activated), frequency selectivity and discrimination are closely linked [2]. For example, the variation of DLFs over frequency is closely connected to the (frequency variation of) critical bandwidths. Analogously, physiological studies of the basilar membrane have shown that equal shift in distance along the membrane results in increasing frequency steps when starting near the helicotrema and moving towards the oval window. In general, pitch discrimination rapidly deteriorates when the effective bandwidths of the stimuli are approaching the critical bandwidth. However, it has been shown that using the same effective bandwidths for signals with different envelopes do not produce similar pitch discrimination results [10]. Instead the "effective duration" of the signal is a better indicator, which means that signals with different envelopes produce similar pitch discrimination results as long as their effective durations are approximately the same.

For pure tones, the position of the maximum excitation on the basilar membrane corresponds to the perceived pitch. This is not the case for more complex sounds, perhaps most clearly illustrated by the "missing fundamental" experiment. For example, if 200 impulses are presented per second, the perceived sound will have a dominating fundamental frequency, f_0, at 200 Hz and harmonics at $n \cdot f_0$. However, high-pass filtering the sound so that f_0 is removed

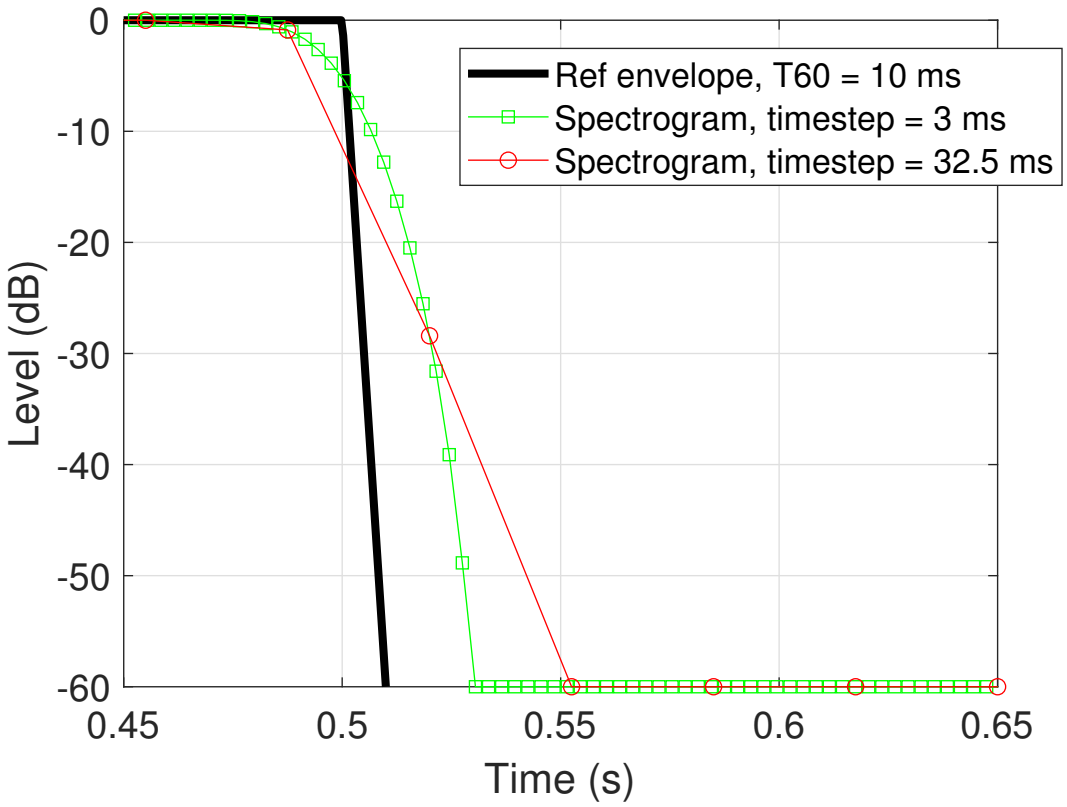

Figure 4.3: Comparison between a reference envelope with a -60 dB decay time of 10 ms applied on a 1 kHz pure tone and spectrogram output at 1 kHz using two different window overlap settings.

does not lead to a change in perceived pitch. Thus, for more complex sounds, the harmonics are crucial for determining the pitch, even if f_0 is present.

4.2 Loudness of short tone bursts - some considerations

In paper A, loudness compensation was implemented for the short duration pure tones with exponential and Gaussian envelopes used during listening tests. The literature is clear on how the compensation should be conducted, assuming that a specific condition is met [2]. In short, the condition concerns the bandwidth of the stimuli, which should be less than the critical bandwidth. If not, "additional bandwidth effects" [2] or "confounding effects of spectral splatter" [11] will influence the resulting loudness, because more than one critical band will be excited. To the author's knowledge, very few, if any, published papers focus specifically on this topic. The reason is probably related to the fact the stimuli lose most of their tonality and turn into "tonal clicks" or clicks having a certain timbre, as the bandwidth is larger than the critical bandwidth. In psychoacoustic research, this represents a gray area, since the tones are somewhere in between a pure click (Dirac delta function) and tone.

It is of interest to investigate the "one critical band limit" discussed above for the stimuli used in paper A. For the more broadband stimuli, it is likely, but still unconfirmed, that the loudness compensation was overestimated. Starting with figure 4.4 and the Gaussian shaped stimuli, the results indicate that tonality was detected for stimuli having almost equal (95 phon at 350, 450, 570 Hz) or smaller than critical bandwidths. This agrees well with the discussion in chapter 4.1. The area below the solid line is the "gray area" mentioned above, i.e. where correct loudness compensation becomes harder to implement.

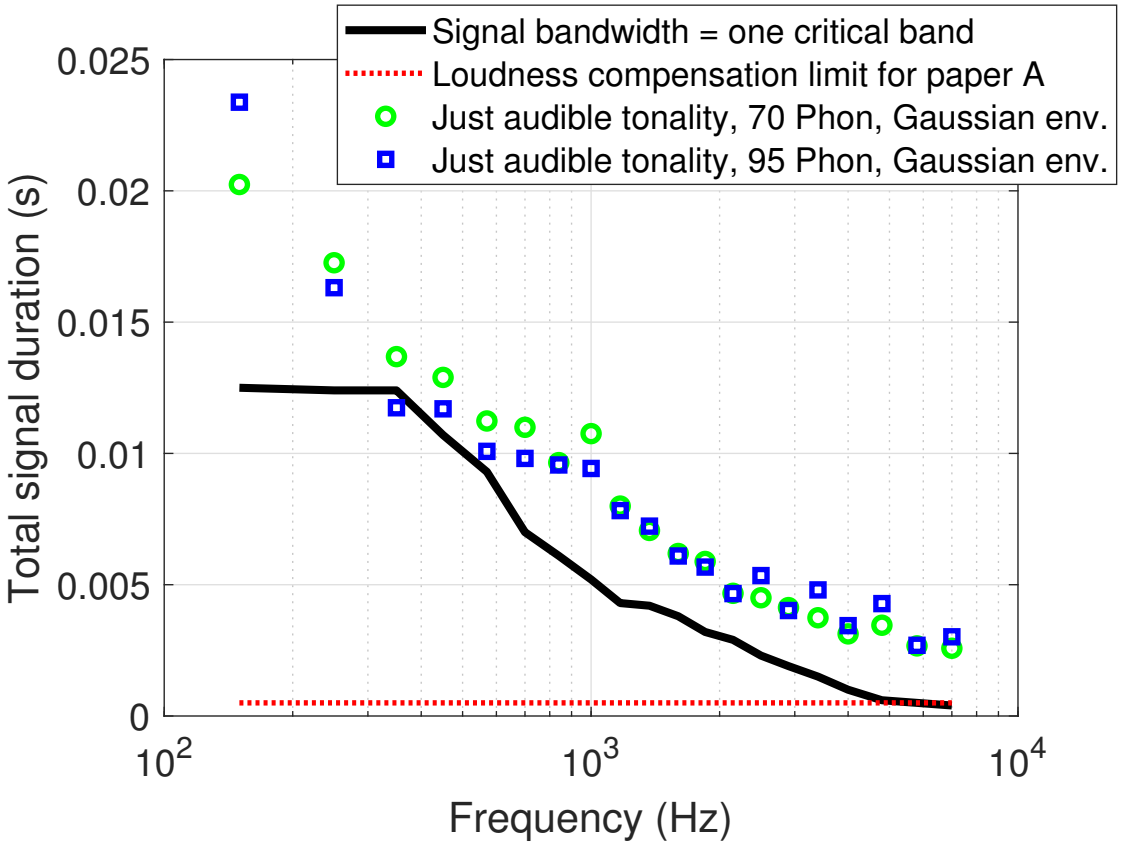

Figure 4.4: Required tone duration for Gaussian envelope stimuli to produce signal bandwidths corresponding to the critical bandwidth (solid line). The dotted line represents the 0.5 ms loudness compensation limit used for all stimuli in paper A. The circles and squares are the JAT (Just Audible Tonality) times from paper A (i.e. without headphone compensations).

For the exponential stimuli displayed in figure 4.5, the results are somewhat similar. However, below 2 kHz, there are more stimuli matching or falling below the critical band threshold (in time). For the 95 phon stimuli, the two largest deviations are around 5-7 ms (at 350 and 570 Hz). As shown in paper A, the exponential stimuli excite headphone resonances more than the Gaussian stimuli. Interestingly, the longest headphone ringing times are observed below 2500 Hz. For the largest deviations, the increase in ringing times is somewhere around $4 - 7$ ms (depending on loudness level) for 350 Hz and $0 - 2.5$ ms for 570 Hz, certainly bringing the data points closer to the critical band limit. However, if the headphone compensations are disregarded, the results indicate

that for around six 95 phon stimuli and equally many 70 phon stimuli from 250 to 840 Hz, the loudness compensation applied to the stimuli is likely somewhat overestimated.

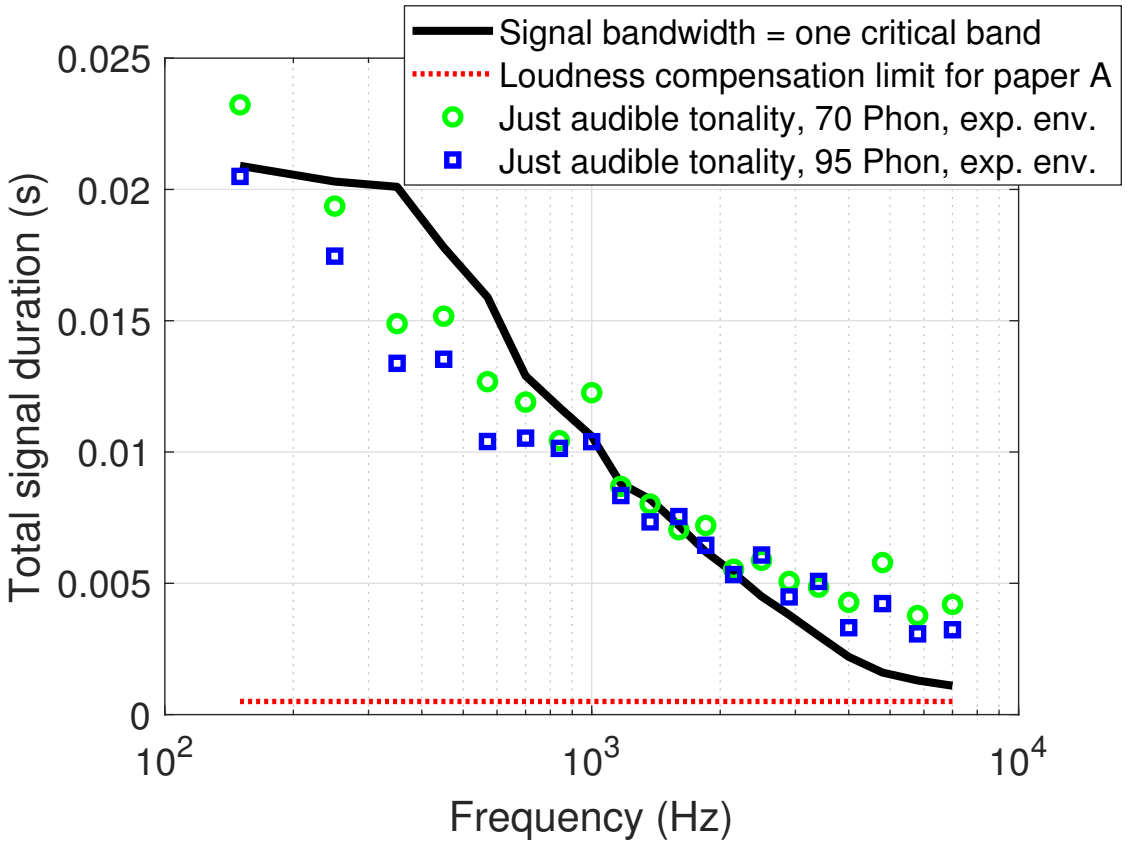

Figure 4.5: Required tone duration for exponential envelope stimuli to produce signal bandwidths corresponding to the critical bandwidth (solid line). The dotted line represents the 0.5 ms loudness compensation limit used for all stimuli in paper A. The circles and squares are the JAT (Just Audible Tonality) times from paper A (i.e. without headphone compensations).

Chapter 5

Time-frequency distributions and the spectrogram

To obtain a useful objective measure, one suggestion is to combine the time and frequency domains [73]. This is exactly what has been done in this thesis. Since many of the pitch detection algorithms operate in the time-frequency (TF) domain, the basis for the coloration detection algorithms proposed here is somewhat similar. In the following text, however, the computations differ since the focus lies on blind detection of sound coloration using the spectrogram. Therefore, a short introduction will follow, where some fundamental properties of the spectrogram are discussed.

A great number of authors have investigated various time-frequency distributions, or TFDs (for example, see [88] and the excellent review by Cohen [89] and their listed references). Here, "distribution" refers to a function of time t and angular frequency ω, $P(t,\omega)$, which is proportional to the intensity (or energy) per unit time per unit frequency.

One can show that an infinite number of distributions can be derived by using the following expression [89]:

$$\begin{aligned} P(t,\omega) = \frac{1}{4\pi^2} \iiint_{-\infty}^{+\infty} e^{-j\theta t - j\tau\omega + j\theta u} \phi(\theta,\tau) \dots \\ \cdot\, s^*\Big(u - \frac{1}{2}\tau\Big) s\Big(u + \frac{1}{2}\tau\Big) \mathrm{d}u\, \mathrm{d}\tau\, \mathrm{d}\theta, \end{aligned} \tag{5.1}$$

where s is a time signal, s^* its complex conjugate and $\phi(\theta,\tau)$ the "kernel". The kernel is an arbitrary function, which will generate different time-frequency distributions depending on how it is defined. The reason for not using time and frequency as variables in the kernel is that it can be a function of other variables, such as the signal itself.

The spectrogram is an *energy* distribution, because it assigns the energy of a signal to certain time and frequency points [88, 90]. Since the energy of

a signal is a quadratic function of the signal, the spectrogram is a so called quadratic or bilinear distribution.

The kernel of the spectrogram is

$$\phi_S(\theta,\tau) = \int_{-\infty}^{+\infty} h^*\Big(t-\frac{1}{2}\tau\Big)e^{-j\theta t}\cdot h\Big(t+\frac{1}{2}\tau\Big)\mathrm{d}t, \tag{5.2}$$

where h denotes a window function. By inserting expression 5.2 into 5.1, one can derive the distribution for the spectrogram as [91]

$$\begin{aligned} P_S(t,\omega) &= |S(t,\omega)|^2 \\ &= \left|\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-j\omega t'}s(t')h(t'-t)\,\mathrm{d}t'\right|^2. \end{aligned} \tag{5.3}$$

Eq. 5.3 means that the spectrogram is the short-time Fourier transform (STFT), with its absolute value squared, of $s(t')h(t'-t)$, which can be described as a locally windowed signal. The window function, $h(t)$, is centered at time instant t, and by altering t, a new part of the signal is "windowed" and transformed. Thus, by moving the window, Fourier transform the windowed signal and calculate the absolute value squared of the transform, a spectrogram will be the result.

