# Development of a modular system for speaker array prototyping

TI Project

Sebastian
Blamberger

Supervisor: Georgios Marentakis

Graz, October 2011

institut für elektronische musik und akustik

**Abstract**

Small speaker arrays provide the opportunity to create acoustic environments with high spatial resolution. Hence, problems due to perceptual instabilities in spatial audio reproduction can be overcome. This pre-condition needs to be fulfilled in order to proceed with the evaluation of auditory direct manipulation designs, as limitations in 3D audio reproduction can influence the usability ratings. To avoid this problem, a modular prototype speaker array consisting of 48 speakers plus corresponding amplification circuits has been developed. This prototype uses small speakers and Class-D amplifiers that offer low power consumption and good efficiency. Furthermore, the modular design allows for creating numerous speaker configurations that may be adjusted depending on the actual application. An 8-channel amplification printed circuit board (PCB) has been developed and evaluated together with 2" broadband speakers. The results show that it is possible to achieve the desired prototyping flexibility together with a reasonable acoustic behavior. Furthermore, the work exemplarily deals with applications of the system, which include audio reproduction and the use as output platform for human computer interaction design.

# Contents

# 1 Introduction

Interaction with computing devices is performed mostly through the visual and to a lesser extent through the haptic and audio modalities. Auditory direct manipulation is a promising way to interact using the auditory channel, as it builds on our ability to interact with objects in space through spatial information obtained through hearing. As most auditory direct manipulation interfaces work using headphones, the potential of using speaker arrays as an output channel for human computer interaction has not been investigated. Speaker arrays have been mainly used to create 3D audio environments for music and entertainment as well as a secondary output channel for virtual reality environments. The task of this project was to provide a platform that will assist in covering the aforementioned gap, by designing and building a new modular platform that can be used to allow for prototyping and testing different speakers arrangements.

Designing such a system involves an iterative process where the steps of designing, prototyping, testing and validating are repeated until the desired outcome is achieved. Small speakers rarely come in the market as active, rather they need to be amplified. For this reason, we initially looked for a way to amplify the speakers. After some research on amplifying techniques, we found that Class D amplifiers could meet the requirements for this system at a low price per channel. This is because they offer very good efficiency and low space requirements as there is no need for heat dissipation due to their high efficiency. Nevertheless, there have been mixed reviews about the performance and for this reason, this was evaluated in the report.

In addition, our target system should provide an easily re-configurable mechanically stable platform, with low building cost and power consumption. Such a requirement requires the use of small components, such as small speakers. This poses a number of constraints on the audio output of the system, that are evaluated in the report. The application area of this system is to provide sound output in sonic interaction design scenarios, installations, performances and 3D audio evaluation experiments. Low power consumption is required to leave the option for portable use open.

The report consists of three parts: In the first part the need for such a system in different contexts and the existing techniques are discussed. The second part describes the used amplification technique, calculation and design of amplifier board, speaker enclosures and mechanical components of the modular speaker array. In the last part the results are discussed and an outlook for future applications is given.

# 2 Literature Review

There have been other attempts to build small dense speaker arrays in the past. Beer et Al. [BMB09] sought a loudspeaker array design that can compensate for the limited frequency response of small speakers panels. The goal was to imitate the response of a flat panel speaker and avoid vibration modes on the thin and large foil membrane resulting from the limited stiffness and weight of these membranes. As is known, when small speakers are used without an enclosure, comb-filtering and acoustic short-circuit are observed. In addition, the resulting lower cutoff frequency is certainly higher than that of speakers with greater membrane area. When, however, the membrane areas of several speakers are combined, a better low frequency response is obtained. To achieve this, the speakers have to be placed in the same enclosure to enlarge the resulting active membrane area. The effective SPL gain depending on the number of speakers can be seen in Table 1. This values are determined using formulas from Zollner and Zwicker [ZZ93]. The proposed speaker feeding strategies improved bass playback and linearized of the frequency response. Furthermore, feeding speaker clusters with different frequency ranges and using Bessel array techniques [Kee90] to avoid problems caused by superposition led to vast improvements regarding the polar magnitude response. See section 4.1.1 for more details.

Such an approach, is however not directly applicable for our goal. This is first, because although a boost in low-frequency response is observed, this is combined with interference as the loudspeakers share the same enclosure. In the case of [BMB09], the effect of interference was limited as loudspeakers played signals of similar amplitude and phase. However, spatial audio algorithms require that speakers play in variable phase and amplitude relations. Especially for high frequencies, this is problematic and can lead to limitations to their reproduction accuracy. To avoid this interference, the air volume behind each speaker membrane has to be separated from others. Second, once combined into a single enclosure, the loudspeakers cannot be rearranged. This causes limitations to the modularity of the system. Consequently, based on the observations made by Beer et Al., we decided to use separate speaker enclosures. This also enabled the speakers to be mounted at variable settings, while their studies headed toward flat panel speakers with square matrix form on a common plane.

| Number of elements | SPL at | |
|---|---|---|
| | 50 Hz | 100 Hz |
| 1 | 57 dB | 69 dB |
| 4 | 69 dB | 81 dB |
| 9 | 76 dB | 88 dB |
| 16 | 81 dB | 93 dB |
| 25 | 85 dB | 97 dB |
| 36 | 88 dB | 100 dB |

Table 1: Calculated SPL of different array sizes at 50 Hz and 100 Hz in 1m distance

## 2.1 Auditory Direct Manipulation

Auditory direct manipulation is an extension of direct manipulation in the auditory domain. Direct manipulation is a fundamental concept within human-computer interaction (HCI) and most graphical user interfaces rely on this concept. Instead of typing complex commands to a command line graphical objects are manipulated directly. As an example, moving an icon from A to B on a computer screen is direct manipulation. This task can be done using a mouse, touch input or any other kind device that supports pointing and selection of input. It has been claimed that direct manipulation makes it is easier to learn basic and new advanced features, improves the user's confidence in using the system and encourages usage by being more enjoying to use [Shn93]. Direct manipulation requires the creation of a model world to interact with. This objects can be files, folders, programs etc. that have an iconic, indexical or symbolic representation that users can understand or learn easily (e.g. moving a file to waste bin deletes it).

Direct manipulation interfaces need to adhere to the following properties [Shn93]: 1. Continuous representation of the object of interest. 2. Physical actions or labeled button presses instead of complex syntax. 3. Rapid incremental reversible operations whose impact on the object of interest is immediately visible. Shneiderman [Shn93] has suggested that direct manipulation systems have the following virtues: 1. Novices can learn basic functionality quickly, usually through a demonstration by a more experienced user, 2. Experts can work extremely rapidly to carry out a wide range of tasks, even defining new functions and features, 3. Knowledgeable intermittent users can retain operational concepts, 4. Error messages are rarely needed, 5. Users can see immediately if their actions are furthering their goals, and if not, they can simply change the direction of their activity.

Support to the auditory direct manipulation attempt is given by the fact that the human auditory system provides both the ability to localize sound and the ability to focus on objects of interest and separate them based on their location. As observed in the cocktail party effect [Aro08], as long as sounds are spatially separated, users can direct their attention to multiple sound sources in an acoustic environment. When creating an auditory direct manipulation interface a sensitive point is how to interact with objects in order to activate or manipulate them when they are not presented visually but as audio. As auditory localization is not as accurate as visual, this question is not easy to answer. It has however been shown that when feedback is provided, selection can be performed successfully in the case of binaural reproduction [MB06].

Most research on Interactive 3D Audio Systems has been done using binaural audio systems because they are much cheaper, and the required hardware technology is already present in most computers/portable devices [CL91], [MB04], [GM99]. Binaural audio is a technique to provide the impression of sounds coming from different directions outside-the-head of the listener though headphones. To achieve this, Head Related Transfer Functions (HRTFs) are applied to the signal we want to spatialize. HRTFs model the way the sound is transformed by the upper body, head, pinna and ear canal, that are different for each person. To obtain HRTFs the transfer functions for various microphone positions for each ear are recorded [ADT01]. As individual HRTF measurements are impractical, there has been an effort to use non-individualized HRTFs. Non-individualized or generalized HRTFs are gained either by using HRTFS from dummy heads, using representative HRTFs from specific persons or averaging the HRTFs of as many subjects as possible. As shown by Wenzel et al. [WAKW93] the use of non-individualized HRTFs, leads to a substantial decrease in localization accuracy and a higher number of front back and elevation confusions in comparison to individual HRTFs. Furthermore, the impact of the headphone system has to be counterbalanced not to influence the system impulse response. This is because even small differences in the headphone placement on the ear lead to a decrease of localization performance because they lead to a change in the HRTFs. When using intra-aural headphones variations in headphone placement are limited and lead to better results [CF95]. In conclusion, although the use of headphones to provide interaction with 3D audio system reduces the system cost factor, the following issues can substantially reduce system usability:

- The ears are covered, so conversation with other users is hindered
- Need for accurate HRTFs to ensure good localization
- Need Headphone Equalization that is not always possible
- Uncomfortable if used for long time
- Need for head tracking to make movement in the virtual world possible

These are important obstacles in the creation of interactive audio systems. Speaker arrays avoid these problems due to the fact that sound is emitting from speakers, therefore problems related to externalization and confusions are largely avoided. Nevertheless, certain issues appear in these systems too that are briefly explained in the following.

## 2.2 Overview over sound source spatialization techniques based on speaker arrays

Historically the following array techniques have been established. See Fig.1 for an overview. Most of the existing speaker array techniques work with regular speaker setups.

### 2.2.1 Regular speaker setup

**Stereo Panning** The basic approach for realizing sound source spatialization with multiple speakers is stereo panning. The speaker pair is fed with signals weighted according to a panning law. As an example the tangent law is given by

$$\frac{tan\Theta_T}{tan\Theta_O} = \frac{g_1 - g_2}{g_1 + g_2} \tag{1}$$

where $\Theta_T$ represents the source position angle and $\Theta_O$ the loudspeaker base angle. The variables $g_1$ and $g_2$ are relative gain factors, the absolute gain factors can be derived for each loudspeaker position [Pul01]. The loudspeaker aperture or base angle is defined as $60°$. The listener is assumed to sit at a point where the distance to each speaker is the same. If this does not happen, localization is not working correctly and sounds are perceived towards the direction of the preceding speaker or directly on it.

