Wilhelm Wimmer

MULTI-SPEAKER SET-UP FOR AUDIOLOGICAL MEASUREMENTS USING GESTURE RECOGNITION

Master thesis

Institute of Medical Engineering
Graz University of Technology
Kronesgasse 5, A-8010 Graz
Head: Univ.-Prof.Dipl.-Ing.Dr.techn. Rudolf Stollberger

ARTORG Center - Artificial Hearing Research
University of Bern
Murtenstrasse 50, CH-3010 Bern
Head: Prof. Dr. ès sc. Christof Stieger

Supervisor: Prof. Dr. ès sc. Christof Stieger

Evaluator: Univ.-Prof.Dipl.-Ing.Dr.techn. Rudolf Stollberger

Bern, October 2011

Acknowledgements

First of all, I would like to thank my supervisor Dr. Christof Stieger for allowing me to join his team and for the continuous support, motivation and enthusiasm. I specially thank him for his infinite patience. The insightful conversations I had with him were invaluable.

I would like to thank Dr. Rudolf Stollberger for his support in Graz and for agreeing to evaluate the thesis.

Financial support was provided by the Faculty of Electrical and Information Engineering at Graz University of Technology and the ERASMUS mobility programme.

I am indebted to many of my colleagues at the ARTORG center in Bern for invaluable assistance, for sharing the literature and for having a great time at work. Especially Jonas Salzmann, who was a great project partner and whose programming skills were priceless. This thesis would not be the same without his many helpful suggestions.

I owe my deepest gratitude to my girlfriend Selina and her parents for their warm welcome and overwhelming support and motivation during my stay in Bern.

Lastly, I am grateful for my family and their endless love and guidance throughout my life.

Statutory Declaration

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

Signature of the author

Abstract

Multi-speaker set-up for audiological measurements using gesture recognition

Spatial hearing measurements conducted under free sound-field conditions require a pointing method which allows the subject to indicate the perceived auditory event. Routine investigations with traditional set-ups often include interactions between the operator and the subject, leading to time-consuming test procedures with complex instructions.

This thesis presents a multi-speaker set-up with a new approach to pointing methods using gesture recognition. An audio processing framework for the software of the measurement set-up (`AIODE`) was implemented to provide the hardware controls and the signal processing routines required for sound localization and speech intelligibility experiments in the horizontal plane. Based on room acoustics measurements and the specifications given by EN ISO 8253-2 and EN ISO 8253-3, the test environment has been characterized. The linear behaviour and acoustic output of the system were validated and a calibration procedure was elaborated. The set-up provides acoustic stimuli with up to 90 dB SPL and a flat frequency response between 0.7 and 12.5 kHz.

The pointer method using automatic gesture recognition was realized with a `MICROSOFT KINECT` sensor. The temporal tracking characteristics and the detection accuracy of the system were analyzed and evaluated. The presented set-up supports the automated detection of indicated directions with 15° accuracy. A practicability test with five volunteers with relation to the institute was conducted, investigating effects on the responses through visual cues given by the speakers. After a short training session (<5 min), test procedures with 36 test steps were performed in less than 5 min. The experiences made in the practicability test promise good applicability of the measurement set-up with gesture recognition for further audiological experiments.

audiology, spatial hearing measurements, multi-speaker set-up, gesture recognition

Zusammenfassung

Multi-Lautsprecher Messsystem für audiologische Untersuchungen mit Gestenerkennung

Audiologische Messungen im freien Schallfeld werden oft durch die Anzeigemethode wahrgenommener Hörereignisse eingeschränkt. Herkömmliche Verfahren zur Routineuntersuchung beziehen meist den Tester und die Testperson ein und benötigen zeitaufwändige Testprozeduren und komplexe Anleitungen.

Diese Arbeit stellt ein Multi-Lautsprechersystem für Schalllokalisations- und Sprachverständlichkeitsexperimente vor, welches die Untersuchungen mithilfe von Gestenerkennung erleichtern kann. Die Hardware-Steuerung der Messanlage und Signalverabeitungsroutinen wurden als Teil der Messsystem-Software `AIODE` implementiert. Die Messumgebung wurde anhand der Richtlinien von EN ISO 8253-2 und EN ISO 8253-3 und raumakustischer Parameter charakterisiert. Ein Kalibrationsverfahren wurde ausgearbeitet und das lineare Systemverhalten sowie akustische Ausgangsgrößen validiert. Mit der Messanlage können akustische Stimuli mit bis zu 90 dB SPL bei einem flachen Frequenzgang zwischen 0.7 und 12.5 kHz erzeugt werden.

Die Gestenerkennung wurde mit einem `MICROSOFT KINECT` Sensor realisiert und stellt einen neuen Ansatz zur Erkennung von angezeigten Hörereignissen dar. Die zeitlichen Verläufe und die Detektionsgenauigkeit des Systems wurden analysiert und ausgewertet. Die vorgestellte Anlage erkennt angezeigte Richtungen mit einer Genauigkeit von bis zu 15°. Es wurde ein Praxistest mit 5 Testpersonen durchgeführt, um den visuellen Einfluss der Lautsprecher auf die Antworten zu untersuchen. Die Testverfahren bestehend aus 36 Testschritten konnten nach einer kurzen Übungsphase (<5 min) in weniger als 5 min durchgeführt werden. Die Erfahrungen versprechen eine gute Anwendbarkeit der Messanlage mit Gestenerkennung für weitere audiologische Experimente.

Audiologie, räumliches Hören, Multi-Lautsprecher Messanlage, Gestenerkennung

Contents

Appendices

List of Acronyms

API	Application programming interface
b	Response bias
d	Sensor distance to the reference point
d_c	Critical distance
dB HL	Decibel hearing level
dB FS	Decibel full scale
dB SPL	Decibel sound pressure level
DSP	Digital signal processing
EDC	Energy decay curve
f_c	Center frequency
Δf	Bandwidth
g	Response gain
g_{fmod}	Software equalizer gain factor
FFT	Fast Fourier transformation
HRTF	Head-related transfer function
iFFT	Inverse fast Fourier transformation
ILD	Interaural level difference
ITD	Interaural time difference
L	Sound pressure level
$\overline{L}$	Sample mean sound pressure level
ΔL	Sound pressure level difference
OLSA	Oldenburger Sentence Test
PCM	Pulse code modulation
p_{ref}	Sound pressure reference
q	Interpolation factor
RTF	Room transfer function
RIR	Room impulse respone
RT_{60}	Reverberation time
r	Radial distance
r^2	Pearson's correlation coefficient
SRT	Speech Reception Threshold
s	Sample standard deviation
THD	Total harmonic distortion
t_d	Delay time of subject's response
t_i	Integration time
t_s	Duration of acoustic stimulus
$\overline{T}_{\mathrm{step}}$	Average duration of test step
$\overline{T}_{\mathrm{tot}}$	Average duration of total test procedure
V	Volume of the room
VOI	Volume of interest
α	Inclination of the sensor
δ	Response elevation
ε	Angular deviation threshold
ϑ_V	Vertical aperture angle of the sensor
ϑ_H	Horizontal aperture angle of the sensor
φ	Response azimuth
$\overline{\varphi}$	Sample mean azimuth error
ϕ	Stimulus azimuth
ϕ_{tot}	Total measurement range
$\Delta\phi$	Speaker spacing

1. Introduction

This chapter specifies the motivation and the objectives of this thesis. Further, an overview of the chapters is presented. The last section introduces terms and concepts of sound localization in the horizontal plane.

1.1. Motivation

Binaural hearing plays a crucial role in everyday life. It facilitates the communication in our social environment and helps to assess potentially dangerous situations. With two functioning ears it is possible to accurately locate sound sources or to detect speech signals in noisy environments. Consequently, it is of great interest for audiologists to be able to determine the binaural hearing abilities of patients.

Sound localization phenomena are often investigated under free sound-field conditions using loudspeaker arrays. These experiments require a so-called pointer method, which allows the subject to indicate the perceived auditory event. Often visual pointer methods, for example pointing toward the auditory event with a finger or with a movable source of light, are used. Traditionally, test procedures are limited by this methods in means of complexity and duration, since interactions between the operator and the subject are required.

For this purpose, a multi-speaker set-up with a new approach to pointing methods using gesture recognition should be developed and validated for audiological applications.

1.2. Objectives

The aim of this work is to make an existing measurement hardware applicable for uncomplicated and fast audiological sound field experiments. The following objectives are defined:

- The test software should support localization tests and speech intelligibility tests. This includes the implementation of software audio processing components.
- The available test environment has to be characterized. This requires the measurement and documentation of room acoustics parameters.
- Test results should be comparable to other measurements. The calibration procedure has to be described and acoustic parameters have to be validated.
- A new pointing method using a time-of-flight camera for gesture recognition has been implemented. The data obtained by skeletal tracking has to be analyzed and validated.
- Finally, the applicability of the set-up has to be evaluated. A test evaluation with healthy subjects should be conducted.

1.3. Overview

The chapter ***Introduction*** specifies the motivation and the objectives of this thesis. The last section introduces terms and concepts of sound localization in the horizontal plane.

The chapter ***System Description*** describes the hardware and software structure of the measurement set-up. The implemented audio processing components are shown and the signal processing elements are depicted.

The chapter ***Test Environment Validation*** embraces the methods and specifications required to characterize the available test environment. Acoustic measurements are performed to specify the sound-field present in the test room. Finally, the application area and the limits given by room acoustics are derived.

The chapter ***Acoustic Validation and Calibration*** describes the concepts and methods used for calibration and validation of the system output. A calibration procedure using both hardware and software equalization is presented. Further, the applicable area of the set-up based on harmonic distortion and linearity measurements is discussed.

The chapter ***Gesture Recognition*** explains the algorithm used for gesture recognition with the `MICROSOFT KINECT` sensor. The determination of the optimal sensor position and a position error correction method are described. Finally, the data obtained by the sensor is analysed and validated.

The chapter ***Practicability Test*** describes the test conducted to evaluate the applicability of the measurement system and reviews existing studies dealing with human sound localization in the horizontal plane.

The chapter ***Conclusion*** summarizes the measurement system capability and gives an outlook for possible improvements of the measurement set-up.

1.4. Principles of Spatial Hearing

This section gives a basic overview of terms and concepts in psychoacoustics, with the main focus put on sound localization in the horizontal plane. For further informations about spatial hearing see Blauert [BLA97].

The concept of *spatial hearing* distinguishes between the physical *sound event* and the perceived, psycho-acoustical *auditory event*. One must keep in mind that auditory events can also occur without corresponding mechanical vibrations or waves (tinnitus) and sound events do not have to cause auditory perception. This relationship is investigated in spatial hearing measurements. An auditory experiment includes interactions between the operator and the testing subject and only subjectively influenced data can be acquired.

Two different units for the sound pressure are used in this thesis. The sound pressure level (L_P in dB SPL) is measured on a logarithmic scale and is defined as

$$L_\mathrm{P} = 20 \cdot \log_{10}\left(\frac{p}{p_\mathrm{ref}}\right) \qquad \text{with} \quad p_\mathrm{ref} = 20\,\mu\mathrm{Pa}. \tag{1.1}$$

The hearing level (L_H in dB HL) is used in audiometry and considers the frequency dependency of perceived loudness. It is defined as the difference between the sound pressure level L_P of a specified signal and the appropriate reference threshold level[1]. Figure 1.1 shows the limits of human sound perception and the areas covered by speech and music.

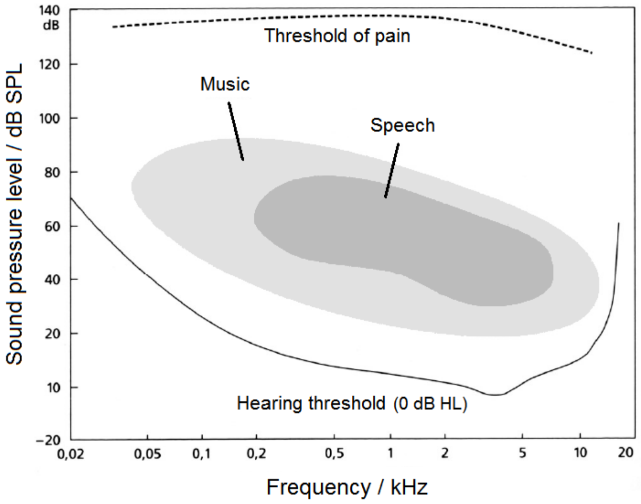


Figure 1.1.: *Auditory thresholds of a human being as well as frequency-level areas used in speech and music. Most audiometric tests are conducted in a frequency range within 125 Hz and 8 kHz, with sound pressure levels up to 100 dB SPL [KUT04].*

Auditory experiments use a system of *head-related coordinates* (figure 1.2).

[1] Reference thresholds under free-field conditions are listed in ISO 389-7 [I389].

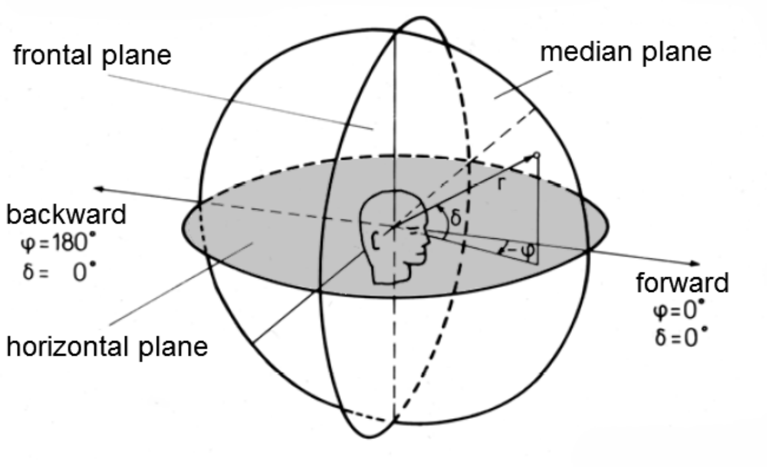


Figure 1.2.: *Head-related coordinate system with distance r, azimuth φ and elevation δ. The horizontal plane is shaded gray [BLA97].*

The origin of the coordinates lies in the middle of the *interaural axis* and is called the *reference point*. The plane formed by the lower margins of the eye sockets and the *interaural axis* is defined as the *horizontal plane*.

Two main attributes for sound evaluation are derived. First, the *interaural time differences* (ITD), which can be further subdivided into the *phase delay* and the *group delay*, describe differences in carrier time shifts and envelope time shifts respectively.
Second, the *interaural level differences* (ILD) are caused by different sound intensities at the eardrums and are evaluated along the entire audible frequency range. The interaction of ITDs and ILDs as a function of frequency is illustrated in figure 1.3.

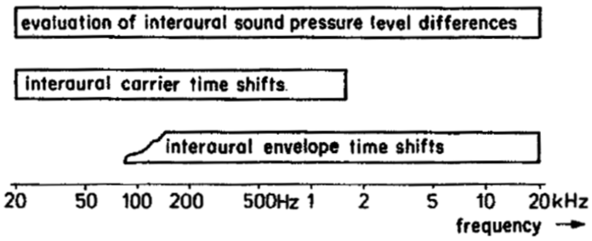


Figure 1.3.: *Evaluation of interaural differences. Differences in carrier time shifts have an effect only below 1.6 kHz [BLA97].*

Sound source locations can not be unambiguously specified with the informations given by ILD and ITD, since various source positions can cause the same interaural differences (*'cone of confusion'*).

The head position plays a very important role in sound localization, as additional informations are gathered through head movement. Ears, head and torso form a movable antenna with directional characteristics and code spatial information to temporal and spectral information[2]. This effect is suppressed when the maximum stimulus length is limited to $T_{\max} = 200\text{-}300$ ms.

Figure 1.4 shows the localization blur for normal hearing and hearing impaired subjects. The minimum localization blur occurs in the forward direction. The maximum spatial resolution of the auditory system is approximately 1°.

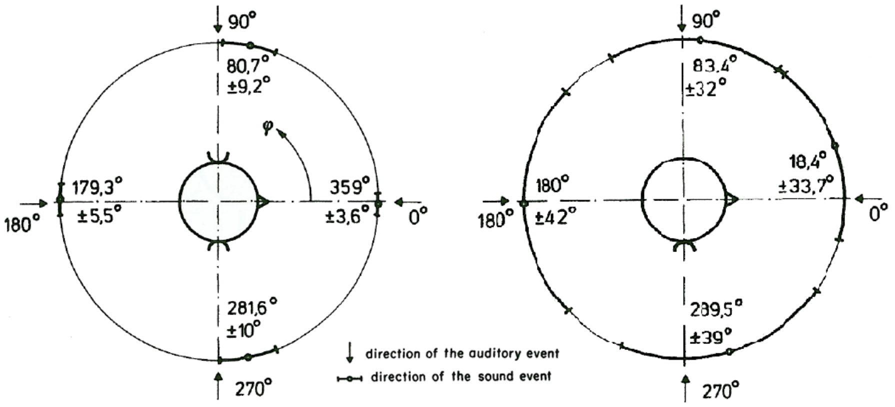


Figure 1.4.: *Localization and localization blur results for normal hearing subjects (left) and subjects with total deafness in the left ear (right) [BLA97].*

When simulating virtual sound sources by multiple sound sources and radiating coherent signals, three different effects occur. The predominating effect mainly is determined by the delay of the arriving wave fronts.

If the delay lies within 1 ms, *summing localization* causes the subject to perceive the auditory event at a position determined by both sound sources. A further delay of radiation leads to the *precedence effect* or even *echoes*. The *precedence effect* describes the phenomenon that the position of the auditory event is determined only by the signal arriving first.

[2] This linear distortion can be described by head-related transfer functions (HRTF). More information can be found in Blauert [BLA97] or Paulus [PAU03].

2. System Description

This chapter describes the hardware and software structure of the measurement set-up. The implemented audio processing components are shown and the signal processing elements are depicted.

2.1. Introduction

As part of this thesis the audio processing framework of the software `AIODE` has to be implemented, which includes the usage of a suitable sound application programming interface (API). The framework should provide the hardware controls and the signal processing routines required for sound-field measurements. Further, the framework should be expandable for future audiological test modules.
In the following sections the existing measurement set-up and the developed software are described.

