

# Investigations into Velocity and Distance Perception Based on Different Types of Moving Sound Sources with Respect to Auditory Virtual Environments

Christian Störig, Christoph Pörschmann

Cologne University of Applied Sciences  
Institute of Communications Engineering  
Laboratory for Acoustics, Audio Technology and Audio Signal Processing  
Betzdorfer Str. 2, 50679 Cologne, Germany

email: christian.stoerig (at) me.com

email: christoph.poerschman (at) fh-koeln.de

www: www.fh-koeln.de/akustik

## Abstract

The characteristics of moving sound sources have strong implications on the listener's distance perception and the estimation of velocity. Modifications of the typical sound emissions as they are currently occurring due to the tendency towards electromobility have an impact on the pedestrian's safety in road traffic. Thus, investigations of the relevant cues for velocity and distance perception of moving sound sources are not only of interest for the psychoacoustic community, but also for several applications, like e.g. virtual reality, noise pollution and safety aspects of road traffic.

This article describes a series of psychoacoustic experiments in this field. Dichotic and diotic stimuli of a set of real-life recordings taken from a passing passenger car and a motorcycle were presented to test subjects who in turn were asked to determine the velocity of the object and its minimal distance from the listener. The results of these psychoacoustic experiments show

that the estimated velocity is strongly linked to the object's distance. Furthermore, it could be shown that binaural cues contribute significantly to the perception of velocity. In a further experiment, it was shown that - independently of the type of the vehicle - the main parameter for distance determination is the maximum sound pressure level at the listener's position.

The article suggests a system architecture for the adequate consideration of moving sound sources in virtual auditory environments. Virtual environments can thus be used to investigate the influence of new vehicle powertrain concepts and the related sound emissions of these vehicles on the pedestrians' ability to estimate the distance and velocity of moving objects.

**Keywords:** Spatial Hearing, Auditory Scene Analysis, Auralization, Moving Sound Source, Doppler shift

## 1 Introduction

Auditory distance and velocity perception of moving sound sources is of great importance in many everyday scenarios. Due to the limited visual field, humans quite often have to rely on the auditory channel in order to avoid collisions, e.g. when participating as a pedestrian or as a cyclist in road traffic [Ker06]. In complex traffic situations different kinds of sound sources (e.g. trucks, motorcycles, passenger cars) with different spectral and temporal characteristics can be perceived. Thus, it seems surprising that the human

### Digital Peer Publishing Licence

Any party may pass on this Work by electronic means and make it available for download under the terms and conditions of the current version of the Digital Peer Publishing Licence (DPPL). The text of the licence may be accessed and retrieved via Internet at <http://www.dipp.nrw.de/>.

distance and velocity perception mechanisms have not yet been examined in great detail for different types of moving sound sources. Only a few studies are available in literature, most of them employing artificial signals as stimuli and thus only considering some of the relevant cues.

When vehicle types with new powertrain concepts like hybrid or electric vehicles (HEV/BEV) have to be considered, a proper simulation of moving sound sources can be especially helpful to predict their implications on pedestrian safety. In the past, several applications in the field of virtual auditory environments were realized for which the distance perception of moving sound sources plays an important role [Str98a, Str98b], e.g. in driving or racing simulators.

It is a well-known fact that the listener's hearing experience dominates distance perception of stationary sound sources (cf. [Col62, Col68]). Both air absorption and sound pressure level are considered as the most relevant factors contributing to the localization of sources in medium and large distances under free-field conditions or environments with a low amount of reverberation (e.g. exterior environments). When the perception of non-stationary sound sources is examined, an analysis of the relevant distance perception cues is far more complex than in the examination of stationary sources. Here, both temporal changes of the localization cues (e.g. Doppler shift, monaural sound pressure level, HRTF resp. ITDs and ILDs) take effect.

In one of our previous studies, that was based on the typical scenario of a passing passenger car, we investigated which of the various cues are the main contributors to the distance and velocity estimates of the subjects. The recordings of a passing passenger car with various vehicle speeds and microphone distances were presented dichotically and diotically to the subjects in three psychoacoustic experiments (cf. [PS09, SP07]). However, it has not been examined in which way these influences vary for different types of sound sources. This is of specific relevance with respect to real world scenarios where a mixture of different types of moving sound sources is typically encountered.

Considering distance and velocity perception of new types of vehicles is an important issue when proceeding towards electromobility. Estimating the implications of these new vehicle sounds might help to prevent accidents caused by misinterpreting their velocities and the distances. The following key questions shall be addressed in the following:

- *Which cues dominate the perception of distance and velocity of moving sound sources?*
- *How do the perceived velocity and the perceived distance vary for different types of moving sound sources in relation to their physical properties (e.g. true velocity and true distance)*

The structure of this paper is as follows: first, a brief overview of human distance and velocity perception is given. Furthermore, the relevant literature is outlined shortly, whereafter the recordings of the moving sound sources which were presented to the test subjects are described. Subsequently, two psychoacoustic experiments are explained in detail; the first one analyses velocity perception, the second one investigates how distance perception is influenced depending on the type of the sound source and its acoustic properties. Moreover, several applications of moving sound sources in audio applications resp. virtual environments are discussed. Finally, it is described in which way a virtual audio environment system can be enhanced so that it allows the adequate simulation of moving sound sources and the performance of psychoacoustic research on this subject.

## 2 Distance and Velocity Perception

In the following section a brief overview on the distance and velocity perception of stationary and moving sound sources is given. A more detailed account of the relevant localization cues and the research in this field can be found in an essay by [PS09]. The reader is encouraged to refer to this article.

Most of the research on distance perception has been performed in respect of stationary sound sources. Estimating the egocentric distance from a sound source is a task which seems to be very difficult for the subjects. Several investigations show that the distance blur is relatively high (e.g. [vB49, Col68, MBL<sup>+</sup>89]), while the directional localization blur is below 1° in the horizontal plane under static conditions [Bla97] in comparison.

The cues which are commonly regarded to be relevant for the distance perception of sound sources are intensity, spectrum and direct-to-reverberant energy ratio (cf. [ZBB05, Zah02]). In case of near-to-head sources dichotic cues are relevant as well [Bla97].

Two more items might be of relevance with respect to psychoacoustic research in this field: Initially, [SS73] and [Hel99] showed that head rotations do not improve

the accuracy of distance judgments in distance perception. Moreover, listeners tend to underestimate the true sound source distance from medium to high ranges, while in contrast the true distance is overestimated in case of nearby sources [Zah02].

In case of non-stationary sound sources, additional cues affect distance perception: The sound pressure level and the influence of air absorption become time-variant, when the distance between the source and the receiver is altered. Furthermore, Doppler shift provides additional information for the listener regarding distance perception. Moreover, dichotic cues potentially gain more influence on distance perception, even for large distances. Binaural cues help to create a spatial impression of the scene and facilitate the estimation of the source's movement angle. Due to the source's movement, all static cues become implicitly time-variant and will be termed 'dynamic cues' in the following, as has been suggested by [SL93]. The listener may use this additional information and may thus potentially enhance the accuracy of the estimate. Table 1 has been generated by [PS09] in order to categorize the various cues that are relevant for distance perception and for the determination of the sound source velocity.

Only a few psychoacoustic experiments dealing with the distance and velocity perception of moving sound sources have been described yet. Most of the studies were performed with artificial stimuli in a simulated environment. For an overview of the relevant literature, please refer to [PS09]. Furthermore, it was shown by [PS09] that the main contributor for the determination of distance is the maximum sound pressure level at the listener's position. However, dichotically presented sound sources were perceived to be closer than sound sources in the corresponding diotic situation which can be explained by the binaural masking level difference (BMLD). In case of the dichotic stimuli, it was assumed that the loudness difference between the moving source and the background noise increases due to binaural masking effects. Here, the source and the present environmental noise can be isolated more easily due to incoherent components that are contained in the ear signals. Hence, the improved separability of the source's signal from the background noise results in an increase of loudness and the source is thus localized in a lower distance, compared with the diotic case (cf. [PS09]). Furthermore, a significant dependency between distance estimates and the velocity of the object was shown for a passenger car.

This influence was clarified by the fact that the sound emissions of the passing vehicles, and thus the static sound pressure level at the position of the listener, increase towards higher velocities. Apart from the static sound pressure level at the listener's position no statistical significance of other dynamic cues on the perceived distance and velocity were proved (cf. [PS09]). However, as the study by [PS09] has been performed only for one specific type of sound source, several open questions remain. The question whether the observed effects are of general nature or specific for the type of the vehicle could not be clarified, especially. In this article it is to be investigated whether systematic effects on the distance and the velocity estimates can be observed for vehicle types with totally different sound characteristics.

### 3 Recording and analysis of moving sound sources

Except for [PS09] and [Hel96] all of the investigations, which have been described in the previous section, made use of simulated scenarios or of synthetic stimuli. The psychoacoustic experiments presented in this article investigate the perception of sound source distance and velocity by applying a set of real-world measurements of a passing passenger car and of a passing motorcycle. Typically, the localization of moving sound sources is of practical importance with respect to the already mentioned scenarios - such as passing vehicles in traffic situations. In addition, the recorded scenarios can be used as a reference for the design and for the assessment of scenarios that have been created inside of a virtual environment.

The concept of using real-world recordings as stimuli within this study makes sense, since one of the key questions of this article is whether the results that have been found in [PS09] can be generalized for different types of sound sources. In this context, the same recording setup has been chosen as in [PS09] in order to avoid unintended effects on the experimental design and distortions of the subject's estimates.

#### 3.1 Recording Procedure

Several conditions for passing vehicles were recorded in various distances between the vehicle and the microphones (resp. the listener) and with various velocities. Two types of microphones were used for

	diotic (monaural)	dichotic (binaural)
static	sound pressure spectral colorizations due to air absorption (relevant for faraway sound sources) direct-to-reverberant energy ratio	incidence direction: HRTF( $\varphi, \delta$ ) respectively: interaural level differences (ILD) interaural time differences (ITD) (relevant for nearby sources)
dynamic	change of sound pressure level over time time resp. frequency scaling (pitch change) due to the Doppler effect ( $S(\tau, t)$ ) changes in spectral colorizations due to air absorption (relevant for faraway sound sources) changes in direct-to-reverberant energy ratio	change of incidence direction over time: HRTF( $\varphi(t), \delta(t)$ ) respectively: ITD( $t$ ) ILD( $t$ )

Table 1: Classification of the various auditory cues into 'diotic' (resp. monaural) and 'dichotic' (resp. binaural) cues vs. 'static' and 'dynamic' cues. In the following paragraphs, the terms 'dichotic', 'diotic', 'static' and 'dynamic' will be used in order to distinguish the different categories

the recordings, in particular one omni-directional microphone (Beyerdynamic MM1) for the diotic stimuli and one artificial head (Head Acoustics HMS III, ID-equalized, calibrated to 100 dB(SPL)  $\equiv$  0 dBFS) for the dichotic stimuli. The microphone signals were recorded simultaneously by means of a multi-track recording front end at distances of 4 m, 8 m, 12 m, 16 m and 20 m and with vehicle speeds of 40 kph, 60 kph, 80 kph and 100 kph. For practical reasons, the distance between the microphones and the sideline of the public street has been set to the described values. As sound sources a passenger car (Opel Astra, service year - 1997, manual transmission, I4-Otto engine (1.6 l cylinder capacity)) and a motorcycle (BMW K1100RS, model year 1994) were used. However, in order to create almost stationary conditions of the exterior noise, the vehicles were driven in the most suitable gear with respect to the required vehicle speeds and without any gear changes. Furthermore, the particular vehicle and engine speeds were held as constant as possible for both vehicles. In case of the motorcycle, the engine speed was held as constant as possible at about 3500 rpm in all recordings. Figure 1 shows the recording setup. Please refer to [PS09] for more details.