For all bilinear distributions, i.e. including the spectrogram, the distribution of the sum of two signals does not equal the sum of the distributions corresponding to each signal [88]. Instead

$$\begin{aligned} P_{S,x+y}(t,\omega) =& P_{S,x}(t,\omega) + P_{S,y}(t,\omega) \\ &+ 2\,\mathrm{Re}\big(P_{S,xy}(t,\omega)\big). \end{aligned} \tag{5.4}$$

Thus, there is an additional interference term, $2\,\mathrm{Re}\big(P_{S,xy}(t,\omega)\big)$ (for more signals than two, eq. 5.4 can be generalized, but for simplicity the case with just two signals is studied). For the spectrogram, the interference is fairly limited if the spectrograms of the two signals do not extensively overlap [90].

For other bilinear distributions, such as the Wigner-Ville distribution,

$$W(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} s^*\Big(t-\frac{1}{2}\tau\Big)e^{-j\tau\omega}s\Big(t+\frac{1}{2}\tau\Big)\,\mathrm{d}\tau, \tag{5.5}$$

the interference is more prominent and can seriously limit the readability of the data [92]. However, because of the interference term, the Wigner-Ville distribution has many desirable properties, such as the ability to recover the instantaneous frequency of a signal. Additionally, the Wigner-Ville distribution perfectly localizes linear chirp signals [88–91].

One way to suppress the artifacts produced by the interference terms of the Wigner-Ville distribution is to use smoothing. The smoothing is implemented using double convolution according to

$$\begin{aligned} &W_{\text{smoothed}}(t,\omega) = \\ &\iint_{-\infty}^{+\infty} L(t-t',\omega-\omega')W(t',\omega')\,\mathrm{d}t'\mathrm{d}\omega', \end{aligned} \tag{5.6}$$

where L is a smoothing function and W is the distribution of the signal [89]. It is interesting to note that if L is the Wigner-Ville distribution of a time window function h, $L = W_h(t, \omega)$, expression 5.6 results in the spectrogram:

$$P_S(t,\omega) = \iint_{-\infty}^{+\infty} W_h(t-t', \omega-\omega') W(t',\omega')\,\mathrm{d}t'\mathrm{d}\omega'. \tag{5.7}$$

This expression can be generalized for other distributions, where the kernels satisfy $\phi(-\theta,\tau)\phi(\theta,\tau) = 1$, according to

$$P_S(t,\omega) = \iint_{-\infty}^{+\infty} P_h(t-t', \omega-\omega') P(t',\omega')\,\mathrm{d}t'\mathrm{d}\omega'. \tag{5.8}$$

Thus, according to eq. 5.7 the spectrogram can be constructed by smoothing the Wigner-Ville distribution. This confirms the fact that the interference terms of the spectrogram are suppressed (due to smoothing), which often results in better readability compared to the Wigner-Ville distribution.

However, the spectrogram suffers from the same limitations as the short-time Fourier transform, which can be described as a trade-off between the time and frequency resolution [88,90]. For adequate time resolution, a short duration time window must be used, but this will result in poor frequency resolution. For a better frequency resolution the time window must be longer, thus making it impossible to achieve adequate time resolution. This time-frequency resolution trade-off is a result of the well-known Heisenberg-Gabor inequality, $T \cdot B \geq 1$, where $T \cdot B$ is the time-bandwidth product [88].

Finally, one can note that for discrete time signals, distortion of the resulting distribution might occur due to aliasing. The aliasing of the discrete-time Wigner-Ville distribution has been extensively studied (see [94] and its listed references). Several "alias-free" distributions have been proposed, but it has been shown that many of them cause aliasing [94]. However, one method proposed by Nutall [95] gives truly alias-free discrete-time Wigner distributions. The discrete-time spectrogram can be considered alias-free if the sampling of the input signal and the time window is carried out using a sufficiently high sampling frequency [93].

5.1 Reassignment of the spectrogram

As discussed previously, the spectrogram suffers from some undesirable properties, especially when compared to the Wigner-Wille distribution. By recalling eq. 5.7, one can conclude that for a given t and for all frequencies, the spectrogram is a sum, or rather an integration with respect to the running time variable t', of

$$B(t,t',\omega) = \int_{-\infty}^{+\infty} W_h(t-t', \omega-\omega') W(t',\omega')\,\mathrm{d}\omega'. \tag{5.9}$$

This can be interpreted as a mean value of all "frequency smoothed" B-terms due to the fact that all distributions $P(t,\omega)$ are defined so that

$$\iint_{-\infty}^{+\infty} P(t,\omega)\,\mathrm{d}\omega\,\mathrm{d}t = \int_{-\infty}^{+\infty} |s(t)|^2\,\mathrm{d}t = 1. \tag{5.10}$$

Thus, the construction of a spectrogram can be interpreted as a process where a number of time averages of "frequency smoothed" B-terms are calculated and assigned to specified values of t. However, the resulting time averages might not be a good indicator of the time localization of the energy contained inside $B(t,t',\omega)$. Instead, one possibility to improve the spectrogram is to calculate the center of gravity of $B(t,t',\omega)$ and use the result to *reassign* the specified values of t to new values, t_R, according to:

$$\begin{aligned} t_R(t,\omega) &= t - \frac{1}{P_S(t,\omega)} \int_{-\infty}^{+\infty} t' B(t,t',\omega)\,\mathrm{d}t' = t - \\ &\frac{1}{P_S(t,\omega)} \iint_{-\infty}^{+\infty} t'\, W_h(t-t',\omega-\omega') W(t',\omega')\,\mathrm{d}\omega'\,\mathrm{d}t'. \end{aligned} \tag{5.11}$$

By using expression 5.7 one obtains

$$\begin{aligned} &t_R(t,\omega) = \\ &t - \frac{\iint_{-\infty}^{+\infty} t'\, W_h(t-t',\omega-\omega') W(t',\omega')\,\mathrm{d}\omega'\,\mathrm{d}t'}{\iint_{-\infty}^{+\infty} W_h(t-t',\omega-\omega') W(t',\omega')\,\mathrm{d}t'\mathrm{d}\omega'}. \end{aligned} \tag{5.12}$$

Instead of interpreting eq. 5.7 as the time averages of $B(t,t',\omega)$, one can claim that it represents the frequency averages of "time smoothed" D-terms, according to

$$P_S(t,\omega) = \int_{-\infty}^{+\infty} D(t,\omega,\omega')\,\mathrm{d}\omega', \tag{5.13}$$

where

$$D(t,\omega,\omega') = \int_{-\infty}^{+\infty} W_h(t-t',\omega-\omega') W(t',\omega')\,\mathrm{d}t'. \tag{5.14}$$

In similar manner as shown above, the center of gravity of $D(t,\omega,\omega')$ can be calculated. This leads to the reassignment of ω to ω_R according to

$$\begin{aligned} \omega_R(t,\omega) &= \omega - \frac{1}{P_S(t,\omega)} \int_{-\infty}^{+\infty} \omega' D(t,\omega,\omega')\,\mathrm{d}\omega' = \omega - \\ &\frac{1}{P_S(t,\omega)} \iint_{-\infty}^{+\infty} \omega'\, W_h(t-t',\omega-\omega') W(t',\omega')\,\mathrm{d}t'\,\mathrm{d}\omega' \\ &= \omega - \frac{\iint_{-\infty}^{+\infty} \omega'\, W_h(t-t',\omega-\omega') W(t',\omega')\,\mathrm{d}t'\,\mathrm{d}\omega'}{\iint_{-\infty}^{+\infty} W_h(t-t',\omega-\omega') W(t',\omega')\,\mathrm{d}t'\mathrm{d}\omega'}. \end{aligned} \tag{5.15}$$

The reassigned spectrogram, $P_{S,R}(t'',\omega'')$ is constructed by summing (by integration) all spectrogram values that have been reassigned to points (t'',ω'').

This can be expressed as

$$P_{S,R}(t'',\omega'') = \iint_{-\infty}^{+\infty} P_S(t,\omega)\,\delta\big(t'' - t_R(t,\omega)\big) \cdot \delta\big(\omega'' - \omega_R(t,\omega)\big)\,\mathrm{d}t\,\mathrm{d}\omega. \tag{5.16}$$

where δ denotes the Dirac impulse.

The resulting expression for $t_R(t,\omega)$ and $\omega_R(t,\omega)$, which are called reassignment operators, may seem a bit complicated to implement for practical computations. However, a first step towards a successful implementation was taken by Kodera et al. [96], where it was shown that the phase information, which is normally not considered when calculating spectrograms, can be used to derive the reassignment operators. By recalling expression 5.3,

$$\begin{aligned} P_S(t,\omega) &= |S(t,\omega)|^2 \\ &= \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-j\omega t'} s(t') h(t'-t)\,\mathrm{d}t' \right|^2, \end{aligned} \tag{5.17}$$

one can see that $S(t,\omega)$ contains phase information. Kodera et al. showed that if $S(t,\omega)$ is expressed in polar coordinates according to

$$S(t,\omega) = |S(t,\omega)|\, e^{j\cdot\arg(S(t,\omega))} = |S(t,\omega)|\, e^{j\Phi(t,\omega)}, \tag{5.18}$$

the reassignment operators can be written as

$$t_R(t,\omega) = -\frac{1}{2\pi}\frac{\partial\,\Phi(t,\omega)}{\partial\,\omega} \tag{5.19}$$

$$\omega_R(t,\omega) = \omega + \frac{1}{2\pi}\frac{\partial\,\Phi(t,\omega)}{\partial\,t}. \tag{5.20}$$

Even though eq. 5.19 and 5.20 are theoretically important, they are not computationally efficient. However, Auger and Flandrin [97] derived exact expressions that were easily implemented for discrete time signals:

$$\begin{aligned} t_R(t,\omega) &= t - \Re\left\{ \frac{\mathrm{STFT}_{\mathcal{T}h}(t,\omega)\cdot \mathrm{STFT}_h^*(t,\omega)}{|\mathrm{STFT}_h(t,\omega)|^2} \right\} \\ &= t - \Re\left\{ \frac{\mathrm{STFT}_{\mathcal{T}h}(t,\omega)}{\mathrm{STFT}_h(t,\omega)} \right\} \end{aligned} \tag{5.21}$$

$$\begin{aligned} \omega_R(t,\omega) &= \omega - \Im\left\{ \frac{\mathrm{STFT}_{\mathcal{D}h}(t,\omega)\cdot \mathrm{STFT}_h^*(t,\omega)}{|\mathrm{STFT}_h(t,\omega)|^2} \right\} \\ &= \omega - \Im\left\{ \frac{\mathrm{STFT}_{\mathcal{D}h}(t,\omega)}{\mathrm{STFT}_h(t,\omega)} \right\}, \end{aligned} \tag{5.22}$$

where $\mathrm{STFT}_h(t,\omega)$ equals $S(t,\omega)$ in eq. 5.3 and $\Re$ and $\Im$ are the real and imaginary parts respectively. $\mathcal{D}h$ simply means that the window function $h(t)$ in eq. 5.3 is replaced by its time derivative

$$\mathcal{D}h(t) = \frac{\partial\,h(t)}{\partial\,t}. \tag{5.23}$$

Similarly, $\mathcal{T}h$ implies that $h(t)$ is modified according to

$$\mathcal{T}h(t) = t \cdot h(t). \tag{5.24}$$

5.2 Reassignment and sound decay

Recalling figure 3.3, where a spectrogram of a speech signal is shown, computing the reassigned spectrogram of the same speech signal will produce the distribution shown in figure 5.1. As can be seen, the reassignment results in more concentrated signal energies, both in time and frequency, which generally means that the readability of the spectrogram is improved. However, the

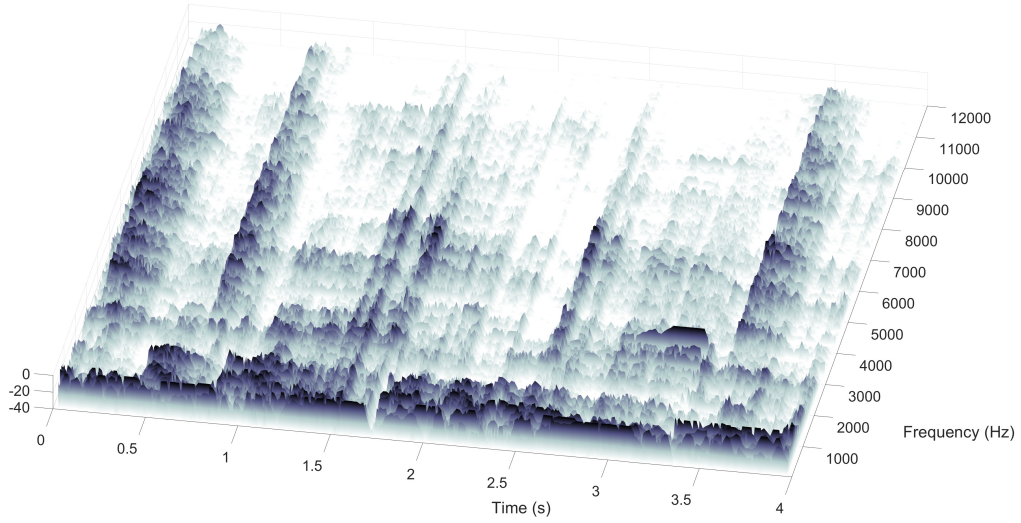

Figure 5.1: Reassigned spectrogram of a male speech signal. The SRS is close to instability.

reassignment process is sensitive to noise, i.e. if there is some noise in a certain time-frequency domain, the reassignment could move the original spectrogram points randomly or erroneously. Thus, for decaying signal components in the presence of noise or "noise-like" components (e.g. "broadband" sound decay in a room), the reassignment could alter the decay times of individual signal components. To study this in more detail, a 1 s synthesized signal with sampling frequency $f_s = 48$ kHz is generated according to

$$y(t) = \frac{y_{\text{wn}}(t)}{10^{(\text{SNR}/20)}} + \sum_{k=1}^{5} e^{-\delta_k t} \sin(2\pi f_k t), \tag{5.25}$$

where $y_{\text{wn}}(t)$ is white noise with the same signal power as the sum of sine signals with $\delta_k = 0$, SNR is the desired signal to noise ratio and δ_k and f_k are

defined according to

$$\begin{cases} \delta_1 = 6.91/0.1 \text{ s}, & f_1 = 2000 \text{ Hz} \\ \delta_2 = 6.91/0.3 \text{ s}, & f_2 = 6000 \text{ Hz} \\ \delta_3 = 6.91/0.5 \text{ s}, & f_3 = 10000 \text{ Hz} \\ \delta_4 = 6.91/0.7 \text{ s}, & f_4 = 14000 \text{ Hz} \\ \delta_5 = 6.91/0.9 \text{ s}, & f_5 = 18000 \text{ Hz}, \end{cases} \tag{5.26}$$

i.e. the -60 dB decay times vary from 0.1 s (for $k = 1$) up to 0.9 s (for $k = 5$) in 0.2 s steps. An example of a spectrogram with SNR = 30 dB is shown in figure 5.2. The reassigned version is shown in figure 5.3. Isolating the relevant

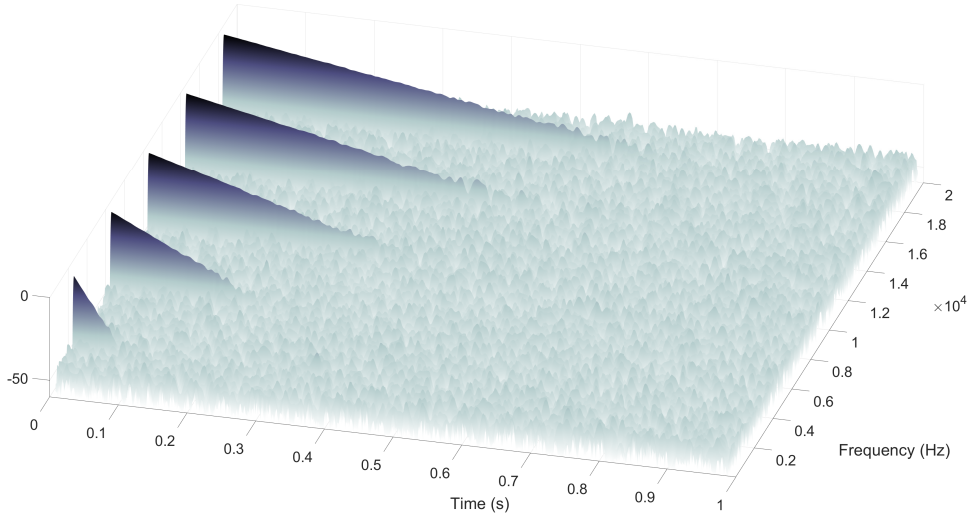

Figure 5.2: Spectrogram of the synthesized signal with SNR = 30 dB.