2.1.1. Existing System

Figure 2.1 outlines the schematic structure of the measurement set-up and its devices. The software (`AIODE`) interacts with the sound card (`TASCAM US-2000`), the switch box/amplifier (`MERZ MEDIZINTECHNIK AUDIOBOX`) and the gesture recognition sensor (`MICROSOFT KINECT`).

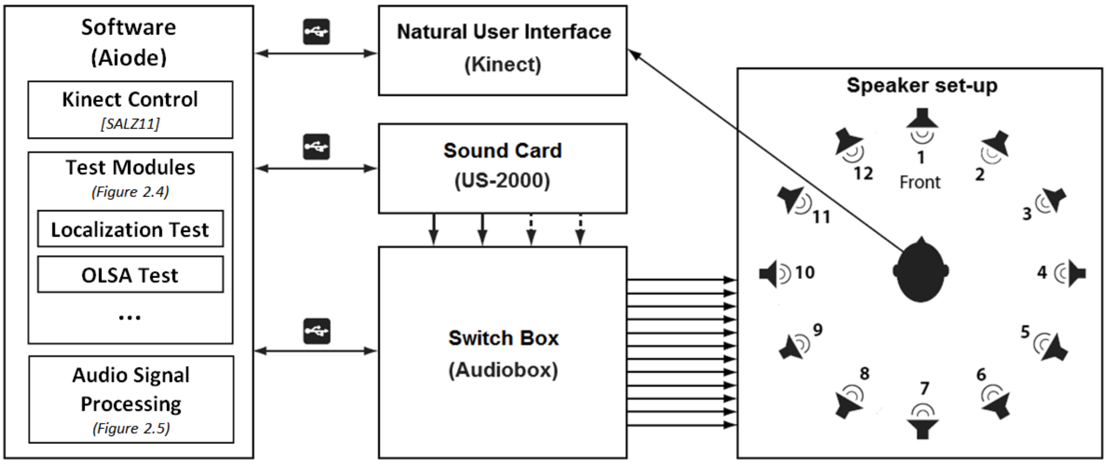


Figure 2.1.: *Components of the measurement set-up, adapted from Berger [BER10].*

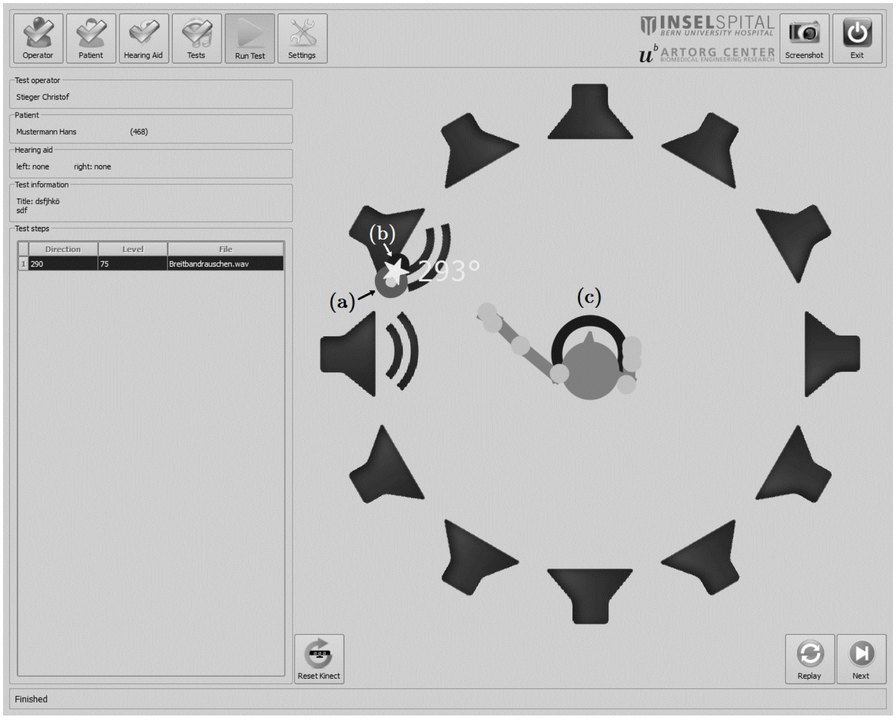
The subject is seated in a chair at the center of 12 circularly arranged loudspeakers (*JBL Control 1 pro*) placed at 30° intervals. The speakers are positioned at ear height (1.2 m) in a distance of 1 m. The whole set-up is built up in a double-walled sound-attenuating chamber (6.0 m × 4.1 m × 2.2 m).

The gesture recognition sensor is located in front of the subject and acts as a *natural user interface*. This means that the pointing gestures of the subject are automatically captured and interpreted as directional indications.

2.1.2. AIODE - Software

AIODE is specially designed for the set-up shown in figure 2.1. It supports sound-field audiometry measurements with automatic gesture recognition. Further, it features patient administration and test result analysis. More details about the functionality of AIODE and the implementation with QT (Nokia Norge AS, Oslo, Norway)[1] can be found in the AIODE software documentation (Salzmann [SAL11]).
The application is built up on a modular basis and is easily expandable. The current software version features two tests which are realized as *test modules*:

The *Localization test* module provides functionality for spatial hearing measurements with automatic gesture recognition. Sound files can be loaded and presented as stimuli from any direction in the horizontal plane, including virtual sound sources. The direction indicated by the subject is automatically detected and stored. A screenshot of the implemented software during a localization test is shown in figure 2.2.



Test operator
Stieger Christof

Patient
Mustermann Hans (468)

Hearing aid
left: none right: none

Test information
Title: dsfjhkö
sdf

Test steps

	Direction	Level	File
1	290	75	Breitbandrauschen.wav

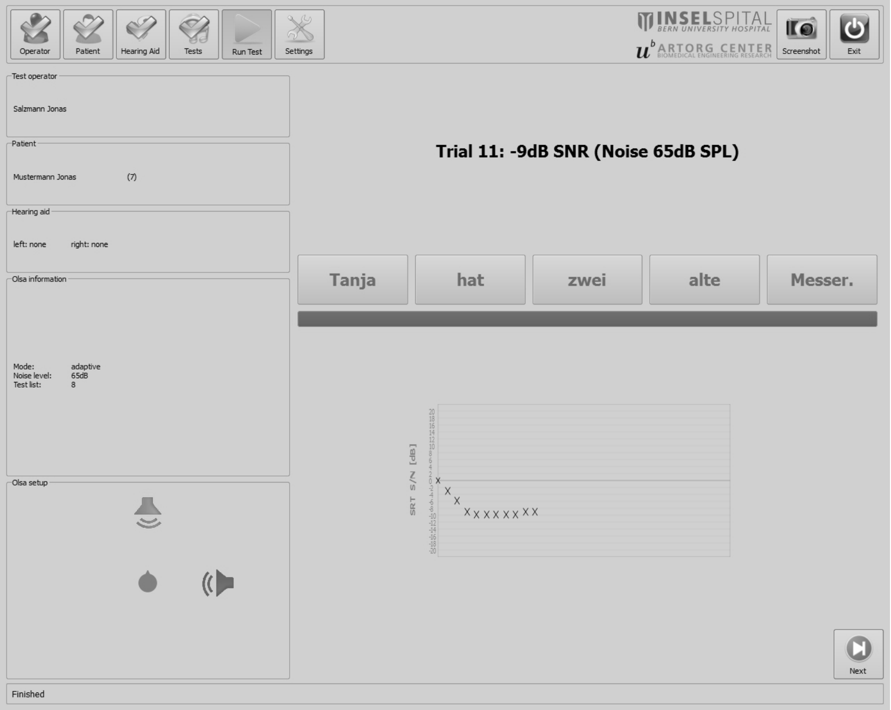
Finished

Figure 2.2.: *Screenshot of AIODE during a localization test. The basic program parts and a top down view of the subject are shown. The actual source location (a), the tracked (b) and the detected direction (star) are marked. Additionally, the position error of the subject with respect to the measurement set-up is indicated (c). The gray dots mark the subject's tracked skeletal joints.*

[1] QT is a framework for advanced C++ GUI-programming, see Blanchette [BLN09].

The *OLSA test* module is the implementation of the Oldenburger Sentence Test (OLSA), a speech intelligibility test which is commonly used within the German speaking area. Usually, the speech reception threshold (SRT) is tested with the standard speaker configuration from 0° and 90°. Depending on the direction of signal and noise, different SRTs are obtained (S_0N_0, S_0N_{90} left/right and $S_0N_{\pm 90}$).

This module provides automated OLSA procedures with arbitrary speaker configurations. Further informations about the test methods and the specifications can be found in the official OLSA handbook [OLS00].



Test operator

Salzmann Jonas

Patient

Mustermann Jonas (7)

Hearing aid

left: none right: none

Olsa information

Mode: adaptive
Noise level: 65dB
Test list: 8

Olsa setup

Trial 11: -9dB SNR (Noise 65dB SPL)

Finished

Figure 2.3.: *OLSA test performed in AIODE. The test signal is represented as a sentence with five words which can be marked if repeated correctly by the subject. The test automatically adjusts the signal level during the test. The progress of the presented signal-to-noise ratio is shown as a diagram. At the bottom left corner the currently used speaker configuration is displayed.*

2.2. Implemented Audio Processing System

2.2.1. Test Modules

Both the *Localization test* module and the *OLSA test* module need access to the sound processing routines as shown in figure 2.4.

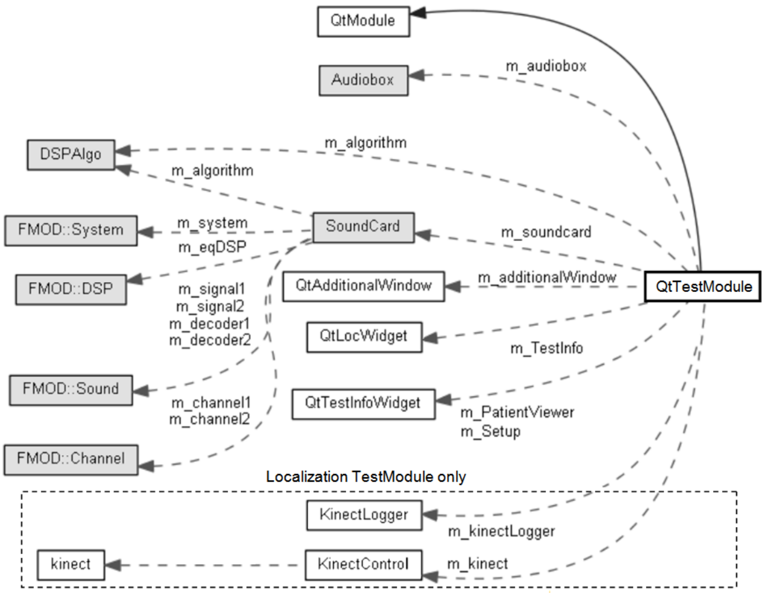


Figure 2.4.: *Dependency diagram of the TestModule-class and relevant object instances. The TestModule-class is derived from the QtModule-class. Gray shaded classes form the sound processing system. Classes with the prefix 'Qt' are part of the graphical user interface. The localization test module additionally requires the `Kinect`-sensor classes for gesture recognition routines.*

The ***QtTestModule*** is a representation of the particular test module and is derived from the *QtModule*-class. The following classes are interacting with a test module:

The ***SoundCard***-class controls the sound card device and uses the `Fmod`-library[2] for file decoding and play-back functionalities. It initializes the `Fmod`-system and the sound device, performs the desired DSP algorithms and generates the stimuli. Since only one virtual sound source at the same time is demanded, 2 output channels are in usage. However, the system could be extended to up to 4 output channels.

The ***Audiobox***-class controls the switch box and sets the calibration values for the maximum level. At program start-up the driver is loaded and the calibration values are read from the configuration file. The input signal is amplified and connected to the desired speaker channel. The stimulus is presented after a fade-in time of about 550 ms, which is the duration of a whole switching operation as declared by the manufacturer.

[2] `Fmod` is an audio API with full C++ support, see the `Fmod`-documentation [FMO11].

The ***DSPAlgo***-class is an abstract base class which provides basic routines for own DSP algorithms and the computation of coefficients. Custom sound processing algorithms are derived from this class. The classes *DSPCrossCanc*, *DSPAmpPan* and *DSPOlsa* were implemented.

The graphical user interface is realized by the classes with the ***'Qt'***-prefix. Included is the graphical output on a second display with feedback information for the subject.

The localization test module additionally requires the `Kinect`-classes for gesture recognition routines.

2.2.2. Digital Signal Processing

Figure 2.5 shows the implemented audio signal processing chain. The software supports the standard CD audio format (44.1 kHz, 16-bit stereo) and 16-bit mono/stereo wave-files with other sampling rates. Other file formats may be supported through the extension with functions provided by `Fmod`.

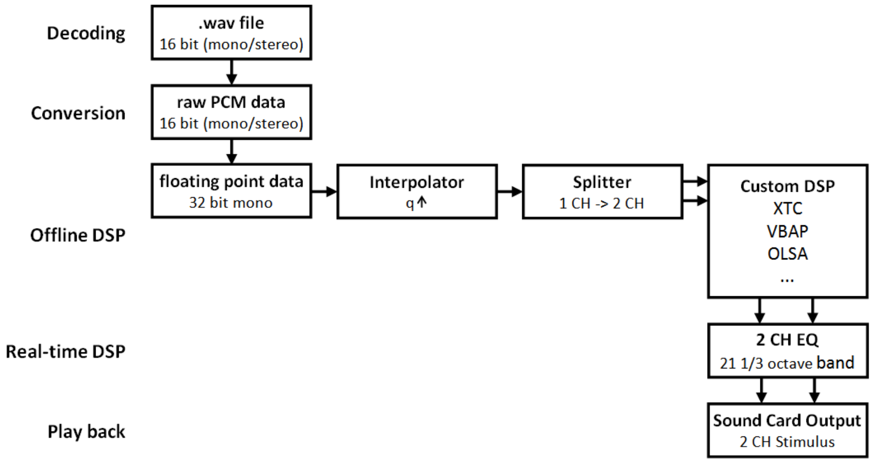


Figure 2.5.: *Implemented audio signal processing stages. Custom algorithms are separately computed for each channel and inserted at the 'Custom DSP'-block.*

The signal processing chain consists of the following stages:

In the ***Decoding*** stage, the file is read into a decoding `Fmod` sound-object to access PCM raw data.

In the ***Conversion*** stage, the samples are converted to 32-bit floating point data. This provides a wider dynamic range and higher precision. With 32-bit data a range of approximately -192 dB FS is available and no data clipping can occur, since the samples are limited to ±1.0.

The ***Offline DSP*** stage represents the signal processing steps that are applied to the samples before the play-back is started.
The *Interpolator* is a fractional delay filter[3] and originally was implemented to achieve better crosstalk filtering results at higher sampling rates. Although no better results were achieved, the interpolator is still useful as it converts other rates to the desired sampling rate (default 44.1 kHz).
The *Splitter* provides two identical mono streams as input for the *Custom DSP*-block. It is intended as an interface for custom algorithms (*DSPAlgo*) or future filter routines.

The ***Real-time DSP*** stage represents the software equalizers. The equalization is performed in real-time using the effects provided by `Fmod`. The `DSP_PARAMEQ`-effect is an adjustable bandpass filter with the parameters center frequency, octave range and gain. The equalizer was realized with 21 1/3-octave band filters at the center frequencies listed in table 3.1.

2.2.3. Filter Algorithms

Three different algorithms were implemented as classes derived from the abstract *DSPAlgo* class. Two algorithms were realized in order to provide virtual sound source functionality. With virtual sound sources, the direction of the presented stimulus can be effected based on psychoacoustic effects. Stimuli can be presented from any direction using a limited number of speakers.
Additionally, an algorithm was implemented for the OLSA test module. In the following, the algorithms are described briefly:

DSPCrossCanc uses crosstalk cancellation which is an effective method for virtual sound source imaging. With crosstalk cancellation, a couple of loudspeakers delivers audio signals at one ear without influencing the sound pressure level at the other ear (crosstalk). Put simply, only signals of the left speaker arrive at the left ear and vice versa. It is possible to simulate virtual sound sources with XTC filters.
The algorithm was implemented on basis of the filter design in Gut [GUT10]. It was adapted to multiple speaker set-ups with variable spacing. A derivation of the adapted calculation of the filter coefficients can be found in appendix C.
A major drawback of the XTC algorithm is the spectral coloration caused by the filter. After tests with higher sampling rates, the coloration did not decrease[4]. Due to the distortion of the stimuli the algorithm is not useful for auditory experiments. A possible solution of the problem using frequency-dependent regularization is elaborated in Choueiri [CHO10].

[3] The interpolation filter was implemented with the `libresample`-library, see Mazzoni [MAZ04].
[4] Gut [GUT10] observed best performance of the filter with an sampling factor $q = 4$.

In ***DSPAmpPan***, vector base amplitude panning (VBAP) is applied for the simulation of virtual sound sources. The position of the sound source is reformulated with vectors and vector bases given by the location of the loudspeakers (Pulkki [PUL97]). This simplifies the algorithm to the calculation of gain factors (ILD) for multiple speaker set-ups.

The ***DSPOlsa*** algorithm provides the functionality needed for OLSA tests. Both, the speech and the noise signal can be played at any speaker position. Further, a mix-down can be performed to present both channels from only one direction. The noise signal has an adjustable offset and fade-in time.

2.3. Discussion

With the implementation of the presented audio processing framework and test modules, the software AIODE provides all functions needed for human sound localization tests and speech intelligibility tests in the horizontal plane.

Spatial hearing experiments can be conducted with the user-friendly *Localization test* module. Additionally, the subject can be tested with predefined automated OLSA test sets, resulting in fast and less complicated OLSA test procedures. OLSA results are automatically evaluated and illustrated in a final report after testing.
Future test functionalities can be added easily due to the modular based software design. This could be the implementation of an audiometric test module for the estimation of hearing thresholds or other speech intelligibility tests.