### 3.2 Signal Analysis

In order to verify that the various conditions were recorded properly, several quantities such as spectrum, short-time spectrum and sound pressure level have been extracted from the audio signals. This was possi-

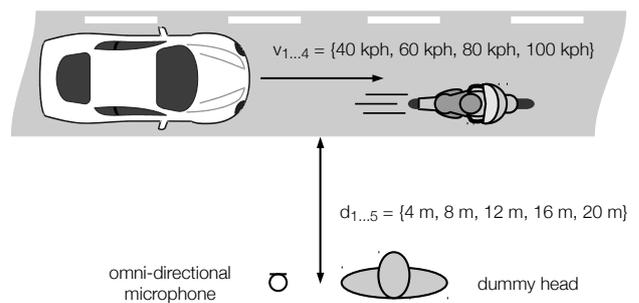


Figure 1: Recording Setup: The recordings of the passenger car and the motorcycle were performed at various distances ( $d_{1...5}$ ) and velocities ( $v_{1...4}$ ). For practical reasons, the distance between the microphones and the outer sideline of the driving track has been set to the annotated values. In addition it is not possible to define a reference point in order to measure the physical distance to the source, because the underlying vehicles feature several contributors, e.g. exhaust system, air induction system and tires. Thus, the sources are neither ideally omni-directional nor feature a spherical geometry like an ideal point source/breathing monopole.

ble, because the digital amplitudes of the recorded signals were calibrated to a reference sound pressure level

of 100 dB(SPL) which corresponds to an amplitude of 0 dBFS. Moreover, the extracted quantities facilitate conclusions with respect to dependencies between the physical properties of the sound sources and the results of the psychoacoustic experiments. The properties resp. the extracted quantities of the recordings of the passenger car and the motorcycle are compared in the following:

**Spectrograms:** As an example, the spectrograms for the passenger car and the motorcycle for condition  $v = 100$  kph and  $d = 4$  m are shown in Figure 2a and Figure 2b. Here, the Doppler shift - due to the translational movement of the vehicles - can easily be observed. Moreover, the differences in the sound character of the passenger car and of the motorcycle are shown.

The spectrogram of the passenger car's exterior sound shows strong contributions of broadband resp. stochastic components (e.g. tire noise) in addition to the harmonic components which correspond to the periodic engine order content that is typically emitted by the intake and exhaust system. In contrast, the spectrogram of the motorcycle's recording shows distinct harmonic components (engine orders) and much less stochastic content up to the higher frequency range. This corresponds well with the pronounced modulations of the time-envelope and thus with the much rougher sound character of the motorcycle.

**Magnitude Spectra (Pass-by-Situation):** Figure 3a and Figure 3b show the magnitude spectra both for the recordings of the passenger car and of the motorcycle for various vehicle speeds ( $v_1, v_2, v_3$  and  $v_4$ ), but with a constant microphone distance of four meters. The spectral progressions have been calculated by means of a FFT-analysis, which was applied to a data window of approx. 750 ms (32768 data points,  $f_s = 44.1$  kHz). Here, the center of the extracted data window has been placed at the audio files' sample position which corresponds to the point in time when the vehicles passed the omni-directional microphone resp. the median plane of the artificial head. In addition, the extracted data vector was multiplied by a window function of the same length (von Hann-window) so as to minimize spectral leakage effects. In order to emphasize the envelope of the magnitude spectra and to show the relative differences between the recorded signals, the obtained FFT-bins were smoothed by using an adaptive third-octave sliding window.

It is remarkable that the progressions of the magnitude spectra of the motorcycle data only show a slight dependency on the vehicle speed. However, in case of  $v = 100$  kph the spectrum shows a strong peak at about 600 Hz which is probably caused by dominant engine orders resp. due to changes in the operating conditions of the engine.

The analysis of the sound pressure levels was conducted as follows: For the calculation of the level-vs.-time curves a weighting filter with a time constant of 125 ms ('fast', exponential averaging) was chosen. In addition, the signals were processed with a spectral weighting filter (A-weighting curve) before the level-vs.-time curves were computed. The settings of the analysis follow a typical procedure for measuring sound pressure levels in noise pollution control, as for example suggested in the German standard DIN 4109 [DIN89] and DIN-EN-61672-1 [DE12].

**Sound Pressure Level (Pass-by-Situation)** The level-vs.-time curves were calculated both for the signals of the artificial head and of the omni-directional microphone. In case of the artificial head, the two level curves - which have been obtained for the ear microphones - were averaged sample by sample, which resulted in a single level-vs.-time vector. Furthermore, the maximum sound pressure levels ( $L_{max}$ ) were extracted for the situation when the passenger car and the motorcycle passed the microphones.

The comparison of the extracted maximum sound pressure levels showed that two recordings in case of the omni-directional microphone (namely the measurements of the passenger car with the conditions  $v = 40$  kph at  $d = 4$  m and  $v = 60$  kph at  $d = 8$  m) were not correctly calibrated. Thus, it was not possible to use these stimuli within the experiments.

In Figure 4a and Figure 4b the influence of the vehicle's distance and velocity on the maximum sound pressure level ( $L_{max}$ ) is shown for each vehicle. Here, the extracted values of the maximum sound pressure level do not exactly follow the distance law of an ideal point source under free-field-conditions as ground reflections interfere partly constructively and partly destructively. Furthermore, due to directivity and size, the vehicles cannot be considered as ideal point sources. Moreover, it is obvious that the influence of the velocity on the peak level is much weaker for the motorcycle than for the passenger car. An explanation for this might be that in case of the motorcycle mainly the engine sound influences the level,

which is only slightly depending on the velocity (while the engine speed was held at approx. 3500 rpm, see above). For the passenger car the level is significantly influenced by tire noise, which depends much more on the velocity of the vehicle.

## 4 Psychoacoustic Experiments

Within this section, two psychoacoustic experiments on velocity and distance perception of linearly translating sound sources are described. Here, the results of the two experiments using various types of sound sources - namely a motorcycle and passenger car - are to be compared. Moreover, systematic influences - with regard to the physical properties of the stimuli on the velocity and distance estimates - are investigated and the salience of the available auditory cues are assessed.

The first experiment analyzes the influence of egocentric distance from a sound source on velocity perception. Here, the phenomenon of the so-called motion parallax can be observed as well. Furthermore, the influence of diotic cues (Doppler shift, dynamic and peak sound pressure level) and dichotic cues (interaural time differences and interaural level differences) on the velocity percepts is investigated.

The second experiment examines how distance perception coincides with the true distance for the situation, that is, when the vehicle passes the listener. Beyond that, the degree to which sound source velocity alters distance perception is going to be discussed.

### 4.1 Experiment 1: Velocity Perception

According to the investigations of [LW99], the Doppler shift dominates velocity perception. In contrast, dynamic changes of the interaural time differences and of the signal amplitude at the ear drums seem to be less important cues. In general, their experiments show that the Doppler cue gains perceptual salience for higher velocities and, therefore, becomes pre-dominant with respect to the remaining cues. This phenomenon can be explained by the increasing frequency changes (span and slew-rate), whereby the detection of pitch changes is facilitated. Experiment 1 investigates and compares the influence of the diotic cues (Doppler shift, sound pressure level and the influence of air absorption) and of the dichotic cues (interaural time differences and interaural level differences) on velocity perception.

In a previous study by [PS09] it has been shown that an increase of distance results in a decrease of the perceived velocity. The mentioned effect is called motion parallax and can also be observed in visual depth perception (cf. [Gol07]). In case of visual perception the distance of a moving object is determined on the basis of its velocity and its displacement on the retina (beside other cues): From the observer's perspective nearby objects seem to move faster than faraway objects (boundary condition:  $v_{source} = \text{const.}$ ). An everyday situation where the effect of motion parallax occurs could, for instance, be the passing landscape, as soon as one looks out of a train's window. Furthermore, in [PS09] it has been shown that for a passing passenger car binaural effects influence the estimation of velocity as well. Within this experiment the following questions are posed:

- *How does the perceived velocity of the sound source vary for different types of moving sound sources?*
- *Which kind of influence do monaural cues (Doppler shift, sound pressure level) and binaural cues (interaural time difference, interaural level difference) have on velocity perception?*

**Method:** Various psychoacoustic methods can be applied in order to determine the differences between diotic stimuli (recordings by means of the omnidirectional microphone) and dichotic stimuli (binaural recordings), which have already been discussed in [PS09] and shall not be repeated here. During the listening tests the recorded stimuli were presented either diotically or dichotically via headphones. The estimates of the vehicle's velocities were rated on a category scale by the subjects. Later, the results for dichotically and diotically presented recordings are compared and thus allow an identification of the influence of binaural cues.

Despite to the fact, that the headphone-presentation of real-world recordings shows inherent limitations, the authors have decided to use these as stimuli for the experiments: In contrast to simulated scenarios, the listener cannot utilize head rotations in order to avoid front-back confusions. Moreover, it is more difficult to create sound events which are localized outside of the head (externalized sources vs. inside-the-head locatedness (IHL), cf. [Bla97]). However, the adequate simulation of such scenarios would be much more complex, compared with the real-world recordings. In

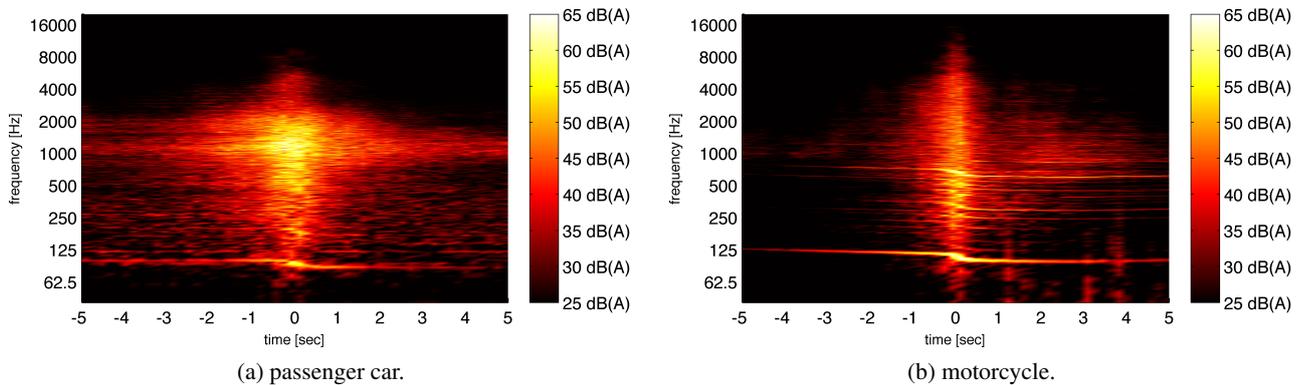


Figure 2: Spectrograms of the vehicles both passing at a distance of 4 m at a velocity of 100 kph. The analysis parameters have been chosen as follows: von Hann-window, window length:  $N_{FFT} = 16384$ , time-overlap:  $L = N_{FFT}/2$ , sample rate:  $f_s = 44.1$  kHz, spectral weighting = “A”.

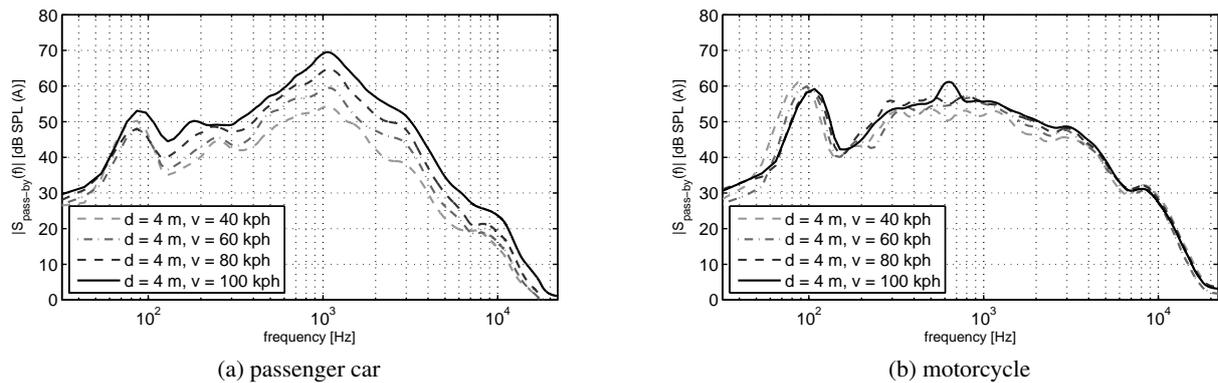


Figure 3: FFT-magnitude spectra (third-octave-smoothed, A-weighted) of the passenger car (left) and the motorcycle (right) when passing at a distance of 4 m with various velocities. The analyzed signal frame had a window-length of about 750 ms ( $N_{FFT} = 32768$ ,  $f_s = 44.1$  kHz).