"frequency slices" and offsetting each decay curve by 0.2 s (to improve the readability of the plot) result in figure 5.4. The differences in the amount of ripple is especially obvious below -20 dB. The same can be seen in the full spectrogram plots shown in figure 5.2 and 5.3.

For a curve fitting procedure, the ripple introduced by the reassignment will in most cases cause an underestimation of the decay time. The result of curve fitting using raw data and Schroeder Backward Integration (SBI), with and without the added "tail", is shown in figures 5.5 - 5.7. Three SNRs are evaluated, 15, 30 and 45 dB. The results confirm that the spectrogram more accurately estimates the decay times for all SNRs, especially if no SBI is used. Therefore, the ordinary spectrogram was chosen for the "coloration detector" computations outlined in paper D.

5.3 All-pole modeling and reassignment

In an attempt to improve the SNR of reassigned spectrograms and therefore reduce the ripple introduced by the reassignment discussed above, all-pole

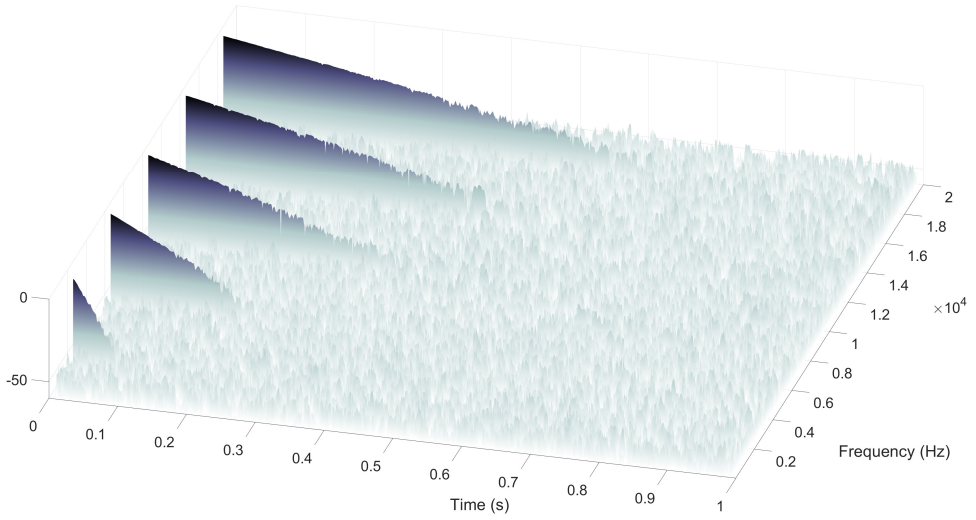

Figure 5.3: Reassigned spectrogram of the synthesized signal with SNR = 30 dB.

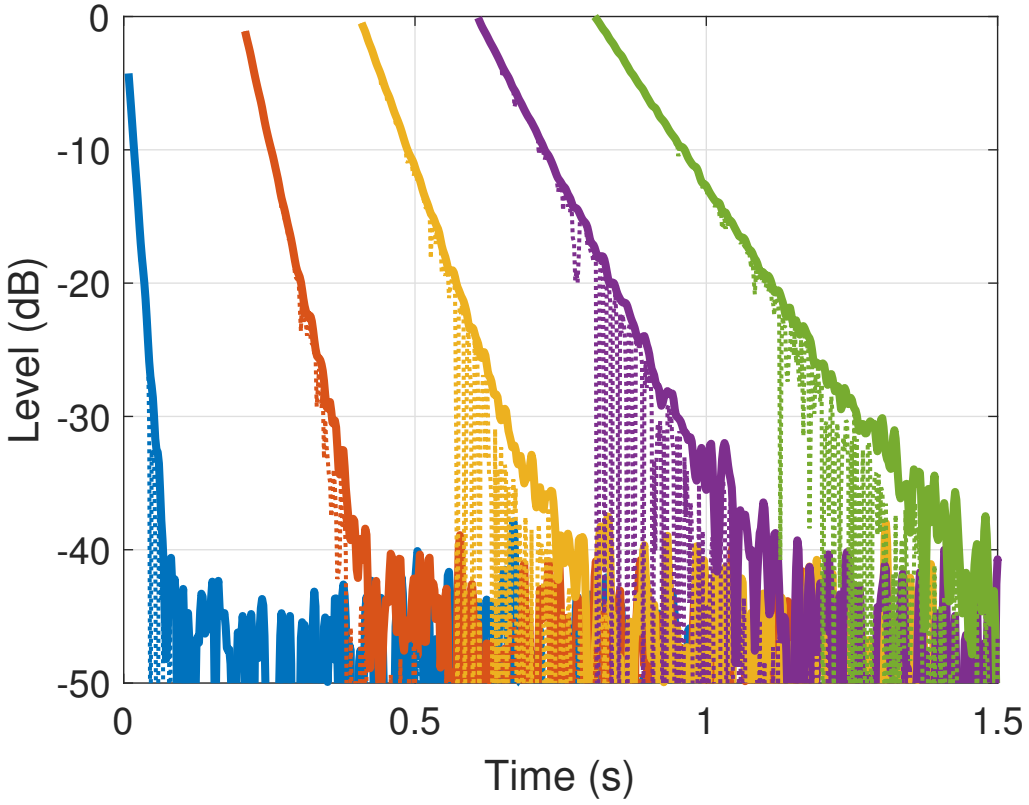

Figure 5.4: Comparisons between the ordinary (solid line) and reassigned (dotted line) spectrogram. The input signal is the synthesized signal with SNR = 30 dB and only the five relevant frequency slices have been plotted. Each decay curve is offset by 0.2 s to improve the readability.

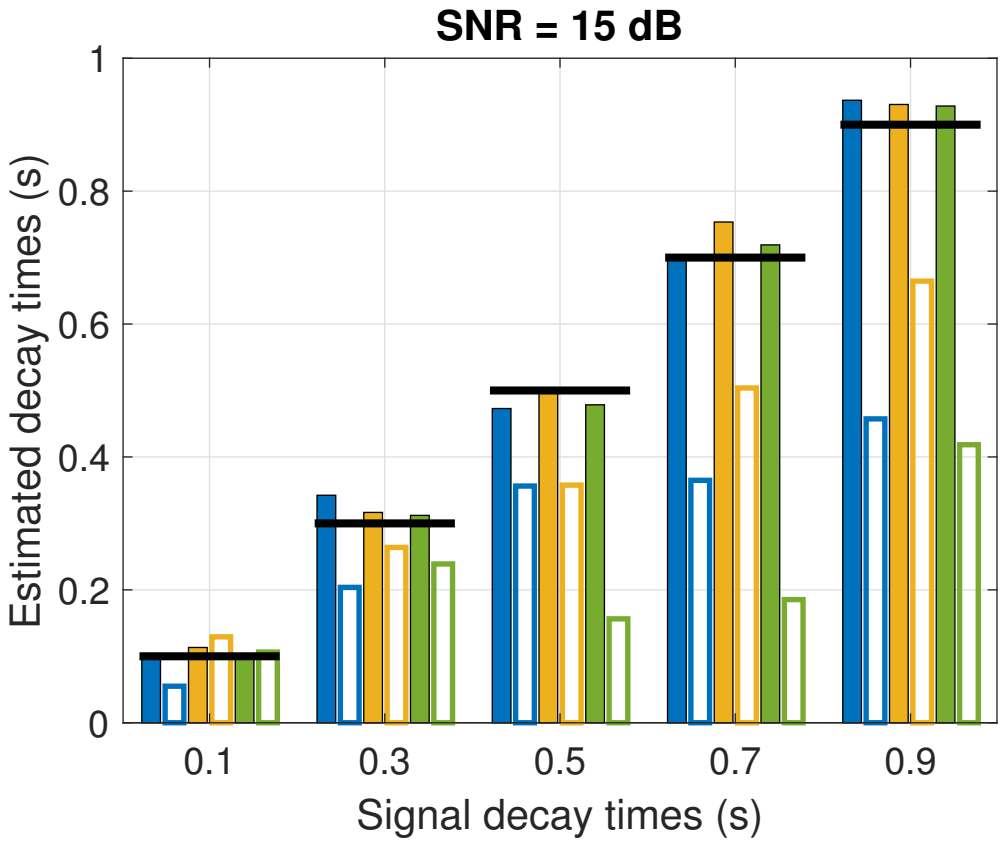

Figure 5.5: Estimated -60 dB decay times for SNR = 15 dB. In each group, from left to right, the bars represent curve fitting using: spectrogram - raw data (solid), reassigned spectrogram - raw data (hollow); spectrogram - SBI data (solid), reassigned spectrogram - SBI data (hollow); spectrogram - SBI data including "tail" (solid), reassigned spectrogram - SBI data including "tail" (hollow).

modeling is implemented in paper B. For SNRs ranging from -15 to 30, a 40 dB improvement in distribution based SNR (SNRTFD) could be achieved by the all-pole modeling. However, computing all-pole models for arbitrary speech and music signal segments could be challenging due to potentially strong modulation effects in both amplitude and frequency. Also, the signal complexity varies heavily over time. This means that estimating the correct model order for each signal segment, which is of fundamental importance, is non-trivial. Thus, for the coloration detector, the all-pole modeling presented in paper B, was abandoned. However, to the knowledge of the author, reassignment operators for all-pole modeled signals have not been published elsewhere. Also, several computational tools for time-frequency analyses are developed, which are discussed more in chapter 5.4.

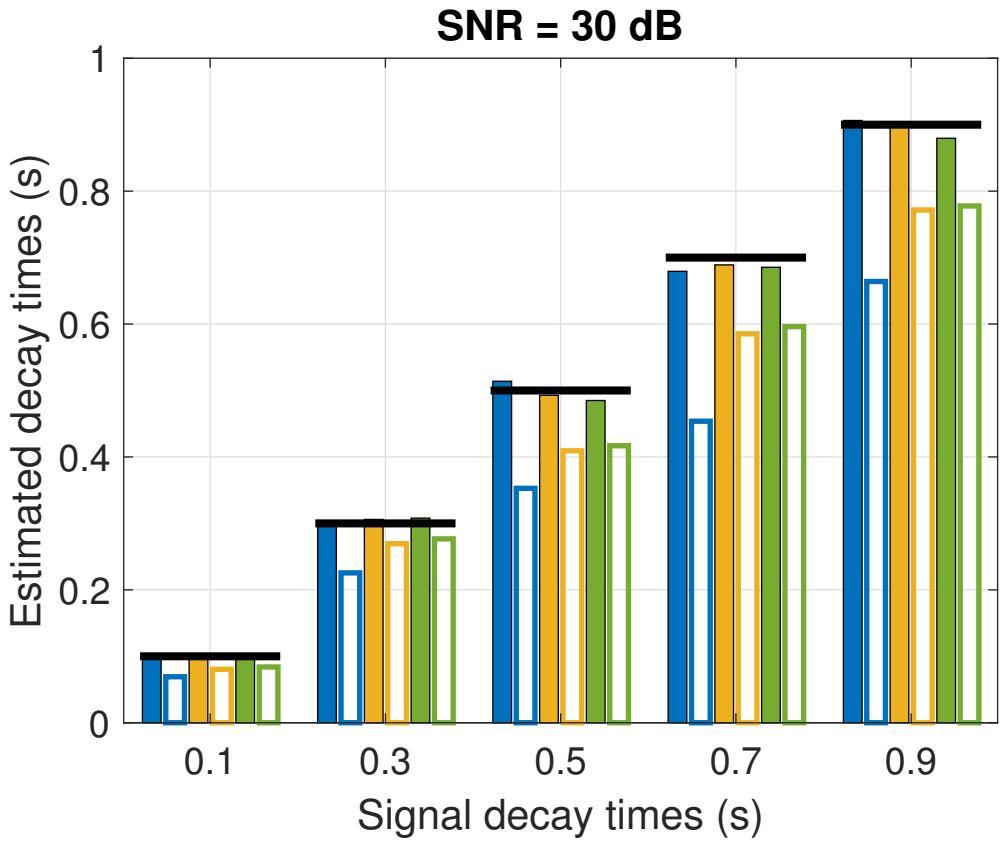

Figure 5.6: Estimated -60 dB decay times for SNR = 30 dB. In each group, from left to right, the bars represent curve fitting using: spectrogram - raw data (solid), reassigned spectrogram - raw data (hollow); spectrogram - SBI data (solid), reassigned spectrogram - SBI data (hollow); spectrogram - SBI data including "tail" (solid), reassigned spectrogram - SBI data including "tail" (hollow).

5.4 Objective measures for analysing time-frequency distributions

In general, time-frequency distributions are plotted to illustrate e.g. specific signal properties or differences between two signals. However, interpreting time-frequency plots is potentially treacherous, since a correct interpretation depends on using optimum "plot floor" (or dynamics), color scale (or viewing angle if 3D plotting), interpolation and time-frequency resolution. Even if all parameters for plotting are optimum, the interpretation of the resulting figure is still subjective and important features can be misinterpreted.

Therefore, using various objective measures to analyze distributions could be helpful. Eight such measures are presented in paper C. The measures are inspired by well-known signal quantifiers such as the SNR and Q-value. A short summary is given here:

1. Valid Peak Point Percentage (VPPP): given a set of reference points, this measure indicates the success of peak detection.

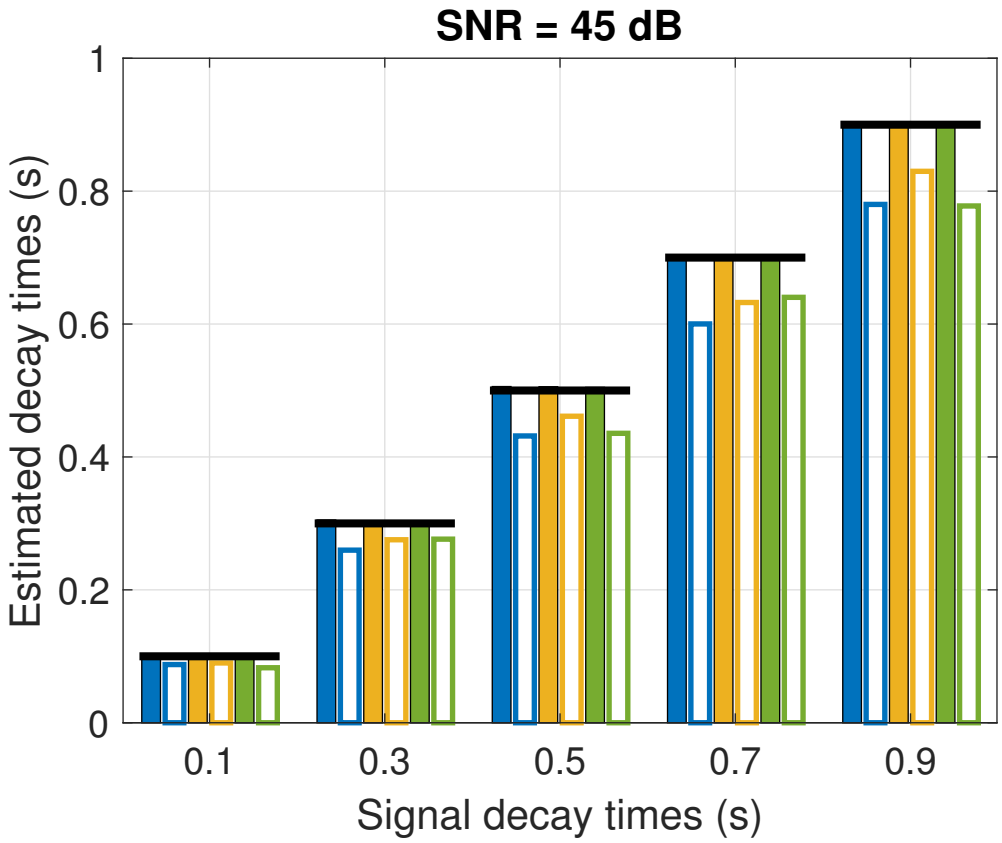

Figure 5.7: Estimated -60 dB decay times for SNR = 45 dB. In each group, from left to right, the bars represent curve fitting using: spectrogram - raw data (solid), reassigned spectrogram - raw data (hollow); spectrogram - SBI data (solid), reassigned spectrogram - SBI data (hollow); spectrogram - SBI data including "tail" (solid), reassigned spectrogram - SBI data including "tail" (hollow).