An audio framework based on the Fmod-library was realized, supporting the standard CD audio wave format. The framework can be adapted for additional file format support. The initially used BASS/BASSasio-libraries were replaced with Fmod due to a better C++-support.
The hardware set-up is able to present acoustic stimuli with a delay time of approximately 550 ms. The delay time is used for multiple audio signal processing routines, including an interpolator, 32-bit floating point DSP filters and a multi-channel parametric equalizer.

Two filters for virtual sound source imaging with variable speaker spacings were implemented as algorithm classes. The *XTC*-algorithm uses crosstalk cancellation for the sound source simulation and was adapted for multiple-speaker arrangements. Hearing evaluation tests showed bad sound quality due to spectral coloration, which makes the filter inapplicable for localization tests. A frequency-dependent regularisation of the filter coefficients could significantly improve the signal quality. As a consequence, the simpler *VBAP*-algorithm was implemented. Since only level differences (ILD) are modulated, the sound quality is not influenced by this filter. Nevertheless, *summing localization* and the *precedence effect* necessitate the validation of both virtual sound source algorithms before applying in spatial hearing experiments.

3. Test Environment Validation

This chapter embraces the methods and specifications required to characterize the available test environment. Acoustic measurements are performed to specify the sound-field present in the test room. Finally, the application area and the limits given by room acoustics are derived.

3.1. Introduction

Free-field audiometric examinations are affected by a number of factors that do not exist in measurements using headphones. Since the position of the ears in the sound field can not be held constant, measurements are affected through the subject's size and body motion during tests. This chapter uses the specifications mentioned in EN ISO 8253-2 [I8253a] to rate the influence of the test environment. The effects of standing waves and reflections can be described by the reverberation time and the type of sound field. In addition, the knowledge of ambient noise levels is required, as test signals may be masked.

3.1.1. Sound Field Characteristics

In order to establish *quasi-free* sound field conditions according to EN ISO 8253-2 [I8253a], the loudspeakers have to be arranged at head-height of the seated subject (about 1.2 m). The distance between the speakers and the reference point should be at least 1 m, so the inverse distance law is applicable.
In the absence of the test subject, sound pressure levels in a distance of 0.15 m to the reference point should not deviate by more than ±2 dB. The sound pressure levels 0.1 m in front and behind the reference point should not deviate more than ±1 dB from the values predicted by the inverse distance law.
Since the signals for testing the sound field shall be the same as those to be used for audiometry, measurements should be carried out for any new signal used as a test stimulus.

The room should have an adequate size (at least 3 m x 3 m) and low reverberation times (250 ms). Further, the layout of furniture, equipment and people in the room during testing has to be defined[1].

3.1.2. Ambient Noise

When testing down to 0 dB HL, the ambient noise levels in the test room should not exceed the levels given in table 3.1. Other hearing level thresholds are obtained by adding the value of the lowest threshold level.

Table 3.1.: *Maximum permissible ambient noise levels in 1/3-octave bands with center frequency f_c when testing down to 0 dB HL. The lowest test tone frequency is assumed to be 250 Hz [I8253a].*

f_c / Hz	100	125	160	200	250	315	400
$L_{\max}$ / dB SPL	32	25	18	12	10	8	6
f_c / Hz	500	630	800	1000	1250	1600	2000
$L_{\max}$ / dB SPL	5	5	4	4	4	5	5
f_c / Hz	2500	3150	4000	5000	6300	8000	10000
$L_{\max}$ / dB SPL	3	1	-1	1	6	12	14

[1] These specifications are taken from the guidelines of the British Society of Audiology [BSA08].

3.2. Materials and Methods

3.2.1. Estimation of Reverberation Times

Reverberation time (RT_{60}) measurements were performed with the set-up in figure 3.1. The measured data was also used to characterize the influence of sound-absorbing curtains on room acoustics. The measuring equipment is listed in appendix B.

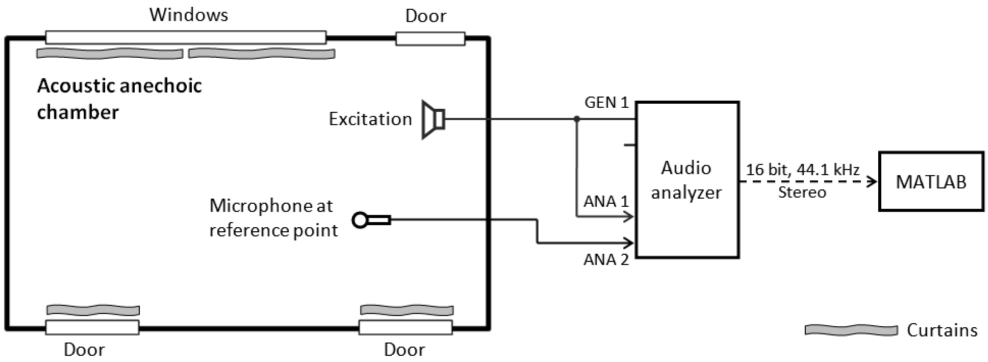


Figure 3.1.: *Measurement set-up for the determination of test room characteristics. The audio analyzer creates the excitation signal at the generator output (GEN1) and captures the signals at the analyzer inputs (ANA1/2). The data is exported as* `.wav`*-file and analyzed in MATLAB (The MathWorks, Inc., Natick, MA, USA).*

A very practical measurement procedure is discussed in Müller [MUE08]. It supports the calculation of the room impulse response (RIR) with a single measurement using non-periodic sweeps. Harmonic distortion products caused by non-linearities of the tested system are discarded automatically. Additionally, neither exact periodical repetitions of the excitation signal nor dynamic range recording limits have to be considered particularly.

Figure 3.2 shows the signal processing script implemented in `MATLAB` (see appendix E).

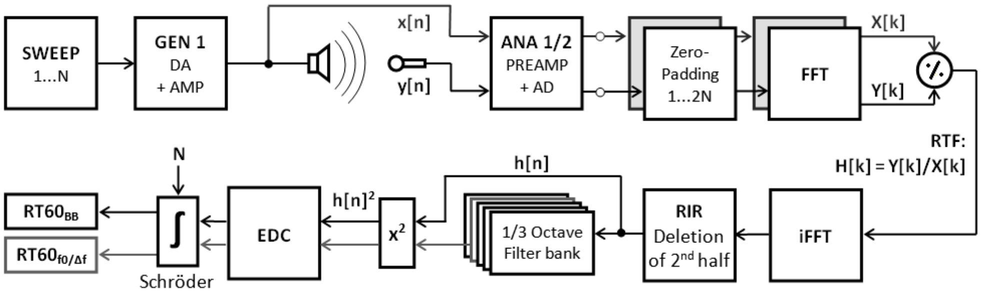


Figure 3.2.: *Determination of the room impulse response (RIR) and reverberation times (RT_{60}) with sweep using linear deconvolution and deletion of distortion products. Adapted from Müller [MUE08] and Karjalainen [KAR01].*

The excitation signal $x[n]$ is generated at the audio analyzer output (GEN1: 0.01 Hz - 20 kHz, 256 log points, dwell = 0.01 s, voltage = 0.4 s and duration = 8 s). The recorded microphone signal $y[n]$ and the excitation signal are recorded (ANA1/2) and exported as `.wav`-files.

The script zero-pads the signals to a block length of $2N$ and performs the FFT. The room transfer function (RTF) is computed by division of the signal spectra ($X[k]$ and $Y[k]$), which corresponds to a linear deconvolution in the time domain. An iFFT yields the RIR without distortion products. The theoretical explanation on the swept-sine technique has been covered extensively in Farina [FAR00].

Narrow-band reverberation times are obtained by filtering the impulse response with the corresponding 1/3-octave band. The filters are designed as third-order Butterworth bandpasses[2] with the center frequencies listed in table 3.1.

A smoothed and monotonic decay curve can be produced by Schroeder backward integration:

$$L[n] = 10 \cdot \log_{10} \left(\frac{\sum_{n}^{N} h^2[m]}{\sum_{0}^{N} h^2[m]} \right) \tag{3.1}$$

When choosing the wrong measurement interval $[0, N]$, included background noise may cause an inaccurate reproduction of the decay ramp[3]. The RT_{60} is obtained with line fitting of the smoothed energy decay curve (EDC) between -5 dB and -25 dB and following extrapolation to -60 dB.

3.2.2. Critical Distance

The critical distance d_c is the point where direct and reverberant sound energies are equal (Kutruff [KUT04]). Since the direct sound energy predominates within the critical distance, it should not be smaller than the distance of the subject to the speakers.

With the reverberation time (in s) and the volume of the room given, the critical distance can be calculated:

$$d_c \approx 0.057 \cdot \sqrt{\frac{V}{RT_{60}}} \qquad \text{with} \quad V = 54.1\,\mathrm{m}^3. \tag{3.2}$$

[2] A very useful 1/3-octave-band filter bank is developed in Couvreur [COU97].

[3] The so-called 'tail problem' of Schröder integration is adressed in Karjalainen [KAR01].

3.2.3. Sound Field Measurements

The sound field was measured with a free-field microphone using three different 1/3-octave noise stimuli (500/1000/3150 Hz). Measurements were taken at the reference point (L_{rp}) and at the positions mentioned in section 3.1.1.

The sound pressure level 0.1 m in front of (L_{f}) and behind (L_{b}) the reference point (L_{rp}) has to be corrected with the level difference predicted by the inverse distance law:

$$\Delta L_{+10} = 20 \cdot \log_{10}\left(\frac{1}{1+0.1}\right) \approx -0.8\,\mathrm{dB} \tag{3.3}$$

and

$$\Delta L_{-10} = 20 \cdot \log_{10}\left(\frac{1}{1-0.1}\right) \approx +0.9\,\mathrm{dB}. \tag{3.4}$$

The deviation of the levels at these points can be calculated with

$$\Delta L_{\mathrm{b}} = L_{\mathrm{rp}} - L_{\mathrm{b}} - \Delta L_{+10}. \tag{3.5}$$

and

$$\Delta L_{\mathrm{f}} = L_{\mathrm{rp}} - L_{\mathrm{f}} - \Delta L_{-10}. \tag{3.6}$$

3.2.4. Ambient Noise Level Measurements

Ambient noise levels were measured with the *UPV audio analyzer* and a free-field microphone (see appendix B). The microphone was calibrated using a pistonphone at a sound pressure level of 124.04 dB SPL (250 Hz). A 16k hanning-windowed FFT was averaged 500 times over a period of 15 min. The rms-levels for each 1/3-octave band were automatically computed via the frequency spectrum. Measurements were taken at 10 a.m. with switched on room lighting, air-conditioning and computer system.

3.3. Results

Figure 3.3 shows the 1/3-octave reverberation times measured in the test room.

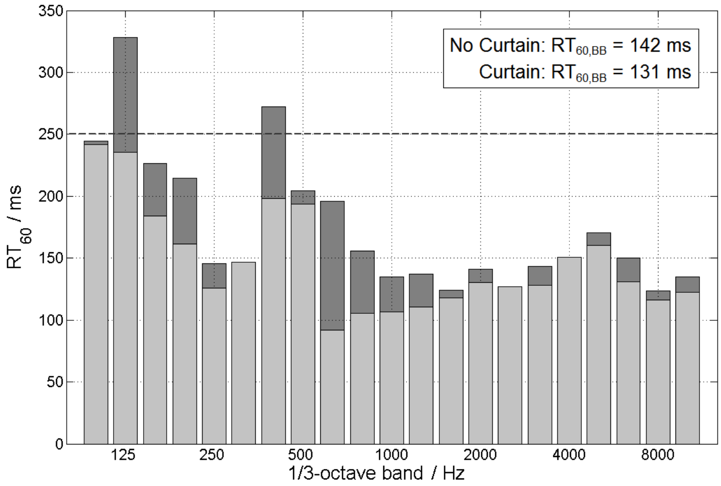


Figure 3.3.: *Narrow-band and broad-band ($RT_{60,\mathrm{BB}}$) reverberation times measured before (dark gray) and after (light gray) fixation of the sound absorbing curtains. The recommended limit (dashed line) is mentioned in [BSA08]. (Room C208.3 - ARTORG Center)*

The critical distance is $d_c = 1.11$ m in the room without sound absorbing curtains. It increases to $d_c = 1.16$ m when the curtains are used.

Ambient noise levels and the maximum permissible values when testing to a certain hearing threshold are illustrated in figure 3.4.

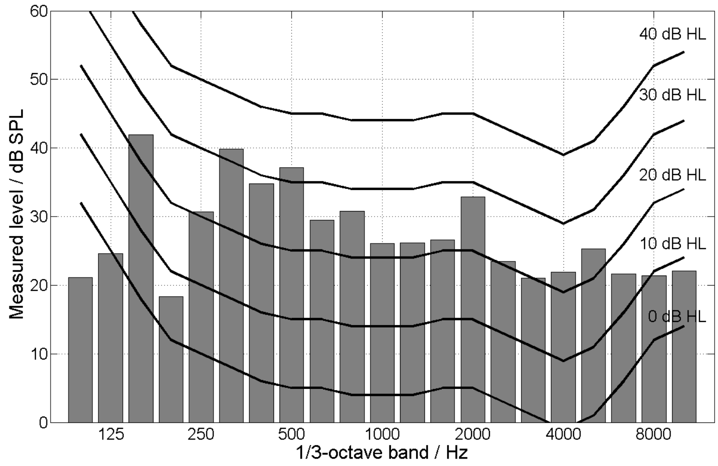


Figure 3.4.: *Ambient noise levels measured over 15 min and 500 averages and maximum permissible levels for hearing levels adapted from table 3.1. (Room C208.3 - ARTORG Center)*

Table 3.2 gives details of the sound field and the permissible deviations around the reference point.

Table 3.2.: *Measured sound pressure level differences around the reference point (L_{rp}). Center frequency (f_c); maximum permissible deviation (ΔL_p) [I8253a]; level difference 0.10 m in front of/behind the reference point ($\Delta L_{\mathrm{f}}/\Delta L_{\mathrm{b}}$) corrected by the inverse distance law values; level difference 0.15 m above/under/left of/right of the reference point ($\Delta L_{\mathrm{u}}/\Delta L_{\mathrm{d}}/\Delta L_{\mathrm{l}}/\Delta L_{\mathrm{r}}$) in dB SPL.*

f_c / Hz	L_{rp}	ΔL_{f}	ΔL_{b}	ΔL_{u}	ΔL_{d}	ΔL_{l}	ΔL_{r}
500	70.1	-0.2	0.0	-1.4	0.3	0.3	0.6
1000	70.2	-0.1	-0.5	-1.8	-0.4	0.9	0.5
3150	69.6	0.1	-0.6	-3.2	-0.5	-0.3	-0.1
ΔL_p		±1.0	±1.0	±2.0	±2.0	±2.0	±2.0

3.4. Discussion

In this chapter the test environment has been characterized based on measurements of room acoustics parameters. The reverberation times shown in figure 3.3 reveal the influence of sound absorbing curtains. The broad-band reverberation time $\mathrm{RT}_{60,\mathrm{BB}}$ is reduced from 142 ms to 131 ms, whereby both values are sufficient for sound field measurements.

Most narrow-band RT_{60} are decreased, especially at 125, 400 and 630 Hz center frequency. The limit of 250 ms given by the British Society of Audiology [BSA08] is exceeded at 125 and 400 Hz 1/3-octave bands, which makes the fixation of sound absorbing curtains recommendable.

The critical distances were calculated for both cases ($d_c = 1.11$ m without and $d_c = 1.16$ m with curtains) and are greater than the distance between the subject and the speakers (1.0 m). This ensures that the major component of broad-band sound energy at the reference point is the direct sound radiated from the sound source.

As shown in figure 3.4, audiometric measurements can be conducted down to 40 dB HL without being influenced through ambient noise masking effects. When testing hearing thresholds down to 30 dB HL masking through ambient noise may be taken into account, especially at frequencies lower than 1 kHz and around 2 kHz. The high ambient noise levels (40 dB SPL) at 160 and 315 Hz frequency bands most probably result from the air-conditioning system. Noise levels at these frequency bands were also measured with the light and the computer system switched off, without having any effects.

The test room and the set-up arrangement (see section 2.1.1) comply with the *quasi-free* sound field conditions mentioned in EN ISO 8253-2 [I8253a]. The results of the sound field measurements (table 3.2) confirm *quasi-free* conditions for 500 and 1000 Hz narrow-band stimuli. The boundaries of the room exert only a moderate effect on the sound waves.

Results are possibly inaccurate when presenting stimuli around 3150 Hz. Since the maximum deviation of the sound pressure level above the reference point (ΔL_u) is exceeded, the sound field is considered as *diffuse*. In a *diffuse* sound field directions of propagation at any point are randomly distributed which may influence spatial hearing cues (ILD, ITD). With the fixation of sound absorbing curtains this effect could be reduced, since the critical distance would be longer.

4. Acoustic Validation and Calibration

This chapter describes the concepts and methods used for calibration and validation of the system output. A calibration procedure using both hardware and software equalization is presented. Further, the applicable area of the set-up based on harmonic distortion and linearity measurements is discussed.

4.1. Introduction

In sound field audiometry, comparable and meaningful test results are only obtained when the system output of each speaker is calibrated. The calibration procedure has to be explained and documented to ensure repeatability for different experiments. Further, the validation of the set-up requires the determination of the system's linearity. This includes the monitoring of attenuation steps and total harmonic distortion estimations.

4.1.1. Sound Field Audiometry Specifications - EN ISO 8253-2/3

According to the specifications of EN ISO 8253-2 [I8253a] and EN ISO 8253-3 [I8253b], comparable audiometric data is obtained with a flat system transfer function between 125 Hz to 8 kHz.

It is recommended to assess and equalize the frequency response prior to the calibration. Besides the basic calibration procedure and subjective hearing evaluation tests, periodic electro-acoustic tests are suggested.

The attenuation steps of the volume control have to be checked for linearity. The signal level should be adjustable at least with a step size of 5 dB and the dynamic range shall at least cover 0 to 80 dB HL[1].

In addition, the total harmonic distortion (THD) should not exceed 3 % at 1 kHz when measured acoustically.