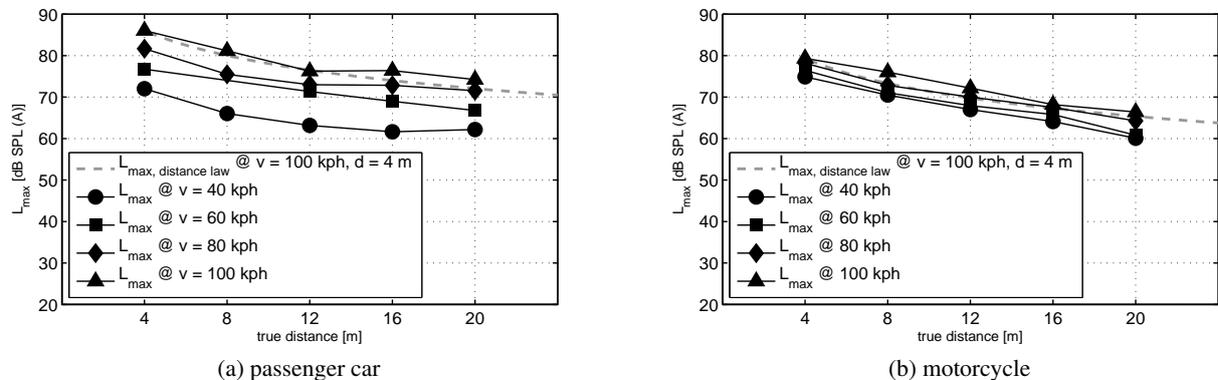


Figure 4: Influence of true distance and velocity on the peak level ( $L_{max}$ ) for the vehicles at the moment of passing. Here, the values of the peak levels have been retrieved from the level-vs.-time vectors by using the ear signals of the artificial head (energetically averaged), for the moment, when the vehicles passed the microphones. In addition, the theoretical curve for the peak level according to the distance law in case of a monopole source under free-field conditions is shown. Moreover, the curve was adapted to the peak level for the condition  $v = 100$  kph and  $d = 4$  m for the particular vehicle (dashed line).

particular, as soon as not only moving sources, but also the environmental resp. background noise, needs to be simulated properly.

**Subjects and Stimuli:** In the experiment the subjects were asked to quantify the velocity of the object. In the first section ('Experiment 1a'), 20 dichotic stimuli of a passenger car were presented in a pseudo-randomized order, which varied among the subjects. Each stimulus was presented repeatedly within the sessions. The data of 22 subjects aged from 22 to 41 was evaluated. In the second part ('Experiment 1b'), 20 diotic stimuli of a passenger car were presented. All other parameters of the experiment remained unchanged. This time the data of 22 subjects, aged between 20 and 33, was used for evaluation.

The same procedure was repeated for the motorcycle. In section 1c (20 dichotic stimuli of a motorcycle), 18 subjects aged from 22 to 32 participated in the experiment and in section 1d (20 diotic stimuli based on the recorded motorcycle), the results of 20 subjects aged from 21 to 30 were collected.

A PC-based graphical user interface was used in order to acquire the judgments. The subjects rated the perceived velocity on a seven-point category scale (*extrem langsam, sehr langsam, langsam, mittel, schnell, sehr schnell, extrem schnell*) which was displayed on a monitor screen by means of setting a slider to the appropriate position. In addition, the subjects were also allowed to rate interim values between the given categories. The equidistance between the categories was underlined by the visual presentation as shown in Figure 5 as suggested by [JM86] and [Mö100].

How did you perceive the velocity of the object?

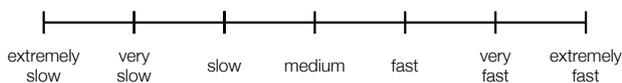


Figure 5: Scale used in order to rate the perceived velocity of the stimulus (the English translation is shown here).

In the initial phase of each session two binaural anchor stimuli were presented to the subjects: One which was introduced as very slow (condition:  $d = 20$  m and  $v = 40$  kph) and one as very fast (condition:  $d = 4$  m and  $v = 100$  kph). These two conditions represented the ones which were rated as slowest or fastest in previous informal listening tests. The listening sessions using the recordings of the motorcycle and of

the passenger car have been conducted separately from each other, whereby only the recordings of the particular type of vehicle were presented as anchor stimuli. Thus, there might be a shift in the rating scale between the ratings for the motorcycle stimuli and for the passenger car stimuli.

**Results and Discussion:** Figure 6a and Figure 6b show that the velocity estimates  $v'$  do not only depend on the particular velocity  $v$  of the source, but also on true distance  $d$ . The same trend can be observed for both vehicle types: With increasing distance to the sound source the vehicles are perceived as being slower (motion parallax).

However, it can be observed that the type of the sound source has a significant influence on velocity perception. Looking at the downward slopes of the regression analysis shown in Figure 7a and Figure 7b, it can be observed that this effect is stronger for the motorcycle than for the passenger car.

Furthermore, both figures show slight differences between the scores for diotic and dichotic presentation. In [PS09], statistically significant differences between the ratings for diotic and the dichotic stimuli have already been observed in case of the passenger car. Here, significance between the diotic and the dichotic presentation has been tested for the motorcycle as well. An analysis of variance (ANOVA) has been applied to the data (cf. [HEK09]) and the tendency of the differences was determined by means of an additional multiple pair comparison (cf. [Beu07]). The results, which are shown in Figure 8a and Figure 8b, indicate that - both for the motorcycle and for the passenger car - there exists a significant but weak influence of the binaural cues. The weakness and the variation of the tendency might be caused by the fact that different cues dominate velocity perception with respect to the motorcycle and the passenger car. For example, due to the distinctive harmonic character of the motorcycle's engine sound, the Doppler shift can be detected much easier than in case of the passenger car. In contrast, changes of the maximum sound pressure level which depend on the true velocity are much stronger for the passenger car.

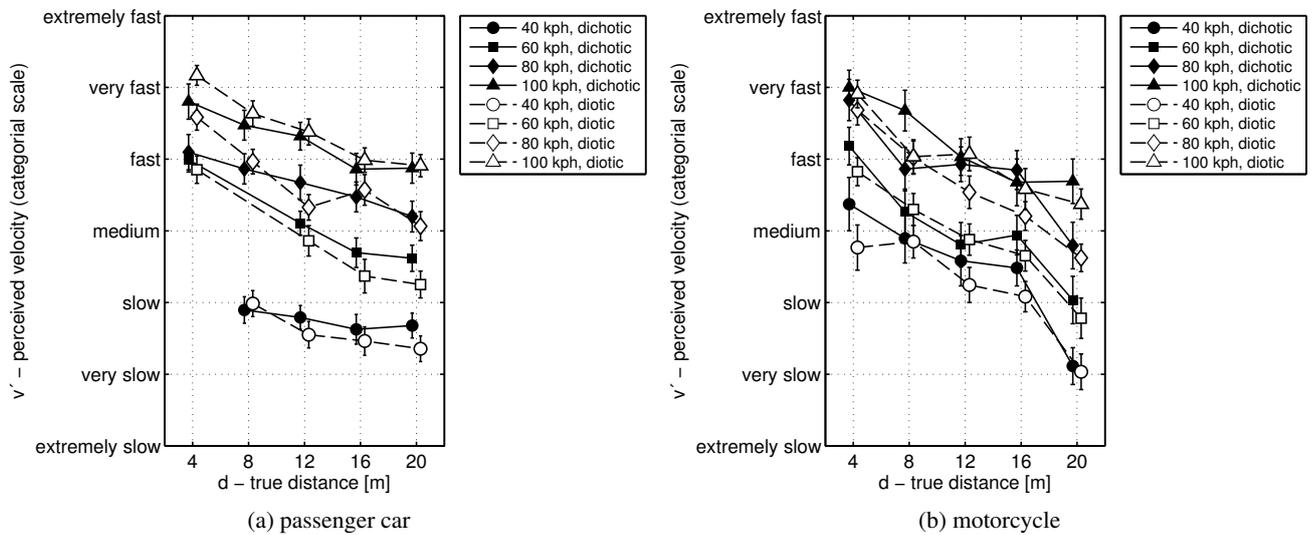


Figure 6: Influence of the true distance  $d$  of the sound source on the perceived velocity  $v'$ . Here, the average scores as well as the 95%-confidence intervals are shown in white for the diotic stimuli and in black for the dichotic ones. Two data points corresponding to the estimates based on the passenger car have been omitted due to a microphone calibration issue during the recordings ( $v = 40$  kph at  $d = 4$  m and  $v = 60$  kph at  $d = 8$  m).

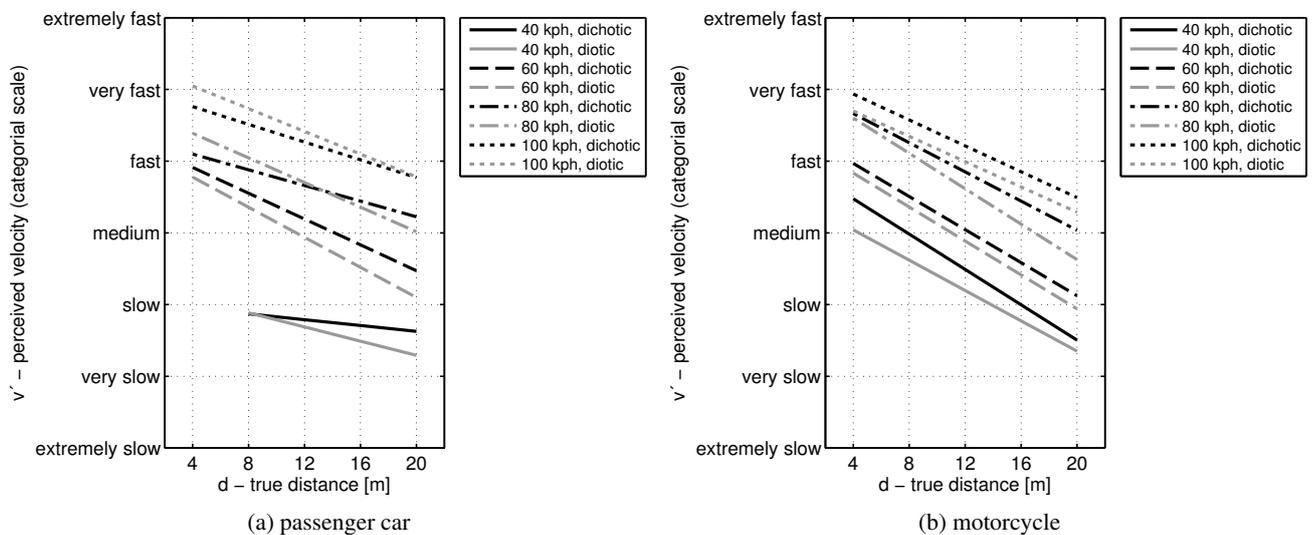


Figure 7: Linear regression analysis: A downward slope of the regression lines for the perceived velocity  $v'' = \beta_0 + \beta_1 d$  can be observed which seems to be stronger for the diotic stimuli (grey lines).

## 4.2 Experiment 2 - Distance Perception

This experiment investigates the way in which the distances of objects passing a listener are determined. For this purpose, various scenarios based on the recordings of the motorcycle and the passenger car were presented to the subjects and their particular distance estimates were evaluated. The scenarios varied with respect to the distance of the passing object and the ve-

locity of the object. The following questions shall be addressed within this experiment:

- How does the perceived distance correlate to the physical parameters of the scenario?
- How do the perceived velocity and the perceived distance vary for different types of moving sound sources?