2. Average Frequency Deviation of a TFD (AFDTFD): estimates the average frequency deviation of points obtained from the peak detection, relative a set of reference points.

3. Reference based Signal-to-Noise Ratio of a TFD (RSNRTFD): using valid peak points (see 1. above), the SNR of the TFD is estimated.

4. Reference based Q-value of a TFD (RQTFD): an estimation of the average Q-value of all TFD signal points using valid peak points.

5. The Signal On-time Difference of a TFD (SODTFD): the difference in time (s) between the total number of reference points and valid peak points.

6. (Absolute) Q-value of a TFD (QTFD): same as 4. above, but only using points obtained from the peak detection (i.e. without reference).

7. (Absolute) Signal-to-Noise Ratio of a TFD (SNRTFD): same as 3. above, but only using points obtained from the peak detection (i.e. without reference).

8. The Signal On-time of a TFD (SOTFD): since the reference points are missing, only the time corresponding to all detected peak points can be computed.

In addition to the measures, a method for peak detection in the time-frequency domain is described. The method is crucial for the operation of the coloration detector.

Chapter 6

Simulating RESs and SRSs

A six channel system has been simulated using the iterative method described in [1]. As previously mentioned, one important difference is that the system simulated in this chapter and paper D is time-invariant and that the entire audible frequency range is used. A simulated system is chosen due to high repeatability, high precision, and obvious practical advantages over a corresponding real system placed inside a large room often used for rehearsals. Apart from the most obvious risk of disturbing the system configuration, including microphone and loudspeaker positions, the acoustical properties of the room are dependent on temperature, the configuration of absorbers and diffusors and the placement of furniture (typically chairs) and other equipment, all of which are difficult and time consuming to set up identically for each new system measurement.

6.1 The transfer function measurements

Before simulations can be performed, a number of transfer functions must be measured in a real room. The room chosen for this task is a "rectangular" symphony orchestra rehearsal room with dimensions 12 $\times$ 20 $\times$ 6.3 m (see figure 6.2 for details). As shown in figure 6.1, several musical instruments and other objects are placed along the room walls. Special drapes are installed to control the reverberation time. During all measurements, the drapes along wall $y = 0$, drawn in figure 6.2, were lowered to minimize the reverberation time. This made it possible to use a high sampling frequency during the measurments.

The x and y coordinates (width and depth) of the system microphones and loudspeakers are randomized according to figure 6.2 (for detailed information, see table 6.1). In order to excite as much reverberation energy as possible, the system loudspeakers and microphones are mounted on high stands. The method resembles the one described in [3]. The active loudspeakers, L1-L5, (Genelec 1029A) are facing the ceiling and are approximately 2.5 m above the ground. LP is facing the listener, i.e. its purpose is to reinforce the direct sound. It is mounted slightly higher than L1-L5.

Figure 6.1: The rehearsal room. The KEMAR dummy head and the drapes covering parts of wall $y = 0$ are visible.

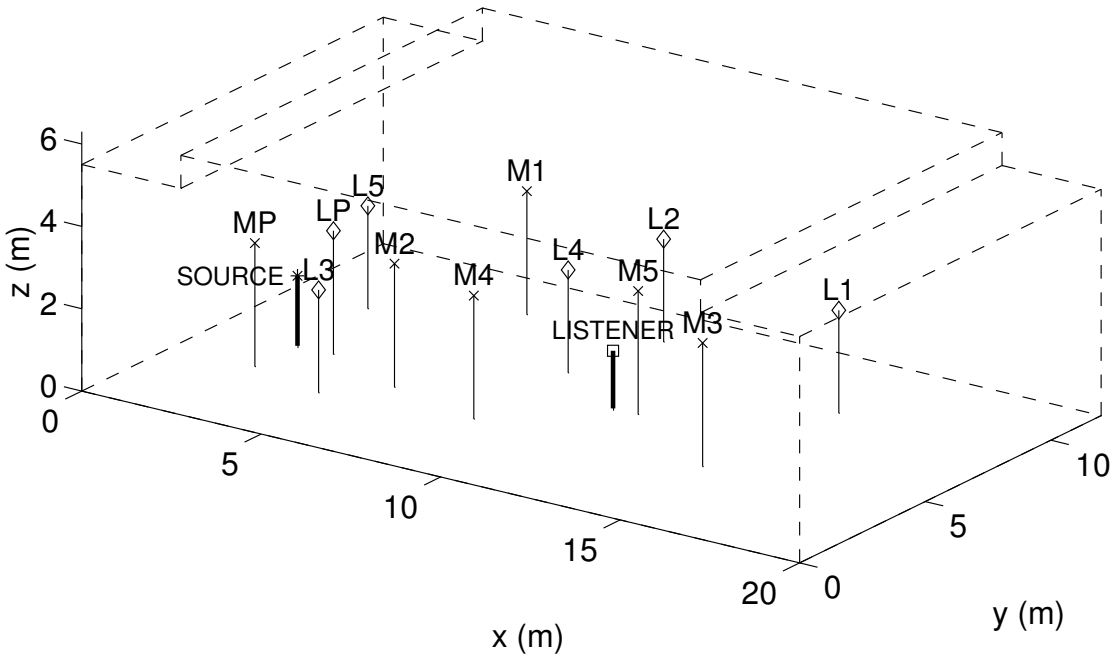

Figure 6.2: The positions of the system loudspeakers (diamonds), microphones (crosses), source (star) and listener (square).

The omni-directional microphones, MP & M1-M5, (AKG C451E) are facing the ground and are positioned about 3 m from the floor surface. The signal

Table 6.1: Coordinates of system components.

System component	x (m)	y (m)	z (m)	Amplified and connected to
M1	5.4	10	3	L1
M2	5.7	4.3	3	L2
M3	14.5	4	3	L3
M4	8.4	3.6	3	L4
M5	11.3	6	3	L5
MP	2.3	3.6	3	LP
L1	15	8.7	2.5	-
L2	9	10.3	2.5	-
L3	4.5	3	2.5	-
L4	8.4	7.36	2.5	-
L5	2.2	8.24	2.5	-
LP	3.3	5.3	3	-
Source	2.3	5.3	1.7, 1.5	-
Listener	10.6	6	1.4	-

from M1 is amplified and sent *to only one loudspeaker*, L1 in this case. All channels are constructed in the same way, with matching microphone and loudspeaker indices.

As seen in the figure 6.2, there is one "source" and one "listener". The listener is a KEMAR dummy head equipped with Sennheiser microphones in its ears, positioned at normal, seated listening height. The source is set up as follows:

Speaker: in an attempt to create a directivity pattern that resembles a human speaker, a Genelec (about 1.7 m above the ground), configured according to figure 6.3, is used. The material covering the woofer is glass fibre wool (1.5 cm thick), with a 3.5 × 2.5 cm opening.

Musical source: a custom built omni-directional loudspeaker (1.5 m above ground) is used for simulating a musical instrument.

The transfer functions were measured using a portable computer and the MLSSA software. An M-Audio DMP3 microphone pre-amplifier supplied phantom power to the microphones. The following transfer functions were measured using a sampling frequency of 60606 Hz, which corresponds to a 20 kHz cutoff frequency:

$H_{\mathbf{LM}ij}(\omega)$ (from loudspeaker i to microphone j)**:** all possible combinations, with i and j having a range of 1-5 (integers),P (a total of $6 \cdot 6 = 36$ transfer functions).

$H_{\mathbf{S}i\mathbf{R}j}(\omega)$ (from source to receiver)**:** the transfer function between both source types (speaker and musical) and the dummy head (using one ear at a time-thus a total of 4 measurements). i and j range from 1 to 2.

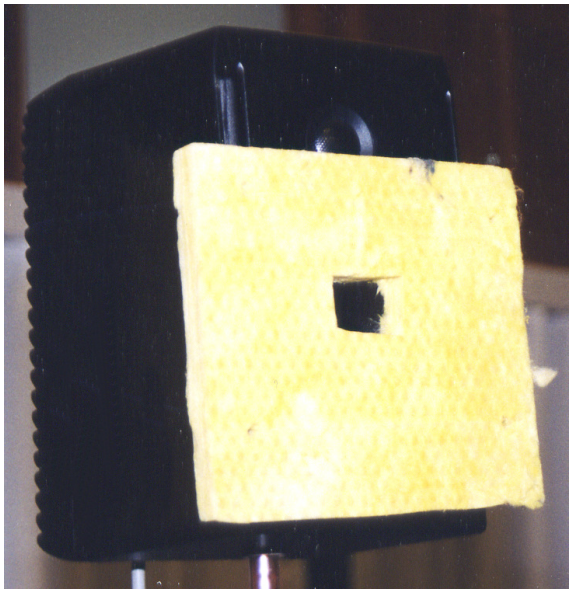

Figure 6.3: Modification of Genelec loudspeaker.

$H_{\mathbf{S}i\mathbf{M}j}(\omega)$ (from source to microphones)**:** the transfer function between both source types (speaker and musical) and the system microphones, with i having a range of 1-2 and j 1-5,P (a total of 12 measurements).

$H_{\mathbf{L}i\mathbf{R}j}(\omega)$ (from loudspeakers to receiver)**:** the transfer function between system loudspeakers and the dummy head using one ear at a time, i.e. a total of 12 measurements. The ranges of i and j are the same as indicated in the above text.

For each transfer function group listed above, the loudspeaker/source volume and microphone pre-amp level were fixed.

6.2 Calculations

Please note that in the following text, MP and LP, will be denoted as M6 and L6. For each channel, the user will specify a desired mean loop gain (MLG). Recalling eq. 3.2, MLG is generally defined as

$$\text{MLG} = \overline{|G(\omega)H_{\text{LM}}(\omega)|^2}. \tag{6.1}$$

Let MLGi denote the desired MLG for channel i in decibels. Next, modify $H_{\text{LM}ii}(\omega)$ according to

$$H_{\text{LM}ii,\text{MOD}}(\omega) = \sqrt{\frac{10^{\text{MLG}i/10}}{\overline{|H_{\text{LM}ii}(\omega)|^2}}} \cdot H_{\text{LM}ii}(\omega) \tag{6.2}$$

With this modification,

$$10 \cdot \log\left(\overline{|H_{\text{LM}ii,\text{MOD}}|^2}\right) = \text{MLG}i \text{ (dB)}. \tag{6.3}$$

Thus, $H_{\mathrm{LM}ii}(\omega)$ has been scaled so that its MLG is MLGi.

After specifying MLGi, the first step of the iteration process can be performed. Let $G_i(\omega)$ represent the amplification of channel i:

$$G_i(\omega) = \sqrt{\frac{10^{\mathrm{MLG}i/10}}{|H_{\mathrm{LM}ii}(\omega)|^2}}. \tag{6.4}$$

The first iteration step, which gives $Y_{0,i}(\omega)$, a part of the total signal fed to loudspeaker L_i, can be written as

$$Y_{0,i}(\omega) = G_i(\omega)X_{0,i}(\omega), \tag{6.5}$$

where $X_{0,i}(\omega)$ is the Fourier Transform of the source signal convolved with H_{SMi} (assume source type S).

The second step represents the feedback path, the sound going from loudspeaker L_i to all microphones, $\mathrm{M}_1 - \mathrm{M}_6$. The signal reaching microphone M_i is calculated as

$$X_{1,i}(\omega) = \sum_{n=1}^{6} Y_{0,i}(\omega)H_{\mathrm{LM}ni}(\omega). \tag{6.6}$$

All 36 H_{LM} transfer functions are used in order to obtain 6 new input signals. This can also be represented by matrices as described in [1].

The next step is to repeat step 1, but with the new input signals, $X_{1,i}(\omega)$:

$$Y_{1,i}(\omega) = G_i(\omega)X_{1,i}(\omega). \tag{6.7}$$

Now, step 2 can be written as

$$X_{2,i}(\omega) = \sum_{n=1}^{6} Y_{1,i}(\omega)H_{\mathrm{LM}ni}(\omega). \tag{6.8}$$

The procedure is repeated until the following expression is true for all 6 channels:

$$10\log\frac{|Y_{k,i}(\omega)|^2}{|Y_{i,\mathrm{TOT}}(\omega)|^2} < L_{\mathrm{stop}}, \tag{6.9}$$

where

$$Y_{i,\mathrm{TOT}}(\omega) = \sum_{k=0} Y_{k,i}(\omega). \tag{6.10}$$

In the time domain, each new term in expression 6.10 represents a shift corresponding to the distance between loudspeakers and microphones. Due to the FFT convolutions in eqs. 6.5 - 6.8, and the fact that the FFT size is fixed at 2^{19} or 2^{20} in all calculations, there is a risk of circular convolutions. In the time domain, a convolution of two (non-zero) signals, M and N samples long respectively, results in a signal of length $P = M + N - 1$. Thus, if P is longer than 2^{20} samples, the FFT operation must "squeeze" all P samples into 2^{20} points, which is impossible, and results in circular convolutions. To avoid this, one can either zero pad or truncate the output after each FFT convolution.

Zero padding is not a good option, since the length of the signals will increase after each convolution, resulting in longer computational times and shortage of system memory. Therefore, a truncation time of 1.5 s, which is far longer than the reverberation times of all impulse responses h_{LMij}, is used. The truncation is carried out in the time domain as described in [1].

6.3 The sound arriving at the dummy head

In order to obtain the signal picked up by the dummy head, each $Y_{i,\mathrm{TOT}}(\omega)$ is first multiplied by the corresponding $H_{\mathrm{L}i\mathrm{R}j}(\omega)$ and then summed. j varies from 1 to 2 and indicates the left and right ear, resulting in 6 transfer function for each ear. Thus, the resulting sum for ears 1 and 2 are

$$Y_{\mathrm{EAR1}}(\omega) = H_{\mathrm{SR1}}(\omega) + b \cdot \sum_{i=1}^{6} Y_{i,\mathrm{TOT}}(\omega) H_{\mathrm{L}i\mathrm{R1}}(\omega) \tag{6.11}$$

$$Y_{\mathrm{EAR2}}(\omega) = H_{\mathrm{SR2}}(\omega) + b \cdot \sum_{i=1}^{6} Y_{i,\mathrm{TOT}}(\omega) H_{\mathrm{L}i\mathrm{R2}}(\omega), \tag{6.12}$$

where $H_{\mathrm{SR}j}$ represents the direct sound and b a constant which is used to vary the energy of the system sound in relation to the direct sound energy.