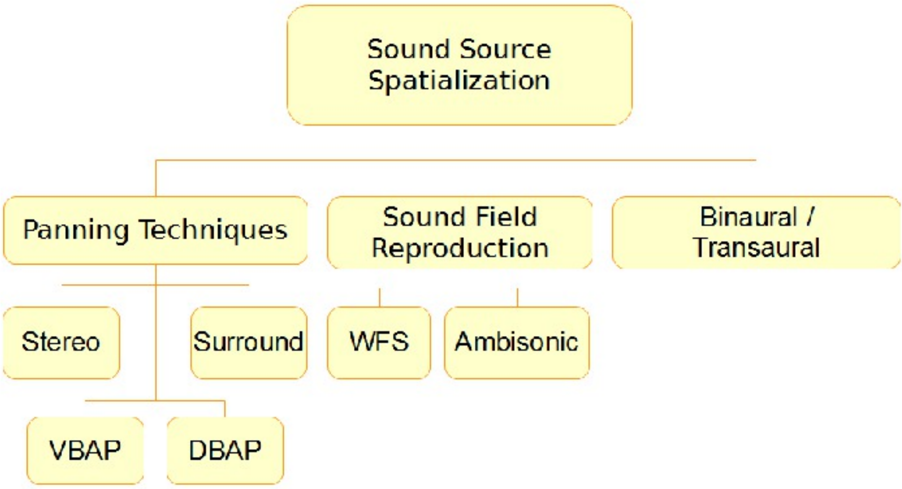



Figure 1: Sound source spatialization methods

**Vector base amplitude panning** Vector Base Amplitude Panning (VBAP) extends stereophony to using more than two speakers. In the 2D case, speakers are placed equidistantly around the listener in regular intervals. In this way, virtual sound sources can appear in a $360°$ range surrounding the listener. By using a Speaker triplet containing elevated speakers, sound sources can be placed on a triangular area, more precisely a spherical segment (See Fig.2). Multiple speaker triplets can be used to extend the area to a half or full sphere. Again the same symmetry requirements exist for VBAP as for stereophony.

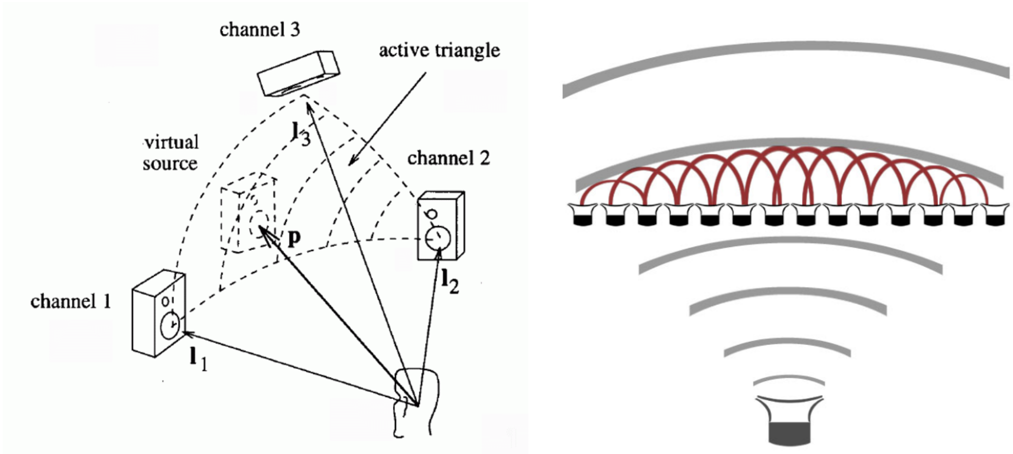



Figure 2: VBAP speaker triplet (left), Superposition of waves (right)

**Holophonic Approaches: Wave field synthesis and Ambisonics** More advanced methods intend to reproduce the recorded or artificially generated sound field. The signals from every independent speaker are superimposed and form a sound front (See Fig.2). As the simulated sound front is as close as possible to the one a real sound from a specific location would create, when it is decoded by the human auditory system, it allows the perception of a sound source that is not physically existent.

WFS is mostly used for one dimensional applications e.g. horizontal line arrays. It results in an area of correct perception that is very large, so that it can be efficiently used for large audiences. Due to the finite length and number of speakers sound perturbations appear at the edges. Because the resulting wavefront is a composite of elementary waves, a sudden change of pressure can occur if no further speakers deliver elementary waves where the speaker row ends.This is called the truncation effect and leads to shadow waves that cause perceptual disadvantages [dVSV94] [SRA08].

The Ambisonics approach uses spherical harmonics to create a sound field that surrounds the user completely. It is not just a technique for reproducing a sound field, it too provides a format that can be played back on different speaker constellations. When playing back Ambisonics audio material the speaker weights have to be calculated respective to the Ambisonics order N and the number of speakers L, the signal has to be decoded. The L loudspeaker signals $y(n)$ are computed from the $O = (N+1)^2$ Ambisonics components $d(n)$ with the decoding matrix $D$ by

$$y(n) = Dd(n) \tag{2}$$

Due to the finite resolution of the array, the acoustic field can only be precisely rendered until a certain limit called "spatial aliasing frequency". This is determined based on the spatial distance $\Delta x$ between the speakers. Spatial aliasing occurs for frequencies above $f_{nyq}$ [BVV93].

$$f_{nyq} = \frac{c}{2 * \Delta x} \tag{3}$$

The effect of spatial aliasing on the rendered acoustic field is an alteration of the shape of the wave front, producing certain degradation on the source localization (especially for pure high-frequency sources) and a coloration of the sound above this frequency due to the comb-filtering effect [LBPE05]. WFS and Ambisonics have two opposing problems. To avoid spatial aliasing the distance between every speaker has to be small, what stands against the requirement for a wide frequency response because smaller distances require smaller speaker diameters. For a frequency of $20kHz$ the wavelength is about $17mm$ and the distance between the speaker centers should be half that value to cover the worst case. Because the human listening is not very sensitive to spatial aliasing the limitations are less severe but the number of speakers required for good sound field reproduction is still quite high. In addition to spatial aliasing, using less speakers in the same environment leads to a degradation of the gainable spatial resolution of the system. The techniques mentioned here are highly dependent on speaker placement and do not work for arbitrary loudspeaker set-ups. This led to the introduction of DBAP.

### 2.2.2 Irregular speaker setup

**Distance-based amplitude panning (DBAP)** DBAP extends the principle of intensity panning from a pair of speakers to a loudspeaker array of any size, with no a priori assumptions about their positions in space or relative to each other. All loudspeakers radiate coherent signals, whereby the underlaying amplitude weighting is based on a distance attenuation model between the position of the virtual sound source and each loudspeaker [LBdlH11]. Nevertheless, its ability to represent uniformly virtual sounds across the speaker array surface has not been evaluated and the system also poses the same symmetry requirements as stereophony.

## 2.3 Synthesis

Stereo, VBAP, DBAP and Ambisonics function well only within the sweet spot area. The sweet spot refers to the area in which the hypotheses of each system still holds. This is usually an area in which the user is symmetrically placed with respect to the speaker array. When the listener position changes or when multiple listeners need to be placed within the system, localization and sound quality problems emerge. The problem is less distinct in WFS because the reproduced sound field is similar to a real one , but problems still occur at the edges. The exact limits of the sweet spot and how it can effect interaction have not been however investigated. Furthermore, although successful directional audio presentation can be achieved, the extent to which this can be done with the uniformity that is required in order to support movement has not been proven.

As shown by Theile and Plenge [The77], localization problems occur when phantom sources are presented at the listeners sides. This was experimentally proved by changing the base center angle of a stereo speaker pair in steps from $0$ to $90°$. The accuracy of the users ability to localize the phantom sources was tested for each step. It was found that even small level differences between the 2 loudspeakers lead to large changes of the perceived phantom source position angle and localization jumps between the loudspeaker at the front and at the back.

Fig.3 shows the results for the phantom source position angle $\varphi$ depending on the level difference $\Delta L$ regarding to the stereo pan law, for a lateral displacement of the base center $\delta$ of $40°$ and $90°$. With $\delta = 40°$ displacement, the localizations works still good. The results for $\delta = 90°$ displacement (lateral base center position) show extremely large variations of $\varphi$. 6 dB level differential and an aperture of $60°$ lead to an angle displacement of over $40°$. For $\delta = 0°$ (frontal base center position) 6 dB level difference causes about $14°$ displacement. The conclusion to be drawn from this is that lateral sources should be represented through real sources to minimize the effect if possible.

The sound field in a room is always influenced by the acoustic properties of the room itself. Reflections from more or less reflective surfaces in the room lead to a sound coloration or even echoes and can influence the performance of source localization. It has been shown by Start et Al. [SRDV97] and Verheijen et Al. [VVTB95] that the reverberation time of a room influences the localization error angle in a negative way.

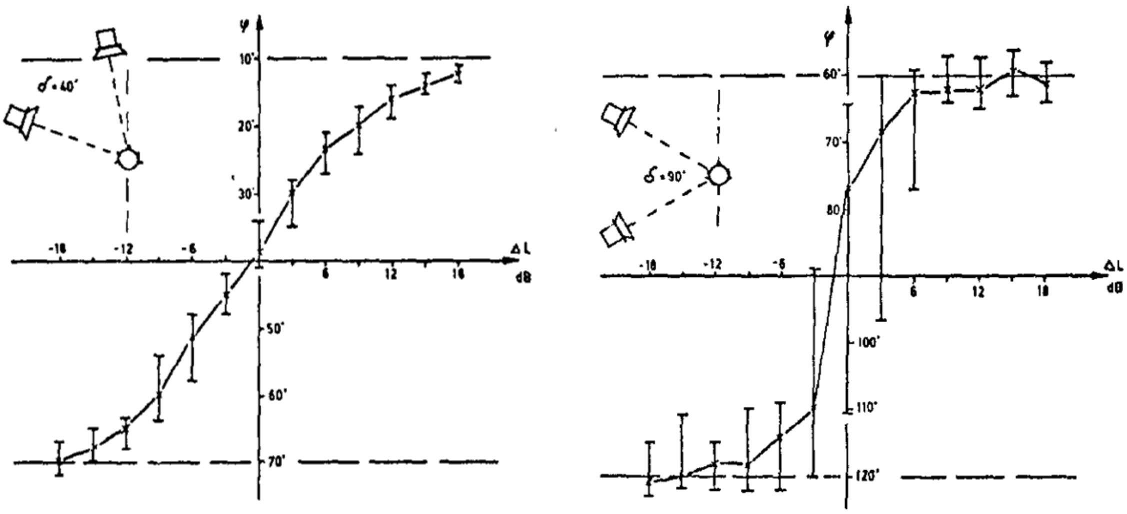



Figure 3: Phantom source position angle $\varphi$ for base angle displacement of $40°$ (left) and $90°$ (right), the x-axis shows the level difference. [The77]

Any spatialization technique is working optimally in anechoic rooms or the free field. Nevertheless this is idealized and not gainable in a majority of playback situations. In general the room of playback should be less echoic than the room of recording [Noi10]. For optimization of the playback room, the room impulse response can be compensated by applying filters to every speaker signal. The compensation is easily achieved for stereo and surround applications, when it comes to WFS or Ambisonics it appears to be a bit more complex. Spors et Al. presented an approach for WFS [SKR03].