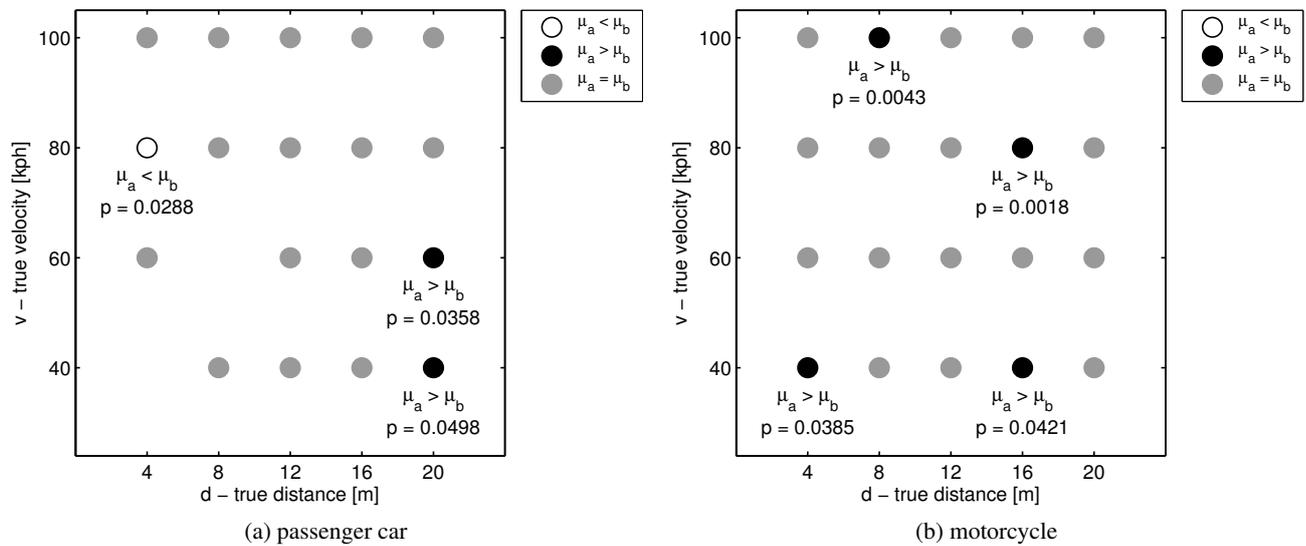


Figure 8: Results of the ANOVA: Significant differences between the velocity judgments  $v'$  over diotic and dichotic stimuli ( $P = 0.95$  (confidence level);  $\mu_a \equiv$  dichotic,  $\mu_b \equiv$  diotic) for both vehicles. The probabilities  $p$  are given in addition, if the zero-hypothesis  $H_0$  has been rejected. In case of the passenger car, the results of the ANOVA corresponding to the conditions  $v = 40$  kph at  $d = 4$  m and  $v = 60$  kph at  $d = 8$  m have been omitted due to microphone calibration errors.

	$v_1 = 40$ kph	$v_2 = 60$ kph	$v_3 = 80$ kph	$v_4 = 100$ kph
$R_{\text{pass. car, dichotic}}$	-0.87	-0.97	-0.99	-0.96
$R_{\text{pass. car, diotic}}$	-0.92	-0.98	-0.91	-0.97
$R_{\text{motorcycle, dichotic}}$	-0.92	-0.93	-0.89	-0.95
$R_{\text{motorcycle, diotic}}$	-0.91	-0.98	-0.99	-0.94

Table 2: Correlation coefficients of the regression analysis for the passenger car and the motorcycle data both for the idiomatic and dichotic presentation.

- *To what extent does the velocity of the object influence the perceived distance?*

**Method:** Again, the recorded stimuli were conveyed via headphones. As in experiment 1, only slight differences between a diotic and a dichotic presentation were observed. As these differences are not the key questions of this experiment, only a dichotic presentation was chosen here. However, in [PS09] the influences of a diotic versus a dichotic presentation have been analyzed. Thus, the stimuli were the same as in the first experiment.

**Subjects and Stimuli:** For the passenger car the results of 31 subjects, aged between 22 and 38 years, were evaluated, for the motorcycle 20 subjects, aged from 22 to 30 years, participated. In the experiment

the subjects were asked to rate how they perceived the distance to the object when passing. By means of the described PC-based user interface the subjects rated the perceived distance on a seven-point category scale (*sehr weit entfernt, weit entfernt, eher weit entfernt, mittel, eher nah, nah, sehr nah*) as shown in Figure 9. In the introductory phase of the particular sessions two anchor stimuli were presented to the subjects. The stimulus corresponding to a distance of four meters and a velocity of 60 kph was introduced as *close*, while the stimulus corresponding to  $d = 20$  m and  $v = 60$  kph was introduced as being *distant*. As in 'Experiment 1', the stimuli for the motorcycle and the passenger car were presented in different sessions of the experiment. Thus, two different anchoring stimuli were used in two sessions, which might result in a slight offset between the rating scales for the motorcycle and

the passenger car.

How did you perceive the distance of the object when passing?

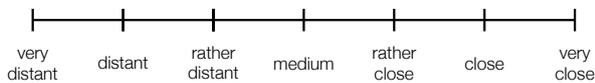


Figure 9: Scale in order to rate the perceived distance when passing (the English translation is shown here).

**Results and Discussion:** Figure 10 shows the average distance estimates  $d'$  versus the true distance  $d$  of the passing vehicle. The data points have been grouped after the actual vehicle speed forming four graphs. In case of the passenger car (Figure 10a) a significant offset between the curves for the different velocities is observable. In contrast to the ratings for the passenger car, the curves based on the motorcycle stimuli nearly overlay. Hence, in case of the passenger car the distance estimates significantly depend on the velocity of the object, which is additionally shown in Figure 11a. An increase of velocity causes the listeners to perceive the vehicle passing at a shorter distance. In [PS09] it has been deduced (applying the stimuli of the passenger car) that this effect is mainly caused by the increase of the sound pressure level towards lower distances. With regard to the motorcycle this effect is much weaker. The perceived distance only slightly depends on true velocity.

However, by looking at the sound pressure levels of the motorcycle these results can be explained: In contrast to the passenger car the influence of velocity on sound pressure at the listener's position is far less distinctive (see Figure 11b). In order to demonstrate the dependency of the sound pressure level on the estimated distance, the maximum sound pressure levels at the moment of passing ( $L_{max}$ ) were extracted from the recordings and compared with the distance judgments.

The perceived distances versus the maximum sound pressure levels are shown in Figure 12. The figure makes clear that both for the passenger car and for the motorcycle the main parameter for the determination of distance is the maximum sound pressure level at the listener's position. Additionally, the stimuli of the motorcycle vary for each specific distance much less than the ones of the passenger car. This corresponds to the peak level analysis which has been shown in Figure 4a and in 4b. In comparison to a passenger car, a to-

tally different relationship between velocity and sound pressure level is given. Thus, it can be expected that the relationship between the estimated distance and the velocity of the vehicle is different as well.

Furthermore, a linear regression analysis was applied to the extracted sound pressure levels and the average distance estimates. In case of the passenger car, the regression coefficients equal  $\beta_0 = 14.9$  and  $\beta_1 = -0.17$ . For the motorcycle, the regression analysis yields the coefficients  $\beta_0 = 16.9$  and  $\beta_1 = -0.21$ . In order to verify the influence of the maximum sound pressure level as a main contributor to distance perception, a correlation analysis of the averaged ratings and the particular sound pressure levels was calculated. The correlation coefficients yield  $R_a = 0.95$  for the passenger car and  $R_b = 0.97$  for the motorcycle. Both conditions show a high correlation between sound pressure level and estimated distance.

## 5 Fields of Application

Further research on human perception of moving sound sources is beneficial not only for scientific research, but also for various fields of application (cf. [SC98]). In practice, there exist many types of scenarios where a plausible presentation of moving sound sources is of relevance, e.g. when pedestrians participate in traffic situations. Several subjects and applications will be discussed in the following.

Within the automotive industry, in-car and exterior driving simulators have become important tools during the last two decades. The adequate simulation of complex scenarios as well as virtual prototyping have become more and more important not only for vehicles with internal combustion engines (ICE-vehicles), but also for the research and development of hybrid or electrical vehicle concepts (HEV/BEV). Here, the investigation of perceptual aspects with respect to pedestrian safety in traffic situations gains importance. Such vehicles are less detectable for pedestrians and thus potentially increase the risk of accidents due to reduced powertrain noise [Ker06, KF08, Zel09, RR09]. Recently, new regulations and requirements about the exterior noise of road vehicles have been discussed and issued [NHT11, NHT13, UNE13, Gov10]. In the near future, this surely will have a high influence on the development work of new vehicle fleets, like e.g. the integration of exterior sound warning devices [UNE10]. A virtual environment generator, which adequately features the auralization of moving

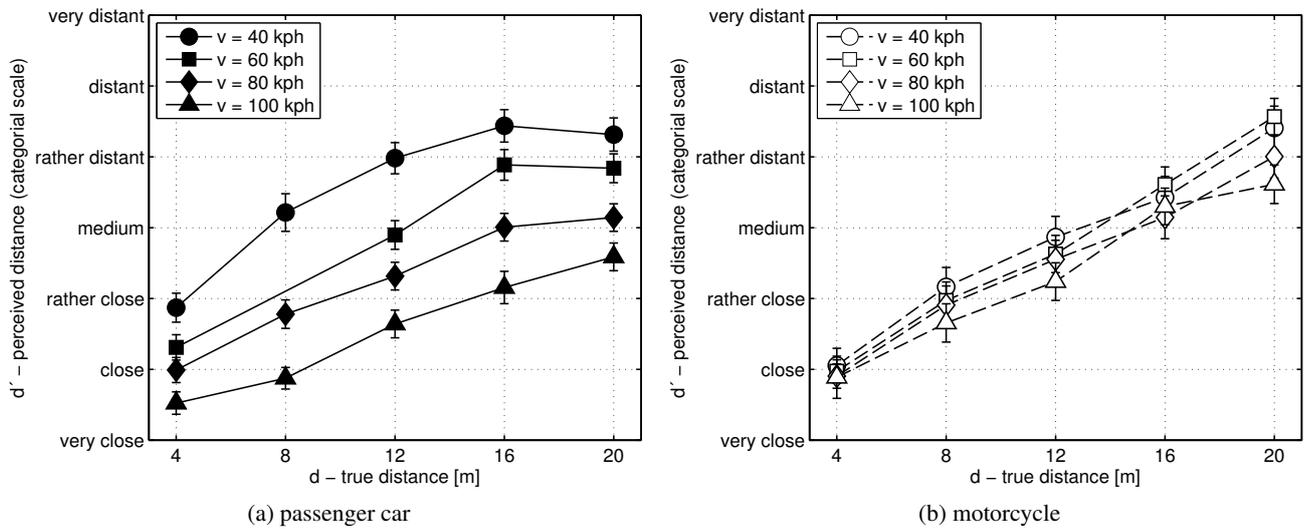


Figure 10: Perceived distance  $d'$  vs. the true distance  $d$  of the source at the moment of passing. Shown are the averaged estimates as well as the 95%-confidence-intervals for the passing vehicles.

objects, could reduce the need for expensive prototype vehicles or facilitate the development process, e.g. in case of attributes like exterior noise legalization, sound quality assessment and brand sound design [BG03, BK12b, BK12a].

In civil engineering, the investigation and avoidance of annoyance and noise pollution effects caused by traffic noise plays an important role [SKG02]: An adequate auralization of non-stationary sound sources in the context of traffic scenarios can be used in order to create soundscapes of urban environments for psychoacoustic studies. Here, the usage of real-world recordings can be complicated, elaborate and - moreover - insufficient. A rendering engine, which allows a flexible control and parameter setting of the synthesized scenarios, can be a powerful tool within this subject.

Moving objects can be a valuable extension to HMI-systems and could therefore serve as an additional information carrier in addition to other modalities. In this context, applications within e.g. auditory assistance and guidance tasks, warning signals in vehicle navigation and in driver assistance systems could be possible. It is referred to [Dür01] for a detailed investigation on the design of auditory displays by using stationary sound sources. Furthermore, an overview of the topic and a discussion on various aspects is given in [CV11].

Besides, the integration of moving sound sources into a toolbox for psychoacoustic research seems valuable.

Moreover, this toolbox could be utilized within audiology for the development of new screening tests or within therapy.

Finally, in case of the composition of computer music, in gaming and sound installations as well as in cinematography moving sources can be an extension to already existing means of expression [Cho77, PMBM08].

## 6 Integration of Moving Sound Objects into Auditory Virtual Environments

As described in the previous section, an integration of the relevant cues for distance perception of moving sound sources into an auditory virtual environment can be useful in various fields of research and development. In the following, it is described how to integrate the relevant cues adequately into a 3-D audio system.