4.1.2. Oldenburger Sentence Test (OLSA) Specifications

Basically, OLSA tests require an audiometer with two channels and a variable step size of at least 1 dB. Loudspeakers should have a flat frequency response within 250 Hz to 6 kHz.

The system has to be calibrated with a speaker height of 1.0 - 1.2 m. The calibration signal is provided by the developers and has to be calibrated to 65 dB SPL[2].

[1] This is equal to a maximum of approximately 83.5 dB SPL in the defined frequency range.

[2] See the official OLSA handbook [OLS00] for further informations.

4.2. Materials and Methods

4.2.1. Calibration Procedure

The measurements were taken with the set-up in figure 4.1 and the equipment mentioned in appendix B.

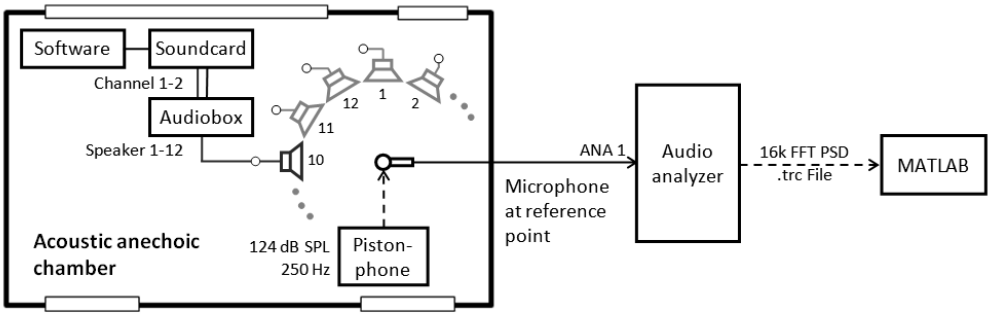


Figure 4.1.: *Calibration and validation measurement set-up using a free-field microphone.*

The calibration was performed with a pink noise signal[3]. A 16k FFT was taken with 48 kHz sampling rate and bandpass filtering (70 Hz - 12.5 kHz). The power spectral density was measured with the UPV audio analyzer and averaged 10 times over a period of 15 s. The free-field microphone was calibrated using a pistonphone.

The data was exported as `.trc`-file and further processed in a `MATLAB` script. Figure 4.2 shows a diagram of the calibration procedure performed for each channel-speaker combination.

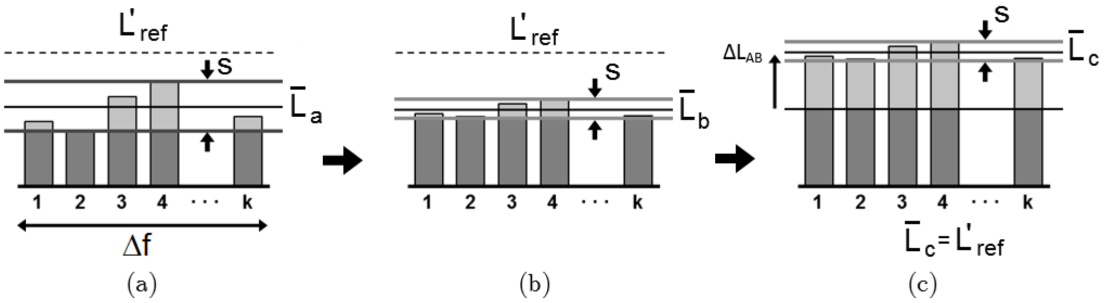


Figure 4.2.: *Calibration steps to obtain desired sound pressure level (L'_{ref}): (a) Spectrum of the uncalibrated system output with narrow-band mean level $\overline{L}_a$ and standard deviation s. The considered bandwidth (Δf) ranges from 70 Hz to 12.5 kHz. (b) Spectrum after software equalization, s should not exceed a threshold of ±1 dB. (c) Finally, the remaining constant level difference ΔL_{AB} is equalized by changing the `AUDIOBOX` calibration values.*

[3] The calibration signal was synthesized as described in [SMI08].

Although the hardware components were chosen for audiological application (Berger [BER10]), a flat frequency response of the whole system is not ensured. Therefore a multi-band software equalizer was implemented (see section 2.2.2).

It would be very time consuming to calibrate each signal path manually. Actually, 528 calibration values would have to be adjusted with the current measurement set-up[4]. For this reason a calibration script was implemented in `MATLAB` (see appendix E). First measurements showed that the `Fmod`-filters attenuate or amplify stronger than theoretically expected, regardless to the center frequency (figure 4.3).

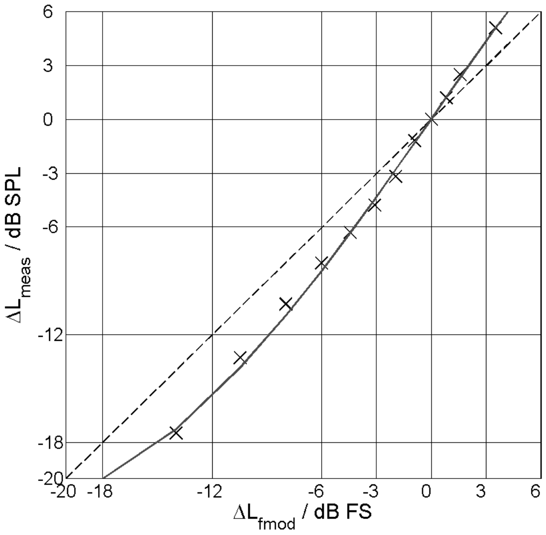


Figure 4.3.: *Level differences of a* `Fmod`*-filter (ΔL_{meas}) with f_c = 1 kHz and Δf = 1/3, data fit (solid line) and ideal curve (dashed line).*

Theoretically predicted level differences can be calculated with

$$\Delta L_{\text{fmod}} = 20 \cdot \log_{10}(g_{\text{fmod}}) \qquad \text{with} \quad 0.05 \leq g_{\text{fmod}} \leq 3. \tag{4.1}$$

The gain factors were corrected using a third-order polynomial fit to compensate this effect:

$$\Delta L_{\text{meas}} = -0.17 \cdot \Delta L_{\text{fmod}}^3 + 0.85 \cdot \Delta L_{\text{fmod}}^2 + 0.27 \cdot \Delta L_{\text{fmod}} + 0.05. \tag{4.2}$$

Further, the `Fmod`-filter does not behave like a filter with constant bandwidth, because filters with adjacent center frequencies are strongly influencing each other. When using constant-Q filters this effect should not occur [BOH86]. The algorithm compensates this influence by matching adjacent gain factors with an empirically found correction value ($c_g = 1.25$).

[4] 12 speakers, 2 channels and 22 parameters per combination ($12 \cdot 2 \cdot (21 + 1) = 528$).

When calibrating with pink noise the signal energy should be equally distributed over each octave. This leads to a lower calibration level for each frequency band. Incoherent sound sources are added with

$$L_\Sigma = 10 \cdot \log_{10}\left(\sum_{i=1}^{k} 10^{\frac{L_\mathrm{i}}{10}}\right) \text{(Sengpiel [SEN11])}. \tag{4.3}$$

The adapted calibration level L'_ref is found assuming that the signal energy is distributed over k incoherent sound sources. Within the investigated frequency range (Δf) about

$$k = 3 \cdot \frac{\log_{10}(12500/70)}{\log_{10} 2} \approx 22.4 \tag{4.4}$$

1/3-octave band sources can be found. The adapted calibration level is calculated with

$$L'_\mathrm{ref} = 10 \cdot \log_{10}\left(\frac{10^{\frac{L_\mathrm{ref}}{10}}}{k}\right) = L_\mathrm{ref} - 10 \cdot \log_{10}(k) \;\approx\; L_\mathrm{ref} - 13.5\mathrm{dB}. \tag{4.5}$$

4.2.2. Linearity

Linearity of the attenuation was measured with the calibration signal and 1/3-octave band noise (f_c = 1 kHz), both with 15 s duration, between 60 - 100 dB SPL (5 dB step width).

Total harmonic distortion (THD) was measured between 90 - 100 dB SPL in order to describe system linearity in the upper working area. The in-built THD-function of the the UPV audio analyzer was used. The test signal was a 1 kHz sine with 44.1 kHz sampling rate.

Additionally, a subjective hearing evaluation was carried out. Five different OLSA speech signals and noise stimuli were presented between 90 - 100 dB SPL.

4.3. Results

Figure 4.4 shows an example of a power spectrum after calibration.

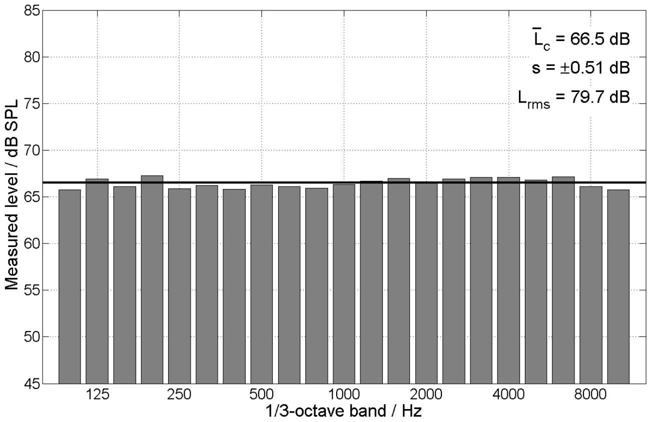


Figure 4.4.: *Spectrum of the signal path between channel 0 and speaker 4, after calibration to* $L'_{\text{ref}} = $ *66.5 dB SPL* ($L_{\text{ref}} = 80$ *dB).*

The calibration results for all channels and speakers are summarized in table 4.1. The calibration values of the `AUDIOBOX` (ΔL_{AB}) for a maximum level of 100 dB SPL are listed in table 4.2.

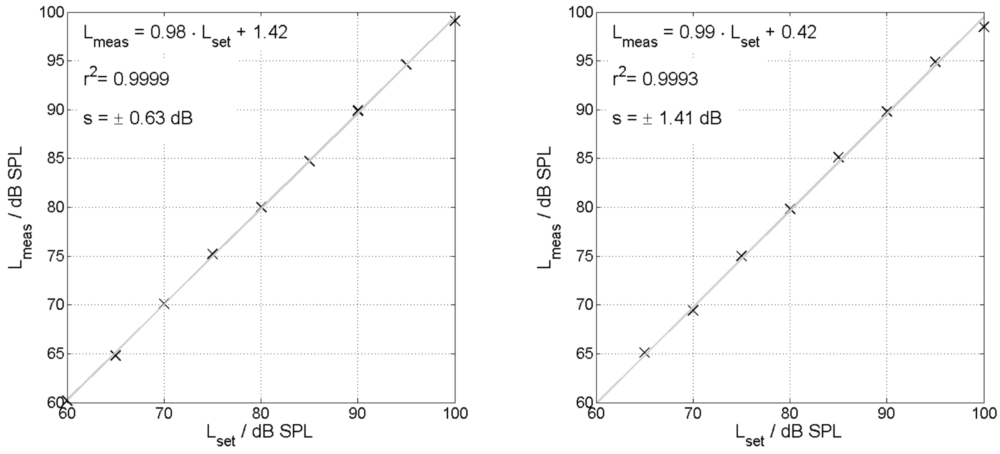
Table 4.1.: *Sound pressure levels at the reference point after calibration for each channel and speaker. Calibration signal presented with* $L'_{\text{ref}} = $ *66.5 dB* ($L_{\text{ref}} = $ *80 dB). Measured rms-value* (L_{rms})*; narrow-band mean level* ($\overline{L}_c$)*; sample standard deviation (s) in dB SPL.*

CH 0	L_{rms}	$\overline{L}_{\mathbf{c}}$	s	**CH 1**	L_{rms}	$\overline{L}_{\mathbf{c}}$	s
1	80.3	66.9	±0.77	**1**	80.4	67.1	±0.79
2	80.3	67.0	±0.99	**2**	79.9	66.6	±0.73
3	80.3	67.0	±0.83	**3**	80.3	67.0	±0.85
4	79.7	66.5	±0.51	**4**	79.8	66.6	±0.50
5	79.9	66.6	±0.54	**5**	80.0	66.7	±0.58
6	80.2	66.9	±0.81	**6**	80.2	66.1	±0.93
7	80.0	66.7	±0.68	**7**	80.3	67.0	±0.74
8	80.3	67.0	±0.73	**8**	80.3	67.0	±0.66
9	80.3	67.1	±0.78	**9**	80.4	67.1	±0.74
10	80.2	66.9	±0.86	**10**	80.2	66.9	±0.89
11	80.2	66.8	±0.89	**11**	80.2	66.9	±0.78
12	80.2	66.9	±0.76	**12**	80.2	66.9	±0.87

Table 4.2.: *AUDIOBOX calibration values (ΔL_{AB}) for 100 dB SPL maximum sound pressure level.*

CH 0	ΔL_{AB} / dB	**CH 1**	ΔL_{AB} / dB
1	0.0	**1**	0.0
2	0.5	**2**	0.0
3	0.5	**3**	0.5
4	0.0	**4**	-0.5
5	1.0	**5**	1.0
6	0.0	**6**	0.0
7	1.0	**7**	1.0
8	1.0	**8**	1.0
9	-0.5	**9**	-0.5
10	0.0	**10**	0.0
11	0.0	**11**	0.0
12	0.0	**12**	0.0

Figure 4.5 shows the recorded attenuation steps for the calibration signal and a 1/3-octave band noise stimulus.



(1) Pink noise

(2) 1/3-octave band noise, $f_c = 1$ kHz

Figure 4.5.: *Sound pressure level (L_{set}) and measured values (L_{meas}) for attenuator steps between 60 and 100 dB SPL. Shown are the linear data fit, the Pearson's correlation coefficient (r^2) and the sample standard deviation (s).*

The OLSA calibration signal was measured with 65.2 - 65.8 dB SPL after calibration with pink noise ($L_{\mathrm{ref}} = 65$ dB SPL).

THD is 3 % at 90 dB SPL and 5 % at 94 dB SPL. The subjective hearing evaluation revealed good signal quality up to 100 dB SPL for OLSA sentences.

4.4. Discussion

With the procedure described in this chapter, calibrated acoustic outputs are ensured for each channel and speaker combination. The narrow-band mean levels and sample standard deviations in table 4.1 verify a flat frequency response ($s < \pm 1$ dB) between 70 Hz and 12.5 kHz. The measured rms-values L_{rms} are not differing by more than ±0.5 dB from the reference level (L_{ref}). The OLSA noise levels measured after calibration are not differing by more than ±1 dB.

The current system settings provide acoustic sound stimuli with up to 100 dB SPL. The maximum level can be adjusted with the corresponding values in table 4.2 and the constant *MAXSPLEVEL* in the *Audiobox*-class. If less than 100 dB SPL maximum level are needed for tests, it can be decreased resulting in a lower noise floor produced by the `AUDIOBOX`.

The attenuation steps measured with the calibration signal and a narrow-band stimulus show linear behaviour (figure 4.5). The Pearson's correlation coefficients of the linear fits are almost 1 and the maximum bias is about 1.4 dB. Further, it can be seen that the narrow-band signal is presented with the right sound pressure level after calibration.

At 94 dB SPL the presented acoustic stimuli turned out to be distorted (5 % THD), which was audible in the signal quality. This suggests to use maximum 90 dB SPL acoustic stimuli for localization experiments. The hearing evaluation tests showed good signal quality up to 100 dB for OLSA sentences, which ensures enough head-room for OLSA tests.

As a consequence of the calibration procedure and the used calibration signal, it has to be ensured that the test signals contain normalized sample data. This should be considered when using self-synthesized audio signals or audio files from other sources.

The implemented calibration script (appendix E) is one way to equalize a frequency spectrum. The whole calibration procedure could be improved with the design of inverse filters based on measured spectra.
The filters could be realized as a custom algorithm class in the DSP chain or as an real-time effect implemented in `Fmod`. The integration of a calibration module in the existing software would provide an automatic calibration procedure. In combination with the UPV audio analyzer the power density spectra could be recorded and transferred via USB for the computation of the filter coefficients.

5. Gesture Recognition

This chapter explains the algorithm used for gesture recognition with the MICROSOFT KINECT sensor. The determination of the optimal sensor position and a position error correction method are described. Finally, the data obtained by the sensor is analysed and validated.

5.1. Introduction

Spatial hearing measurements require a method to indicate the perceived auditory event. Indications can be made verbally, via touch-screen or by measurement of head movements (Altmann [ALT09]).

The presented measurement set-up uses gesture recognition for the capturing of indicated directions. A `MICROSOFT KINECT` sensor was adapted to the system to take benefit of natural user interfaces (NUI) which provide intuitive test procedures. This could reduce the whole test procedure duration, since no complex indication methods are necessary and test steps are proceeded automatically.

The optimal position, the temporal tracking characteristics and the angular resolution of the sensor should be estimated and analyzed by appropriate methods.

5.2. Materials and Methods

5.2.1. Skeletal Tracking and Optimal Sensor Position

During the test procedure, the patient naturally points his arm towards the direction of the perceived auditory event. Gesture recognition is performed by a skeleton tracking method which is part of the `Kinect SDK` [KIP11].

The sensor automatically detects the position of the subject's wrists, hands, hips and so on (*skeletal joints*). The detection is based on a IR depth image and RGB image, captured with approximately 30 frames per second (figure 5.1).