Overall, it seems that there is a requirement for the experimental validation of how different algorithms perform when used in different speaker array designs in conjunction with interactive settings. This is difficult in practice as most loudspeaker systems are already installed in fixed locations. A modular platform can give the possibility to realize different set-ups quickly, in order to perform measurements as well as perceptual experiments, in particular related to interaction. In order to come up with our target modular system, an iterative process of designing and prototyping was followed [Bux07]. Prototyping helps to understand how the final product will look like or work. The nature of design is to create ideas and explore different approaches to meet the given requirements. This leads to an expansion of possible concepts where not all of them can be used in a final product and need to be narrowed down again. Prototyping is a contracting process, it shows if ideas are possible to realize. Before we proceed with the system presentation, we present some initial specifications/requirements that based on the literature review were set as follows:

– Class D amplification
– 48 channels / speakers - resulting from the plan to create a square pyramid shape with 12 speakers on each side
– Modularity to allow different array shapes

- Low power consumption, small size, ease of configuration and transportation
- Good audio quality (Distortion, Frequency Response)

# 3 The Modular System

In this section the modular system is described part by part. As it mainly consists of amplifiers and speakers, some basics about class D amplification are given. Later the design of the printed circuit board is described and tested. At last the design of the speaker enclosure and the mechanical components such as mounting platform and rack construction is presented.

## 3.1 Class D Amplification Basics

Class D Amplifiers are Pulse Modulation Amplifiers (PMA). That means that the power stage of the amplifier uses pulse width modulation (PWM) techniques. The basic structure of a class D amplifier is shown in Fig.4.

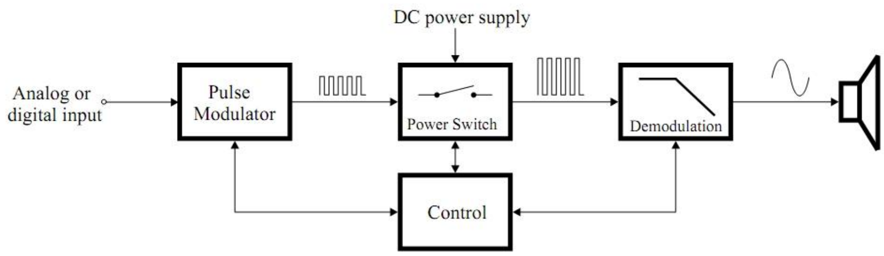



Figure 4: General PMA with Class D Power Stage [Nie98]

The input to the amplifier has to be converted to a pulse modulated signal. A PWM signal is usually generated by comparing the input signal with a triangle waveform as shown in Fig.5 and Fig.6. The triangle wave defines both the switching frequency and input amplitude for full modulation. The switching frequency of the output FETs must be higher than that of the maximum input frequency. Following Nyquist theorem, we need at least twice that frequency, but low distortion designs use higher factors (typically 5 to 50). The reference triangle signal amplitude influences the dynamic range of the amplifier system as it sets the maximum for the input voltage. The lower the threshold, the narrower the dynamic range is.

Class-D amplifiers can also accept digital input. In such cases the digital signal has to be supplied in the appropriate format. Pulse density modulated (PDM) bit stream is a widely used encoding for sending and receiving serial audio streams. The availability of a digital signal allows the integration of digital audio processing tasks, such as volume control and equalization, into the amplifier. The part where these tasks are performed is called the "modulator", at its end a PWM signal is created to feed the power stage. Digital input Class D amplifiers are used in Mobile Phones, PDAs, Portable Multimedia Players, Notebooks, etc. To use such a technique in this project, an easy way to connect 48 speakers to the USB or Firewire port of a personal computer had to be devised. The effort to build a USB or firewire interface requires excessive driver software programming and for this reason was not realized in this project. Instead, analog input Class D

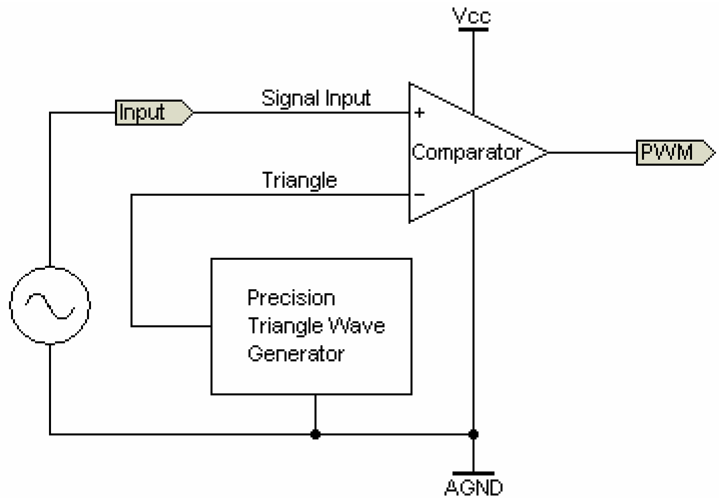



Figure 5: PWM generation with comperator [Mor05]

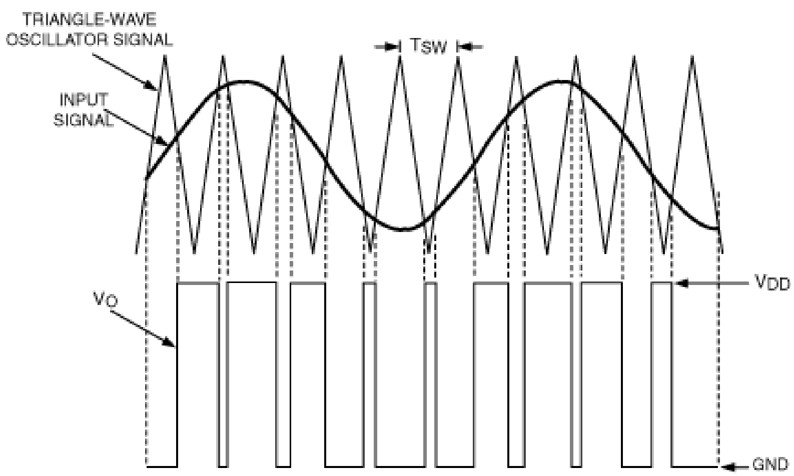



Figure 6: PWM generation: input signal, triangle reference signal with signal period $T_{SW}$ , corresponding PWM signal $V_O$, $V_{DD}$ is the supply voltage [Max07]

amplifiers were used, that were receiving input from two RME Fireface 800 interfaces and two RME M-16 D/A converters are used, that can provide the required number of 48 channels, using the 8 analog outputs and 2 times 8 ADAT outputs of each Fireface 800. The direct utilization of an ADAT port of a sound card was also considered but rejected because digital input class D amplifiers lack of support for ADAT interfaces. The creation of an interface to Digital Input Class-D Amplifiers, using a firewire or USB output, may be considered in future projects.

The main differences between class A and D amplification are explained here. Analog class A amplifiers typically use transistors in linear mode as output devices to create an output voltage that is a scaled copy of the input voltage. In this case the output devices are continuously conducting for the entire sinusoidal cycle. That means that there is a physical connection between input and load all the time. The transfer function of class A amplifiers is linear in a wide frequency range and therefore suitable for good audio quality and low distortion. The problem of this design is that power is dissipated because a large DC bias current flows over the resistor R (see Fig.7) without being delivered to the speaker. Therefore, class A amplifiers have a typical efficiency of about 20 to 25%. There have been improvements to this kind of amplifiers. Namely class B, class AB and some other approaches that are capable of gaining better efficiency up to 50% but the improvement in power dissipation mostly comes along with a degradation in linearity.

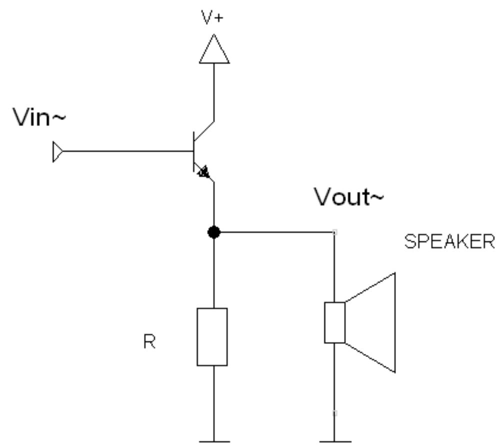



Figure 7: Simplest form of class A linear output stage

Instead of conducting through 360 degrees of the input signal cycle, a PMA amplifier switches the signal ON and OFF to full power or zero with a high switching frequency. The switches are typically MOSFET transistors and the fact that there is no permanent connection to the output load results in less power dissipation. Amplification happens when the power switch converts the incoming small-signal pulse modulated waveform to power levels. This conversion is the process of amplification itself and happens inside the so called power stage. The power delivered to the speaker is dependent on the relation between the time the switch is On or Off. The power stage consists of two MOSFET transistor in the form of a half H-bridge and a switch control logic (also referred to as gate driver). That means that one FET is a high side switch (pulls output to high level) and one is a low side switch (pulls output to ground level). Both switches must never be active at the same time because the supply voltage would then be short circuited to ground. The switch control logic ensures that the switches are activated in a correct manner (See Fig.8).