For a certain system gain setting, MLGi, and N active channels, the sound energies

$$E_{\mathrm{dir}} = \sum_{n} \left| \sum_{j=1}^{2} H_{\mathrm{SR}j}(\omega_n) \right|^2, \tag{6.13}$$

$$E_{\mathrm{sys}} = \sum_{n} \left| \sum_{j=1}^{2} \sum_{i=1}^{6} Y_{i,\mathrm{TOT}}(\omega_n) H_{\mathrm{L}i\mathrm{R}j}(\omega_n) \right|^2 \tag{6.14}$$

and

$$\frac{E_{\mathrm{sys},pr}}{E_{\mathrm{dir},pr}} = \frac{N S_{\mathrm{avg}}^2}{1 - N S_{\mathrm{avg}}^2}, \tag{6.15}$$

where $S_{\mathrm{avg}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} 10^{\mathrm{MLG}i/10}}$, are determined, so that b can be written as

$$b = \sqrt{\frac{E_{\mathrm{sys}}}{E_{\mathrm{dir}} \cdot (E_{\mathrm{sys},pr}/E_{\mathrm{dir},pr})}} \cdot 10^{(G_{\mathrm{sys}}/20)}. \tag{6.16}$$

Eq. 6.15 represents the predicted (highlighted by pr in the subscripts) energy according to diffuse field theory of the system relative to the direct sound [1]. Thus, eq. 6.16 allows a scaling of the system sound energy so that $E_{\mathrm{sys}}/E_{\mathrm{dir}} = E_{\mathrm{sys},p}/E_{\mathrm{dir},p}$. Additionally, this ratio can be modified by the "system sound gain", G_{sys}, which is important for the PA channel (channel 6). Its loudspeaker is facing the listener, i.e. eq. 6.15 is underestimating the system energy. This means that a $G_{\mathrm{sys}} = 3$ dB compensation is used when the PA channel is active

and has the same mean loop gain as the other channels. In general, the scaling that eq. 6.15 introduces is important when creating sounds for listening tests. If a compensation is not done by using b, i.e. if $b = 1$ in eq. 6.11 and 6.12, the level of the direct sound in relation to the system sound will be unrealistic. This is because different microphone preamplifier gain settings had to be used during the measurements of $H_{\mathrm{L}i\mathrm{R}j}$ and $H_{\mathrm{SR}j}$ and the gain settings could not be recorded with sufficient precision.

Chapter 7

Some coloration detection attempts

Before arriving at the level of refinement presented in paper D, extensive work was required to develop and evaluate several different coloration detection algorithms and coloration measures. The later attempts incorporate the concept of damping distributions, where the damping constants of the room are estimated using the source material and compared to a reference distribution (see paper D, sections II and III for more details). For example, in one of the methods, $H_{\text{est}}(k_h, t_m)$ (see eq. 22 in paper D) is fitted to the chi-square distribution, $p_\chi(x \mid \overline{\bar{\delta}(\omega_{k_r}, t_m)})$ for each envelope maximum at a fixed system gain setting. Here, $\overline{\bar{\delta}(\omega_{k_r}, t_m)}$ is the mean value of all valid decay constants belonging to a certain envelope maximum. Then the approximation error is computed as a function of system gain setting. The hypothesis is that the approximation error would increase with increasing system gain. The results, however, indicate no such relationship. This is explained by the fact that using $\overline{\bar{\delta}(\omega_{k_r}, t_m)}$ in the chi-square distribution effectively minimizes the approximation error, regardless of gain settings and envelope maxima.

As a second attempt, $H_{\text{est}}(k_h, t_m)$ is computed for each envelope maximum followed by averaging of all distributions within a certain time interval according to eq. 23 in paper D. The chi-square distribution is then fitted to the resulting average distribution, $\overline{H_{\text{est}}(k_h)}$, by computing the median value, $\mu_{1/2}$, of both distributions. Since the interval of the decay constants is limited, the resulting distributions are "truncated". The median values are computed by numerical integration of the distributions. The truncated chi-square distribution is shifted so that its median value matches that of $\overline{H_{\text{est}}(k_h)}$ *at the lowest system gain.* An example is shown in figure 7.1. The dashed blue line represents a stable system at MLG $= -60$ dB with a median value of $\mu_{1/2} = 12.97$. The solid black line is the chi-square distribution with a matching 12.97 median value. Now, the aligned chi-square distribution is considered as the reference distribution and the median value, denoted $\mu_{1/2}^{\text{ref}}$ is a key component for the next computational step.

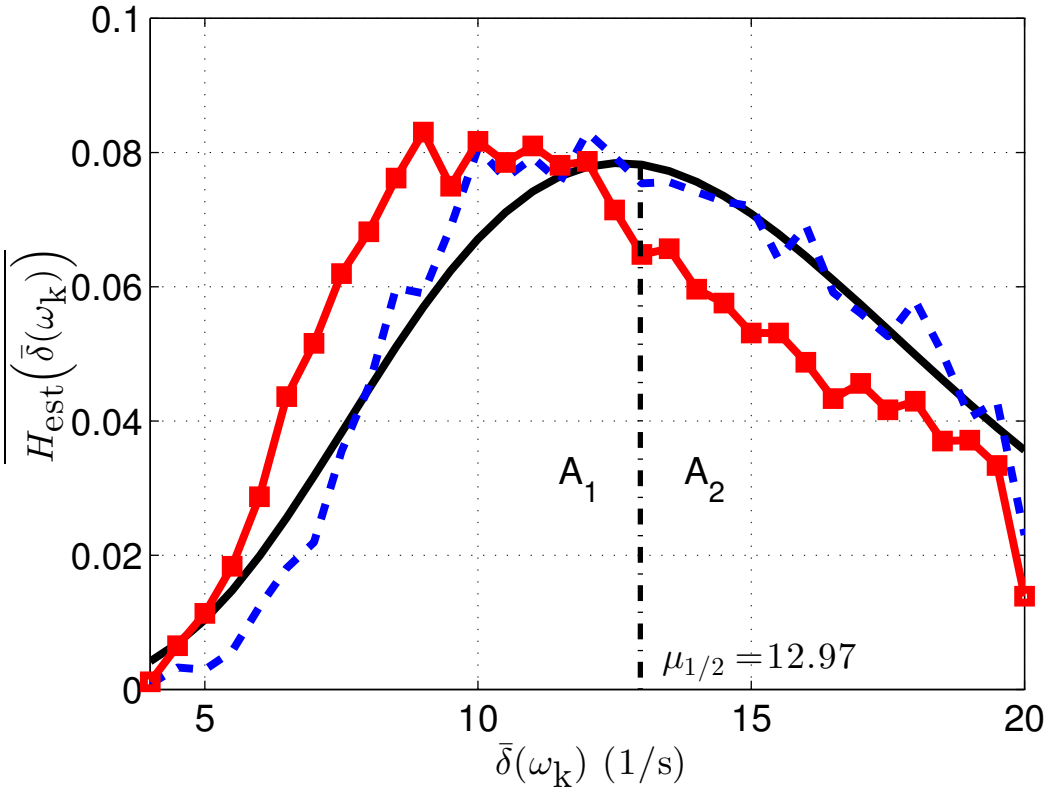

Figure 7.1: An example of how the subareas A_1 and A_2 are defined using $\overline{H_{\text{est}}(\bar{\delta}(\omega_k))}$ and the median value corresponding to a stable system. The dashed blue line represents $\overline{H_{\text{est}}(\bar{\delta}(\omega_k))}$ for a stable system at MLG $= -60$ dB, the red solid-square line a system close to instability and the black solid line the chi-square distribution with the same median value as found for the stable system (12.97 in this example).

For increasing system gain settings, each resulting $\overline{H_{\text{est}}(k_h)}$ is integrated numerically (by summation) below and above $\mu_{1/2}^{\text{ref}}$, resulting in two subareas A_1 and A_2 as shown in figure 7.1:

$$A_1 = \frac{1}{N_{\mathbf{\Psi}}} \sum_{k \in \mathbf{\Psi}} \overline{H_{\text{est}}(k_h)} \tag{7.1}$$

and

$$A_2 = \frac{1}{N_{\mathbf{\Gamma}}} \sum_{k \in \mathbf{\Gamma}} \overline{H_{\text{est}}(k_h)}, \tag{7.2}$$

where the sets of indices Ψ and Γ are defined so that

$$\bar{\delta}(\omega_{\mathbf{\Psi}}) \leq \mu_{1/2}^{\text{ref}} \tag{7.3}$$

and

$$\bar{\delta}(\omega_{\mathbf{\Gamma}}) > \mu_{1/2}^{\text{ref}} \tag{7.4}$$

hold true. For the reference distribution and for $\overline{H_{\text{est}}(k_h)}$ at minimum system gain, $\mu_{1/2}^{\text{ref}}$ represents the point of equal areas, which implies that $A_1 = A_2$. The

idea for the sound coloration measure proposed here is to use the same median value, $\mu_{1/2}^{\text{ref}}$, to define the subareas of $\overline{H_{\text{est}}(k_h)}$ as the system gain increases, even though the actual median value of each new distribution is different from $\mu_{1/2}^{\text{ref}}$. This means that A_1 will differ more and more from A_2 as the system gain increases. The red solid-square line in figure 7.1 represents a system close to instability and as can be observed, A_1 is now substantially larger than A_2. By plotting $10 \cdot \log_{10}(A_1/A_2)$ as a function of system gain, the ratios between the subareas can be studied, starting from 0 dB at low system gains. As the gain increases, the amount of damping (or attenuation) in the system will decrease causing an increasing number of components with longer decay times. Thus, in general, it is expected that $A_1 \geq A_2$ and that the ratio will increase with increasing system gain. For paper D, however, the concept of subareas was omitted, since the distribution median turned out to be simpler and more robust (for the median based coloration measure, refer to eq. 24 in paper D).

A third attempt incorporates exploring the distribution of the sound pressure amplitude measured in an arbitrary room (sinusoidal excitation), which follows the Rayleigh distribution and is independent of several key room properties: acoustical qualities, volume, shape and type [4]. Thus if an SRS alters the impulse response so that it becomes "unnatural", the response will deviate from the Rayleigh distribution. Thus, it is of interest to try this approach for coloration detection.

Based on the estimated decay constants, the goal is to estimate the distribution of the sound pressure amplitude, i.e. the FFT of the impulse response. Here Parseval's theorem is useful:

$$\int_{-\infty}^{\infty} |x(t)|^2 \, \mathrm{d}t = \int_{-\infty}^{\infty} |X(\omega)|^2 \, \mathrm{d}\omega, \tag{7.5}$$

because $|x(t)|^2$ is estimated in eq. 9 in paper D. Thus, for one frequency slice,

$$|X(k)|^2 = \int_0^{\infty} B_k^2 \mathrm{e}^{-2\delta_k t} \, \mathrm{d}t = \frac{B_k^2}{2\delta_k} \tag{7.6}$$

or

$$X(k) = \frac{B_k}{\sqrt{2\bar{\delta}_k}}, \tag{7.7}$$

where B_k and δ_k define the amplitude and decay constants of the decay curve within bin k that gives the best match to the actual decay curve within bin k. By expressing the frequency dependence of B_k and δ_k as $B(\omega)$ and $\delta(\omega)$, eq. 7.7 leads to the following estimate of the sound pressure amplitude:

$$P_{\text{est}}(\omega) = \frac{B(\omega)}{\sqrt{2\delta(\omega)}}. \tag{7.8}$$

The next step is to compare the distribution of $P_{\text{est}}(\omega)$ to the Rayleigh distribution, which is defined as follows:

$$P_{\text{RL}}(z) = \frac{\pi z}{2} e^{\frac{-\pi z^2}{4}}, \tag{7.9}$$

where z is the absolute value of $P_{\text{est}}(\omega)$ divided by its frequency average. The distribution of $P_{\text{est}}(\omega)$ is computed using the histogram function in Matlab:

$$P_{\text{HIST}} = \text{hist}\Big(\frac{P_{\text{est}}(\omega)}{\bar{P}_{\text{est}}(\omega)}\Big), \tag{7.10}$$

where $\bar{P}_{\text{est}}(\omega)$ is the frequency average of $P_{\text{est}}(\omega)$. An example of the Rayleigh distributions discussed here is shown in figure 7.2. As seen in the figure, the distribution based on the estimated decay constants and amplitudes matches the corresponding FFT distribution. The theory behind the Rayleigh distributed sound pressure amplitudes will be discussed in more detail in the following text.

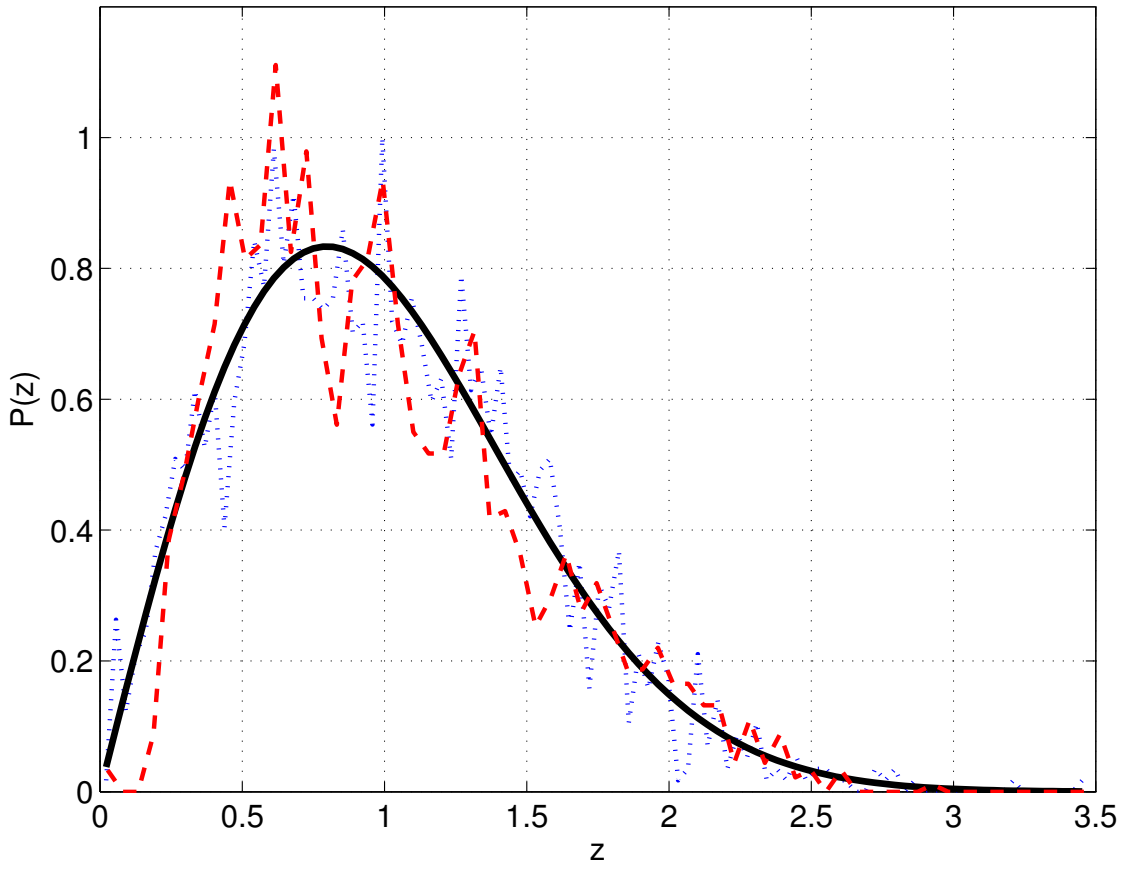

Figure 7.2: The theoretical Rayleigh distribution (solid line), the distribution based on the FFT of a measured room impulse response (dotted line) and the distribution based on the estimated amplitudes and mean decay times of the same measured room impulse response as for the dotted line (dashed line).

7.1 The Central Limit Theorem and time domain windowing

The fact that the sound pressure amplitude measured in a room (sinusoidal excitation) is Rayleigh distributed is a result of the Central Limit Theorem (CLT). In

essence, the CLT states that the sum of a large number of mutually independent random variables will be normally distributed [6]. If $x_1(k), x_2(k), \ldots, x_N(k)$ denote N such random variables, the sum random variable, $s(k)$, in

$$s(k) = \sum_{i=1}^{N} c_i x_i(k) \tag{7.11}$$

will be normally distributed when $N \to \infty$. The CLT holds even if the individual distributions of the random variables are unspecified and different and weighted by the fixed arbitrary constants c_i.