### 6.1 Technical Requirements and System Architecture

The integration of moving objects has a strong effect on the system design and architecture of such an auditory virtual environment generator. In comparison with AVEs, which only feature stationary sources, the employed signal processing algorithms and sound field models become even more complex and computation-

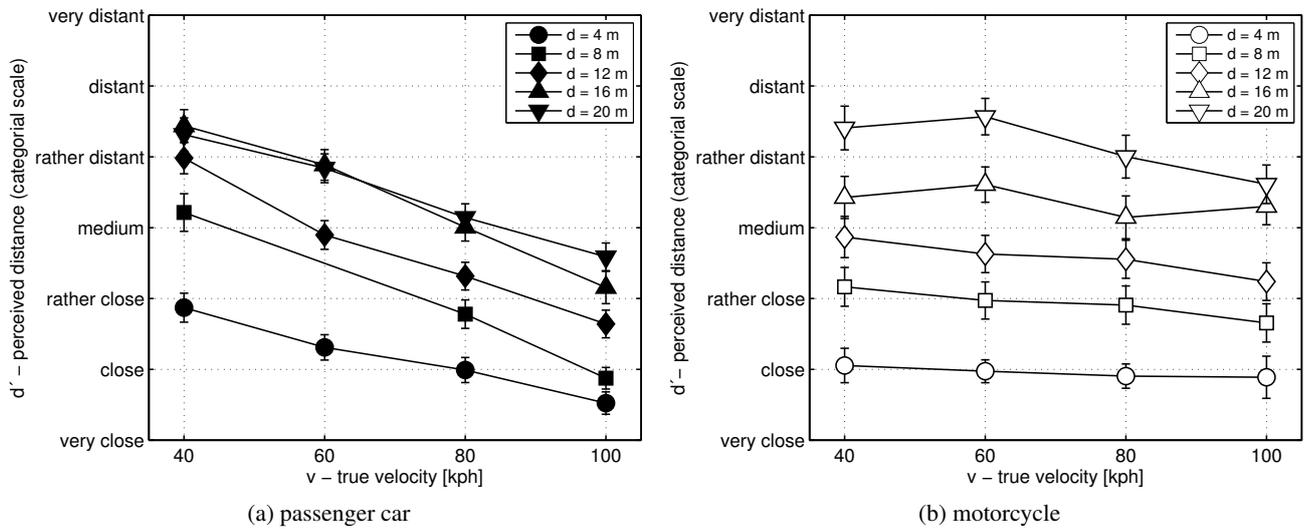


Figure 11: Perceived distance  $d'$  depending on the true velocity  $v$ . Shown are the average estimates as well as the 95%-confidence-intervals for the passenger car and the motorcycle.

ally expensive (cf. [Dür01, vK08, Str00]). Not all architectures of virtual environment generating systems are capable of handling moving sound sources. This holds true especially for systems applying measured or pre-simulated head-related impulse responses of the complete room (e.g. the BRS-System [MFT+99]).

Building a physically appropriate model for auralizing moving sound sources in the context of Wave Field Synthesis-Systems (WFS) has been investigated in [AS08a, AS08b, FGKS07]. However, the realization of such a model in a real-time system might be too elaborate for many applications and it might be advantageous to consider only specific cues of the moving sound source.

Figure 13 shows a typical structure of the system architecture of an auditory virtual environment generator. The geometrical and acoustical data of the room is being used to calculate the sound field model at the user's actual position in the virtual environment appropriately. In the signal processing engine a convolution of the audio data with the respective HRTFs for the direct sound and each reflection needs to be calculated. Applying additional filters and propagation delays allows auralizing the direct sound and the reflections in a manner as it would be in the respective real environment. The auralization can be performed binaurally via headphones [Bla97, Str00] or via loudspeakers [BR09, Bla08].

For loudspeaker-based solutions, Higher Order Am-

bisonics (HOA), trans-aural methods with dynamic crosstalk cancellation, Vector-based Amplitude Panning (VBAP) or near-speaker panning (cf. [SH07]) can be used. Nowadays Wave-Field-Synthesis systems can be applied as well in order to auralize complex scenes [Ber88] and several investigations with respect to the auralization of moving objects have been conducted [AS08a, FGKS07].

With regard to the auditory virtual environment, mainly the sound field model and the signal processing engine need to be modified and adapted when integrating moving sound sources into the virtual environment generating system. For further discussion on the system architecture and structures of AVEs the reader is referred e.g. to [BLSS00, Nov05, Len07, SB00, Tsi01, Tsi98, Vor08].

## 6.2 Sound Field Model

When enhancing the existing sound field model of virtual auditory environment generators, it has to be carefully investigated which of the cues of moving sound sources need to be adapted in order to give a plausible presentation to the listener. The results of the psychoacoustic experiments have shown that not all of the cues (cf. Table 1) have the same impact on perception. Depending on computation power and the requested quality not all of the cues need to be considered and thus allow to be rendered less precisely.

A physical exact sound field model is not always ap-

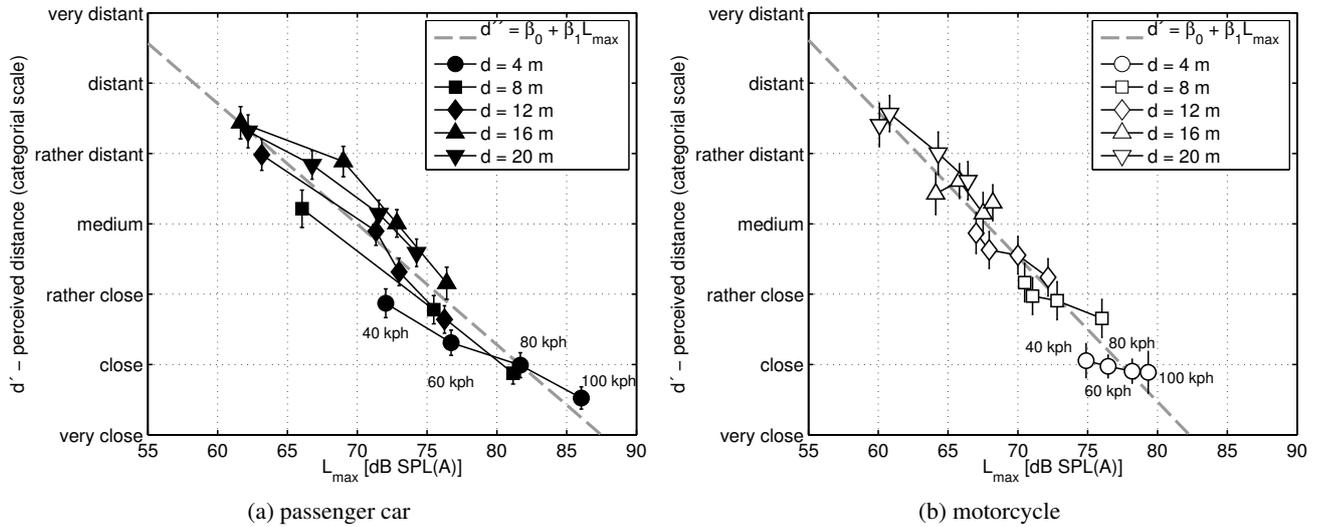


Figure 12: Influence of  $L_{max}$  on the distance-estimates  $d''$  in case of the passenger car and the motorcycle. Given are the mean values and the 95%-confidence-intervals. Additionally the result of a linear regression is shown. The coefficients of the regression line  $d'' = \beta_0 + \beta_1 L_{max}$  have been calculated to  $\beta_0 = 15$  and  $\beta_1 = -0.17$  in case of the passenger car and  $\beta_0 = 19.96$  and  $\beta_1 = -0.21$  in case of the motorcycle. The correlation coefficients equal  $R_a = 0.95$  (passenger car) and  $R_b = 0.97$  (motorcycle).

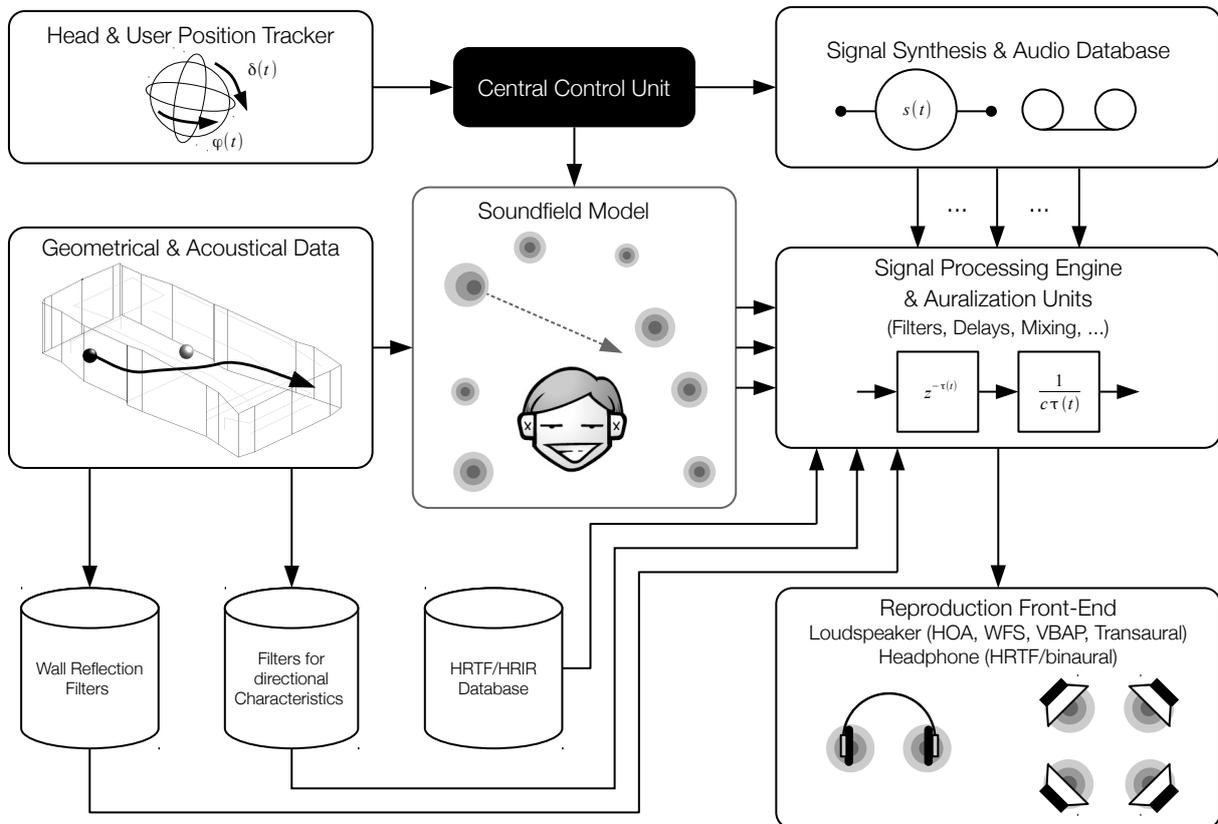


Figure 13: System architecture of an auditory virtual environment generator; cf. [Str00].

appropriate for the simulation. The replacement or extension with a sound field model - which has been created on the basis of psychoacoustic insights - can possibly bring some benefits: reduced computation load, improved scalability of the employed auralization units and in consequence less hardware-effort. These benefits are extremely relevant when the simulation of complex scenarios in real-time is required. However, when implementing a tool for further psychoacoustic research on moving sound sources, then a physically exact model needs to be used as a first step. As a second step, the exact model could then be utilized in order to create adapted sound field models according to psychoacoustic effects.

In order to create a plausible presentation of the virtual world and to care for (head) movements of the listener an update of the sound field model needs to be performed in real-time. Figure 14 shows an example of a room impulse response that can be divided into three sections: direct sound, early reflections and late reverberation (cf. [JCW97, Vor08]). In case of a relative movement between source and receiver, all of these sections become time-variant and thus are affected by pitch changes due to the Doppler effect (cf. [Str00]). Both for the direct sound and the early reflections the incidence direction, the change in sound pressure level due to the distance law and the propagation delays need to be updated in real-time [BLSS00]. Furthermore, the pattern of the reflections depends on the source and listener position due to room dimensions or obstacles inside of the simulated environment. Figure 16b shows the 'time-variant reflectrogram' of a simulated room impulse response based on a rectangular room (Figure 16a) with a moving source and stationary receiver as an example.

The time-variant convolution-integral is described by the following equation:

$$g(t) = \int_{-\infty}^{\infty} h(\tau, t) s(t - \tau) d\tau \quad (1)$$

with:  $s(t)$  := input signal  
 $h(\tau, t)$  := time-variant impulse response  
 $g(t)$  := output signal

Consequently, the relation between time-variant impulse response and the time-variant transfer-function can be described by the Fourier-Transform (cf.