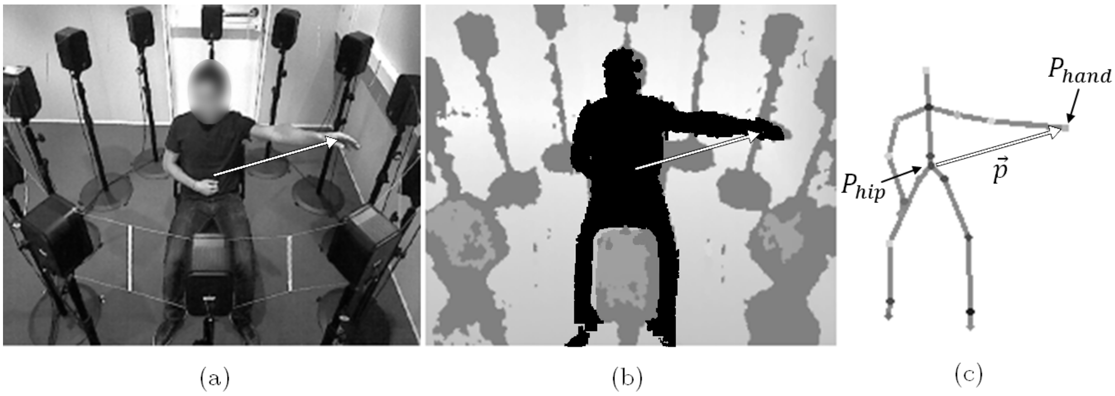


Figure 5.1.: *Image data provided by the `Kinect` sensor and the `Kinect SDK`. (a) RGB image; (b) Depth image with the detected person (black area); (c) Tracked skeletal joints. The indication vector $\vec{p}$ is found by subtraction of the hand joint (P_{hand}) and the hip center joint (P_{hip}). 'SkeletalViewer'-software [KIP11]*

Figure 5.2 shows a schematic representation of the set-up. The obtained coordinates (x_s, y_s, z_s) are expressed in meters and are oriented in the so-called *skeleton-space*, a right-handed coordinate system with the positive z_s axis extending in the sensor direction[1]. Since the sensor is tilted with α, the *skeleton-space* coordinates have to be transformed with a rotation matrix:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\sin(\alpha) & -\cos(\alpha) \\ 0 & \cos(\alpha) & -\sin(\alpha) \end{pmatrix} \cdot \begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} \tag{5.1}$$

The optimal skeletal tracking distance range is specified between 1.2 m and 3.5 m [KIP11]. The indicated direction of the subject is computed via vector subtraction of two joint positions. The position of the *hand joint* and the *hip center joint* (figure 5.1c and 5.2). The algorithm starts the detection if one hand is higher positioned than the *spine joint*. A special case occurs when indicating the rear speaker ($\varphi = 180°$), since the pointing arm is obscured by the head and torso of the subject. In this case the algorithm uses the corresponding *shoulder joint* position as reference point. Therefore subjects are instructed to point over their shoulder when indicating the backward direction[2].
After the indicated direction is detected, the software plays a short beep-signal as confirmation.

The optimal position of the `Kinect` sensor has to be a trade-off between various factors. Better tracking results and higher resolution can be achieved through a sensor placed near the subject, while a higher distance to the set-up leads to a bigger field of view. All speakers should be covered as far as possible. Additionally, the frontal speaker array and the subject itself decrease the vision through shadowing effects.
These factors basically depend on two variables, the horizontal distance to the reference point (d) and the inclination of the sensor (α). Within the given limits of the test room, a numerical simulation of the field of view, approximated by a defined volume of interest (VOI), as a function of d and α has been computed. The parameters used for the simulation are shown in figure 5.2.

[1] The skeleton space is defined in the official programming guide [KIP11].
[2] A more detailed description of the algorithm can be found in Salzmann [SAL11].

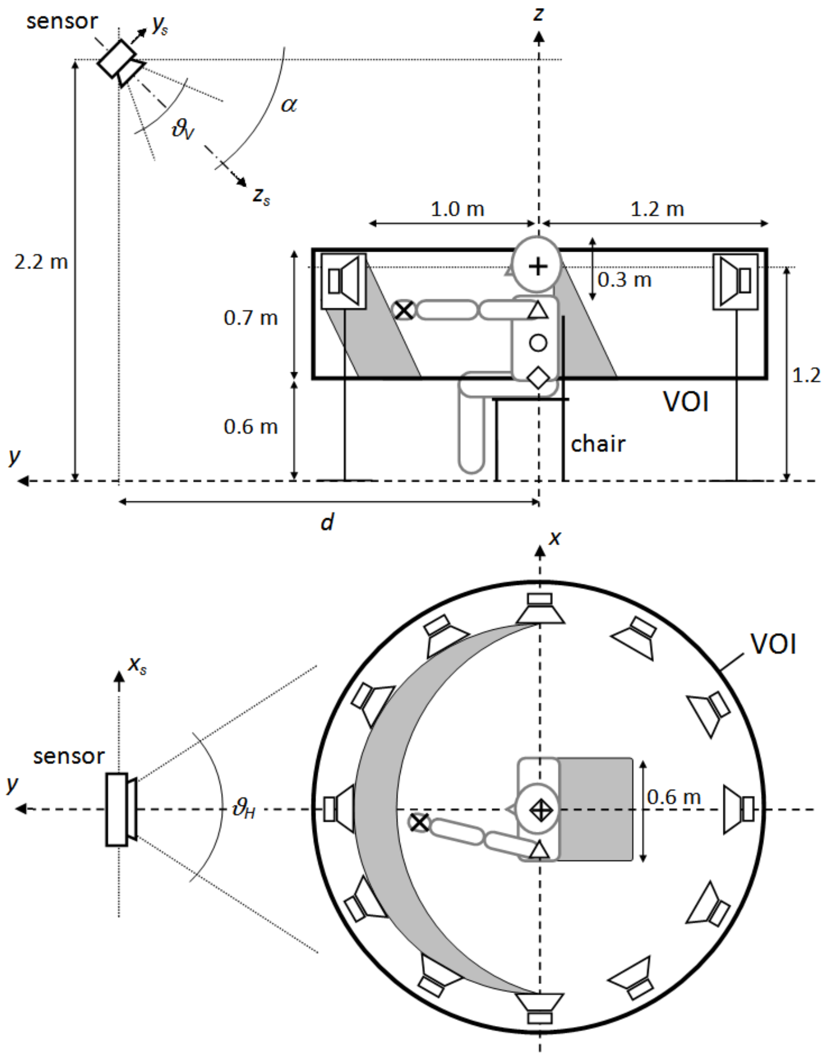


Figure 5.2.: *Subject's position and the reference point (+) in a schematic representation of the measurement set-up (top: side view, bottom: top view). The loudspeakers and the joint positions needed for gesture recognition lie within the volume of interest (VOI), which should be covered by the sensor as far as possible. In addition, shadowing effects (gray shaded areas) caused by the frontal speakers and the head/torso are taken into account. The aperture angles of the sensor are $\vartheta_H = 57°$ and $\vartheta_V = 43°$, as mentioned in the official* `Kinect` *programming guide [KIP11].*
If one hand ($\times$) is positioned higher than the spine joint ($\circ$), the algorithm computes the indicated direction. Reference points are the hip center position ($\diamond$) and the ipsilateral shoulder position ($\triangle$). Note that the room coordinate system (x, y, z) is different to the coordinate system used by the sensor (x_s, y_s, z_s).

5.2.2. Position Error of the Subject

During the test, the angle detection is highly dependent on the subject's center position. Since measurements are done without mechanical fixations, movements cause a translation of the hip and the hand joint position vectors.

Figure 5.3 shows a representation of the problem. The true direction $\vec{r}$ can be described with the error vector $\vec{e}$ and the indication vector $\vec{p}$:

$$\vec{r} = \vec{e} + \vec{p}. \tag{5.2}$$

The vector $\vec{p}$ is written in parametric representation

$$\vec{p} = \lambda \cdot \vec{u} \qquad \text{with} \quad \vec{u} = \frac{\vec{p}}{\|\vec{p}\|}, \tag{5.3}$$

which assures that increasing values for λ are extending in the positive direction ($\vec{u}$) of the vector. The condition $\|\vec{r}\|^2 = 1$, given by the set-up geometry, describes a circle equation. After intersection with $\vec{p}$ and solving of the quadratic equation, the greater value for λ is:

$$\lambda_{max} = -\langle \vec{u}, \vec{e} \rangle + \sqrt{\langle \vec{u}, \vec{e} \rangle^2 - \|\vec{e}\|^2 + 1}. \tag{5.4}$$

Finally, φ can be calculated with

$$\varphi = \arctan\left(\frac{r_x}{r_y}\right). \tag{5.5}$$

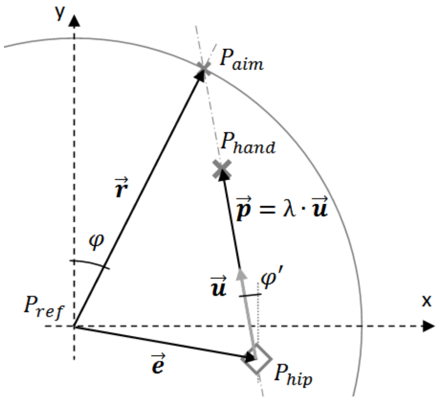


Figure 5.3.: *Sketch of a situation with a wrong center position of the subject. The error vector $\vec{e}$ depicts the deviation of the actual center position (P_{ref}), which results in an incorrect angle estimation ($\varphi \neq \varphi'$). The intersection of the indication vector $\vec{p}$ and the circle equation given by $\|\vec{r}\|^2 = 1$ yields the target point P_{aim}.*

5.2.3. Temporal Tracking Characteristics

First measurements of the tracked data showed a typical response. Basically, the automatic detection is effected through three parameters (figure 5.4): The delay time (t_d), the integration time (t_i) and the deviation threshold (ε). The algorithm waits for the patient to react (t_d) and computes the mean angle during the integration time with the time derivative not exceeding ε.

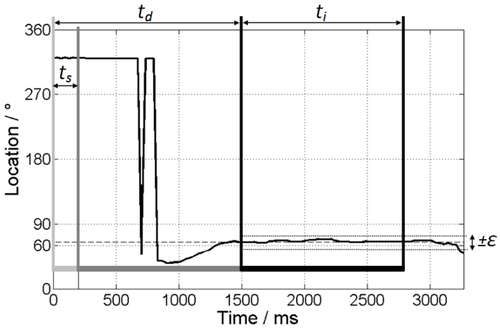


Figure 5.4.: *Example of a subject's response to an acoustic stimulus (t_s). After a certain delay time (t_d), the curve stabilizes at the indicated direction (dashed line). During integration time (t_i) the mean value of the tracked data is computed ($n = 30$). The time derivative must not exceed the defined deviation threshold (ε).*

The algorithm was analyzed using the responses of 10 subjects after an acoustic stimulus ($t_s = 200$ ms). The parameters were set to $t_d = 500$ ms, $t_i = 1000$ ms and $\varepsilon = \pm 1°$. With this setting about 30 values are used for the detection.

5.2.4. Angular Resolution

The angular resolution was determined in a validation test with 10 subjects of our research group (2 female/8 male) in the age between 25 and 39 years. The body height (158 - 188 cm), the height of the shoulders when sitting (97 - 108 cm) and the arm length (46 - 55 cm) were noted.

The subjects were instructed to directly point at the speakers which were marked with an indicator and label. An acoustic broad-band impulse was randomly presented at 24 positions ($\Delta\phi = 15°$). The procedure was repeated 3 times for each subject which leads to a sample size of 30 values per direction. Since the data should not be influenced by auditory effects, test steps with wrongly indicated directions were notified and repeated.

Additionally, the three-dimensional stability of the detection algorithm was measured at 12 directions ($\Delta\phi = 30°$). The lower detection limit z_L (given by the height of the *spine joint*) and the upper detection limit z_H (deviation error smaller than 1 %) in the z-direction were noted.

The sensor and the speakers were aligned with a laser pointer mounted to a pivoting angle gauge. The accuracy of this alignment method is approximately ±1°. Measurements for each speaker were plotted in a *QQ*-plot and the *Shapiro-Francia* test was performed to check whether the measured data was normally distributed[3].

5.3. Results

5.3.1. Sensor Position

Figure 5.5 shows the percentage of the covered VOI as a function of the distance d and the inclination α. The sensor was mounted at $d = 1.65$ m with an inclination of $\alpha = 41°$.

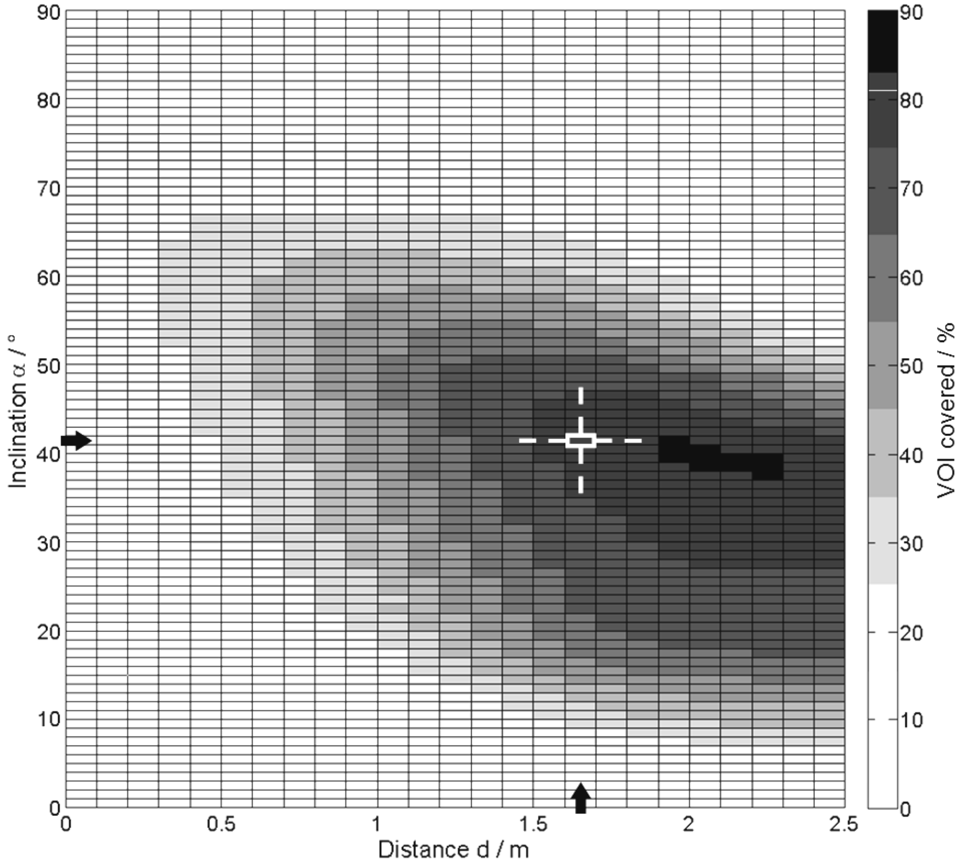
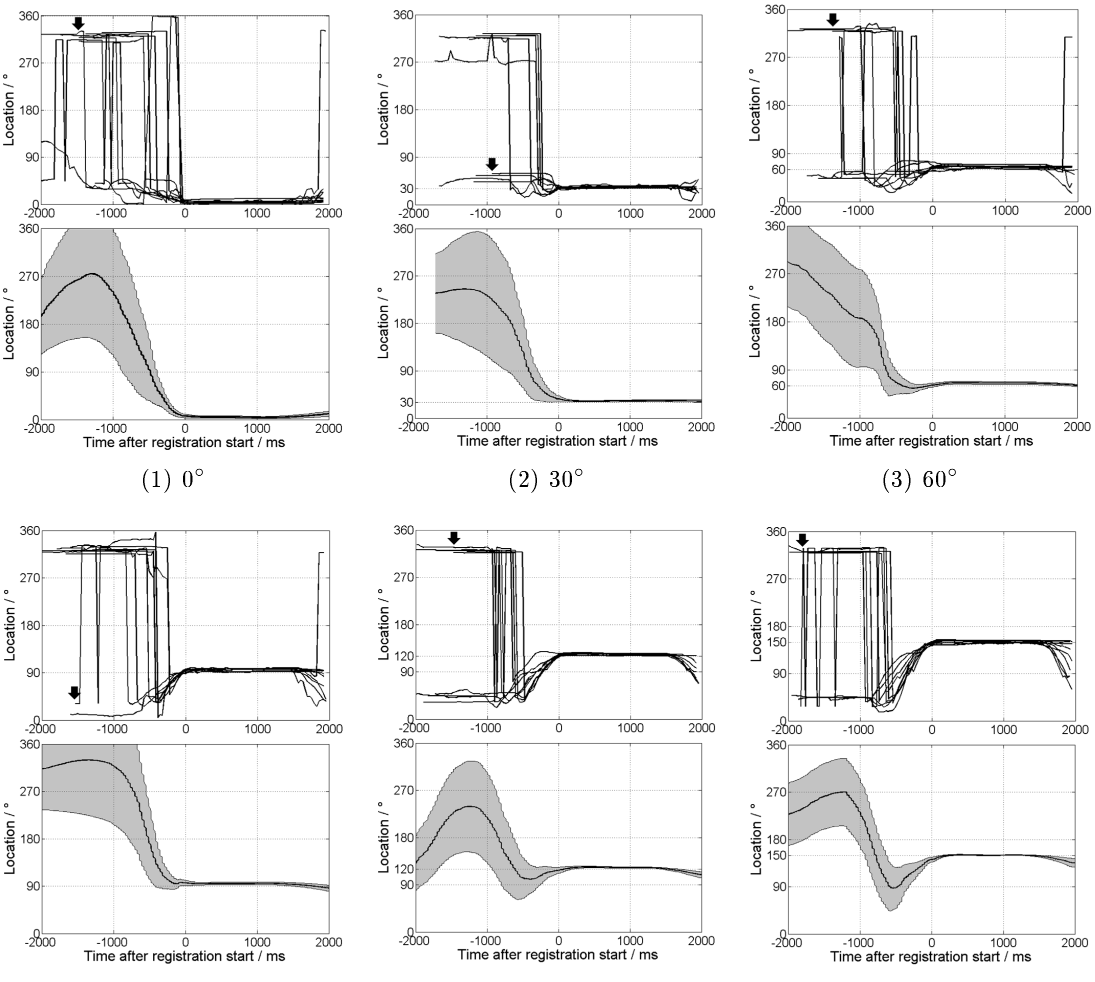


Figure 5.5.: *Numerical simulation of the covered volume as a function of distance d and sensor inclination α. At the actually chosen position (arrows) theoretically around 80 % of the volume are covered by the sensor.*

[3] The tests for normality were carried out as described in Hüsler [HUE06].