In the frequency domain, the sound pressure amplitude is expressed as [4]

$$P(\omega) = \sum_{n} \frac{A_n}{\omega^2 - \omega_n^2 - 2j\delta_n\omega_n}, \tag{7.12}$$

where A_n are functions of ω, the source position and the receiver position. Viewing eq. 7.12 purely mathematically, it is clear that it can be expressed as

$$P(\omega) = (a_1(\omega) + a_2(\omega) + \ldots + a_N(\omega)) + j(b_1(\omega) + b_2(\omega) + \ldots + b_N(\omega)), \tag{7.13}$$

i.e. the sum of a large number of complex numbers, $a + jb$. According to [4], the magnitudes of ω_n and δ_n are changing in such an irregular manner between the current and next eigenfrequency that each term in eq. 7.12 or eq. 7.13 can be considered as mutually independent. Thus, the CLT can be applied, i.e.

$$\text{hist}\Big[\sum_{i=1}^{N} a_i(\omega)\Big] \to N(\mu_a, \sigma_a^2) \tag{7.14}$$

and

$$\text{hist}\Big[\sum_{i=1}^{N} b_i(\omega)\Big] \to N(\mu_b, \sigma_b^2). \tag{7.15}$$

If the real and imaginary parts of $P(\omega)$ are normally distributed, one can show that $|P(\omega)|$ is Rayleigh distributed.

Since the computation of the damping constants is based on spectrograms, it is of interest to investigate if the CLT still holds for a windowed impulse or frequency response. During the computation of the spectrogram, the signal is windowed in the time domain by multiplying the signal, $s(t')$, with a window function shifted by time t, $h(t' - t)$: $s(t') \cdot h(t' - t)$. In the frequency domain, the time shift of the window function corresponds to multiplying the Fourier transform of the window function, $H(\omega)$, by $e^{-j\omega t}$, i.e. $h(t' - t)$ corresponds to $e^{-j\omega t} \cdot H(\omega) = H_t(\omega)$. Also, multiplication in the time domain corresponds to convolution in the frequency domain. In fact, circular convolution will be used so that the length of the resulting vector will be the same as the frequency response vector (the actual computations are done in the discrete frequency domain). Now, let us assume that the Fourier transform of the signal, $S(\omega)$, is described by eq. 7.12. Then, the frequency domain equivalent of the spectrogram of an impulse response can be written as

$$P_S(t, \omega) = |S(t, \omega)|^2 = |H_t(\omega) \circledast P(\omega)|^2. \tag{7.16}$$

The circular convolution in eq. 7.16, represented by the "$\circledast$", will be computed by frequency domain windowing and integration (summation), where the shifted (by time t) window function, $H_t(\omega)$, is multiplied by the portion of the frequency response covered by $H_t(\omega)$ (typically, the bandwidth of the window function, $H_t(\omega)$, is much smaller than the bandwidth of $P(\omega)$). The "scaled" frequency response values are then integrated (summed) followed by a frequency shift of $H_t(\omega)$ after which the multiplication and integration are repeated. To illustrate this in the discrete frequency domain, a window consisting of three frequency points, $[H_t(1), H_t(2), H_t(3)]$, and a frequency response consisting of five points, $[P(1), P(2), P(3), P(4), P(5)]$, are convolved in table 7.1. From table 7.1, it

Table 7.1: Example of circular convolution.

Freq. bin	1	2	3	4	5
Window	$H_t(1)$	$H_t(2)$	$H_t(3)$		
Freq. resp.	$P(1)$	$P(2)$	$P(3)$	$P(4)$	$P(5)$
Freq. resp.	$H_t(1) \cdot \mathbf{P\ (1)}$	$H_t(1) \cdot P(2)$	$H_t(1) \cdot P(3)$	$H_t(1) \cdot P(4)$	$H_t(1) \cdot P(5)$
Shift: 0 bins	+	+	+	+	+
Freq. resp.	$H_t(2) \cdot P(5)$	$H_t(2) \cdot \mathbf{P\ (1)}$	$H_t(2) \cdot P(2)$	$H_t(2) \cdot P(3)$	$H_t(2) \cdot P(4)$
Shift: 1 bin	+	+	+	+	+
Freq. resp.	$H_t(3) \cdot P(4)$	$H_t(3) \cdot P(5)$	$H_t(3) \cdot \mathbf{P\ (1)}$	$H_t(3) \cdot P(2)$	$H_t(3) \cdot P(3)$
Shift: 2 bins					
Column sums:	$= S(t,1)$	$= S(t,2)$	$= S(t,3)$	$= S(t,4)$	$= S(t,5)$

is clear that the spectrogram at time t, is computed by frequency domain windowing and summation. The windowing is essentially carried out by scaling certain frequency response points with the window. The number of terms in the resulting sums is defined by the bandwidth of the window function (three frequency points in the above example). In the above text, it was shown that the real and imaginary parts of $P(\omega)$ are normally distributed. Is this still the case for the real and imaginary parts of $S(t,k)$? Assuming M terms, each sum leading to $S(t,k)$ has the form

$$S(t,k) = \sum_{i=1}^{M} H_t(i) \cdot P(\mathbf{I}_k(i)), \tag{7.17}$$

where the notation $\mathbf{I}_k$ represents the indices valid for frequency bin k (there are three terms and indices in the above example). Since both $H_t(k)$ and $P(k)$ are complex, eq. 7.17 can be written as

$$\begin{aligned} S(t,k) = \sum_{i=1}^{M} \Bigg[& \Re\Big[H_t(i)\Big] \cdot \Re\Big[P(\mathbf{I}_k(i))\Big] - \Im\Big[H_t(i)\Big] \cdot \Im\Big[P(\mathbf{I}_k(i))\Big] \\ & + j\left(\Re\Big[H_t(i)\Big] \cdot \Im\Big[P(\mathbf{I}_k(i))\Big] + \Im\Big[H_t(i)\Big] \cdot \Re\Big[P(\mathbf{I}_k(i))\Big] \right) \Bigg], \end{aligned} \tag{7.18}$$

where the real and imaginary parts of the complex variables are denoted by $\Re$ and $\Im$ respectively. Assuming that the $\Re\Big[P(\mathbf{I}_k(i))\Big]$ and $\Im\Big[P(\mathbf{I}_k(i))\Big]$ terms in eq. 7.18 form $2M$ independent and normally distributed random variables (the

distributions may differ) and that $\Re\Big[H_t(i)\Big]$ and $\Im\Big[H_t(i)\Big]$ are "constants" in eq. 7.11 , applying the CLT to $\Re\Big[S(t,k)\Big]$ and $\Im\Big[S(t,k)\Big]$ leads to the answer that both quantities are normally distributed. In other words, when computing the spectrogram of an impulse response, the frequency data in each "time slice" should be Rayleigh distributed - similar to the FFT of an entire impulse response. It is also assumed that the width of $H_t(k)$ is around 15 frequency bins or more so that the CLT is applicable. This means that N in eq. 7.11 is at least 30 ($M = 15$, $2M$ independent random variables).

Chapter 8

Implementing the final coloration detector

The final coloration detector is described in more detail in paper D, section III. Thus, only a short summary will be given here. Starting with the general structure of the detector, which is shown in figure 8.1, one can note that there are eight main blocks:

1. The input signal is defined. In practice, a microphone is placed where the sound field is highly diffuse.

2. A TFD, or Time-Frequency Distribution, of choice is computed. Here, both the ordinary and reassigned spectrogram have been considered.

3. For each spectrogram frequency bin, RMS filtering (sliding window method) is applied to smooth the decay curve ripples, which simplifies the subsequent decay detection.

4. For each spectrogram frequency bin, decays are detected and validated. For example, decays with insufficient dynamics are discarded.

5. For all valid decays, curve fitting is applied to the Schroeder backward integrated decays. The slopes of the fitted lines result in a set of damping constants, generally one set per bin.

6. Histogram analysis of decay onset times is applied to find "broadband decays", which are analyzed for harmonic contents. When detected, all harmonically related components are removed. The underlying idea is that sound decays from musical instruments or other sources should be removed from the following damping distribution analyses, because the goal is to analyze the acoustics of the room, not its sources.

7. The mean damping distribution is computed using the distributions corresponding to each "broadband decay".

8. The sound coloration is estimated based on the mean damping distribution and a reference distribution (chi-square in this case). Both the median value and shape of the distribution are considered.

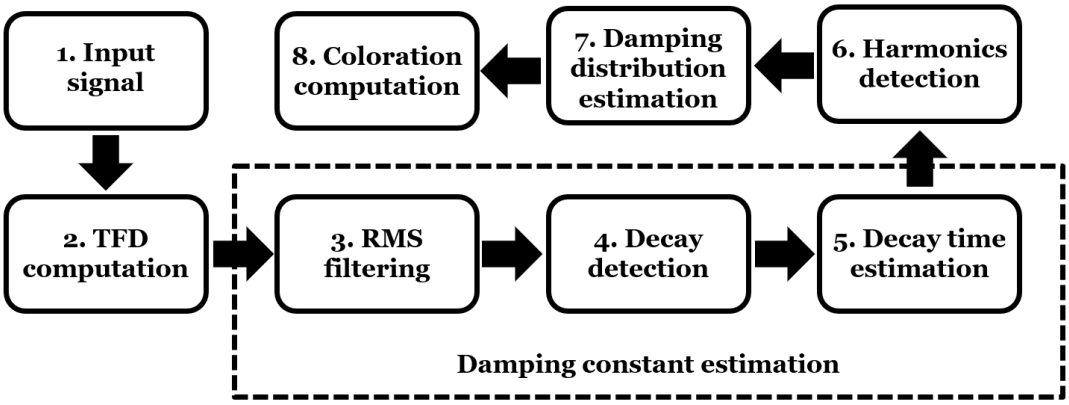

Figure 8.1: The general layout of the coloration detector.

Currently, the code for each block is not optimized when it comes to computational efficiency or future DSP implementations. The detector will probably not operate in real-time within a foreseeable future (due to numerous complex computations), but true real-time operation is not needed if the detector works as intended. A well-functioning coloration detector can sample the program material quite sporadically, say every 10-20 s, and inform the sound engineer about the current status. Therefore, a first prototype of the detector could apply post-processing to a recently recorded signal and repeat the process continuously. Also, true real-time operation is questionable, since all damping constant estimations belonging to a certain envelope maximum must be computed before any coloration data can be produced. This, of course, depends on the reverberation time of the auditorium, which means that "real-time" would translate to "the time instant after a set of damping constants have been computed".

Most computations presented in this thesis are based on a simulated RES, using measured RIRs. The microphones picking up the signal for the coloration detector are the KEMAR dummy head microphones. It is assumed, however, that an ordinary measurement microphone would work equally well in a practical application. In most cases, the positioning of the microphone will be more critical than the microphone type. As previously mentioned, one of the key assumptions used when deriving the damping distribution is a highly diffuse sound field. Therefore, the microphone must be placed away from room surfaces, main PA loudspeaker arrays, direct sound sources, etc.

For the simulated system, the coloration detector works for both speech and music. However, this will not always be the case. The more "impulse like" the source is, the better. Also, the time interval between these impulses must allow for some room reverberation and sound decay so that the damping constants can be computed. If either condition is not fully met, the performance of the

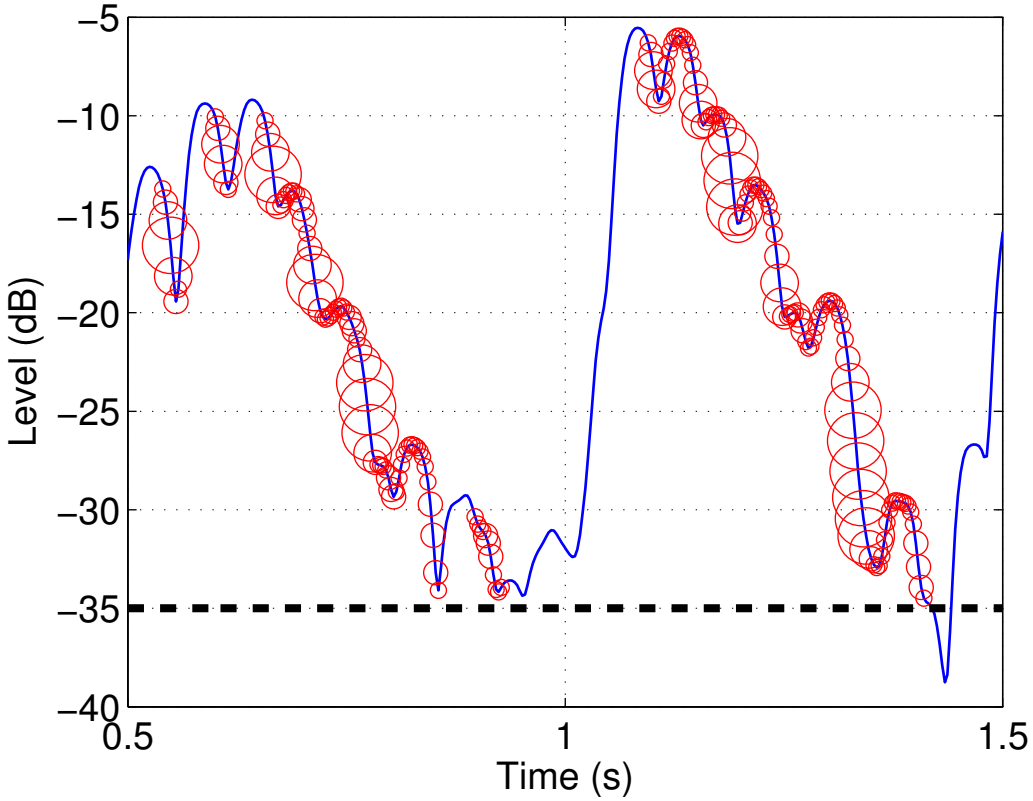

Figure 8.2: Decay detection of a TFD frequency slice at $f = 1464.2$ for a speech signal.

detector will degrade. Some examples are:

- Musical passages with virtually no dynamics.
- Musical passages containing few notes.
- Quiet (or muffled) and very fast speech.

It is important that the source is positioned correctly and powerful enough to excite the reverberant sound field of the room and produce a tolerable SNR. On the other hand, if the source is too dominating, problems will also occur.

For the decay detection in block 4, which is based on a purely numerical procedure, it is interesting to note the resulting redundancy. An example of the decay detection is shown in figure 8.2. The circles illustrate the stored decay indices and the radii of the circles show the occurrences of each index. Hence, it is obvious that there will be some redundancy for the decay detection and its stored indices. Only unique indices are kept and sequences of consecutive integers are identified. Each sequence is assumed to represent a decay with arbitrary dynamics. The beginning and end of each sequence are stored and decays with less than 10 dB dynamics are omitted.

Consecutive decays which are close in time and level are omitted to avoid erroneous multiple slope decays (due to e.g. ripples). In this case, only the first decay, having the strongest start and stop levels, is saved. Finally, only

broadband decays are saved, i.e. for a certain decay in one frequency bin, there must be several "parallel" decays in other bins with roughly similar start times. The criterion for this is computed using the histogram of all remaining start indices.

8.1 Real rooms

Worth mentioning is that a real SRS was set up in a lecture hall in order to record sound samples for coloration analyses. The same four sound samples as in paper D (section III D) were used and fed to a small active loudspeaker, which acted as the sound source. However, instead of the six independent and randomly positioned loudspeaker-amplifier-loudspeaker channels of the simulated system, just one such channel was used for the measured SRS. In other words, the sound from the "source loudspeaker" was picked up by just one microphone, amplified, and sent to two PA loudspeakers mounted on medium height stands. The "listener microphone" (i.e. the microphone responsible for recording the sound samples used for coloration analyses) was positioned in the audience area at sufficient distance from the nearest surfaces.