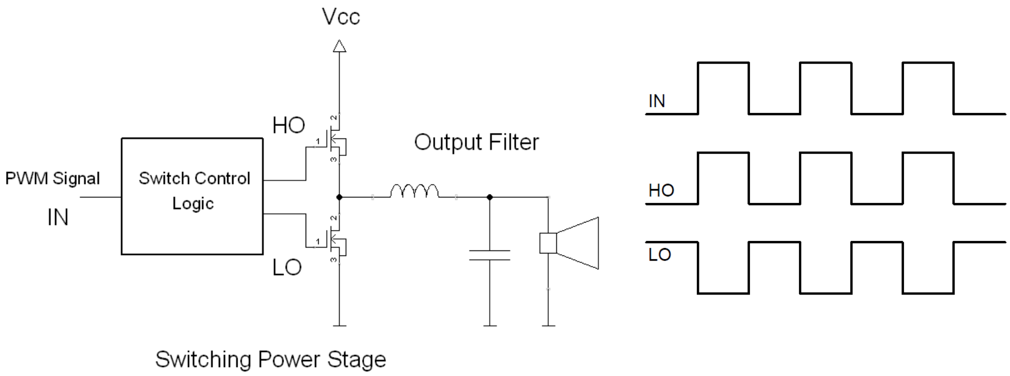



Figure 8: Half H-bridge layout (left) and Timings of the high side (HO) and low side switch (LO) according to the given input signal IN (right)

After passing the power stage the signal is still pulse modulated, and has to be demodulated by low-pass filtering. This filter is also called reconstruction filter, because its purpose is to demodulate the PWM signal to a common analog audio signal. In addition to demodulation the low-pass filter avoids high frequency disturbances originating from the high switching frequency. The filter is designed as a 2$^{nd}$ order Butterworth filter, it will be further described in section 3.2. The output after demodulation equals the average value $\bar{y}$ calculated as follows. If we consider a pulse waveform $y(t)$ with a low value $y_{min}$, a high value $y_{max}$ and a duty cycle $D = \frac{\tau}{T}$ where $\tau$ is the duration of the function value at $y_{max}$ and $T$ is the period of the function (See Fig.9 for explanation), the average value of the waveform is given by:

$$\bar{y} = \frac{1}{T} \int_0^T y(t)dt \tag{4}$$

As $y(t)$ is a pulse wave, its value is $y_{max}$ for $0 < t < D \cdot T$ and $y_{min}$ for $D \cdot T < t < T$. The above expression then becomes:

$$\begin{aligned}\bar{y} &= \frac{1}{T}\left(\int_0^{DT} y_{max}dt + \int_{DT}^{T} y_{min}dt\right)\\ &= \frac{D \cdot T \cdot y_{max} + T\,(1-D)\,y_{min}}{T}\\ &= D \cdot y_{max} + (1-D)\,y_{min}\end{aligned}$$

This latter expression can be fairly simplified in many cases where $y_{min} = 0$ as

$$\bar{y} = D \cdot y_{max} \tag{5}$$

From this, it is obvious that the average value of the signal ($\bar{y}$) is directly dependent on the duty cycle D.

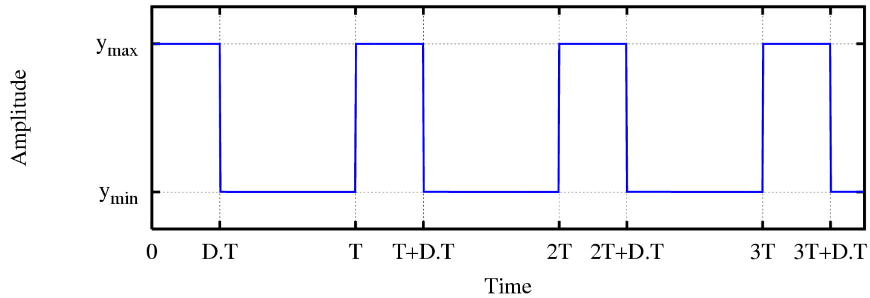



Figure 9: Pulse waveform

This design leads to very high power efficiency. The theoretical maximum efficiency of Class-D designs is 100%, and over 90% is attainable in practice. The PMA's high power efficiency translates into less power consumption for a given output power but, more important, it reduces heatsink requirements and space requirement of the IC drastically. [Mor05]

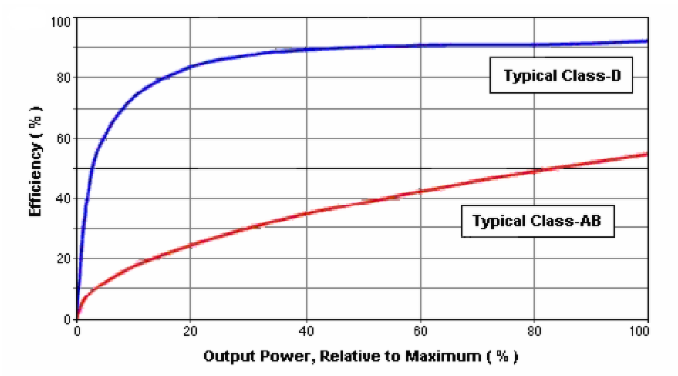



Figure 10: Power Efficiency of Class D and Class AB Amps [Mor05]

### 3.1.1 Error Sources

In theory, the power conversion within a switching power amplification stage has 100% efficiency. In practice the power stage has limited efficiency and can contribute with significant distortion and noise. The reasons for the imperfection are (See Fig.11 for localization):

1. Nonlinearity in the PWM signal: The pulse signal ideally should be a perfect square wave with vertical switching edges as shown in Fig.6. The deviations from the ideal case are caused by limited resolution (quantization) and/or jitter in timing.

2. Timing errors added by the switches, such as dead-time (when switching between low

side and high side transistor at the zero-crossing), turn-on/turn-off delay, and rise/fall-times.
3. Unwanted characteristics in the switching devices, such as finite drain/source ON-resistance $R_{DS}(on)$, finite switching speed and body diode characteristics (parasitic diodes in the structure of the semiconductor).
4. Parasitic components in the microchip, mainly expressed as capacitances, that cause ringing on transient edges of the pulse waveform.
5. Power supply voltage fluctuations due to its finite output impedance $Z_o$ and reactive power flowing through the DC bus (Bus-pumping).
6. Non-linearity of inductance and capacitance in the output low pass filter, DC-Resistance (DCR).

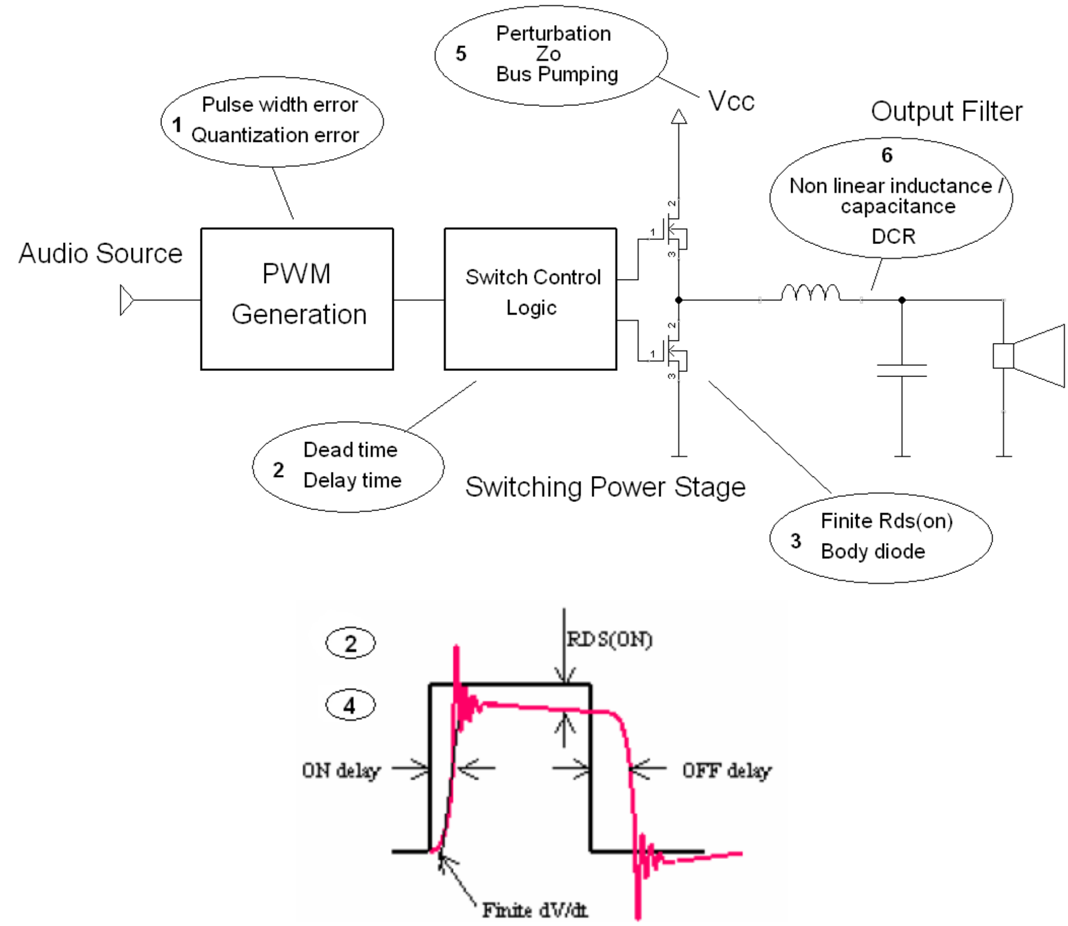



Figure 11: Location of error sources in the circuit layout (top) and error sources regarding finite switching speed and finite $R_{DS}(on)$ of the FET switches (bottom) [HA05]

## 3.2 Amplifier Circuit

After browsing a number of amplifiers available on the market we narrowed down the candidates to the Texas Instruments TPA3122 and the Analog Devices SSM2305. This devices were chosen by evaluating the following parameters.

- *Output power:* A reasonable power output to drive speakers with an SPL sufficient for small room reproduction
- *Supply voltage:* To allow portable use the voltage must not exceed 12V. The requirement for multiple supply voltages has to be avoided.
- *External components:* How many external components are needed. This affects the size of the circuit board.
- *Complexity of prototyping:* Chip Packaging, soldering options etc.

The TPA3122 comes in a Dual in-line package (DIP or DIL) package which makes it easy to use on breadboards or hole matrix boards. The SSM2305 is shipped as MSOP8 (Mini Small Outline Package) or LFCSP(Lead Frame Chip Scale Package) with a outer dimension of 3mm. Because soldering SMD (Surface Mount Device) chips is problematic without having special equipment a ready made evaluation board was purchase for testing this amplifier chip. Building the TPA3122 prototype on a breadboard led to instability because the quality of the contacts was not sufficient. Soldering the components on a hole matrix board was more expedient. Fig.12 shows an early prototype of the Texas Instruments TPA3122 on a hole matrix board (left) and the Analog Devices SSM2305 evaluation board (right).