5.3.2. Temporal Tracking Characteristics

Figure 5.6 shows the measured tracking responses ($n = 10$) and the averaged curve for each speaker.



(1) 0°

(2) 30°

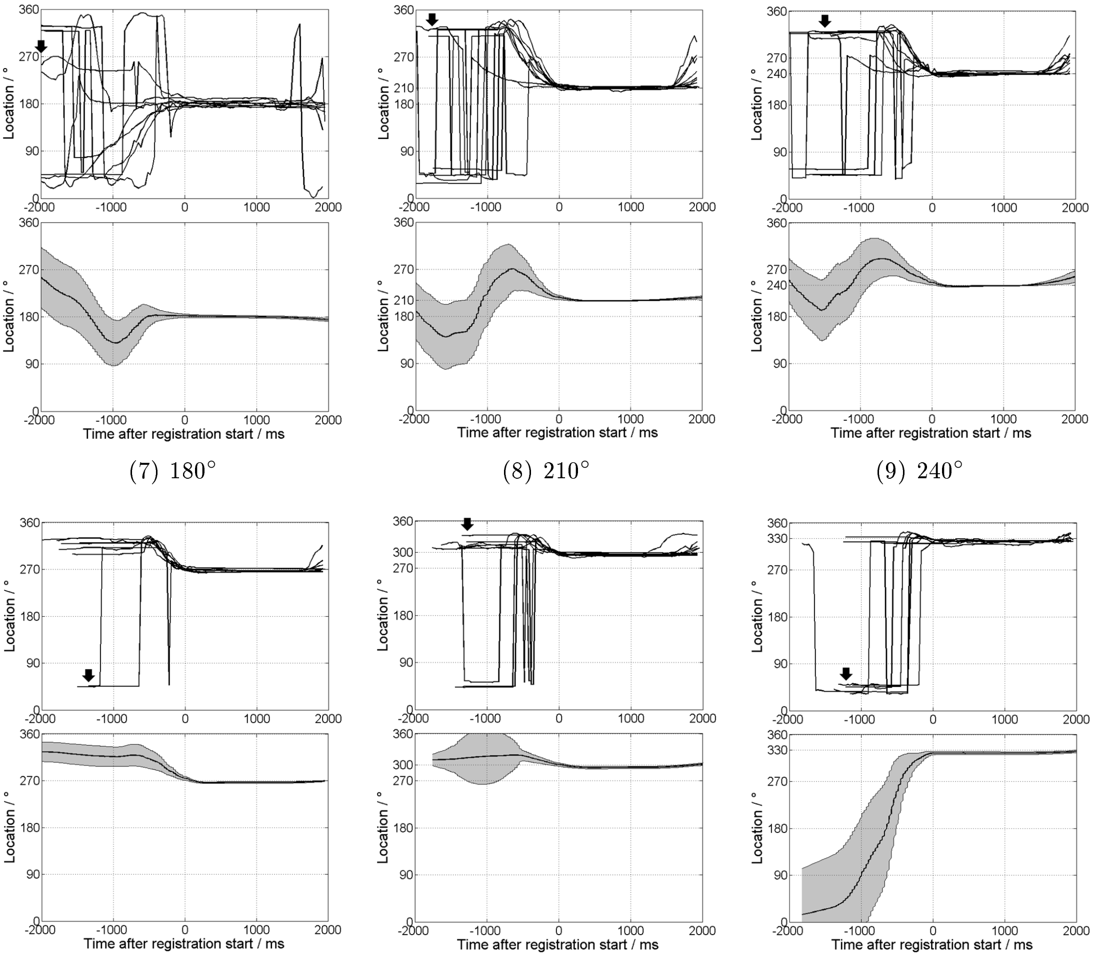
(3) 60°

(4) 90°

(5) 120°

(6) 150°

Figure 5.6.: *Top: Tracked data ($n = 10$) for each speaker ($\Delta\phi = 30$ °), synchronized to the start of registration ($t = 0$ ms). The shortest delay times after the tracking start (t_d) are marked with black arrows and can be found between -900 ms and -2000 ms. The responses were measured with $t_i = 1000$ ms and $\varepsilon = \pm 1°$. Bottom: The averaged response (thick line) and sample standard deviation (gray shaded), smoothed with a moving average filter.*



(7) 180°

(8) 210°

(9) 240°

(10) 270°

(11) 300°

(12) 330°

Figure 5.6.: *(continued)*

5.3.3. Angular Resolution

The statistical parameters of the investigated directions are listed in table 5.1. The parameters were used to illustrate the angular resolution of the measurement set-up (figure 5.7).

Table 5.1.: *Statistical parameters and angular resolution in the horizontal plane assuming normally distributed data from 10 subjects ($n = 30$). Direction (ϕ); mean error (μ); standard deviation ($\pm\sigma$); 95 % confidence interval ($\pm 2\sigma$). Lower detection limit (z_{L}) and upper detection limit (z_{H}) in the z-direction with respect to the horizontal plane ($z = 120$ cm).*

Speaker	ϕ	μ / °	σ / °	2σ / °	z_{L} / cm	z_{H} / cm
(1)	0°	2.3	±4.6	±9.1	-5	+60
	15°	2.7	±2.8	±5.5		
(2)	30°	2.4	±2.6	±5.3	-50	+60
	45°	2.5	±2.8	±5.7		
(3)	60°	5.3	±3.0	±6.1	-50	+40
	75°	2.4	±2.4	±4.7		
(4)	90°	3.6	±2.3	±4.6	-50	>+60
	105°	0.2	±2.2	±4.3		
(5)	120°	1.8	±1.8	±3.7	-50	>+60
	135°	0.0	±1.5	±3.0		
(6)	150°	-0.1	±2.0	±4.0	-50	+60
	165°	-0.6	±3.5	±7.0		
(7)	180°	1.7	±6.2	±12.4	-5	+40
	195°	-0.7	±2.7	±5.5		
(8)	210°	-1.0	±1.8	±3.7	-50	+60
	225°	-1.3	±2.6	±5.3		
(9)	240°	-2.4	±2.4	±4.7	-50	>+60
	255°	-2.0	±1.8	±3.7		
(10)	270°	-4.3	±2.7	±5.4	-50	>+60
	285°	-4.6	±2.9	±5.8		
(11)	300°	-5.3	±2.2	±4.4	-50	+40
	315°	-4.4	±2.4	±4.7		
(12)	330°	-5.7	±3.7	±7.4	-50	+60
	345°	-3.7	±3.3	±6.7		

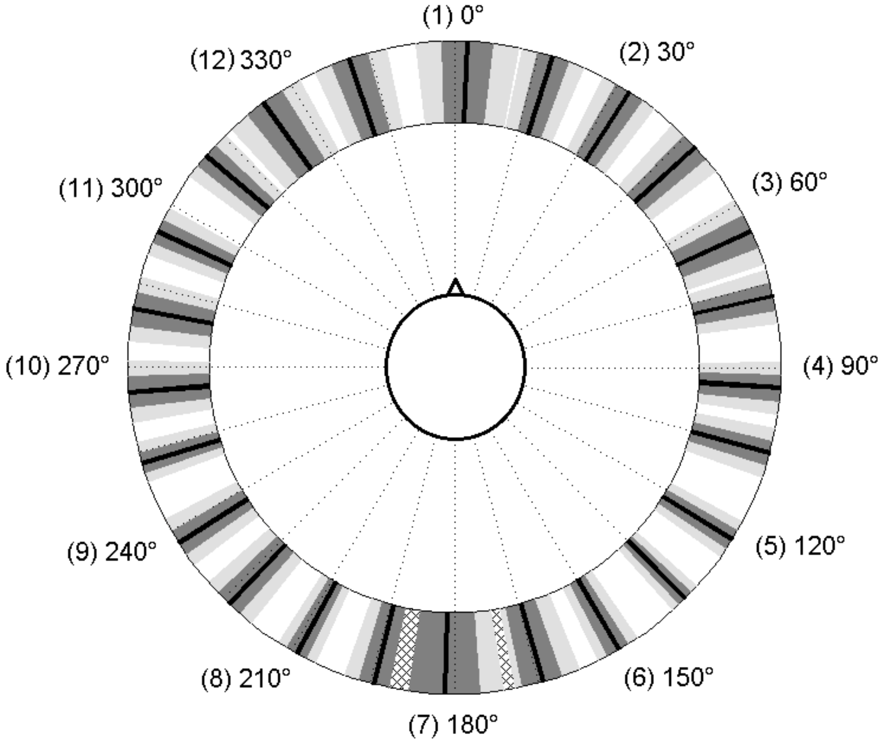


Figure 5.7.: *Mean errors μ (thick lines), standard deviations σ (dark gray shaded) and 95% confidence intervals (light gray shaded) for 15° spacing in the horizontal plane. Overlapping areas are crosshatched. Calculated from test data of 10 subjects ($n = 30$).*

5.4. Discussion

In this chapter, the optimal sensor position and tracking parameters for the given speaker arrangement and subject position have been estimated. Further, the tracked sensor-data has been analyzed and validated for applicability in spatial hearing measurements.

The chosen sensor position is a compromise between the given room and set-up geometry and the tracking quality. First, the theoretically optimal position based on field of view and shadowing effects was simulated. Second, the tracking quality was empirically evaluated at different sensor positions. The closer the sensor was positioned to the reference point, the better tracking results were observed. The nearest possible position with the inner speaker radius covered was found at a distance of 1.65 m to the reference point with an inclination of 41°. The sensor position is marked in figure 5.5 and lies nearby the maximum area. Tracking within the specified distance range is provided. About 80% of the VOI are covered at the chosen position, higher coverages can be only achieved through greater sensor distances. The remaining 20 % lie outside of the field of view or are obscured. At the current position speakers at 60° (3) and 330° (11) are not completely covered. This should be considered when using image data for further processing. Many advantages arise from the implementation of the automatic position error correction. No fixation of the subject is needed during the experiment. Measurements can be started without complicated position calibration procedures and sitting posture instructions. The subject can naturally react and fully concentrate on the given tasks, which additionally improves the detection quality.

Synchronisation of the tracking characteristics to the registration start reveals a similar temporal behaviour for all investigated directions (figure 5.6). During the delay time (t_d) the tracked direction mainly jumps between the resting positions of the hands (around $\varphi = \pm 45°$). It can be clearly seen that the curve has to stabilize ($\varepsilon = \pm 1°$) before the algorithm starts the detection ($t = 0$ ms). The fastest responses are expected at the directions closest to the hand resting positions. The measured delay times range between 900 ms and 2000 ms. Measurements at 180° take longer due to the different indication method. The delay time for the algorithm was set to $t_d = 1.5$ s. This ensures that a large part of data without evaluable information is discarded. The qualitative progression of the averaged curves in figure 5.6 underlines the stability of the measured data. It can be seen that the deviation decreases to a minimum during integration time t_i (between $t =$ 0 ms and $t = 1000$ ms).

The statistical parameters in table 5.1 were obtained under assumption of normally distributed data. The *Shapiro-Francia* test and the QQ-plots determined normal distribution for all directions except 30° (2), 150° (6) and 330° (12). The distributions of (2) and (12) had positive skewness values (2.2 and 0.9), whereas data at (6) showed a negative skew (-1.0). Since the standard deviations at the mentioned positions are small compared to other directions, parameters of an ideal normal distribution are used to analyze the detection accuracy at all directions.

A first look at figure 5.7 shows that the accuracy of the detection is better for backward directions with the exception at 180°, where the values are scattered through the different indication method. A comparison with the sensor field of view in figure 5.1a indicates an influence of perspective factors. This could be a residual perspective distortion caused by the inclination and the subsequent correction of the values obtained by the skeletal tracking algorithm. Errors caused by a wrongly configured value of the sensor inclination are unlikely, since this would lead to distortions in the whole horizontal plane. A further analysis of the skeletal tracking algorithm provided by the `Kinect SDK` could give some additional clues. The perspective distortion could be corrected by factors determined through field of view measurements or minimized by using a second `Kinect` sensor.

The directions 0° and 180° have the largest standard deviations. At 0° the tracking algorithm is not able to detect the skeletal joints with the same precision as for other directions, which results in higher detection errors. The large deviation at 180° results from pointing backwards without visual informations for the subject. The mean errors in the forward and backward direction indicate a slightly twisted fixation of the sensor. The mean error at 300° lies outside of the 95% confidence interval. This characterizes a systematic error, most probably caused by a wrong position of the speaker (11).

The angular resolution is evaluated with the standard deviations (σ) and the 95 % confidence intervals (2σ) listed in table 5.1. Since the 95% confidence intervals do not overlap or exceed ±15° at all, experiments using the discrete speaker positions only ($\Delta\phi = 30°$) can be conducted without restrictions. Measurements with a spacing of $\Delta\phi = 15°$ can be performed considering the overlapping regions at 180° (figure 5.7).

The validation of the limits in the z-direction showed good detection results between -50 cm and +40 cm with respect to the horizontal plane (z = 120 cm) for all directions except 0° and 180°, where shadowing effects limit the lower detection level to -5 cm. Within these ranges, it can be assumed that usable detection data is obtained.

6. Practicability Test

This chapter describes the test conducted to evaluate the practicability of the measurement system and reviews existing studies dealing with human sound localization in the horizontal plane.

6.1. Introduction

After the implementation and validation of the measurement set-up, a practicability test should be conducted to check the applicability of the system. The aim of this test is to investigate whether visual cues given by the speakers have an influence on sound localisation test results. Further, the total duration of the test procedure should be monitored.

6.2. Materials and Methods

6.2.1. Studies on Human Sound Localization

During the preparation phase of the thesis, studies dealing with human sound localization in the horizontal plane were reviewed with a focus on the used test stimuli and data analysis (table 6.1). The used acoustic stimuli are listed in column *stimulus*. Mainly, noise stimuli were presented to avoid standing waves which cause unpredictable sound pressure fluctuations within the room. Narrow-band stimuli were used to separately investigate ITD/ILD processing of the auditory system. Most signals had Gaussian shape or were filtered to 1/3-octave bandwidth. Further, broad-band noise or speech stimuli were presented to simulate more realistic situations.
In most cases, the signals were not longer than 300 ms to prevent head movements. Stimuli were mainly presented around 65 dB SPL and had on- and offset ramps.

The column *set-up* describes the number of used speakers and their arrangement around the subject. Besides set-ups with fixed speaker positions, movable speaker arrays with up to 58 speakers were built up. In many cases, experiments were carried out in completely dark rooms to avoid visual cues.

Roving is a method to minimize the influence of cues given by different frequency responses of the speakers by slightly changing stimuli levels for different test steps.

Most studies were conducted without a *fixed* head. Alternatively, head or chin rests were adjusted to maintain the subject in a stable and upright position during the tasks.

In some studies, *subjects* were assessed before testing to ensure normal bilateral audiometric thresholds (within 20 dB HL).

A frequently used method for *data analysis* is linear fitting in the least mean-square error sense. The stimulus-response relationship is described with the parameters response bias (b), response gain (g) and Pearson's correlation coefficient (r^2). Further, the mean absolute error (MAE) and the number of correct answers (CA) were evaluated for each direction.

Table 6.1.: *Selection of reviewed sound localization studies. Narrow-band noise (NBN); broad-band noise (BBN); range in horizontal plane (ϕ_{tot}); speaker spacing ($\Delta\phi$); number of speakers (n); normal hearing (NH); hearing impaired (HI); unilateral (u); least-square data fitting (LSF); mean absolute error (MAE); minimal audible angle (MAA); number of correct answers (CA)*

Author	**Stimulus**			**Set-up**	**Roving**	**Fixed**	**Subjects**	**Data analysis**	
	Type	Length	Level	n/ϕ_{tot}/$\Delta\phi$					
Agterman	1/3-NBN: 0.5/3 kHz	1 s and 150 ms, 5 ms ramps	40 - 70 dB SPL	7/180°/30°	-	-	12: uHI	LSF, MAE (°)	[AGT11]
Kumpik	White-BBN, rand. filtered: 0.2-20 kHz	300 ms, 5 ms ramps	50 dB A	12/360°/30°	yes	Chin	20: NH	LSF, CA (%)	[KUM10]
V. Grootel	White-BBN: 0.2 - 20 kHz	150 ms	60 dB A	58/360°/<0.1°	-	Neck	6: NH	LSF, MAE (°)	[VGR10]
Altmann	NBN: 0.25 - 4 kHz	250 ms	78 dB A	Earphones	-	-	18: NH	EEG-analysis	[ALT09]
Populin	BBN: 0.1 - 25 kHz	150 ms, 10 ms ramps	50 - 53 dB SPL	32/360/var.	0-3 dB	-	9: NH	MAE (°)	[POP08]
Buschermöhle	1/3-NBN: 0.3/3.4 kHz	-	-	2/60°/60°	-	-	41: NH, 62: HI	MAA (°)	[BUS08]
Eneman	1/3-NBN: 0.5/3.15 kHz, BB: telephone	1s and 200 ms, 50 ms ramps	65 dB SPL	13/180°/15°	-4-0 dB	-	5: NH	LSF, MAE (°)	[ENE06]
V. Wanrooij	BBN: 1 - 20 kHz, NBN: 3 kHz	150 ms, 0.5 ms ramps	30 - 60 dB A	58/180°/var.	-	-	5: NH	LSF, MAE (°)	[VWA06]
V. d. Bogaert	1/3-NBN: 0.5/3.15 kHz, BB: telephone	1s and 200 ms, 50 ms ramps	65 dB SPL	13/180°/15°	-4-0 dB	-	10: NH, 10: HI	MAE (°)	[VBO06]
Hol	1/3-NBN: 0.5/3 kHz	1 s	65 dB SPL	9/180°/30°	-	-	29: uHI	CA (%)	[HOL05]
György	BBN: 20 Hz - 22 kHz, NBN: 1.5/7 kHz	300 ms	-	Earphones	-	-	50: NH	MAA (°)	[GYO03]

6.2.2. Test Design

The influence of the set-up appearance was investigated using two test cases. In the *Blind* test case, subjects were blindfolded, in the *See* test case they were allowed to use visual cues. The chair in the middle of the speaker array was rotated by 7.5 degrees, leading to a shift of directions.