[Lin04]):

$$h(\tau, t) \xrightarrow{\tau} H(f, t) = \int_{-\infty}^{\infty} h(\tau, t) e^{-j2\pi f\tau} d\tau \quad (2)$$

and

$$H(f, t) \xrightarrow{f} h(\tau, t) = \int_{-\infty}^{\infty} H(f, t) e^{j2\pi f\tau} df \quad (3)$$

For more details in terms of time-variant systems, the reader is referred to [Lin04, Rup93] in the subject of wireless communication systems namely Doppler/fading channels.

The signal block diagram of an auralization unit in order to render early reflections - based on the image source model - is shown in Figure 15. Moreover, the block diagram takes into account the directional pattern of the source, the time delays of the propagation paths, wall reflection filters and binaural rendering by using head related transfer functions (cf. [Str00]).

In order to simplify the sound field model, it can be considered to apply only a static rendering of some sound field parameters. For example, only the amplitudes and probably the incidence directions of the early reflections are updated, but the sound field is not completely re-calculated in each time frame. This strategy results in a decreased computation effort and a reduction of the complexity of the implementation.

Additionally, a model-based adaptation of the diffuse sound field resp. the late reflections within the room impulse response might be considered. Strauss [Str00] investigated the influence of a moving sound source on the statistical properties of late reverberation (e.g. average Doppler frequency scaling factors of the late reverberation tail). Here, the statistical model can be used in order to approximate the properties of the late reflections of the room impulse response by means of suitable reverb- and convolution-algorithms (cf. [Gar94, MT99, Str00, Wef08, Zöl05, Zöl11]).

One further step can be taken to simplify the model: It might be appropriate for many scenarios to modify only the direct sound depending on the movement of the sound sources and to remain the reflections unchanged. Moreover, it is possible to render only the reflections which are relevant in order to create a proper distance percept (cf. [Dür01, Dür00]). One further strategy could be to apply an amplitude

threshold and to render only the image sources that exceed the threshold. However, trade-offs like an altered spatial impression of the rendered environment due to omitted image sources have to be taken into account (cf. [Dür00]).

As incidence direction, amplitude and pitch of the reflections have minor impact on perception, this seems to be acceptable for a plausible representation of the real world. However, it can be complicated to find simplified strategies which allow a perceptually robust and plausible presentation of moving objects, due to the time-variant properties of the regarded systems.

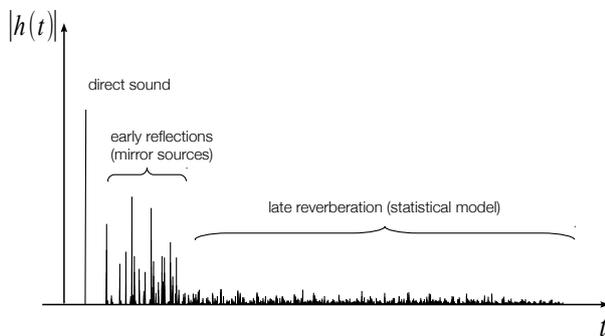


Figure 14: Room Impulse Response (RIR) with direct sound, early reflections (image sources) and the late reverberation tail (which can be described by a statistical model): Each section of the room impulse response is time-variant due to the relative movement of the source/receiver and furthermore is more or less affected by Doppler shift. The absolute values of the impulse response vector  $h(t)$  have been plotted versus time in order to give a more clear illustration of the various sections.

### 6.3 Signal Processing Engine and Auralization Units

As mentioned in the previous section, the sound field model has to be extended with regard to time-variant systems, as soon as movements of the source and the receiver in the virtual environment need to be auralized (cf. [Lin04, Rup93]). As soon as the filters and parameters for direct sound and reflections are handled frame-based, an adequate interpolation between the different time frames needs to be applied. The temporal variation of the amplitude, the change of the interaural parameters and the pitch changes need to be interpolated. Several methods, in particular with respect to HRTF/HRIR interpolation algorithms have been developed in the past and can be applied here (e.g. [FBD02, HP00, LB00, KD08]). The majority

of these methods have already been implemented in state-of-the-art virtual environment generating systems. Even though the parameters need to be carefully adjusted no extra effort has to be taken here in the implementation.

Different methods can be applied in order to auralize pitch changes (cf. [Zöl05, Zöl11]): On the one hand frequency-domain methods directly adapt the pitch according to the movement of the listener and the sound source. These methods are easy to implement and are part of a standard repertoire of signal processing algorithms. The required computation power is low and thus the method can be easily used for a plausible auralization.

One disadvantage of frequency domain-based pitch shift algorithms is that these do not correctly modify the phase information of the original signal. Moreover, the variation of the delay due to the modified distance between sound source and receiver is not modeled appropriately. However, pitch shifting by means of frequency-domain algorithms might be interesting for psychoacoustic research as it offers the possibility to control the signal-duration and the frequency-scaling factors independently from each other.

Time-domain based methods can be used which modify the length of the audio signal and by that indirectly modify pitch. When appropriately realized, these methods are a correct representation of the real world, at least from a physical point of view. However, accounting for the modified sampling rate fractionally addressed delay lines need to be applied (cf. [Smi10, Roc00]) and non-harmonic distortions due to interpolation algorithms (jitter) need to be avoided (cf. [Bor08, Fra11a, Fra11b, Smi10, Str00, Zöl05]).

Principally, both frequency-based and time-based methods have to be applied separately to the direct sound and to each of the reflections. However, as has already been described above, it might also be possible to modify the direct sound only and to let the reflections remain unchanged.

### 6.4 State-of-the-Art: Available Audio-Applications and Frameworks

A great variety of technical solutions and virtual environment generating systems exists. Depending on the sound field model the complexity of an adequate consideration of moving sound sources in the system

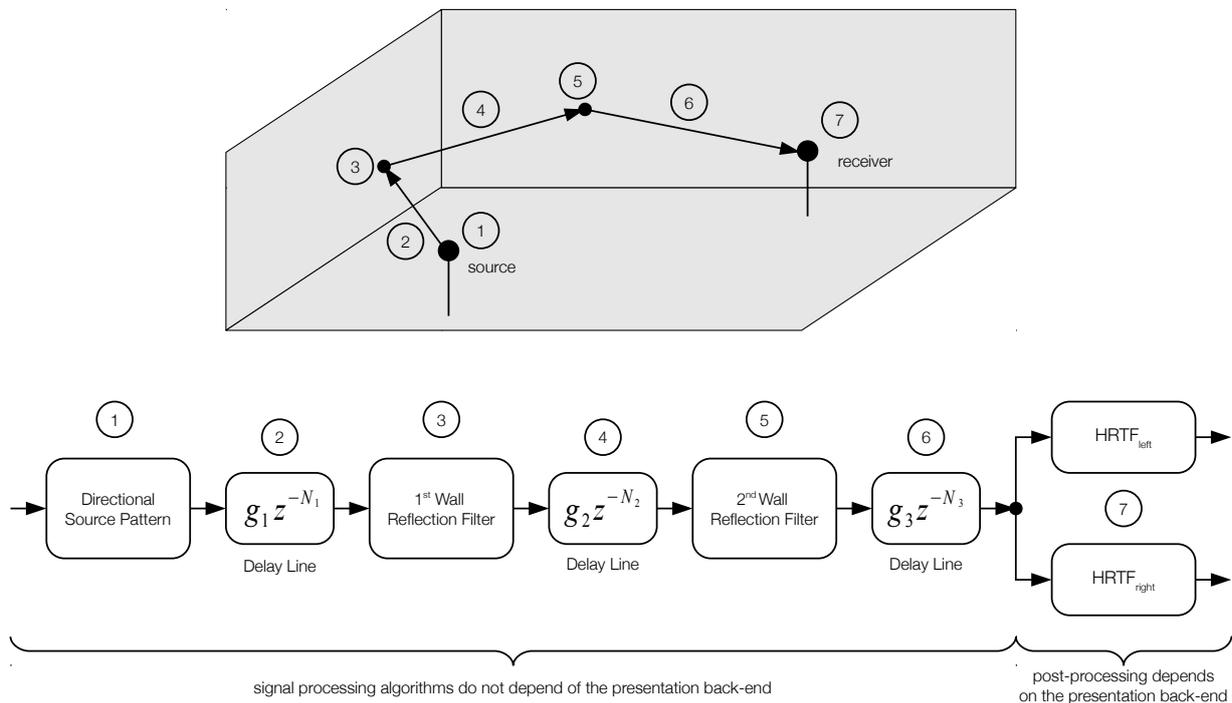


Figure 15: Block diagram: Example of a single auralization unit of an early reflection ( $2^{nd}$  order) based on the image source model. Moreover, the directional characteristics of the source is taken into account. The upper image shows the corresponding elements of the image source resp. sound field model, namely the directional characteristics of the source (#1), the propagation paths (#2, #4 and #6), the wall reflections (#3 and #5) and the receiver (#7) (after [Str00]).

architecture varies significantly. In the following only a selection of a few AVE-systems shall be named as examples:

One commercially oriented tool for spatial modeling of sound sources is the Spatial Audio Workstation by IOSONO<sup>1</sup> [MSPB10]. In combination with the specific Spatial Audio Processor [IOS13b] it supports headphone-based and loudspeaker-based auralization. So far, this system is capable of simulating moving sources under free field conditions but without featuring the simulation of reflections/reverberation resp. of time-variant room impulse responses (according to [IOS13a]).

As a commercial product designed for applications in the automotive industry the Bruel&Kjaer NVH Vehicle Simulator and PULSE - Exterior Sound Simulator (ESS)<sup>2</sup> are interesting tools in order to simulate and auralize scenarios containing moving vehicles (cf. [BK12b, BK12a]). Here, it is possible to create sce-

narios from the perspective of a pedestrian who is able to move freely through traffic situations, e.g. in order to evaluate the exterior warning sounds of electric vehicles. In addition, it is possible to create scenarios from the driver's perspective. In this case, passing vehicles can be auralized and realistic traffic scenarios can be created; similar to the scenarios which have been investigated in the psychoacoustic experiments within this article.

The freely available SoundScapeRenderer [AGS08, GAM<sup>+</sup>07] is a tool which is capable of auralizing sound fields for different reproduction set-ups. Both loudspeaker-based and headphone-based systems can be supported. An enhancement of the SoundScapeRenderer towards moving sound sources is comparably easy as the code has been published by the development team<sup>3</sup>. However, the SoundScapeRenderer is not capable of simulating reflections in the virtual environment. Thus possibly a simplified sound field model can be applied which only modifies the direct sound.

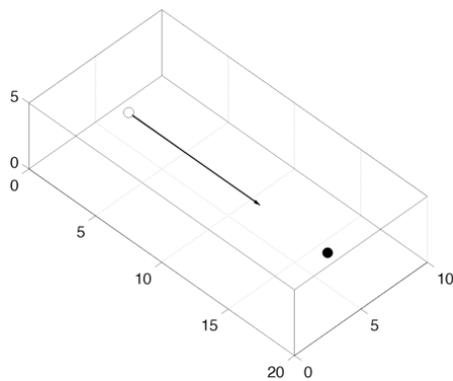
Furthermore, a freely available toolbox called Audio-

<sup>1</sup><http://www.iosono-sound.com/>

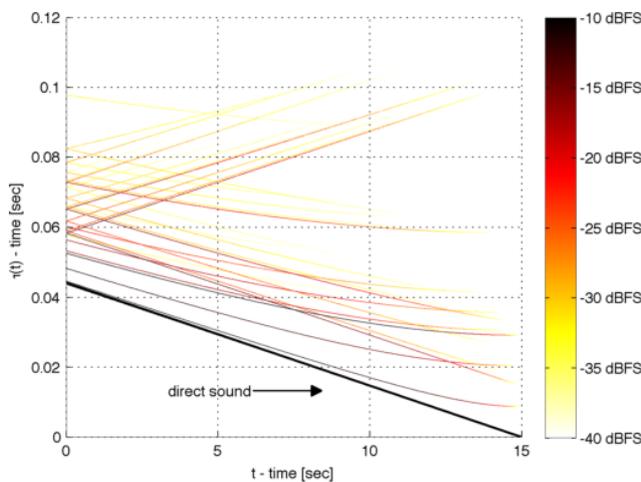
<sup>2</sup><http://www.bksv.com/products/analysis-software/acoustics/nvh-simulators.aspx>

<sup>3</sup><http://soundscaperenderer.github.com>

## 7 Summary and Conclusion



(a) Rectangular room with point source moving with constant speed towards a stationary receiver.