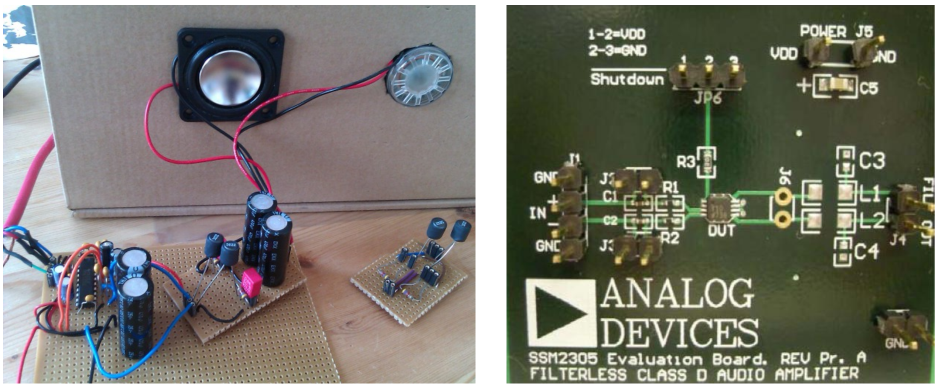



Figure 12: Texas Instruments TPA3122 matrix hole board prototype (left) and Analog Devices SSM2305 evaluation board (right)

### 3.2.1 TPA 3122 Peripheral Components

The Texas Instruments TPA3122 amplifier needs some peripheral components to work properly in combination with the Peerless PLS-P830983 $4\Omega$ speaker. These are input capacitors, power supply decoupling, bootstrap capacitors, gate voltage clamp capacitors, bypass capacitors and output filter components. These components can be changed to modify the frequency response, power supply rejection ratio and optimize the resulting total harmonic distortion.

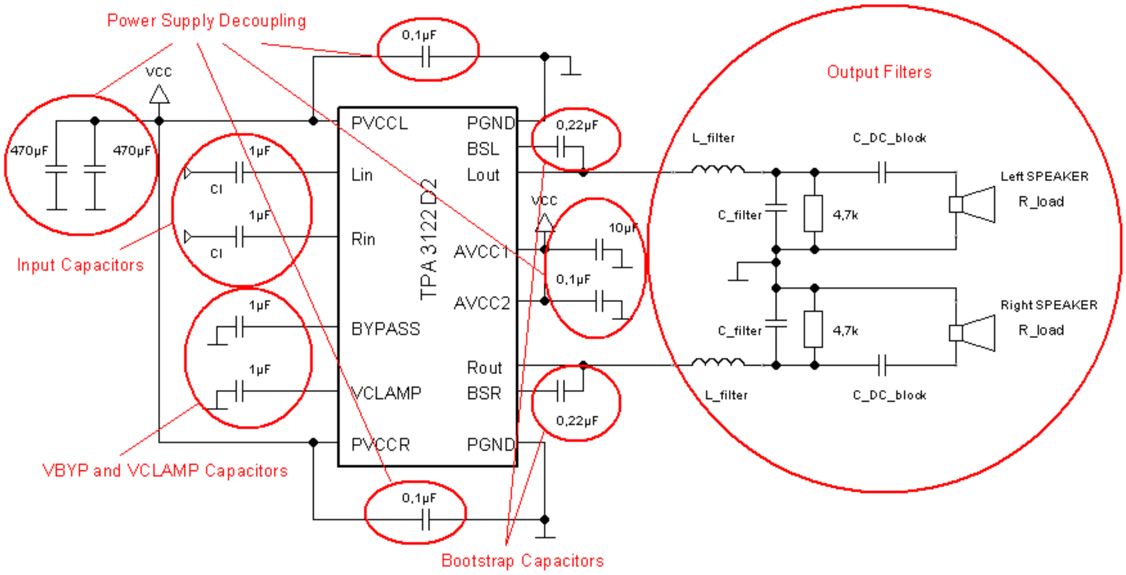



Figure 13: Amplifier with minimum necessary components and single ended filter configuration [Ins07]

*Input Capacitor* $C_I$: This capacitor is required to allow the amplifier to add an optimal DC bias to the input signal for optimum operation. The value of $C_I$ is important, as it directly affects the bass (low-frequency) performance of the circuit since it forms a high pass filter together with the input resistance of the amplifier IC, where $f_c$ represents the $-3dB$ cutoff frequency and $Z_I$ the input resistance (See equation 6).

$$f_c = \frac{1}{2\pi Z_I C_I} \tag{6}$$

$$C_I = \frac{1}{2\pi Z_I f_c} \tag{7}$$

$Z_I$ depends on the gain setting of the amp which in this case is $20dB$. The settings are accomplished by connecting Pins *Gain 0* and *Gain 1* to ground or positive supply voltage (See Table 3 for configuration details) and lead in this case to a typical input resistance $Z_I$ value of $60k$. The value for $C_I$ was chosen $1\mu F$ as suggested in the Texas Instruments evaluation board user guide. This value gives us a cutoff frequency of around $2,6Hz$ what is sufficient to block DC.

*Power Supply Decoupling Cs*: As described in the datasheet of the DPA 3122 some decoupling capacities are needed for the power supply to ensure that the output total harmonic distortion (THD) is as low as possible. Power supply decoupling ensures that disturbances of the supply voltage are smoothed. It also prevents oscillations for long lead lengths between the amplifier and the speaker.

| Name | Amount | Capacity | Location |
|---|---|---|---|
| Input Capacitor $C_I$ | 2 | $1\mu F$ | at each input |
| Power Supply Decoupling | 2 | $470\mu F$ | PVCCR & PVCCL |
| | 2 | $0,1\mu F$ | PVCCR & PVCCL |
| | 1 | $0,1\mu F$ | AVCC |
| | 1 | $10\mu F$ | AVCC |
| Bootstrap Capacitors | 2 | $0,22\mu F$ | BSR<->Rout, BSL<->Lout |
| VCLAMP Capacitor | 1 | $1\mu F$ | VCLAMP |
| VBYP Capacitor | 1 | $1\mu F$(same as $C_I$) | VBYP |
| Filter Capacitance $C_{Filter}$ | 2 | $680nF$ | Output Filter |
| Filter Inductance $L_{Filter}$ | 2 | $22\mu H$ | Output Filter |
| Discharge Resistor | 2 | $4,7k\Omega$ | Output Filter |

Table 2: List of peripheral components

| Gain 1 | Gain 0 | Amplifier Gain $[dB]$ Typical | Input Impedance $[k\Omega]$ Typical |
|---|---|---|---|
| 0 | 0 | 20 | 60 |
| 0 | 1 | 26 | 30 |
| 1 | 0 | 32 | 15 |
| 1 | 1 | 36 | 9 |

Table 3: Amplifier gain settings

*Bootstrap Capacitors* : The half H-bridge output stages use only NMOS transistors instead of NMOS and PMOS transistor as in full bridge designs. Therefore, they require bootstrap capacitors for the high side of each output to turn on correctly. The component values were taken from the datasheet.

*VCLAMP Capacitor*: To ensure that the maximum gate-to-source voltage $U_{GS}$ for the NMOS output transistors is not exceeded, one internal regulator clamps the gate voltage utilizing a Zener-diode and a resistor. See figure 14.

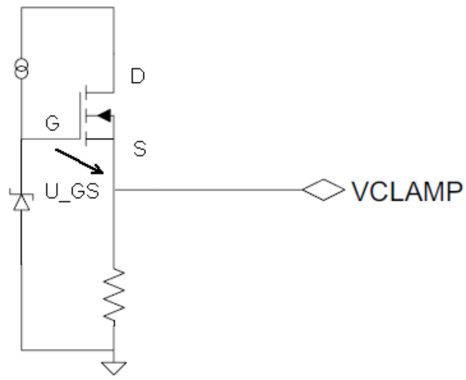



Figure 14: Limitation of $U_{GS}$ utilizing a Zener diode and a resistor

*VBYP Capacitor*: The internal bias generator (VBYP) nominally provides a 1.25-V internal bias for the preamplifier stages. The external input capacitors and this internal reference allow the inputs to be biased within the optimal common-mode range of the input pre-amplifiers. The selection of the capacitor value on the VBYP terminal is critical for achieving the best device performance. During power up or recovery from the shutdown state, the VBYP capacitor determines the rate at which the amplifier starts up. When the voltage on the VBYP capacitor equals VBYP, the device starts a timer. When this timer completes, the outputs start switching. A secondary function of the VBYP capacitor is to filter high-frequency noise on the internal 1.25-V bias generator. For the best power-up and shutdown pop performance, the VBYP capacitor should be greater than or equal to the input capacitors.

*Output Filter*: For Stereo configuration a single ended filter configuration is used (See Fig.15). The DC blocking capacitor $C_{DC}$ forms a high pass filter with the speaker impedance.

$$f_c = \frac{1}{2\pi C_{DC} Z_{load}} \tag{8}$$

With a $C_{DC}$ value of $470\mu F$ and the Speaker DC impedance of $4\Omega$ the resulting cutoff frequency is $84,65Hz$.

The Reconstruction filter itself is a 2nd order Butterworth filter formed by $C_{filter}$, $L_{filter}$ and $R_{load}$. The cutoff frequency $f_c$ is recommended to be at $40kHz$ by Texas Instruments. The filter components for a speaker with DC resistance $R_{load}$ are calculated as follows:

$$C_{filter} = \frac{1}{2\pi f_c \cdot R_{load} \cdot \sqrt{2}} \qquad L_{filter} = \frac{R_{load} \cdot \sqrt{2}}{2\pi f_c} \tag{9}$$

For a speaker with a DC resistance of $4\Omega$ the resulting component values are:

$$C_{filter} = 680nF \qquad L_{filter} = 22\mu H$$

In addition do the filter capacitance and inductance there is a $4,7k\Omega$ resistor placed parallel to $C_{filter}$ to allow discharging when the device is not operating. There is a possibility to build a Bridge Tied Load (BTL) filter (See Fig.15) with both output channels of each amp to get a maximum output power of 45W if needed. For this project the maximum gained 15W were sufficient because of the low power of the speakers.