The hypothesis proposes that the rotated chair will cause a uniform difference of the mean error values between the two test cases. A power analysis was conducted to calculate the required number of subjects for a significance level of 5%. For a verifiable shift of 7.5° with a mean standard deviation of ±4° (table 5.1), five subjects are necessary.

Subjects

Five healthy volunteers (4 female/1 male), aged 23-54, who had normal hearing abilities as determined by self-report, participated in the experiment. Two subjects had low experience with the measurement set-up. All subjects were naive about the purpose of the test. During the experiment, the subjects were not fixated and instructed to make themselves comfortable in a upright sitting position.

Experimental set-up

The experiment was conducted in the test environment described in the previous chapters. The measurement set-up was used with $\Delta\phi = 30°$ spacing between the speakers and with the specifications given in appendix A. The average background noise level was 38.5 dB SPL.

Stimuli

The test was carried out with broad-band noise stimuli, since frequency-dependent differences in binaural processing were not of interest. The stimuli consisted of 0.1-20 kHz white noise with 44.1 kHz sampling rate.

Since the subjects were not fixated, the test signal duration was limited to 200 ms (10 ms cosine ramps). The stimuli were presented with 65 dB SPL and were randomly changed in 1 dB steps in the range of -3 to 0 dB for different test steps (*roving*).

Test procedure

First, subjects were asked to sit down and put on the blindfold. They were introduced to the pointing method used with the set-up. During a short training session with six test steps, the subjects were instructed to indicate the directions of the auditory events with their outstretched arm or by pointing over their shoulder. Although the subjects could not see anything, the introduction phase did in no case take longer than 5 min.

During the tests, the subjects did not get any feedback about their performance. In both test cases, stimuli were presented 3 times for each direction ($\Delta\phi = 30°$), leading to 36 stimuli in total. The subjects were informed that stimuli could come from any direction, including from virtual sound sources. The elapsed time for a whole test case was noted (T_{tot}) for each subject.

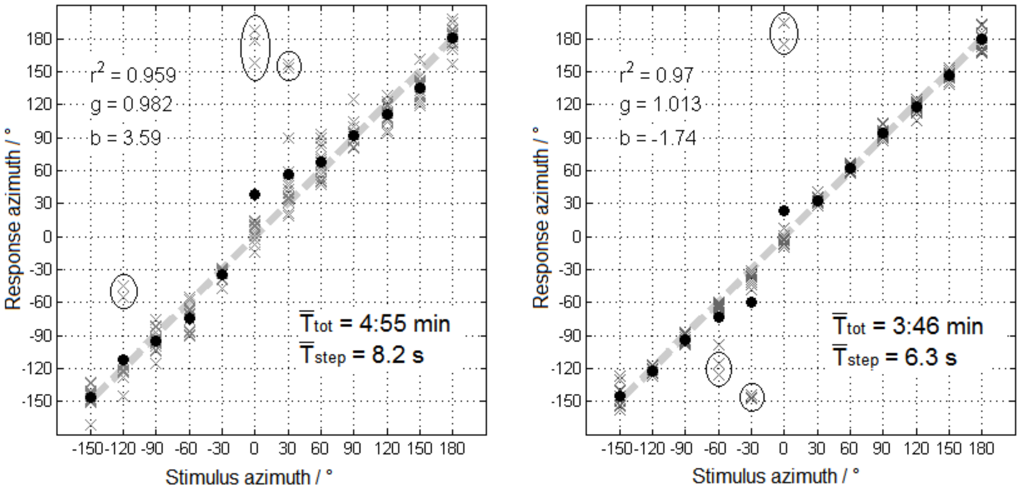
Data analysis

Data analysis was performed with a linear regression fit. For a better clarity, the data range of the results was wrapped from 0°/330° to $-150°/180°$. Since the stimuli are very short, front-back confusions can occur (*'cone of confusion'*). To avoid effects on the results, data was corrected from errors larger than ±30°.

The datasets of both test cases were compared using a two-sided t-test for paired samples.

6.3. Results

The responses of the practicability test are illustrated in figure 6.1. Additionally, errors caused by front-back confusions and outliers were discarded for reasons of data analysis. The corrected results are shown in table 6.2 and in figure 6.2.

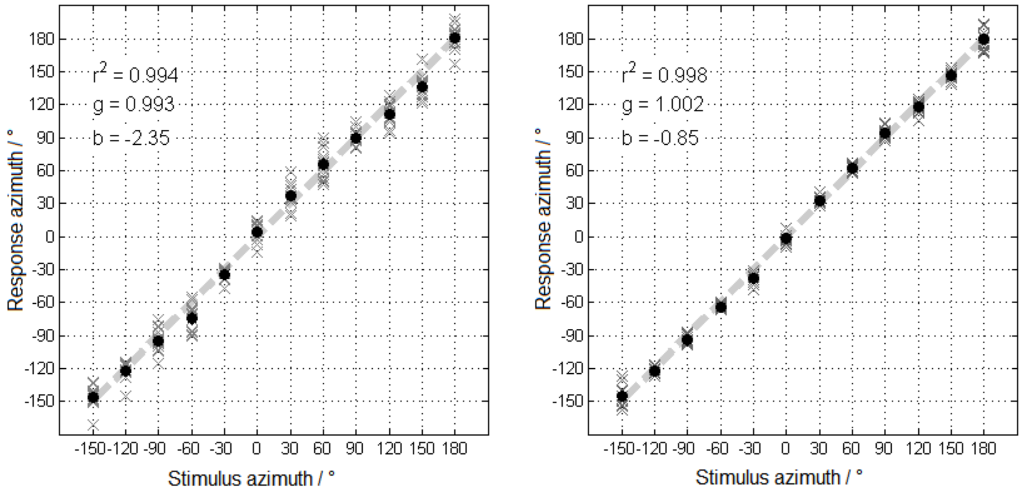


(1) Test case *Blind*

(2) Test case *See*

Figure 6.1.: *Measured responses of 5 healthy subjects ($n = 15$). Wrong indicated directions caused by front-back confusions are marked with circles. The mean errors of each direction are shown as black dots. Pearson's correlation coefficient (r^2); response gain (g); response bias (b); mean total duration of test procedure ($\overline{T}_{\text{tot}}$); mean duration of test step ($\overline{T}_{\text{step}}$); ideal response (dashed line).*

The paired t-test rejects the hypothesis with p = 0.1188 (5% significance level) and the confidence interval -0.75° to 6.52° for a sample size of $n = 5$.



(1) Test case *Blind*

(2) Test case *See*

Figure 6.2.: *Corrected results of the practicability test (n = 15). The mean errors of each direction are shown as black dots. Pearson's coefficient (r^2); response gain (g); response bias (b); ideal response (dashed line).*

Table 6.2.: *Statistical parameters of the corrected data for both test cases (Blind/See). Sample mean error ($\overline{\varphi}$); sample standard deviation (s).*

	Blind		*See*		*Blind - See*
Direction	$\overline{\varphi}$ / °	s / °	$\overline{\varphi}$ / °	s / °	$\Delta\overline{\varphi}$ / °
-150°	3.9	±8.8	4.6	±8.6	-0.7
-120°	-2.0	±8.2	-2.3	±3.2	0.3
-90°	-4.8	±10.1	-3.9	±3.6	-0.9
-60°	-14.1	±11.4	-3.9	±2.2	-10.2
-30°	-5.1	±5.5	-8.3	±5.1	3.2
0°	4.3	±8.6	-1.9	±5.1	6.2
30°	7.4	±11.2	2.5	±3.4	4.9
60°	5.6	±13.5	2.7	±2.6	2.9
90°	-0.3	±7.4	4.3	±4.1	-4.6
120°	-8.8	±9.5	-1.7	±5.0	-7.1
150°	-13.0	±9.9	-3.7	±4.0	-9.3
180°	0.7	±10.2	-0.6	±8.6	1.3

6.4. Discussion

In this chapter a practicability test with healthy subjects was conducted to evaluate the measurement set-up. The main purpose of this evaluation was to examine whether visual cues given by the set-up appearance have an effect on the results. Further, the duration of the test procedures was noted.

The short stimulus length lead to front-back confusions in both test cases (11 in the test case *Blind* and 8 in the test case *See*). This effect frequently occurs in spatial hearing measurements and is described in section 1.4. Confused responses are marked in figure 6.1 and can be found at directions mirrored with respect to ±90°. For better clarity and more meaningful results, those responses were discarded.

The following conclusions can be made with the corrected results shown in figure 6.2 and table 6.2:

No significant difference between the two test cases can be seen. However, this does not reject the hypothesis that the position of the speakers give visual clues for indication. A better experiment to investigate the influence of the speaker positions could be established with virtual sound sources or movable speakers. What the results do show, is that the standard deviations decrease if people are allowed to have a look. This indicates that the speaker positions affect the decisions made by the subjects.

The indication method (outstretched arm/pointing over shoulder) has an great impact on the accuracy of the detection algorithm. This can be seen at -150°, which was partially considered as backward direction caused by the rotation of the chair. The standard deviation for -150° and 180° is high for both test cases.

The evaluation test showed that a profile of spatial hearing abilities of an untrained subject can be measured in about 5 min (*Blind* test case). The shorter test duration in the second test case (*See*) could result from learning effects and visual cues.

7. Conclusion

This chapter summarizes the measurement system capability and gives an outlook for possible improvements of the measurement set-up.

7.1. Measurement System Capability

The presented measurement set-up demonstrates a practicable approach to gesture recognition in spatial hearing measurements. Audiological experiments such as speech recognition in noisy environment or localization tests are reduced in complexity through predefined test functions and automatic response detection.

A major advantage of this measurement system is that during a test, the patient can comfortably sit and intuitively respond to the presented stimuli. It is possible to quickly assess spatial hearing abilities at different angles and sound pressure levels with test step durations less than 10 s. In most cases, a short training session (5 min) is sufficient to familiarize the patient with the used gesture recognition concept. Without having to follow complex instructions, the patient can fully concentrate on the given tasks.

The software `AIODE` provides the functions needed for speech intelligibility tests and human sound localization tests in the horizontal plane. The current version features a *Localization test* module and an *OLSA test* module in German language. The software can be easily expanded through the implementation of additional modules and custom DSP algorithms. Besides the provided broad-band/narrow-band noise stimuli and OLSA sentences, custom stimuli can be loaded from `.wav` files.

As part of this work, an audio processing framework for the software of the measurement set-up was implemented.

Further, the test environment was described based on room acoustics parameters. The measurement limits given by the test environment were evaluated using EN ISO 8253-2 and EN ISO 8253-3.

A calibration method was implemented and the acoustic system output was validated. The system was checked on linear behaviour in the ranges commonly used for audiological measurements.

The tracking characteristics of the implemented gesture recognition routines were analyzed and the possible area of application in spatial hearing measurements was evaluated.

Finally, a practicability test was conducted to determine the applicability of the measurement set-up. The experiences made in the test promise good applicability for further audiological measurements with the specifications given in appendix A.

7.2. Insights and Outlook

7.2.1. Virtual Sound Sources

Virtual sound sources were realized using two different filtering algorithms. The VBAP algorithm simply modulates the amplitude of two channels causing a shift of the perceived auditory event (ILD). This method provides good results in the forward and backward range. Still, the filter can only be applied in audiological tests if the generated auditory events are validated at directions near the interaural axis.

The XTC filter generates two signals which produce a defined sound field at the ears of the subject. Auditory events can be simulated more exactly because ILD and ITD cues are reproduced more accurately. After the first implementation of the XTC filter severe spectral colorations were observed. These signal distortions make the filter inapplicable for speech recognition tests.

Several efforts were made to improve the performance of the implemented XTC filter. Different filter structures (real-time and off-line) were realized. Additionally, the filter was tested with signals at higher sampling rates using an interpolator. The spectral coloration was not reduced using this methods since they are caused by the ill-conditioned inversion of the system's transfer function (see Choueiri [CHO10]).

For this reason, a simple scaling factor (*'reg'*) of the recursive filter part was added, which allows to reduce the spectral coloration effect. Spectral analysis of the filter output showed less distortion, but the performance of the filter with regularized coefficients was not validated. The XTC algorithm may be improved by frequency-dependent regularization as described in Choueiri [CHO10].

7.2.2. Audio Processing and OLSA Test Functionality

As mentioned above, two audio processing structures were implemented. At the beginning, a real-time filtering structure was implemented by adapting callback routines of the `Fmod`-API. This processing structure was discarded for two reasons.

First, the advantage of faster audio processing has no impact, since sound presentation is delayed by the switching time of the `AUDIOBOX` which is approximately 550 ms. With the given stimulus duration of the applied audiological tests (less than 5 s) no difference can be noticed.

Second, OLSA test functionality is easily provided with off-line filtering routines. Basically, a speech and a noise signal are mixed for each OLSA test step. The noise fade-in time and the speech presentation delay can be adjusted. These functionalities were quickly implemented using off-line processing structures.

7.2.3. Improvements

Further work mainly involves the realization of higher resolutions in data acquisition and stimulus presentation (virtual sound sources). The following improvements of the presented measurement set-up are suggested:

- A feasibility study could investigate the implementation of additional `Kinect` sensors. The questions of the controllability of multiple instances and mutual influences of the sensors have to be clarified. The second sensor could significantly improve the detection results, especially at 0°/180°. Further, perspective errors could be minimized.

- The calibration procedure may be automated and improved by the implementation of an interface between `AIODE` and the UPV audio analyzer. Automatic equalization could be achieved by the realization of inverse filters. The filter coefficients may be computed from the spectra recorded with the audio analyzer.

- Alternatively, a zooming objective (`NYKO ZOOM`) could be applied to increase the field of view of the sensor. With the sensor fixated straight above the subject and speakers the detection results should be improved as well.

- The arrangement of the speakers may be verified via an automatic position detection using the RGB image and IR depth data of the `Kinect` sensor. This could avoid measuring errors caused by wrongly arranged speakers.

- An audiometric measurement tool may be implemented to support hearing level determinations. This includes a concept to reduce ambient noise levels to provide a wider measuring range.

- More complex test situations (*'cocktail-party effect'*) may be tested when using all 4 output channels of the sound card. This could be easily achieved, because only little software expansion is necessary.

- A wizard for localization tests could be implemented. After the experiment, the results could be automatically analysed and summarized.

- The audio processing framework may be improved by the support of different file formats and the usage of `Fmod`-effects.

Appendices

A. Measurement System Specifications

- **Maximum stimulus level**
  90 dB SPL, THD $< 3\%$

- **Minimum testable hearing thresholds**
  70 Hz - 12.5 kHz, 40 dB HL

- **Maximum testable OLSA SNR**
  $+25$ dB with 65 dB SPL noise level

- **Attenuation steps**
  1 dB, ±0.5 dB

- **Angular detection resolution**
  15° with $\phi_{\text{tot}} = \pm 165°$ and 30° with $\phi_{\text{tot}} = \pm 180°$

- **Controllable channels/speakers**
  2/12, $\phi_{\text{tot}} = \pm 180°$

- **Frequency response**
  70 Hz - 12.5 kHz, ±1.0 dB

- **Channel switching time**
  550 ms per channel

- **Test modules**
  Localization tests and OLSA tests

- **Supported file format**
  16 bit mono/stereo PCM WAVE

- **DSP**
  32-bit floating point processing
  interpolator
  21 1/3-octave band multi-channel equalizer

B. Measuring Equipment

- **Rohde & Schwarz UPV** Audio analyzer
  DC - 250 kHz
  SN: 101296

- **Norsonic 116** Integrating-averaging sound level meter
  SN: 20330

- **Brüel & Kjaer 4228** Pistonphone
  124.04 dB SPL at 250 Hz
  SN: 1504075

- **Brüel & Kjaer DP 0776** 1/2" pistonphone adapter

- **Brüel & Kjaer 2829** 4 Channel microphone power supply
  SN: 2716168

- **Brüel & Kjaer 2639** Preamplifier
  SN: 1202302

- **Brüel & Kjaer 4133** 1/2" Free-field condenser microphone cartridge
  SN:400674

- **Montarbo MT 160 D** Processor controlled powered monitor
  70 Hz - 19.5 kHz: ±2 dB
  SN: 0931261

C. XTC Filter Coefficients

The coefficients of a crosstalk-cancellation filter for a multi-speaker set-up are derived for N regularly arranged speakers with spacing $\Delta\phi$. The algorithm can be adapted to different configurations by setting $\Delta\phi$ and N.

The sketch in figure C.1 represents the XTC situation for an arbitrary pair of loudspeakers (S_i, S_j). A rotating coordinate system (x', y') is used, where y' is always aligned with the direction of the speaker positioned more left (S_i).