(b)  $L_{h_{RIR}}(\tau, t) = 20 \log_{10}(|h_{RIR, norm}(\tau, t)|)$

Figure 16: Example: Reflectogram of an normalized time-variant room impulse response of a rectangular room resulting by a moving point source with constant speed ( $v_x = 1$  m/s) and a stationary receiver: The source approaches the receiver on a straight line and arrives after 15 seconds.

Here, only the direct sound and the early reflections (image sources) are considered. Furthermore, Doppler shift was not taken into account (low velocity of the source). Moreover, the level of each element of the impulse response vectors has been calculated ( $L_{h_{RIR}}(\tau, t) = 20 \log_{10}(|h_{RIR, norm}(\tau, t)|)$ ) in order to show the relation and progression of the direct sound and the reflection patterns corresponding to the simulated image sources. The x-axis shows the time vector and the y-axis shows the sample-index of the impulse response, while the z-axis shows the colormap corresponding to the audio level  $L_{h_{RIR}}$  [dBFS] of the reflections.

ProcessingFramework - which has been partly created by the developers of the SoundScapeRenderer - could be extended by suitable algorithms in order to simulate moving objects (cf. [GS12]). Then, the AudioProcessingFramework could serve as a basis for the development of new 3-D audio applications featuring moving objects.

Two experiments have been conducted in order to understand distance perception and velocity perception of different types of moving sound sources. The sound sources were moved on a rectilinear path and passed the listener at a specific distance. The subjects were asked to determine both the velocity of the object (cf. Experiment 1) and the minimal distance from the listener (cf. Experiment 2).

The results of the first experiment can be summarized as follows: The perceived velocity strongly depends on the minimum distance of the sound source's trajectory. The same trend can be observed for both vehicle types: An increase of distance results in a decrease of the perceived velocity. The strength of this effect seems to be signal-depending. In this experiment the observed effect is stronger in case of the motorcycle than for the passenger car. Binaural effects which have been proven to be significant for the passenger car (cf. [PS09]) can be observed for the motorcycle as well. However, significance can only be proved for some of the stimuli

In Experiment 2 the parameters relevant for distance perception of moving sound sources have been analyzed. Both for the motorcycle and for the passenger car the main parameter for the determination of distance is the maximum sound pressure level at the listener's position. With regard to the perceived egocentric distance, it has been shown that the dependency on the velocity of the object is much lower for the motorcycle than for the passenger car. A reason for this is that the passenger car's sound is much stronger affected by the true velocity (e.g. by tire noise). A regression and a correlation analysis have shown that the sound pressure level is the main parameter for distance estimation in an environment with low reverberation.

On the one hand, human distance perception is not independent of velocity. On the other hand it cannot be deduced directly from the velocity. The results of experiment 2 show that the distance between a moving object and the listener can be controlled by a simple adaptation of the sound pressure level. However, in [PS09] a significant decrease in plausibility is obtained by a manipulation of the sound pressure level compared to a scenario where other dynamic and static cues are adapted as well.

Different applications have been identified and discussed which can benefit from an adequate understanding of the perception and presentation of mov-

ing sound sources. A variety of possible applications exists in industrial applications, e.g. the automotive industry, civil engineering and human-machine interfaces. Furthermore, such systems can be applied within psychoacoustic research and audiology.

The results of these experiments are useful for many applications in the area of virtual auditory environments. Distance perception of moving sound sources plays an important role in the creation of virtual environments in which the listener and the sound-radiating objects can move freely [Str00, Str98a, Str98b]. Driving simulators can serve as a typical example, where a proper auralization of moving sound sources may be required.

Finally, it has been described in which way systems that generate virtual auditory environments need to be extended and which elements in the system architectures need to be modified. It has been shown that, depending on the required accuracy, already existing virtual auditory environment generators can be easily adapted to handle moving sound sources.

One of the biggest challenges is the creation of plausible and perceptually robust rendering strategies. For instance, one approach could be to render the most important cues with higher precision and to handle the remaining cues with less priority. As a consequence, it might be possible to reduce computational effort. However, further psychoacoustic research needs to be done in this subject in order to validate such strategies.

## 8 Acknowledgements

The recordings and experiments have been conducted within several diploma theses at the Cologne University of Applied Sciences. The authors would like to thank Nicolas Ritschel who contributed to this paper by performing the psychoacoustic experiments based on the recordings of the motorcycle within his diploma thesis. Moreover, the authors would like to thank Holger Strauss who clarified our questions on the sound field model via e-mail correspondence. Last but not least, the authors would like to thank the anonymous reviewers for their comments and ideas in improving the article.

## References

[AGS08] J. Ahrens, M. Geier, and S. Spors, *The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods*, 124th

- AES Convention, Amsterdam, The Netherlands, Audio Engineering Society, 2008, paper no. 7330.
- [AS08a] J. Ahrens and S. Spors, *Reproduction of Moving Virtual Sound Sources with Special Attention to the Doppler Effect*, 124th AES Convention, Amsterdam, The Netherlands, Audio Engineering Society, 2008, paper no. 7363.
- [AS08b] J. Ahrens and S. Spors, *Reproduction of Virtual Sound Sources Moving at Supersonic Speeds in Wave Field Synthesis*, 125th AES Convention, Amsterdam, The Netherlands, Audio Engineering Society, 2008, paper no. 7557.
- [Ber88] A. J. Berkhout, *A Holographic Approach to Acoustic Control*, *Journal of the Audio Engineering Society* **36** (1988), no. 12, 977–995, ISSN 1549-4950.
- [Beu07] O. Beucher, *Wahrscheinlichkeitsrechnung und Statistik mit MATLAB: Anwendungsorientierte Einführung für Ingenieure und Naturwissenschaftler*, Springer Verlag GmbH, Berlin, 2007, ISBN 978-3-540-72155-0.
- [BG03] M. Blommer and J. Greenberg, *Realistic 3D Sound Simulation in the VIRTTEX Driving Simulator*, *Driving Simulator Conferences (DSC) North America Proceedings*, Dearborn, Michigan, USA, University of IOWA, 2003.
- [BK12a] Bruel and Kjaer, *Product information: Dts exterior sound simulator - type 8601-t*, <http://www.bksv.com/products/pulseanalyzerplatform/pulsesolutionsoverview/acousticapplications/-nvhvehiclesimulator.aspx>, 2012, last access on: December 30th, 2012.
- [BK12b] Bruel and Kjaer, *Product information: Nvh vehicle simulator - type 3644*, <http://www.bksv.com/products/pulseanalyzerplatform/pulsesolutionsoverview/acousticapplications/-nvhvehiclesimulator.aspx>, 2012, last access on: December 30th, 2012.
- [Bla97] J. Blauert, *Spatial Hearing - The Psychophysics of Human Sound Localization*, revised ed., MIT Press, Cambridge, Massachusetts, 1997, ISBN 0-262-02413-6.
- [Bla08] J. Blauert, *3-D Lautsprecher-Wiedergabemethoden*, *Fortschritte der Akustik - DAGA Tagungsband*, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 2008, pp. 25–26.
- [BLSS00] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss, *An Interactive Virtual-Environment Generator for Psychoacoustic Research. I: Architecture and Implementation*, *Acta Acustica united with Acustica* **86** (2000), no. 1, 94–102, ISSN 1610-1928.
- [Bor08] C. Borß, *Efficient Asynchronous Resampling of Audio Signals for Spatial Rendering of Mirror Image Sources*, *Fortschritte der Akustik - DAGA Tagungsband*, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 2008, pp. 903–904.
- [BR09] J. Blauert and R. Rabenstein, *Providing Surround Sound with Loudspeakers: A Synopsis of Current Methods*, *Archives of Acoustics* **37** (2009), no. 1, 5–18, ISSN 0137-5075.

- [Cho77] J. M. Chowning, *The Simulation of Moving Sound Sources*, Computer Music Journal **1** (1977), no. 3, 48–52, ISSN 0148-9267.
- [Col62] P. D. Coleman, *Failure to Localize the Source Distance of an Unfamiliar Sound*, Journal of the Acoustical Society of America **34** (1962), no. 3, 345–346, ISSN 0001-4966.
- [Col68] P. D. Coleman, *Dual Role of Frequency Spectrum in Determination of Auditory Distance*, Journal of the Acoustical Society of America **44** (1968), no. 2, 631–632, ISSN 0001-4966.
- [CV11] M. Cohen and J. Villegas, *From Wherever to Whence- and Whitherware: Augmented Audio Reality for Position-Aware Services*, 2011 IEEE International Symposium on VR Innovation (ISVRI), 2011, pp. 273–280, ISBN 978-1-4577-0055-2.
- [DE12] DIN-EN-61672-1:2010-12, *Electroacoustics - Sound level meters - Part 1: Specifications (IEC 29/724/CDV:2010) - German version prEN 61672-1:2010*, 2012.
- [DIN89] DIN 4109, *Sound insulation in buildings; requirements and testing*, 1989.
- [Dür00] B. Dürrer, *Gestaltung von Reflexionen in Auditory Displays*, Fortschritte der Akustik - DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 2000.
- [Dür01] B. Dürrer, *Untersuchungen zum Entwurf von Auditory Displays/Investigations Into the Design of Auditory Displays*, Ph.D. thesis, Ruhr-University Bochum, Institute of Communication Acoustics (ICA), 2001.
- [FBD02] F. P. Freeland, L. W. P. Biscainho, and P. S. R. Diniz, *Efficient HRTF Interpolation in 3D Moving Sound*, AES 22nd International Conference: Virtual, Synthetic, and Entertainment Audio, Audio Engineering Society, 2002, paper no. 000232.
- [FGKS07] A. Franck, A. Gräfe, T. Korn, and M. Strauss, *Reproduction of Moving Sound Sources by Wave Field Synthesis - An Analysis of Artifacts*, AES 32nd International Conference: DSP for Loudspeakers, Audio Engineering Society, 2007, paper no. 10.
- [Fra11a] A. Franck, *Performance Evaluation of Algorithms for Arbitrary Sample Rate Conversion*, 131st AES Convention, New York, USA, Audio Engineering Society, 2011, paper no. 8552.
- [Fra11b] A. Franck, *Efficient Algorithms for Arbitrary Sample Rate Conversion with Application to Wave Field Synthesis*, Ph.D. thesis, Technical University Ilmenau, Faculty of Electrical Engineering and Information Technology, 2011.
- [GAM<sup>+</sup>07] M. Geier, J. Ahrens, A. Möhl, S. Spors, J. Loh, and K. Bredies, *The SoundScape Renderer: A Versatile Framework for Spatial Audio Reproduction*, Proceedings of 1st DEGA Symposium on Wave Field Synthesis, Ilmenau, Germany, 2007.
- [Gar94] W. Gardner, *Efficient Convolution without Input-Output Delay*, 97th AES Convention, Audio Engineering Society, 1994, paper no. 3897.
- [Gol07] E. B. Goldstein, *Wahrnehmungspsychologie: Der Grundkurs (Sensation and perception)*, 7th ed., Spektrum Akademischer Verlag, Berlin, 2007, ISBN 978-3-8274-1766-4.
- [Gov10] U. S. Government, *Pedestrian safety enhancement act of 2010*, <http://www.govtrack.us/congress/billtext.xpd?bill=s111-841>, 2010, last access on: January 6th, 2013, p. 841.
- [GS12] M. Geier and S. Spors, *An Open-Source C++ Framework for Multithreaded Realtime Multichannel Audio Applications*, Proceedings of the Linux Audio Conference 2012, 2012.
- [HEK09] J. Hartung, B. Elpelt, and K.-H. Klöserer, *Statistik: Lehr- und Handbuch der angewandten Statistik*, 15th ed., Oldenbourg Wissenschaftsverlag, München, 2009, ISBN 978-3-486-59028-9.
- [Hel96] A. Hellmann, *Auditory Perception of Approaching Sound Sources, and Time-of-arrival Judgements*, Ph.D. thesis, Ruhr-University Bochum, Faculty of Psychology, Germany, 1996.
- [Hel99] A. Hellmann, *Auditive Distanzlokalisierung und der Effekt von Kopfbewegungen*, Fortschritte der Akustik - DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 1999.
- [HP00] K. Hartung and R. S. Pellegrini, *Vergleich verschiedener Interpolationsalgorithmen für Außenohrübertragungsfunktionen in dynamischen auditiven Umgebungen*, Fortschritte der Akustik, DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 2000.
- [IOS13a] IOSONO, *IOSONO Corp. - Erfurt/Germany - Operation Manual: Spatial Audio Workstation 2 (Version 2.1)*, <http://iosono-sound.com/assets/files/press/Products/Documents/-Spatial Audio Workstation/Spatial Audio Workstation Operation Manual 2.1.pdf>, 2013, last access on: February 20th, 2013.
- [IOS13b] IOSONO, *Iosono corp. - erfurt/germany - product information: Spatial audio workstation 2 (version 2.1)*, [http://www.iosono-sound.com/assets/files/IOSONO\\_IPC100\\_brochure.pdf](http://www.iosono-sound.com/assets/files/IOSONO_IPC100_brochure.pdf), 2013, last access on: February 20th, 2013.
- [JCW97] J.-M. Jot, L. Cerveau, and O. Warusfel, *Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model*, 103rd AES Convention, Audio Engineering Society, 1997, paper no. 4629.
- [JM86] B. L. Jones and P. R. McManus, *Graphic Scaling of Quantitative Terms*, SMPTE journal, Society of Motion Picture and Television Engineers **95** (1986), no. 11, 1166–1171, ISSN 0036-1682.
- [KD08] F. Keyrouz and K. Diepold, *A New HRTF Interpolation Approach for Fast Synthesis of Dynamic Environmental Interaction*, Journal of the Audio Engineering Society **56** (2008), no. 1/2, 28–35, ISSN 1549-4950.
- [Ker06] S. Kerber, *Das Fahrzeugaußengeräusch im urbanen Verkehr und seine Bedeutung für die Interaktion Fußgänger - Fahrzeug*, ATZ Automobiltechnische Zeitschrift **108** (2006), no. 7-8, 602–609, ISSN 0001-2785.