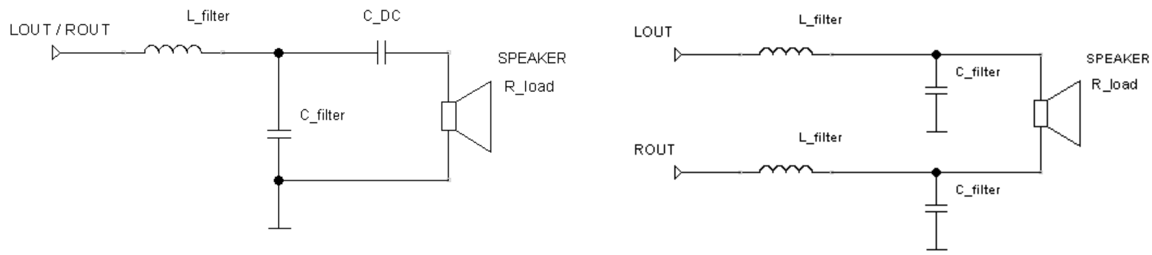



Figure 15: Single ended filter (left) and BTL filter configuration (right)

## 3.3 Testing

### 3.3.1 Frequency Response of the Amplifier

To test the quality of the amplifier more precisely the prototype was inspected with the Audioprecision measurement tool. Frequency responses (Fig.16) and total harmonic distortion (THD) (Fig.17) of the chip on a hole matrix board were measured.

The different frequency responses in Fig.16 originate in different modes of the amplifier. The TPA3122 can be used either in single ended mode or bridge tied load (BTL) mode. In BTL mode the two outputs are combined to a full H bridge circuit and thus can deliver more power to the speaker. Using the bridge tied load mode the power of both output stages is combined. In this mode the amplifier reaches a maximum power output of 45W. Especially at low frequencies the higher power output leads to an improvement of linearity of the frequency response. To be able to use both available channels each chip offers it has to be operated in single ended mode.

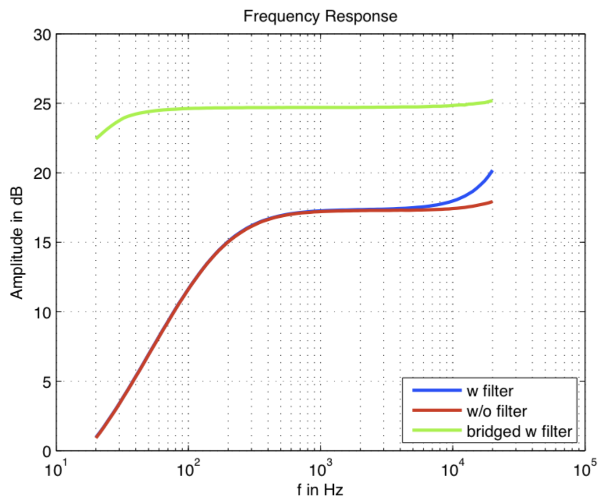



Figure 16: Frequency response of the breadboard prototype with one TPA3122 amplifier, 12V supply voltage and 774,5mV RMS input voltage

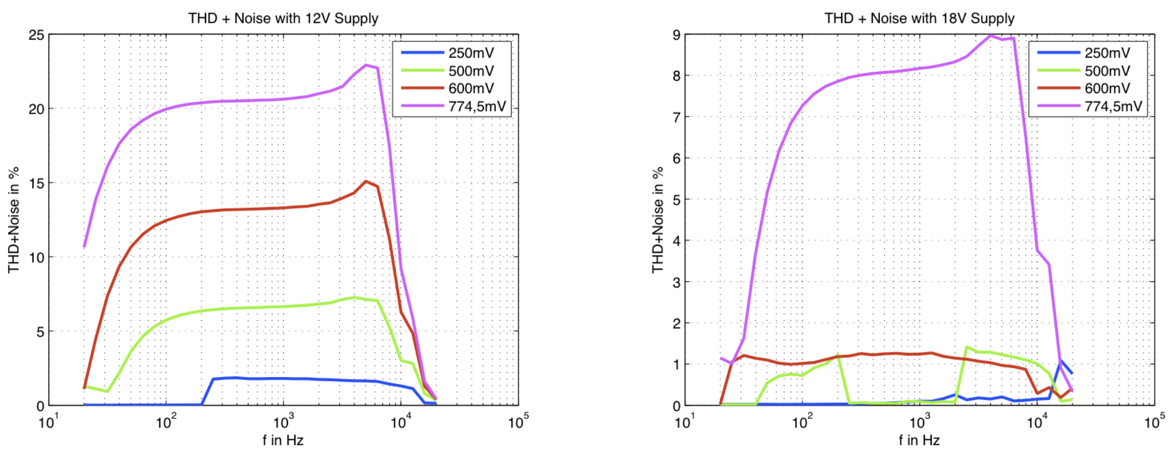



Figure 17: THD + Noise with 12V (left) supply and 18V (right)

Fig.17 shows the THD + Noise ratings against frequency. The voltage ratings in the legend describe the input level. The TPA3122 does not offer volume control, so the output level can only be controlled by the input level. 774,5mV equals 0dBu and represents full line level. As you can see in these figures the THD is highly dependent on the supply voltage. The supply voltage represents the highest voltage at the output Stages. With higher voltages the MOSFET transistors get to saturation at higher output power. This effect is apparent up till the maximum rated supply voltage of 27 volts. To get the best possible result regarding to audio quality the voltage should be chosen as high as possible.

### 3.3.2 Quantitative Initial Test

In the very beginning of the project both amplifiers were tested together with different speakers to find out the how they perform with speakers of variable size. The tests were performed with the sine sweep method to get qualitative results [Far00]. The speakers were mounted in a cardboard wall to avoid the acoustical short circuit. For frequency response results see Fig. 18

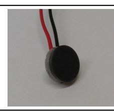
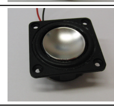
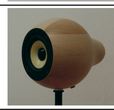
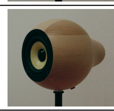
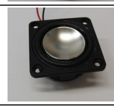
| Nr. | Amp | Name | Membrane size | Impedance | |
|---|---|---|---|---|---|
| 1 | SSM2305 | Miniature speaker | 8mm | 8 Ω |  |
| 2 | | Knowles subminiature spk. | 1mm | 150 Ω |  |
| 3 | | Transparent miniature speaker | 24mm | 56 Ω |  |
| 4 | | Tang Band W1-1070SE 1 | 1 ” | 8 Ω |  |
| 5 | | Vifa 10 BGS Drop Speaker | 3,5 ” | 8 Ω |  |
| 6 | TPA3122 | Vifa 10 BGS Drop Speaker | 3,5 ” | 8 Ω |  |
| 7 | | Transparent miniature speaker | 24mm | 56 Ω |  |
| 8 | | Tang Band W1-1070SE 1 | 1 ” | 8 Ω |  |

Table 4: Tested amplifier + speaker combinations

The insufficient bass playback of the small speakers and the huge non-linearities in the frequency responses lead to the decision to use bigger speakers. The speaker of choice

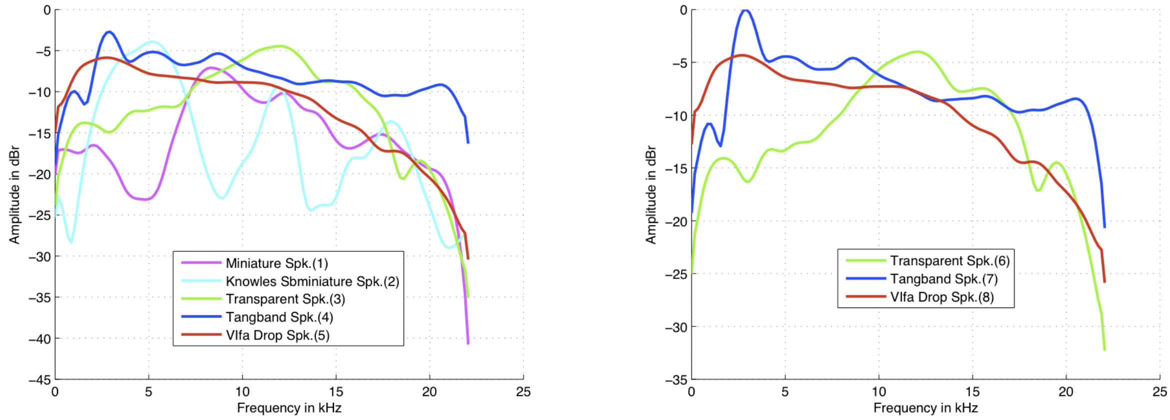
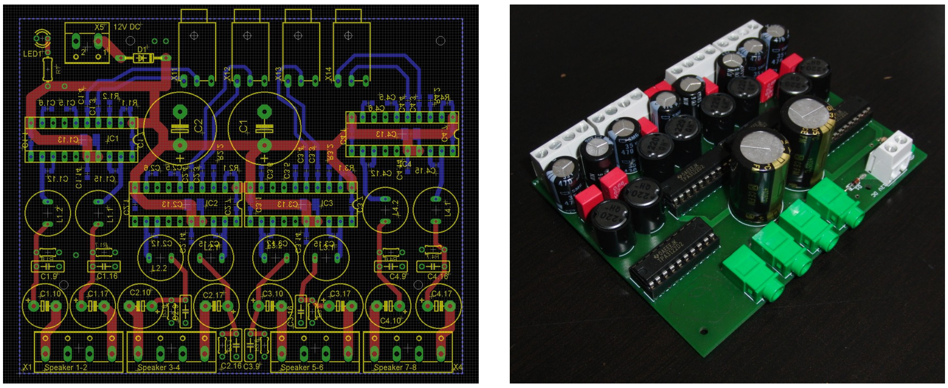



Figure 18: SSM2305 (left) and TPA3122 (right) with different speakers, the dBr scale is related to the maximum output levels in all measurements

was the Peerless model PLS-P830983. This speaker has very good characteristics and reproduces very low frequencies for its small size of 2". The frequency response of the Peerless Speaker in combination with the final board design can be seen in Fig.25.

## 3.4 Printed Circuit Board

After the prototyping phase a Printed Circuit Board (PCB) had to be made. Cadsoft's Software EAGLE was used to create the board layout (See Appendix B for the circuit diagram). This software is free but limited to a board size of 100x80mm. This specific board size is also known as half eurocard format. The Texas Instruments evaluation board circuit was modified, so that one board would hold 4 TPA3122 amplifiers giving a total number of 8 channels per board. Each board is equipped with 4 3,5mm headphone jacks at the input, connecting 2 unbalanced channels each. At the output screw terminals are used to connect the speaker cables to the board. For protection against power supply polarity failures a diode is positioned at the power screw terminal. A LED is placed on the board too, to provide information about the power state. The board was then manufactured by Conrad PCB service. 6 Boards were equipped with components to get a total number of 48 channels.