As for the simulated system, equalization is applied so that the system sounds uncolored at low gain settings. This is verified by performing measurements and frequency response analyses using white noise. All signals are routed through a Tascam FW-1804 audio interface and controlled by a DAW software (Magix Samplitude). For each source sound, the GBI is determined by listening to the system and carefully adjusting the system gain until the ringing artifacts slowly fade out. Finding the actual GBI, where the ringing should be sustained at a constant level, proved to be very sensitive in practice. For example, restoring the GBI setting after a few recordings at lower gain often lead to an unstable system. A similar gain range as for the simulated system, i.e. 42 dB, is used. For each source sound, a total of 17 measurements are performed with decreasing gain steps as the system approaches instability.

With this quite basic one channel setup, the reverberant sound field was poorly excited, resulting in less successful estimations of $\overline{H_{\text{est}}(k_h)}$. Also, the signal to noise ratio of the recorded sound files turned out to be a major issue and the use of noise reduction algorithms was discarded due to the unknown effects they would introduce to the coloration detection.

Chapter 9

General discussion

The focus of this thesis is to develop sound coloration measures which are independent of measured references such as impulse responses or source sounds. Instead, it relies on known reference distributions of damping constants. One can argue that the reference distribution must be measured in some cases, because the theoretical distributions (e.g. the chi-square or gamma distribution) can not cover all different room types. In practice, however, the process of estimating the reference distribution will be the result of merely running the coloration detector at low system gains with normal program material (e.g. speech or music) as input.

Since all aspects of the coloration detector have been developed from square one, several key questions had to be answered during the development process. This led to detailed explorations in fields such as psychoacoustics, signal processing and time-frequency analysis accompanied by many years of trial and error and testing new computational strategies. Therefore, the work presented in this thesis may appear somewhat shifting. However, as the following text will show, the main goal has always been clear.

The required time resolution of a coloration detector is one key question which is addressed in paper A, since published research discussing this particular topic remains unknown to the author. The time resolution can, however, be linked to the audibility of tonality in pure tones, which has been studied by numerous authors (see e.g. [12] - [15]). More specifically, unmasked decaying pure tones should correspond to single feedback components in an SRS close to instability, which, arguably, is the best case scenario for detecting sound coloration in real SRSs. The hypothesis is that the resulting signal durations, which correspond to the just audible tonality (JAT), will lead to the required time resolution.

In paper A, the trends observed in Fig. 7, i.e. increasing JAT times for decreasing frequencies, are also reported in Doughty and Garner [14]. The reason for the observed behaviour at lower frequencies is the fact that at least two periods of the signal must be presented to the listener ("click-pitch") [14]. At low frequencies, the JAT times are around 20–23 ms, which corresponds to signal decay times of about 13–16 ms. At higher frequencies, the JAT times

decrease, while the number of required signal periods increases substantially (refer to figure 9 in paper A). The lowest times of around 3 ms are observed for the Gaussian signals.

The shorter JAT times for the Gaussian signals can be explained by their comparatively narrower bandwidths. For all JAT times displayed in paper A, Fig. 7, the fast Fourier transforms (FFT) of the corresponding stimuli have been derived. By estimating the -3 dB bandwidths from the spectra, the Q-values have been computed and the result is shown in paper A, Fig. 8. For the exponential signals, the Q-values are generally lower for the 95 phon signals. Similar results are reported in Doughty and Garner [14] for pure tone bursts (rectangular window). For essentially all frequencies, the Gaussian signals produce higher Q-values compared to the exponential signals. At 4800 and 7000 Hz, however, the Q-values of only the 95 phon Gaussian signals are higher.

In general, tonality is related to signal Q-values [16], which are based on FFTs of the entire time signals. For short duration signals, high Q-values, which means small bandwidths, imply more tonality when the slope of the spectrum becomes comparable to the slope of the excitation pattern [17]. This slope varies depending on level and excitation type. However, the steepness of the slope generally decreases with increased signal level. This might empirically explain why the Q-values are lower for test 1 (exponential signals, 95 phon max loudness) compared to test 2 (exponential signals, 70 phon max loudness). At higher signal levels, the Q-values are allowed to be somewhat wider due to the broadening of the excitation pattern and decreased slope steepness.

As pointed out in chapter 4.2, it would be interesting to dive deeper into the "gray area" related to loudness, i.e. the loudness of short duration pure tones having bandwidths larger than one critical band. Following such research, the loudness compensation implemented in paper A could be modified and the listening tests repeated. Also, it would be interesting to re-estimate the headphone compensation once more due to the constantly improving quality of headphones, digital-to-analogue converters, amplifiers and measurement systems including dummy heads. Naturally, this also implies using more modern equipment for all listening tests.

Following the estimation of the coloration detector time constant, the attention turned to time-frequency analyses. At an early stage during the development of the coloration detector, it became clear that it was necessary to track signals both in time and frequency. Since time-frequency reassignment had been reintroduced and made computationally effective by Auger and Flandrin [97], reassignment seemed to be an interesting option. However, the reassignment is sensitive to low SNRs. Therefore, in paper B, the first natural step is to investigate the noise sensitivity and possible improvements, especially by all-pole modeling. Additionally, in order to conduct such an investigation, several objective measures for time-frequency analysis are suggested.

In general, all-pole modeling can improve the SNRTFD (signal to noise ratio of a TFD) of the reassigned spectrogram by up to 40 dB. Although a promising result, the signal model order is hard to estimate, rendering all-pole modeling unusable for the signals studied in paper B. Additionally, as shown in

chapter 5.2, time-frequency reassignment is not suitable for signal decay time estimation. In spite of this, the objective measures proposed in paper B turned out to be robust and correlated well with the visual impressions of the TFDs. Thus, in paper C, the number of measures are expanded to a total of eight (from three in paper B) in two groups (absolute and relative measures) and each measure is systematically tested. The computational examples display how the suggested objective measures perform and also demonstrate how the measures can be used to analyze TFDs and different signal components within the TFDs. Overall, the proposed measures perform as expected and appear robust when using a (noise contaminated) synthesized signal. The analyses show that certain features of the TFDs, not easily distinguished by visual inspection, are revealed by the measures. Naturally, the more general characteristics, easily detected by visual inspection, are also quantified using the measures.

Based on the above explorations into time-frequency analysis, the foundation of the coloration detector presented in paper D is defined, i.e. the required time resolution and choice of TFD. Despite the promising properties of reassignment and (all-pole) signal modeling, neither of them are used in the detector. Instead, the "ordinary" spectrogram is chosen. For the coloration detector, robustness is challenging due to the vast array of computational parameters. Since the chi-square distribution was found to be a suitable reference for low system gains, it is natural to begin tuning computational parameters of the implementation to obtain the best fit to the distribution. For the computations presented in paper D, the TFD window length and overlap are important parameters as both will introduce shifts in distribution median values. Another key parameter, which will cause similar shifts, is the rms averaging filter length of the time signal contained within each frequency bin as mentioned in Sec. III.A, step 3. Setting a value somewhat lower or higher than 30 TFD time steps, corresponding to 90 ms, results in minor shifts in the distribution median values and slight deviations from the chi-square shape. The estimation of the slopes of the decaying signals also introduces some uncertainty depending on the complexity of the signal envelopes. Some estimations simply fail if the signal envelopes are far from linear, if plotted as level (dB) vs time (s).

For the assumptions mentioned in paper D, Sec. II and lightly damped systems with real-valued mode shapes, Burkhardt and Weaver [7] found theoretically that the damping distribution should be a chi-square distribution. Their numerical simulations of a vibrating membrane with viscous dampers in distinct nodes resulted in deviations from the chi-square shape for cases with high modal overlap. However, as long as the assumptions, and especially the approximation of real-valued mode shapes, are valid, there are no apparent theoretical limitations of their model. In fact, the studied rehearsal room produces a chi-square damping distribution when the modal overlap is significant.

For the case of higher damping, resulting in the necessity to assume complex mode shapes, the gamma distribution model by Schroeder [8] could be a potential candidate. However, the studied rehearsal room showed a poorer fit to the gamma distribution than the chi-square distribution - even when tuning computational parameters to favor the gamma distribution (e.g. shorter rms averaging filter length). For rooms not satisfying the necessary assumptions for

a chi-square or gamma distribution, other distributions might be applicable. For the general case, a reference distribution could be found by using the result of the method at low system gain. However, for the rehearsal room in this paper, it is clear that the chi-square distribution works well. Interestingly, a method to estimate distributions of damping constants in rooms and displaying excellent fit to a theoretical distribtion, has not been published elsewhere (to the author's knowledge).

As shown in paper D, the suggested coloration detector works well using a simulated SRS. However, it is clear that additional real systems and rooms should be tested. Given the underlying theory and method, described in section II and III of paper D, the detector is expected to perform well for other rooms and systems. In fact, it is difficult to argue for a complete failure of the detector. Altering the natural distribution of the sound decay, i.e. the distribution of damping constants, in a room by introducing acoustic feedback should always register in the detector if its hardware and software are set up and implemented correctly. Thus, the more relevant research questions should focus on detection sensitivity and robustness, which relate to "tuning" the detector by altering various computational parameters and/or steps. When optimally tuned (as in paper D), the coloration detector is very sensitive and robust. For real systems and rooms, however, the performance will depend on the retuning of the detector.

Unique to the suggested detector, and one of its major advantages, is that no measured reference is required. Eliminating (continuous) measurements of source sounds or impulse responses will improve the robustness of the suggested coloration detector, because measured quantities always introduce some level of uncertainty while being more or less intrusive during live performances. It is clear that the possibility to use theoretical reference distributions is superior and will improve the robustness of the coloration classification. Even if a theoretical distribution is less suitable for the room in question, the suggested detector can always compute the reference distribution at low system gains.

Naturally, the ultimate extension of the work presented in this thesis would be to implement the "real-time" system mentioned earlier. Then, it would be possible to test the coloration detection in real systems and rooms using just a single microphone, computer, sound card and software. Also, it would be possible to optimize or "tune" the computational steps and parameters to improve the robustness of the detector for various real systems and rooms. Included in the "tuning" is to investigate the reference distributions of various real rooms and how they relate to the theoretical distributions. As a final step, the detector could be used by sound engineers as a complementary (and highly experimental) tool for sound quality and coloration monitoring during live performances. Although a complex task, building a "real-time" detector would be feasible. Matlab Simulink, Max MSP and Python are some examples of suitable software. Hardware wise, the requirements would be a fairly powerful laptop, a microphone with high SNR and a soundcard with low noise preamplifiers. In other words, nothing out of the ordinary would be required.

Another interesting extension of the work would be to link the audible

sound coloration to the coloration measures. Different SRSs, rooms and source signals should be considered, thus making this a complex task. The results, however, would be valuable for sound engineers using the detector, since the risk of audible sound coloration could be estimated.

Chapter 10

Conclusions

Starting with the time constant of the coloration detector, it is shown in Paper A that tonality is just audible for total signal durations as low as 2.6 ms. However, JAT times around 3 ms only exist for high frequencies (above 3400 Hz). For lower frequencies, the JAT times increase, from around 5 ms at 3400 Hz to 20–23 ms at 150 Hz. The Gaussian stimuli produce the lowest JAT times. This is explained by the time-varying Q-values, which are higher for the Gaussian stimuli, over all time instants, than for the exponential stimuli. Thus, for a Gaussian signal, the auditory system is excited by a high-Q signal during a longer effective duration. The use of an attack function, which generally reduces the levels of clicks, also contribute to higher time-varying Q-values. This partly explains why the JAT times presented in paper A are lower than the corresponding results in related papers.

In paper B, where the main goal is to improve the readability of noise contaminated reassigned spectrograms, it is shown that improvements are feasible using all-pole modeling. Using two test signals, one speech and one synthesized (i.e. the sum of four sinusoids with different properties), the apparent visual improvements are verified objectively using the proposed SNRTFD, which indicates improvements up to 40 dB for the synthesized signal and at least 40 dB for the speech signal. However, as shown in chapter 5.2, the reassigned spectrogram is found to be less accurate for estimating decay times, which is a fundamental part of the coloration detector. As a consequence, the ordinary spectrogram is selected for the coloration detector.

The measures developed in paper B are refined further in paper C, where a set of "intuitive" objective measures for basic analyses of time-frequency distributions is proposed. Two additional measures are introduced in paper C, the Q-value of a TFD and the signal on-time difference. The measures reveal and confirm a number of interesting differences between the two distributions, some of which are virtually impossible to detect and/or quantify by visual inspection alone. For example, the signal-to-noise ratios of the spectrogram (S) and reassigned spectrogram (RS) are very similar, the RS has problems with the AM signal components, the RS exhibits superior frequency accuracy for SNRs between -5 and 20 dB and the Q-values of the TFD signal components

are approximately doubled for the RS.

The contributions from papers A-C are vital for the core operation of the coloration detector, which is the main contribution of this thesis. In paper D, the possibility to blindly compute coloration measures based on estimated damping distributions is explored. Two measures are proposed, one shaped based, which has the chi-square distribution as a reference, and one median based, which has the distribution median of a system with MLG = −60 dB as reference. Both measures clearly indicate sound coloration as the system gain increases, typically 12–22 dB from system instability depending on source sound. In general, speech results in the highest absolute coloration measure values. Comparably, initial listening tests reported in paper D suggest that coloration is audible at around 5.7 dB (average over all sounds) from system instability.

Being able to extract the distributions just by analyzing speech or music signals (i.e. live program material) picked up by a microphone placed in the audience area offers several advantages. Perhaps the most obvious one is the existence of theoretical distributions, which enables the classification of "coloration strength" and, as mentioned several times in the above text, should eliminate the need for other references (at least for certain room types). For the simulated SRS and the four source signals studied in paper D, the method achieves an almost perfect fit to the chi-square distribution for the lowest MLG. For increasing MLGs, the damping distributions of all sounds tend to deviate further and further from the reference distribution, which clearly indicates sound coloration. Additionally, the deviations occur for a relatively large range of MLGs, for which sound coloration is inaudible, suggesting great potential for professional audio applications such as hidden RESs. Thus, it is concluded that the suggested approach and its implementation successfully detect sound coloration for the studied cases.

It is emphasized that all computational steps used for the coloration estimation in this paper have parameters that are carefully selected. Thus, future work could focus on optimizing the computations and the parameter selection to make the coloration estimation less sensitive. Also, several different system and room types should be evaluated, both simulated and measured. Reference distributions should be studied in more detail, i.e. how and why they change depending on typical room types. A database of reference distributions of typical real rooms would improve the accuracy of the coloration estimation.

Bibliography

[1] P. Svensson: On Reverberation Enhancement in Auditoria, PhD Thesis, Department of Applied Acoustics, Chalmers (1994).

[2] E. Zwicker, H Fastl: Psychoacoustics: facts and models, Springer Science & Business Media, 2nd edition (1999).

[3] F. Kawakami, Y. Shimizu: Active Field Control in Auditoria, Applied Acoustics 31(1-3), pp. 47-75 (1990).

[4] H. Kuttruff: Room Acoustics, 3rd edition, Elsevier, London/New York (1994).

[5] J.-D. Polack: Modifying chambers to play billiards: the foundations of reverberation theory, Acta Acustica united with Acustica, 76(6), pp. 257-272 (1992).

[6] J. S. Bendat, A. G. Piersol: Random data. Analysis and Measurment Procedures, 2nd edition, John Wiley & Sons Inc (1986).

[7] J. Burkhardt, R. L. Weaver: Spectral statistics in damped systems. Part I. Modal decay rate statistics, J. Acoust. Soc. Am., 100(1), pp. 320-326 (1996).

[8] M. R. Schroeder: Some new results in reverberation theory, Proc. 5th International Congress on Acoustics, Liege G31, pp. 1-4 (1965).

[9] B.C.J. Moore: An Introduction to the Psychology of Hearing, 5th edition, Elsevier (2003).

[10] D. A. Ronken: Some effects of bandwidth-duration constraints on frequency discrimination, J. Acoust. Soc. Am., 49(4B), pp. 1232–1242 (1971).

[11] S. Buus, M. Florentine, T. Poulsen: Temporal integration of loudness, loudness discrimination, and the form of the loudness function, J. Acoust. Soc. Am., 101(2), pp. 669-680 (1997).