- [KF08] S. Kerber and H. Fastl, *Prediction of Perceptibility of Vehicle Exterior Noise in Background Noise*, DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin - Germany, 2008, pp. 623–624.
- [LB00] E. H. A. Langendijk and A. W. Bronkhorst, *Fidelity of Three-dimensional Sound Reproduction Using a Virtual Auditory Display*, Journal of the Acoustical Society of America **107** (2000), no. 1, 528–537, ISSN 0001-4966.
- [Len07] T. Lentz, *Binaural Technology for Virtual Reality*, Ph.D. thesis, Technical University Aachen, Institute of Technical Acoustics, 2007.
- [Lin04] J. Lindner, *Informationsübertragung: Grundlagen der Kommunikationstechnik*, ch. 5: Übertragungskkanäle, Kanalmodelle, pp. 251–266, Springer Verlag, Berlin Heidelberg, 2004.
- [LW99] R. A. Lutfi and W. Wang, *Correlational Analysis of Acoustic Cues for the Discrimination of Auditory Motion*, Journal of the Acoustical Society of America **106** (1999), no. 2, 919–928, ISSN 0001-4966.
- [MBL<sup>+</sup>89] D. H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan, *Effects of Room Reflectance and Background Noise on Perceived Auditory Distance*, Perception **18** (1989), no. 3, 403–416, ISSN 0301-0066.
- [MFT<sup>+</sup>99] P. Mackensen, U. Felderhoff, G. Theile, U. Horbach, and R. S. Pellegrini, *Binaural Room Scanning - A New Tool for Acoustic and Psychoacoustic Research*, Journal of the Acoustical Society of America **105** (1999), no. 2, 1343–1344, ISSN 0001-4966.
- [Mö100] S. Möller, *Assessment and Prediction of Speech Quality in Telecommunications*, Kluwer Academic Publ., Boston MA, 2000, ISBN 0-7923-7894-6.
- [MSPB10] F. Melchior, C. Sladeczek, A. Partzsch, and S. Brix, *Design and Implementation of an Interactive Room Simulation for Wave Field Synthesis*, 40th AES Conference, Audio Engineering Society, 2010, paper no. 7-5.
- [MT99] C. Müller-Tomfelde, *Low-Latency Convolution for Real-Time Applications*, AES 16th International Conference on Spatial Sound Reproduction, Audio Engineering Society, 1999, paper no. 16-040.
- [NHT11] NHTSA - National Highway Traffic Safety Administration, *Nhtsa studying environmental impact of 'quieter cars'*, www.nhtsa.gov/About+NHTSA/Press+Releases-/2011/NHTSA+Studying+Environmental+Impact+of+'Quieter+Cars', 2011, last access on: February 20th, 2013.
- [NHT13] NHTSA - National Highway Traffic Safety Administration, *Notice of intent to prepare an environmental assessment for pedestrian safety enhancement act of 2010 rulemaking (docket no. nhtsa-2011-0100)*, www.nhtsa.gov/staticfiles/-rulemaking/pdf/QuietCar-noi-draft.pdf, 2013, last access on: February 20th, 2013.
- [Nov05] Pedro Novo, *Communication acoustics*, ch. 11: Auditory Virtual Environments, pp. 277–298, Springer Verlag Berlin Heidelberg, 2005.
- [PMBM08] N. Peters, T. Matthews, J. Braasch, and S. McAdams, *Spatial Sound Rendering in MAX/MSP with ViMiC*, Proceedings of the International Computer Music Conference (ICMC), Belfast - Northern Ireland, International Computer Music Association (ICMA), 2008.
- [PS09] C. Pörschmann and C. Störig, *Investigations Into the Velocity and Distance Perception of Moving Sound Sources*, Acta-Acustica united with Acustica **95** (2009), no. 4, 696–706, ISSN 1610-1928.
- [Roc00] D. Rocchesso, *Fractionally-addressed delay lines*, IEEE Transactions on Speech and Audio Processing **8** (2000), no. 6, 717–727, ISSN 1063-6676.
- [RR09] R. L. Robart and L. D. Rosenblum, *Are hybrid cars too quiet?*, Journal of the Acoustical Society of America **125** (2009), no. 4, 2744, ISSN 0001-4966.
- [Rup93] W. Rupperecht, *Signale und übertragungssysteme. modelle und verfahren für die informationstechnik*, ch. 13: Lineare zeitvariante Übertragungssysteme, pp. 251–266, Springer Verlag, Berlin, 1993.
- [SB00] J. Sahrhage and J. Blauert, *Zur Architektur interaktiver Systeme zur Erzeugung virtueller Umgebungen*, Fortschritte der Akustik - DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 2000.
- [SC98] B. Shinn-Cunningham, *Applications of Virtual Auditory Displays*, 1105 – 1108, ISSN 1094-687X.
- [SH07] B. U. Seeber and E. Hafter, *Perceptual Equalization in Near-speaker Panning*, Fortschritte der Akustik - DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 2007, pp. 375–376.
- [SKG02] R. Sottek, W. Krebber, and K. Genuit, *Simulation des Außengeräusches in Innenstadtstraßen*, Fortschritte der Akustik - DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin - Germany, 2002, pp. 166–167.
- [SL93] J. M. Speigle and J. M. Loomis, *Auditory Distance Perception by Translating Observers*, Proceedings of the IEEE Symposium on Research Frontiers in Virtual Reality, 1993, pp. 92–99, ISBN 0-8186-4910-0.
- [Smi10] J. O. Smith, *Physical Audio Signal Processing*, W3K Publishing, 2010, ISBN 978-0-9745607-2-4, Online available at <https://ccrma.stanford.edu/jos/pasp/>, Last accessed November 6th, 2013.
- [SP07] C. Störig and C. Pörschmann, *Psychoakustische Experimente zur Untersuchung der Wahrnehmung Doppler-behafteter Signale*, Fortschritte der Akustik - DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 2007, pp. 377–378.
- [SS73] W. E. Simpson and L. D. Stanton, *Head movement does not facilitate perception of the distance of a source of sound*, American Journal of Psychology **86** (1973), no. 1, 151–159, ISSN 0002-9556.
- [Str98a] H. Strauss, *Realzeitsimulation von zeitvarianten Schallfeldern in virtuellen auditiven Umgebungen*, Fortschritte der Akustik - DAGA Tagungsband,

- [Str98b] H. Strauss, *Implementing Doppler Shifts for Virtual Auditory Environments*, 104th AES Convention, Audio Engineering Society, 1998, paper no. 4687.
- [Str00] H. Strauss, *Simulation instationärer Schallfelder für auditive virtuelle Umgebungen*, Ph.D. thesis, Ruhr-University Bochum, Institute of Communication Acoustics (IKA), 2000.
- [Tsi98] N. Tsingos, *Simulation of high quality sound fields for interactive graphics applications (in french)*, Ph.D. thesis, Université Joseph Fourier Grenoble, Sciences et Géographie, U.F.R. D'Informatique et de Mathématiques Appliquées, 1998.
- [Tsi01] N. Tsingos, *A versatile Software Architecture for virtual Audio Simulations*, Proceedings of the 2001 International Conference on Auditory Display (ICAD), Espoo - Finland, 2001, Available online at <http://www.acoustics.hut.fi/icad2001/proceedings/-index.htm>, last accessed November 6th, 2013, pp. 01–38–01–43, ISBN 951-22-5520-0.
- [UNE10] UNECE - United Nations Economic Commission for Europe, *Press Release: UNECE Works on Standard for Sound Device to Equip Silent Electric Vehicles*, [http://www.unece.org/press/pr2010/10trans\\_p12.html](http://www.unece.org/press/pr2010/10trans_p12.html), 2010, last access on: January 6th 2013.
- [UNE13] UNECE - United Nations Economic Commission for Europe, *World Forum for Harmonization of Vehicle Regulations (WP.29) - Working Party on Noise (GRB) - Informal group on Quiet Road Transport Vehicles (QRTV)*, [http://www.unece.org/ru/trans/main/wp29/-meeting\\_docs\\_grb.html](http://www.unece.org/ru/trans/main/wp29/-meeting_docs_grb.html), 2013, last accessed January 6th, 2013.
- [vB49] G. von Békésy, *The Moon Illusion and Similar Auditory Phenomena*, *American Journal of Psychology* **62** (1949), no. 4, 540–552, ISSN 0002-9556.
- [vK08] W.-C. von Karstedt, *Ein hybrides Verfahren zur Berechnung von Schallfeldern infolge bewegter Quellen*, Ph.D. thesis, Technische Universität Hamburg-Harburg - Institut für Modellierung und Berechnung, 2008.
- [Vor08] M. Vorländer, *Auralization: Fundamentals of Acoustics*, Springer Verlag, Berlin, 2008, ISBN 978-3-540-48829-3.
- [Wef08] F. Wefers, *Optimierung der Filtersegmentierung für die Echtzeitfaltung*, Fortschritte der Akustik - DAGA Tagungsband, Deutsche Gesellschaft für Akustik (DEGA e.V.), Berlin, Germany, 2008, pp. 901–902.
- [Zah02] P. Zahorik, *Assessing auditory distance perception using virtual acoustics*, *Journal of the Acoustical Society of America* **111** (2002), no. 4, 1832–1846, ISSN 0001-4966.
- [ZBB05] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, *Auditory Distance Perception in Humans: A Summary of Past and Present Research*, *Acta-Acustica united with Acustica* **91** (2005), no. 3, 409–420, ISSN 1610-1928.
- [Zel09] P. Zeller (ed.), *Handbuch der Fahrzeugakustik: Grundlagen, Auslegung, Berechnung, Versuch*, ch. 14.3: Außengeräusch - Akustische Wahrnehmbarkeit, pp. 273–279, Vieweg + Teubner, GVV Fachverlage, 2009.
- [Zöl05] U. Zölzer, *Digitale Audiosignalverarbeitung*, 3rd ed., Teubner, Stuttgart, 2005, ISBN 3-519-26180-4.
- [Zöl11] U. Zölzer, *DAFX - Digital Audio Effects*, 2nd ed., Wiley, New York, 2011, ISBN 978-0-470-66599-2.

Citation
<p>Christian Störig and Christoph Pörschmann  <i>Investigations into Velocity and Distance Perception Based on Different Types of Moving Sound Sources with Respect to Auditory Virtual Environments</i>,  <i>Journal of Virtual Reality and Broadcasting</i>,                      10(2013), no. 4, November 2013, urn:nbn:de:0009-6-38009, ISSN 1860-2037.</p>