(a) EAGLE layout of the PCB (b) Assembled PCB

Figure 19: Printed Circuit Board

After assembling the board the circuit was measured again with the Audio Precision tool. Frequency response (Fig.20 left) and total harmonic distortion (Fig.20 right) and their behavior with different supply voltages were investigated.

According to the the frequency response measurements in Fig.20 left, there was almost no influence of the supply voltage. THD+N was measured at variable supply voltages and input signal levels to identify the optimum operation range of the amplifier. In general, higher supply voltages result in lower distortion. Input voltages above full line level, respectively $774,5mV$ RMS, should be avoided as they yield THD+N values over 1%. The implementation on a PCB overall improved THD+N in comparison to the hole matrix board prototype, (compare Fig.17 and Fig.21 right).

When testing the amplifiers circuit in a real use scenario as part of an installation at the "ESC-Labor", an audible noise floor was observed, even when there was no signal input. This was recorded and analyzed with the Audio Precision tool and is presented in Fig.21 left. The analysis shows that tonal disturbances occur when supply voltages above 15V are used. One possible reason for the peaks in the noisefloor was considered to be the interference between the multiple amplifiers on the PCB. To prove this, a board was equipped with only one amplifier and tested again. The result is represented in Fig.21 right. A very similar to the multi chip PCB noise floor was observed, ruling out this possibility. Nevertheless, the higher frequency noise is summed up when multiple chips are placed on one board. Also the power supply was checked for being the cause of the noise but was confirmed to not be the source of it. Using different supplies did not lead to any improvement. The reason for the peaks in the noisefloor could not be resolved. It seems that this kind of class D amplifiers do produce a noise floor.

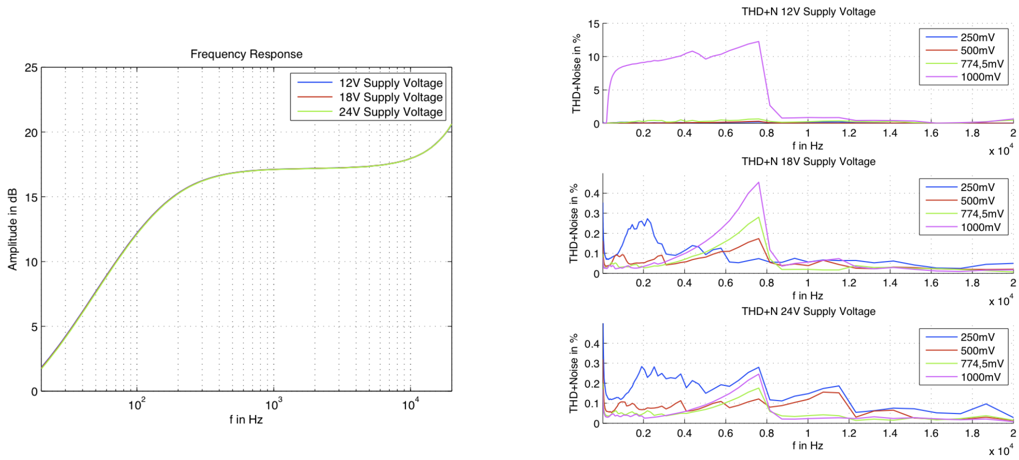



Figure 20: Frequency responses with different supply voltages (left) and THD + Noise measurement with different supply voltages (right)

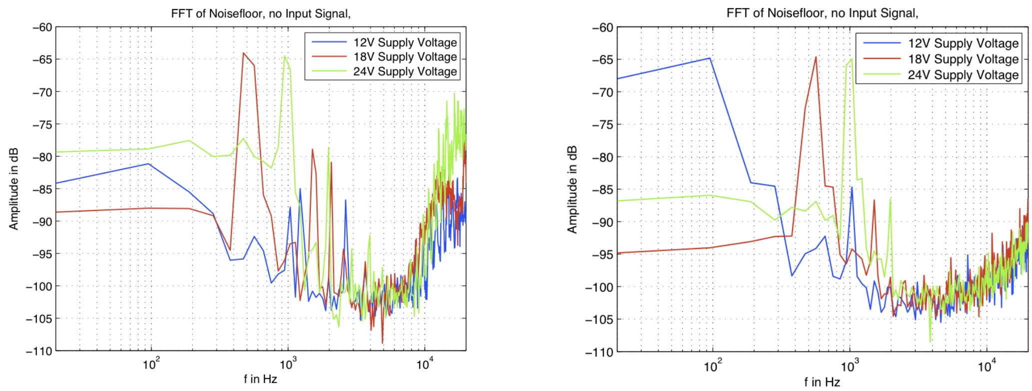



Figure 21: FFT of the noisefloor with different supply voltages on single chip PCB

As an attempt so improve the noise behavior some capacitors that were mentioned in the data sheet to be critical for optimal THD and power supply rejection were changed to better quality X7R replacing Y5V components that were initially used to lower the construction cost. The X7R capacitors are better regarding to capacitance tolerance. While Y5V capacitors have a tolerance of $-20\%$ to $+80\%$ of the nominal value, X7R capacitors have a range of $\pm 15\%$. Namely the changed input capacitors were: $C_I$, the VCLAMP and the VBYP capacitors (See Table 2 and Fig. 13 for placement and component values). Replacing this components changed the noise behavior but did not improve it. The subjective results were the same, the measurements even revealed a degradation in quality. The THD +N and Noisefloor levels were higher than before. The results can be seen in Fig.22 left and Fig.22 right.

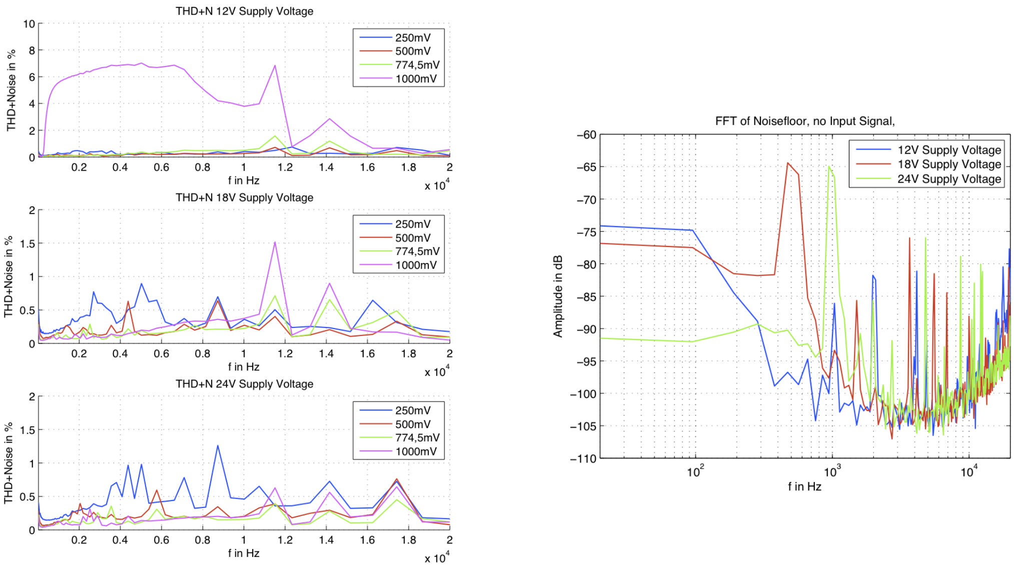
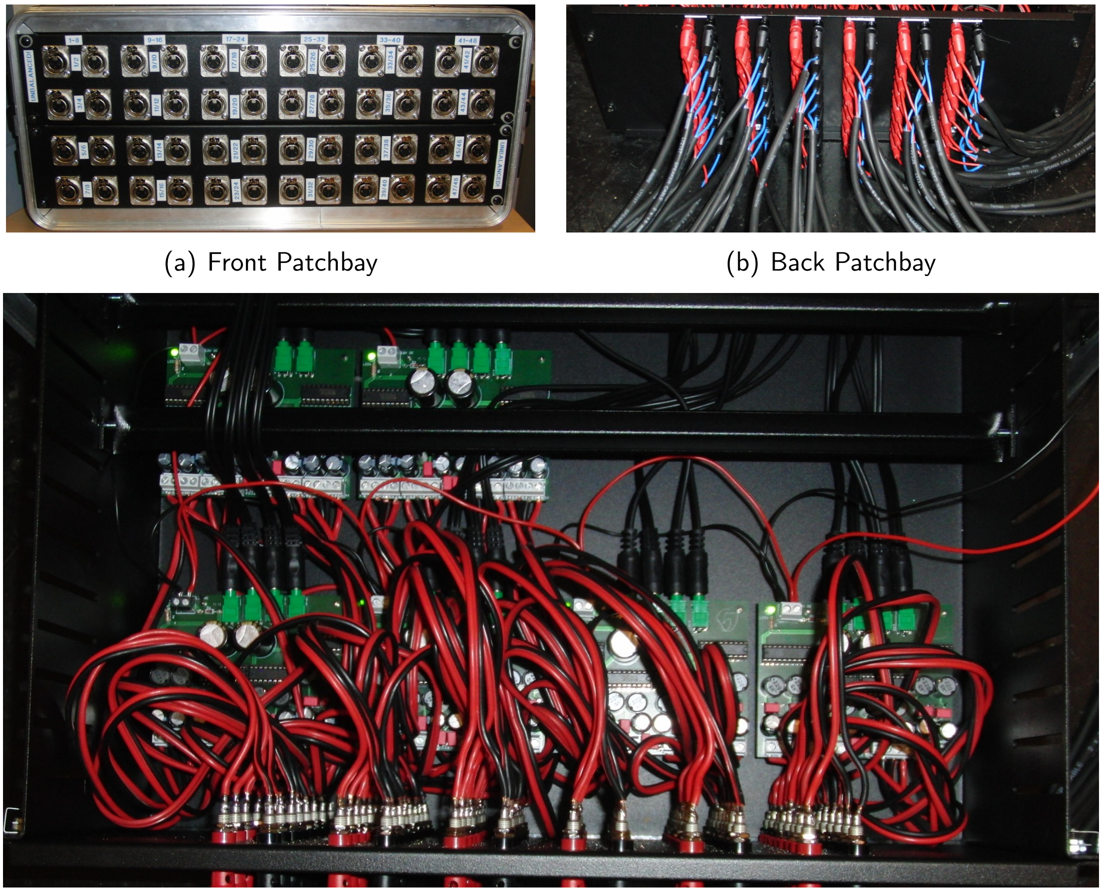



Figure 22: THD + Noise measurement (left) and FFT of the noisefloor with different supply voltages on multi chip PCB (right) and changed capacitors

## 3.5 Amplifier Rack

To make the system portable and protect the electronics a housing was designed. It holds a power supply and 6 amplifier boards and offers patch bays for input and output connections. On the input side 48 unbalanced XLR connectors can be addressed, the outputs are realized as Banana Jacks. See Fig.23 in the appendix.