[12] R. D. Patterson: The sound of a sinusoid: Spectral models, J. Acoust. Soc. Am., 96(3), pp. 1409–1418 (1994).

[13] T. Irino and R. D. Patterson: Temporal asymmetry in the auditory system, J. Acoust. Soc. Am., 99(4), pp. 2316–2331 (1996).

[14] J. M. Doughty and W. R. Garner: Pitch characteristics of short tones. I. Two kinds of pitch threshold, J. Exp. Psychol., 37(4), pp. 351–365 (1947).

[15] W. R. Garner and G. A. Miller: The masked threshold of pure tones as a function of duration, J. Exp. Psychol., 37(4), pp. 293–303 (1947).

[16] K. N. Stevens: Frequency discrimination for damped waves, J. Acoust. Soc. Am., 24(1), pp. 76–79 (1952).

[17] B. C. J. Moore: Frequency difference limens for short-duration tones, J. Acoust. Soc. Am., 54(3), pp. 610–619 (1973).

[18] M. R. Schroeder: Modulation transfer functions: definitions and measurement, Acta Acustica united with Acustica, 49(3), pp. 179-182 (1981).

[19] J. L. Nielsen: Detection of colouration in reverberation enhancement systems, INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Institute of Noise Control Engineering, 1995(5), pp. 1213-1222 (1995).

[20] X. Meynial, O. Vuichard: Objective measure of sound colouration in rooms, Acta Acustica united with Acustica, 85(1), pp. 101-107 (1999).

[21] T. Watanabe, M. Ikeda: Objective detection method of sound coloration in electroacoustic enhancement system, ICA 2016, Proc. of the 22nd International Congress on Acoustics (2016).

[22] M. A. Poletti: Colouration in assisted reverberation systems, Proceedings of ICASSP 94. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2, pp. II/269-II/272 (1994).

[23] N. Korany: Factors affecting sound coloration perceived due to room reverberation, 25th National Radio Science Conference, pp. 1-8 (2008).

[24] P. Rubak: Coloration in room impulse responses, Joint Baltic-Nordic Acoustics Meeting 2004, pp. 1-14 (2004).

[25] J. Y. C. Wen, P. A. Naylor: Objective measurement of colouration in reverberation, 15th European Signal Processing Conference, pp. 1615-1619 (2007).

[26] T. van Waterschoot, M. Moonen: Fifty Years of Acoustic Feedback Control: State of the Art and Future Challenges, Proc. IEEE, 99(2), pp. 288-327 (2011).

[27] A. Spriet, K. Eneman, M. Moonen, J. Wouters: Objective measures for real-time evaluation of adaptive feedback cancellation algorithms in hearing aids, Proc. 16th Eur. Signal Process. Conf. IEEE, pp. 1-5 (2008).

[28] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, C. A. McGonegal: A Comparative Performance Study of Several Pitch Detection Algorithms, IEEE Trans. Acoustics, Speech and Signal Processing, 24(5), pp. 399-418 (1976).

[29] L. R. Rabiner: On the Use of Autocorrelation Analysis for Pitch Detection, IEEE Trans. Acoustics, Speech and Signal Processing, 25(1), pp. 24-33 (1977).

[30] W. H. Tucker: A Pitch Estimation Algorithm for Speech and Music, IEEE Trans. Acoustics, Speech and Signal Processing, 26(6), pp. 597-604 (1978).

[31] D. Liu, C. Lin: Fundamental Frequency Estimation Based on the Joint Time-Frequency Analysis of Harmonic Spectral Structure, IEEE Trans. Speech and Audio Processing, 9(6), pp. 609-621 (2001).

[32] T. Abe, T. Kobayashi, S. Imai: Robust Pitch Estimation with Harmonics Enhancement in Noisy Enviroments Based on Instantaneous Frequency, Proceedings ICSLP 96, Fourth International Conference on Spoken Language Processing, pp. 1277-1280 (1996).

[33] F. J. Charpentier: Pitch detection using the short-time phase spectrum, ICASSP 86, Tokyo, pp. 3.9.1-3.9.4 (1986).

[34] D. H. Friedman: Instantaneous-frequency distribution vs. time: an interpretation of the phase structure of speech, Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1121-1124 (1985).

[35] T. Tanaka, T. Kobayashi, D. Arifianto, T. Masuko: Fundamental frequency estimation based on instantaneous frequency amplitude spectrum, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. I/329-I/332 (2002).

[36] Y. Medan, E. Yair, D. Chazan: An Accurate Pitch Detection Algorithm, Proceedings - International Conference on Pattern Recognition, pp. 476-480 (1988).

[37] T. R. Black, K. D. Donohue: Pitch Determination of Music Signals Using the Generalized Spectrum, Conference Proceedings - IEEE SOUTHEASTCON, pp. 104-109 (2000).

[38] M. Petroni, A. S. Malowany, C. C. Johnston, B. J. Stevens: A crosscorrelation-based method for improved visualization of infant cry vocalizations, Canadian Conference on Electrical and Computer Engineering, pp. 453-456 (1994).

[39] D. Pang, P. V. Sankar, L. A. Ferrari: A Modified MUSIC Algorithm for Detecting Sinusoids, Conference Record - Asilomar Conference on Circuits, Systems & Computers, pp. 586-588 (1989).

[40] Y. Tadokoro, W. Matsumoto, M. Yamaguchi: Pitch detection of musical sounds using adaptive filters controlled by time delay, Proceedings 2002 IEEE International Conference on Multimedia and Expo, pp. 109-12 (2002).

[41] M. S. Andrews, J. Picone, R. D. Degroat: Robust pitch determination via SVD based cepstral methods, ICASSP 90, 1990 International Conference on Acoustics, Speech and Signal Processing, pp. 253-256 (1990).

[42] T. S. Verma, T. H. Y. Meng: An analysis/synthesis tool for transient signals that allow a flexible sines+transients+noise model for audio, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 3573-3576 (1998).

[43] E. Barnard, R. A. Cole, M. P. Vea, F. A. Alleva: Pitch Detection with a Neural-Net Classifier, IEEE Trans. Signal Proc., 39(2), pp. 298-306 (1991).

[44] A. Belouchrani, M. G. Amin: Time-Frequency MUSIC, IEEE Signal Proc. letters, 6(5), pp. 109-110 (1999).

[45] P. McLeod, G. Wyvill: Visualization of Musical Pitch, Proceedings Computer Graphics International 2003, pp. 300-303 (2003).

[46] A. M. Noll: Cepstrum Pitch Determination, J. Acoust. Soc. Am., 41(2), pp. 293-309 (1966).

[47] A. S. Durey, M. A. Clements: Direct estimation of musical pitch contour from audio data, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 561-564 (2003).

[48] M. Karjalainen, P. A. A. Esquef, P. Antsalo, A. Mäkivirta, Vesa Välimäki: Frequency-Zooming ARMA Modeling of Resonant and Reverberant Systems, J. Audio Eng. Soc., 50(12), pp. 1012-1029 (2002).

[49] J. Makhoul: Spectral Analysis of Speech by Linear Prediction, IEEE Trans. Audio and Electroacoustics, 21(3), pp. 140-148 (1973).

[50] P. M. T. Broersen: Automatic Spectral Analysis With Time Series Models, IEEE Trans. Instrumentation and Measurement, 51(2), pp. 211-216 (2002).

[51] K. B. Eom: Time-Varying Autoregressive Modeling of HRR Radar Signatures, IEEE Trans. Aerospace and Electronic Systems, 35(3), pp. 974-988 (1999).

[52] P. Ha, S. Ann: Robust time-varying parametric modelling of voiced speech, Signal Processing, 42(3), pp. 311-317 (1995).

[53] B. Koo, J. D. Gibson: Filtering of colored noise for speech enhancement and coding, IEEE Transactions on Signal Processing, 39(8), pp. 1732-1742 (1991).

[54] M. G. Hall, A. V. Oppenheim, A. S. Willsky: Time-varying parametric modeling of speech, Signal Processing, 5(3), pp. 267-285 (1983).

[55] Y. Grenier: Time-Dependent ARMA Modeling of Nonstationary Signals, IEEE Trans. Acoustics, Speech and Signal Proc., 31(4), pp. 899-911 (1983).

[56] F. Kozin, F. Nakajima: The Order Determination Problem for Linear Time-Varying AR Models, IEEE Trans. Automatic Control, 25(2), pp. 250-257 (1980).

[57] K. S. Nathan, Y. Lee, H. F. Silverman: A Time-Varying Analysis Method for Rapid Transitions in Speech, IEEE Trans. Signal Proc., 39(4), pp. 815-824 (1991).

[58] Y. Lee, H. F. Silverman: On a general time-varying model for speech signals, ICASSP 88: 1988 International Conference on Acoustics, Speech, and Signal Processing, pp. 95-98 (1988).

[59] A. Kacha, F. Grenez, K. Benmahammed: Time-frequency analysis and instantaneous frequency estimation using two-sided linear prediction, Signal Processing, 85(3), pp. 491-503 (2005).

[60] J. J. Rajan, P. J. W. Rayner: Generalized Feature Extraction for Time-Varying Autoregressive Models, IEEE Trans. Signal Proc., 44(10), pp. 2498-2507 (1996).

[61] Y. Grenier: Time-frequency analysis using time-dependent ARMA models, Proceedings - ICASSP 84, IEEE International Conference on Acoustics, Speech and Signal Processing, 9, pp. 270-273 (1984).

[62] M. Aboy, O. W. Marquez, J. McNames, R. Hornero, T. Trong, B. Goldstein: Adaptive Modeling and Spectral Estimation of Nonstationary Biomedical Signals Based on Kalman Filtering, IEEE Trans. Biomedical Engineering, 52(8), pp. 1485-1489 (2005).

[63] C. Sodsri: Time-varying autoregressive modelling for nonstationary acoustic signal and its frequency analysis, PhD Thesis, The Pennsylvania State University (2003).

[64] M. P. Lewis, T. J. Tucker, D. M. Oster: Method and apparatus for adaptive audio resonant frequency filtering, United States Patent, no. 5,245,665 (1993).

[65] C. Chen: Automatically tunable notch filter and method for suppression of acoustical feedback, United States Patent, no. 4,091,236 (1978).

[66] S. H. De Koning, A. Verwijmeren: Amplifier with automatic gain control, United States Patent, no. 4,817,160 (1989).

[67] J. R. Cox: Analog signal translating system with automatic frequency selective signal gain adjustment, United States Patent, no. 4,602,337 (1986).

[68] S. Muraoka, M. Sakamoto: Anti-howl back device, United States Patent, no. 4,493,101 (1985).

[69] E. T. Patronis Jr.: Acoustic feedback detector and automatic gain control, United States Patent, no. 4,079,199 (1978).

[70] E. T. Patronis JR: Electronic Detection of Acoustic Feedback and Automatic Sound System Gain Control, J. Audio Eng. Soc., 26(5), pp. 323-325 (1978).

[71] J. Wei, L. Du, Z. Chen, F. Yin: A new algorithm for howling detection, Proceedings - IEEE International Symposium on Circuits and Systems, 4, pp. IV-IV (2003).

[72] S. Ibaraki, H. Furukawa, H. Naono: Pre-howling howlback detection method, ICASSP 86, IEEE International Conference on Acoustics, Speech and Signal Processing, 11, pp. 941-944 (1986).

[73] J. L. Nielsen: Control of stability and coloration in electroacoustic systems in rooms, PhD thesis, The Department of Telecommunications, Norwegian University of Science and Technology (1996).

[74] X. Meynial, O. Vuichard: Objective measure of sound colouration in rooms, Acta Acustica united with Acustica, 85(1), pp. 101-107 (1999).

[75] B. H. Hutchins: An Adaptive Delay Comb Filter for the Restoration of Audio Signals Badly Corrupted with a Periodic Signal of Slowly Changing Frequency, J. Audio Eng. Soc., 30(1/2), pp. 24-27 (1982).

[76] H. Nyquist: Regeneration Theory, Bell System Tech. J., 11, pp. 126-147 (1932).

[77] M. Kleiner, P. Svensson: Review of active systems in room acoustics and electroacoustics, INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Institute of Noise Control Engineering, 1995(5), pp. 39-51 (1995).

[78] P. H. Parkin, K. Morgan: Assisted Resonance in the Royal Festival Hall, London, J. Sound Vib., 2(1), pp. 74-85 (1965).

[79] P. H. Parkin, K. Morgan: "Assisted Resonance" in the Royal Festival Hall, London: 1965-1969, J. Acoust. Soc. Am., 48(5A), pp. 1025-1035 (1970).

[80] B. Tunbridge et al.: The Interaction of an Acoustic Feedback Channel with the Transient Field in a Rectangular Enclosure, Acta Acustica united with Acustica, 31(5), pp. 271-280 (1974).

[81] G. Curtis: An Analysis of the Regenerative Reverberation Effects of Acoustic Feedback in Rooms, Acta Acustica united with Acustica, 20(3), pp. 119-133 (1968).

[82] G. Dodd: The Stability of Room Transmission Response and the Related Effects on Assisted Resonance Systems, J. Sound Vib., 36(4), pp. 443-471 (1974).

[83] A. J. Jones: The History and Application of Assisted Resonance, AIRO Res. Summary (1982).

[84] J. Bradley: The In-Channel Response of an Electroacoustic Feedback Channel, Acta Acustica united with Acustica, 33(1), pp. 1-12 (1975).

[85] J. Bradley: The Response of an Electroacoustic Feedback Channel with a Remote Source or a Remote Receiver, Acta Acustica united with Acustica, 33(1), pp. 13-22 (1975).

[86] J. G. Charles, J. Miller, H. Gwatkin: Assisting the Assisted Resonance at the Central Hall, York, UK, Appl. Acoust., 21(3), pp. 199-223 (1987).

[87] M. Poletti: A unitary reverberator for reduced colouration in assisted reverberation systems, INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Institute of Noise Control Engineering, 1995(5), pp. 1223-1232 (1995).

[88] P. Flandrin: Time-Frequency/Time-Scale Analysis, Academic Press (1998).

[89] L. Cohen: Time-Frequency Distributions-A Review, Proc. IEEE, 77(7), pp. 941-981 (1989).

[90] F. Auger, P. Flandrin, P. Gonçalvés, O. Lemoine: Time-Frequency Toolbox Tutorial, Centre National de la Recherche Scientifique, http://crttsn.univ-nantes.fr/∼auger/tftb.html (1997).

[91] T. A. C. M. Claasen and W. F. G. Mecklenbräuker: The Wigner distribution, a tool for time-frequency analysis, Part I: Continuous-time signals, 35(3), pp. 217-250; Part II: Discrete-time signals, 35(4-5), pp. 276-300; Part III: Relations with other time-frequency signal transformations, Phillips J. Res., 35(6), pp. 372-389 (1980).

[92] F. Plante, G. Meyer, W. A. Ainsworth: Improvement of Speech Spectrogram Accuracy by the Method of Reassignment, IEEE Trans. Speech and Audio Processing, 6(3), pp. 282-287 (1998).

[93] J. M. Morris: On Alias-Free Formulations of Discrete-Time Cohen's Class of Distributions, IEEE Trans. Signal Processing, 44(6), pp. 1355-1364 (1996).

[94] A. H. Costa, G. F. Boudreaux-Bartels: An Overview of Aliasing Errors in Discrete-Time Formulations of Time-Frequency distributions, IEEE Trans. Signal Processing, 47(5), pp. 1463-1474 (1999).

[95] A. H. Nuttall: Alias-free Wigner distribution function and complex ambiguity function for discrete-time samples, Tech. Rep. 8533, Naval Underwater Syst. Cent. (NUSC), New London, CT (1989).

[96] K. Kodera, R. Gendrin, C. Villedary: Analysis of Time-Varying Signals with Small BT values, EEE Trans. Acoust., Speech, Signal Processing, 26(1), pp. 64-76 (1978).

[97] F. Auger, P. Flandrin: Improving the Readability of Time-Frequency and Time-Scale Repesentations by the Reassignment Method, IEEE Trans. Signal Processing, 43(5), pp. 1068-1089 (1995).