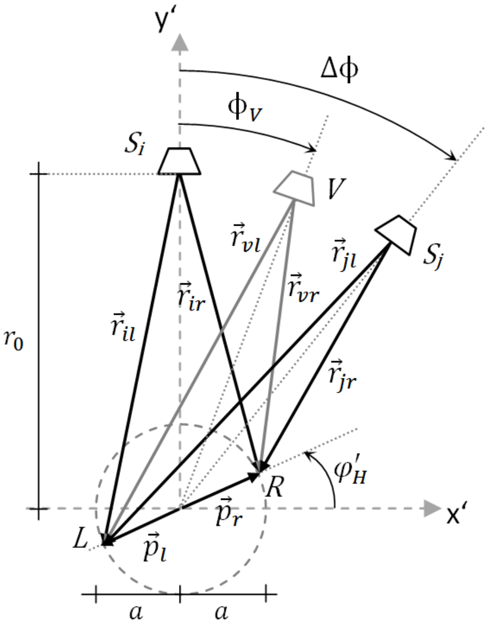


Figure C.1.: *Crosstalk-cancellation situation for a speaker pair in a multi-speaker arrangement*

The head orientation is determined through the chosen pair of speakers:

$$\varphi_H(i) = (i-1)\cdot\Delta\phi \qquad \text{with } i = 1\dots N. \tag{C.1}$$

After a rotation of ±90° the head orientation (φ_H) has to be corrected, since the system is mirrored:

$$\varphi'_H = \begin{cases} \varphi_H & \text{for } -90^\circ \le \varphi_H \le 90^\circ \\ \varphi_H + 180^\circ & \text{otherwise} \end{cases} \tag{C.2}$$

The position of the ears ($\vec{p_r}$,$\vec{p_l}$) and the speakers ($\vec{s_i}$,$\vec{s_j}$) are

$$\vec{p_r} = \begin{pmatrix} a \cdot \cos\varphi'_H \\ a \cdot \sin\varphi'_H \end{pmatrix} = -\vec{p_l} \qquad \vec{s_i} = \begin{pmatrix} 0 \\ r_0 \end{pmatrix} \qquad \vec{s_j} = \begin{pmatrix} r_0 \cdot \sin\Delta\phi \\ r_0 \cdot \cos\Delta\phi \end{pmatrix}. \tag{C.3}$$

The virtual sound source ($\vec{v}$) can be described by

$$\vec{v} = \begin{pmatrix} r_0 \cdot \sin\phi_V \\ r_0 \cdot \cos\phi_V \end{pmatrix} \qquad \text{with } 0 \leq \phi_V \leq \Delta\phi. \tag{C.4}$$

Next, the distance vectors between the sound sources and the ears are determined:

$$\begin{aligned} \vec{r_{ir}} = \vec{s_i} - \vec{p_r} &= \begin{pmatrix} -a \cdot \cos\varphi'_H \\ r_0 - a \cdot \sin\varphi'_H \end{pmatrix} \\ \vec{r_{il}} = \vec{s_i} - \vec{p_l} &= \begin{pmatrix} a \cdot \cos\varphi'_H \\ r_0 + a \cdot \sin\varphi'_H \end{pmatrix} \end{aligned} \tag{C.5}$$

$$\begin{aligned} \vec{r_{jr}} = \vec{s_j} - \vec{p_r} &= \begin{pmatrix} r_0 \cdot \sin\Delta\phi - a \cdot \cos\varphi'_H \\ r_0 \cdot \cos\Delta\phi - a \cdot \sin\varphi'_H \end{pmatrix} \\ \vec{r_{jl}} = \vec{s_j} - \vec{p_l} &= \begin{pmatrix} r_0 \cdot \sin\Delta\phi + a \cdot \cos\varphi'_H \\ r_0 \cdot \cos\Delta\phi + a \cdot \sin\varphi'_H \end{pmatrix} \end{aligned} \tag{C.6}$$

$$\begin{aligned} \vec{r_{vr}} = \vec{v} - \vec{p_r} &= \begin{pmatrix} r_0 \cdot \sin\phi_V - a \cdot \cos\varphi'_H \\ r_0 \cdot \cos\phi_V - a \cdot \cos\varphi'_H \end{pmatrix} \\ \vec{r_{vl}} = \vec{v} - \vec{p_l} &= \begin{pmatrix} r_0 \cdot \sin\phi_V + a \cdot \cos\varphi'_H \\ r_0 \cdot \cos\phi_V + a \cdot \cos\varphi'_H \end{pmatrix} \end{aligned} \tag{C.7}$$

After calculation of the norm, the distances can be simplified to:

$$\begin{aligned} \|\vec{r_{ir}}\| &= \sqrt{r_0^2 - 2a \cdot r_0 \sin\varphi'_H + a^2} \\ \|\vec{r_{il}}\| &= \sqrt{r_0^2 + 2a \cdot r_0 \sin\varphi'_H + a^2} \\ \|\vec{r_{jr}}\| &= \sqrt{r_0^2 - 2a \cdot r_0 \sin\left(\Delta\phi + \varphi'_H\right) + a^2} \\ \|\vec{r_{jl}}\| &= \sqrt{r_0^2 + 2a \cdot r_0 \sin\left(\Delta\phi + \varphi'_H\right) + a^2} \\ \|\vec{r_{vr}}\| &= \sqrt{r_0^2 - 2a \cdot r_0 \sin\left(\phi_V + \varphi'_H\right) + a^2} \\ \|\vec{r_{vl}}\| &= \sqrt{r_0^2 + 2a \cdot r_0 \sin\left(\phi_V + \varphi'_H\right) + a^2} \end{aligned} \tag{C.8}$$

Finally, the filter coefficients are obtained by inserting the distances in the equations described in Gut [GUT10]. The constants of the XTC algorithm are set to $\Delta\phi = 30°$, $N = 12$, $r_0 = 1.0$ m and $a = 0.17$ m .

D. Audio Processing Classes

AUDIOBOX-Class

`AUDIOBOX()`
Constructor.

`virtual ~AUDIOBOX()`
Destructor.

`int loadDLL()`
Loads the AUDIOBOX-DLL and returns the error code.

`void getConfig()`
Reads in the calibration values of the configuration file.

`static float* getEQParam(int index)`
Static getter to access the EQ parameters.

`bool InitDevice()`
Initializes the AUDIOBOX and writes the new calibration values if necessary.

`bool IsInit()`
Returns the initialization state of the AUDIOBOX.

`void SetChannel(unsigned char InputChannel, unsigned char OutputSpeaker)`
Connects the input channel with the desired output channel.

`void MuteAll()`
Mutes all channels.

`bool ShowError(UINT64 e)`
Error handling method.

`static float m_paramEqAB[24][21]`
Calibration values for the software EQ

`float m_deltaAB[24]`
Calibration values for the AUDIOBOX

`bool m_DeviceInit`
Initialization state

DSPAlgo-Class

`DSPAlgo()`
Constructor.

`virtual ~DSPAlgo()`
Destructor.

`bool calcCoeffs(int angle, int* speak)`
Computes filter coefficients for the algorithm depending on the direction.

`virtual void startDSP(float* data, int channels, float rate, unsigned int ratio, unsigned int length, float* left,float* right)`
Performs the particular DSP algorithm.

`virtual void reSample(float* input, unsigned int inlength, float* output, unsigned int outlength, double ratio)`
Changes the sampling rate of the data.

`int wrapIndex(int buffer_position, int index)`
Wraps the index of the circular buffer.

`float circ_buf[BUFSIZE]`
Circular buffer

SoundCard-Class

`SoundCard()`
    Constructor.
`SoundCard(int max_channels, char* device_name)`
    Copy-Constructor.
`virtual ∼SoundCard()`
    Destructor.
`bool Init()`
    Initializes the `Fmod`-system and the sound device.
`bool isInit()`
    Returns the initialization state of the sound system.
`void Close()`
    Closes and releases the `Fmod`-system.
`unsigned int Play(char *file, int volume, int ref_vol, int* speakers)`
    Decodes the file, performs DSP for localization tests and plays the stimulus.
`unsigned int PlayOlsa(char *speech, char *noise, int delta_vol_speech, int vol_noise, int ref_vol, int fading, int offset, bool mix, int* speakers)`
    Decodes the file, performs DSP for OLSA and plays the stimulus.
`void setDSP(DSPAlgo *algorithm)`
    Sets the desired algorithm.
`bool startDSP(int angle, int* speakers)`
    Starts the DSP algorithm.
`void MuteAndFree(bool muteval)`
    Mutes and releases all playback channels and `Fmod`-objects.
`void ERRCHECK(FMOD_RESULT result)`
    Error handling method.

---

int m_maxchannels
    Number of used channels.
char m_name[255]
    Name of the sound device.
float* m_paramEq
    Parameters for the software EQ.
bool m_isinit
    Initialization state
DSPAlgo* m_algorithm
    DSP algorithm object
FMOD::System *m_system, m_result, *m_channel1, *m_channel2
    `Fmod`-system objects
FMOD::System *m_decoder1, *m_decoder2, *m_signal1, *m_signal2
    `Fmod`-sound objects
FMOD::DSP *m_eqDSP[42]
    2 channel `Fmod` parametric EQs

E. MATLAB Scripts

computeRT60.m

This script computes the reverberation times based on 16-bit wav-data recorded with the UPV audio analyzer.

- **Lines 8-21:** The recorded .wav-file is loaded, scaled, truncated and windowed.
- **Lines 23-27:** The data is zero-padded to double length.
- **Lines 29-49:** The system input and output are plotted.
- **Lines 51-59:** The FFT is performed and the RTF/RIR are computed.
- **Lines 61-63:** The second half of the RIR is discarded.
- **Lines 65-77:** The EDC is computed and plotted.
- **Lines 79-85:** The Schröder backward integral is computed.
- **Lines 87-103:** The broad-band RT_{60} is computed via interpolation.
- **Lines 105-126:** The narrow-band RT_{60}s are computed via RIR filtering.
- **Lines 128-141:** The results are plotted.

calibration.m

The calibration script computes the required gain factors and the hardware calibration offsets for one channel based on power spectra recorded with the UPV audio analyzer (.trc-data).

- **Line 10:** The desired calibration rms-value is set with DBCAL.
- **Lines 13-16:** The frequency and time properties are set.
- **Lines 19-27:** The power spectral densities are loaded from the .trc-files.
- **Lines 29-54:** The 1/3-octave band spectrum is computed for each speaker.
- **Line 57:** The measured total rms-value is computed.
- **Lines 59-74:** The 1/3-octave band spectra are plotted.
- **Lines 76-97:** The Fmod-gain parameters are calculated using the polynomial fit.
- **Lines 99-145:** The Fmod-gain parameters are corrected with the empirically found factors.
- **Lines 147-154:** The calibration file config.cfg is written. A single AUDIOBOX-calibration value (relative difference in dB) and 21 fmod-gain parameters are generated in separated lines for each channel and speaker.

pinknoise.m, noisecreator.m and sinecreator.m

These scripts generate 16 bit `.wav`-signals which are applicable for the measurement set-up.

kinectpos.m

This script computes the numerical simulation of the `Kinect` position as shown in figure 5.5.

- **Lines 8-22:** The geometrical parameters and the resolution are set.
- **Lines 29-41:** The loops for the different distances and inclinations are started.
- **Lines 43-115:** The algorithm computes the overlapping area of the field of view and the volume of interest for each step on the z-direction. Shadowing effects caused by the speakers are calculated at lines 46-52. The effect of head/torso shadowing is approximated in line 99.
- **Lines 116-120:** The hit points are summed up and saved for each direction and inclination.
- **Lines 122-128:** The results are normalized with the maximum volume.
- **Lines 130-137:** A 3D-surface is plotted to illustrate the results.

evaluation.m

This script processes the results of the practicability test. The measured data has to be extracted from the `AIODE`-database. The script is described for one test case (*blind*).

- **Lines 6-64:** The measurement results are loaded into a dataset from separate data blocks of each subject. The starting and end time of the test procedure are extracted and the durations are computed.
- **Lines 66-72:** For a better clarity, the data range is wrapped from $0°/330°$ to $-150°/180°$.
- **Lines 75-84:** The results are corrected from errors larger than $30°$.
- **Lines 86-86:** The linear data fit and the Pearson's correlation coefficient are computed.
- **Lines 91-96:** The ideal response or the regression line are plotted.
- **Lines 98-115:** The responses are plotted and the average test duration is computed and displayed.
- **Lines 120-128:** The mean values and the standard deviations are computed and plotted.

List of Figures

List of Tables

Bibliography

[ALT09] Altmann CF, Wilczek E et al.: *Processing of Auditory Location Changes after Horizontal Head Rotation.* The Journal of Neuroscience 29(41), Frankfurt: 13074 - 13078 (2009).

[AGT11] Agterberg MJH, Snik AFM et al.: *Improved Horizontal Directional Hearing in Bone Conduction Device Users with Acquired Unilateral Conductive Hearing Loss.* Journal of the Association for Research in Otolaryngology 12, Nijmegen: 1 - 11 (2011).

[BER10] Berger M, Schweingruber T: *Entwicklung einer Richtungshöranlage.* Bachelor's Thesis, Bern University of Applied Sciences (2010).

[BLA97] Blauert J: *Spatial Hearing - The Psychophysics of Human Sound Localization.* The MIT Press, Cambridge (1997).

[BLN09] Blanchette J, Summerfield M: *C++ GUI PRogramming with Qt 4 - Second Edition.* Prentice Hall, Boston (2009).

[BOH86] Bohn DA: *Constant-Q Graphic Equalizers.* AES Reprint, 79th Convention, New York (1986).

[BSA08] British Society of Audiology: *Guidelines on the Acoustics of Sound Field Audiometry in Clinical Audiological Applications.* `www.guymark.com/filestore/BSA_soundfieldguidelinesfeb2008.pdf`, accessed May 2011 (2008).

[BUS08] Buschermöhle M, Van Esch T et al.: *Richtungshörtests für die rehabilitative Audiologie - Ergebnisse des HearCom-Projektes.* 39. DGMP Tagung, Oldenburg (2008).

[CHO10] Choueiri EY: *Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers.* `www.princeton.edu/3D3A/Publications/BACCHPaperV4d.pdf`, accessed July 2011 (2010).

[COU97] Couvreur C: *Implementation of a One-Third-Octave Filter Bank in MATLAB.* `citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.5728`, accessed July 2011 (1997).

[ENE06] Eneman K, Van den Bogaert T et al.: *Specification spatial-hearing test.* Hearcom - Integrated Project D-7-3 (2006).

[FAR00] Farina A: *Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique.* AES Reprint, 108th Convention, Paris (2000).

[FMO11] Firelight Technologies: *FMOD Ex Interactive Audio Middleware - Documentation.* `www.fmod.org`, accessed July 2011 (2011).

[GUT10] Gut I, Schwander D: *Demonstrationsanlage einer virtuellen Schallquelle.* Bachelor's Thesis, Bern University of Applied Sciences (2010).

[GYO03] György W: *Psychoakustische Messungen auf dem Gebiet der menschlichen Lokalisationsschärfe.* `heja.szif.hu/ELE/EE-020110-A/ee020110a.pdf`, accessed June 2011 (2003).

[HAF92] Hafter ER, Saberi K et al.: *Localization in an Echoic Environment.* Pergamon Press, Advances in the Biosciences Vol. 83, Berkeley: 555 - 561 (1992).

[HOL05] Hol MKS, Bosman AJ et al.: *Bone-Anchored Hearing Aids in Unilateral Inner Ear Deafness: An Evaluation of Audiometric and Patient Outcome Measurements.* Otology & Neurotology 26, Nijmegen: 999 - 1006 (2005).

[HUE06] Hüsler J, Zimmermann H: *Statistische Prinzipien für medizinische Projekte.* Verlag Hans Huber, Bern (2006).

[I389] EN ISO 389-7: *Acoustics - Reference zero for the calibration of audiometric equipment - Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions.* (1998).

[I8253a] EN ISO 8253-2: *Acoustics - Audiometric Test Methods - Part 2: Sound field audiometry with pure tone and narrow-band test signals.* (1998).

[I8253b] EN ISO 8253-3: *Acoustics - Audiometric Test Methods - Part 3: Speech Audiometry.* (1998).

[KAR01] Karjalainen M, Antsalo P et al.: *Estimation of Modal Decay Parameters from Noisy Response Measurements.* AES Reprint, 110th Convention, Amsterdam (2001).

[KIP11] Microsoft Research: *Kinect for Windows SDK beta - Programming Guide* Beta 1 Draft Version 1.0a, `research.microsoft.com/en-us/um/redmond/projects/kinectsdk/guides.aspx`, accessed June 2011 (2011).

[KUM10] Kumpik DP, Kacelnik O et al.: *Adaptive Reweighting of Auditory Localization Cues in Response to Chronic Unilateral Earplugging in Humans.* The Journal of Neuroscience 30(14), Oxford: 4883 - 4894 (2010).

[KUT04] Kuttruff H: *Akustik - Eine Einführung.* Hirzel Verlag, Stuttgart (2004).

[MAZ04] Mazzoni D: *libresample 0.1.3 - Real-time Library Interface* `www-ccrma.stanford.edu/~jos/resample/`, accessed Jul 2011 (2004).

[MUE08] Müller S: Measuring Transfer-Functions and Impulse Responses. In: Havelock D, Kuwano S, Vorländer M (Ed.): *Handbook of Signal Processing in Acoustics. Volume 1.* Springer: 66 - 85, New York (2008).

[OLS00] HörTech gGmbH: *Oldenburger Satztest - Handbuch und Hintergrundwissen.* `www.hoertech.de/web/dateien/HT.IE.007-Handbuch_und_Hintergrundwissen_OLSA.00.1.pdf`, accessed September 2011 (2000).

[PAU03] Paulus E: Sound Localization Cues of Binaural Hearing. In: *Laryngo-Rhino-Otologie; 82.* Georg Thieme Verlag Stuttgart, New York: 240 - 248 (2003).

[POP08] Populin LC: *Human Sound Localization: Measurements in Untrained, Headunrestrained Subjects Using Gaze as a Pointer.* Exp Brain Res. 190(1), Madison: 11 - 30 (2008).

[PUL97] Pulkki V: *Virtual Sound Source Positioning Using Vector Base Amplitude Panning.* J. Audio Eng. Soc., Vol. 45. No. 6, Helsinki: 456 - 466 (1997).

[SAL11] Salzmann J: *AIODE - Software Documentation*, ARTORG Center - Artificial Hearing Research, University of Bern (2011).

[SEN11] Sengpiel E: *Tontechnikrechner - sengpielaudio* `http://www.sengpielaudio.com/Rechner-spl.htm`, Homepage, accessed Jul 2011 (2011)

[SMI08] Smith OJ: *Spectral Audio Signal Processing, October 2008 Draft.* `ccrma.stanford.edu/~jos/sasp/`, Online Book, accessed Aug 2011 (2008).

[VBO06] Van den Bogaert T, Klasen TJ et al.: *Horizontal localization with bilateral hearing aids: Without is better than with.* J. Acoust. Soc. Am. 119, Leuven: 515 - 526 (2006).

[VGR10] Van Grootel TJ, Van Opstal AJ: *Human Sound-Localization Behavior Accounts for Ocular Drift.* J Neurophysiol 103, Nijmegen: 1927 - 1936 (2010).

[VWA06] Van Wanrooij MM, Van Opstal AJ: *Sound Localization Under Perturbed Binaural Hearing.* J Neurophysiol 97, Nijmegen: 715 - 726 (2006).