(a) Front Patchbay

(b) Back Patchbay

(c) Inside

Figure 23: Amplifier Rack

## 3.6 Design of the modular speaker array

### 3.6.1 Speaker enclosure

Speaker enclosures are essential for electrodynamic transducers to avoid the acoustic short circuit and maximize the sound power emitted to the air. Closed box and vented box are the two basic approaches for realizing speaker boxes.

For the calculation of speaker enclosures Thiele Small parameters of the transducer are required.

| Parameter | Value | Description |
|---|---|---|
| $f_s$ | $176Hz$ | Resonant frequency of the driver |
| $V_{as}$ | $0,16l$ | Equivalent Compliance Volume |
| $R_e$ | $3,6\Omega$ | DC resistance of the voice coil |
| $Q$ | - | A unitless measurement, characterizing the electrical, mechanical or combined damping of the driver at $f_s$ |
| $Q_{ms}$ | $5,6$ | Mechanical $Q$ of the driver |
| $Q_{es}$ | $0,81$ | Electrical $Q$ of the driver |
| $Q_{ts}$ | to be calculated | Combined electrical and mechanical $Q$ of the driver |
| $Q_{tc}$ | to be calculated | Total $Q$ of the driver with enclosure |

Table 5: Thiele Small parameters

*Closed-box:* The first step is to calculate the combined Qts. Usually the DC resistance of the speaker, cable and amplifier has to be included in this calculation but was neglected for simplicity.

$$Qts = \frac{Q_{ms} \cdot Q_{es}}{Q_{ms} + Q_{es}} = \frac{5,6 \cdot 0,81}{5,6 + 0,81} = 0,707 \tag{10}$$

Following equation 11 the net volume for an enclosure according to total $Q_{ts}$ and the variable $Q_{tc}$ can be calculated. $Q_{tc}$ influences the behavior of the frequency response at the cutoff frequency. Higher $Q_{tc}$ values lead to steeper drops below the cutoff frequency but cause a peak in the response (See Fig. 24).

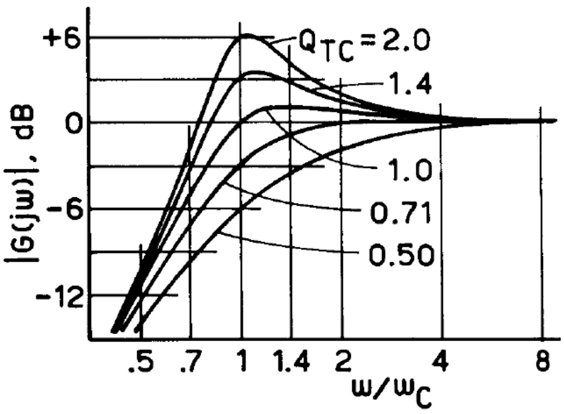



Figure 24: Normalized amplitude vs. normalized frequency response of closed-box loudspeaker system for several values of a total system $Q$ [Sma72]

$$V_{ab} = \frac{V_{as}}{\left(\frac{Q_{tc}}{Q_{ts}}\right)^2 - 1} \tag{11}$$

The geometrical restraints allowed only a small enclosure. A nett volume of $0,2l$ was chosen. The chosen $V_{ab}$ leads to a $Q_{tc}$ of $0,95$ (Equation 12). Values between $0,8$ and

$1,0$ are recommended, where lower values are said to sound more "detailed" and higher values to sound "warmer".

$$Q_{tc} = Q_{ts} \cdot \sqrt{\frac{V_{as}}{V_{ab}} + 1} = 0,95 \tag{12}$$

*Vented-box*: The vented box approach was calculated using the software "WinISD beta" by Juha Hartikainen. The vented box supports lower frequencies by adding a Helmholz resonator at a certain frequency. This frequency is mainly influenced by the dimension of the ventilation port. The air volume inside the port has a certain mass and the remaining volume of the enclosure acts as a spring. Fig.25 top left shows the simulation of the closed compared to the vented box design. While the closed box approach drops below -3dB at 190Hz the -3dB drop can be shifted down to 125Hz with use of a ventilation port.

After building prototypes of both speaker box concepts, the amplifier and speaker ensembles were tested at the IBK Studio with ARTA acoustic software. The results of the frequency response measurement were compared to the outcome of the simulation. As you can see in Fig.25 the gain of low frequencies is not as high as predicted by the simulation (22Hz vs 65Hz gain at -3dB).

Despite the fact of better bass playback with the vented box design we came to the decision to use the closed box because the better bass was not worth the cost of the additional space requirement. See Fig.26 and Fig.33 to get an idea of the additional space requirement. In Addition to the larger space consumption another argument against the vented box was that the bass frequencies were too present and kind of blurry and inexact.

As you can see, the bass reflex concept changes the linearity of the response drastically. This is probably because of the not ideal form of the ventilation port that leads to resonances. Typically this port is of round or rectangular shape and does not have any turns. Because of the small size requirement the turns are necessary. The form of the port can be seen in Fig.26(a).

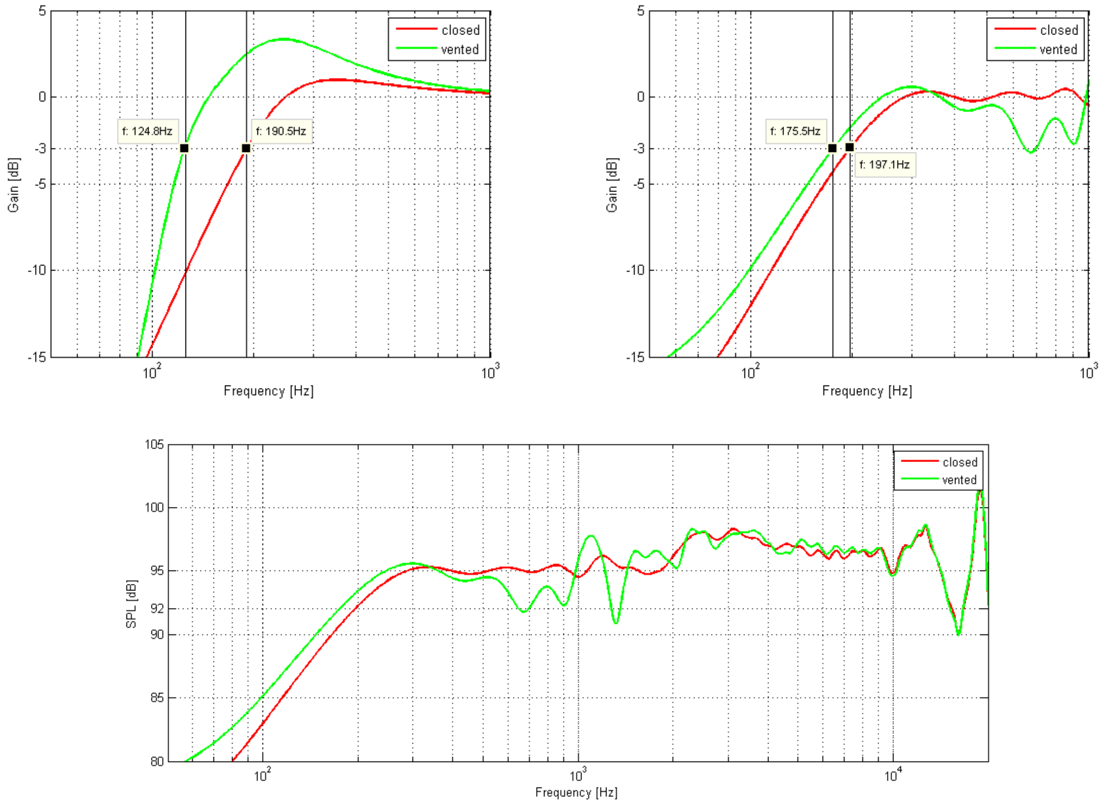
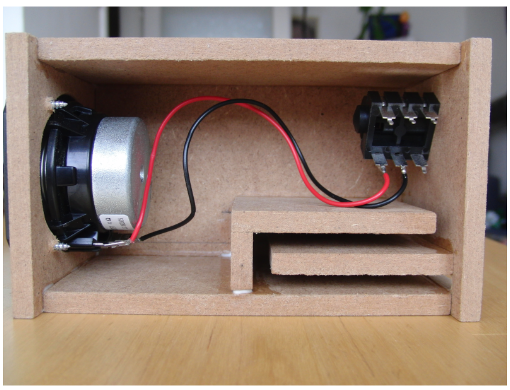
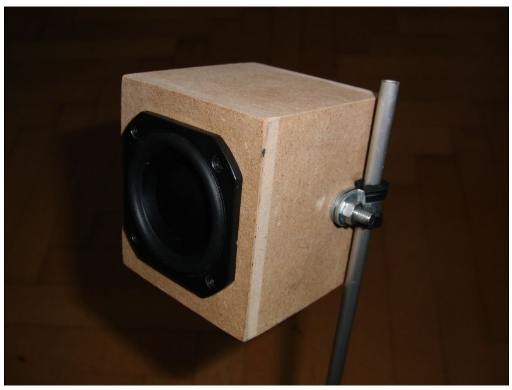



Figure 25: Comparison of closed and vented speaker box. Simulation with WinISD(top left), Measurement (top right), Measured total frequency response (bottom)



(a) Vented speaker enclosure



(b) Closed speaker enclosure

Figure 26: Prototypes of speaker enclosures

### 3.6.2 Modular Platform

The modular system was intended to provide a platform so that different loudspeaker arrangements could be constructed and tested. The arrangements could be placed on a table or the floor for that purpose. To come to the final design we followed a process of prototyping, sketching and drafting. An initial idea was to create forms using rods connected with elastic rubber parts, Fig.27. This approach was rejected because of problems with stability. Another dropped approach was to imitate a microphone stand and mount the speakers with rubber straps. See Fig.28.

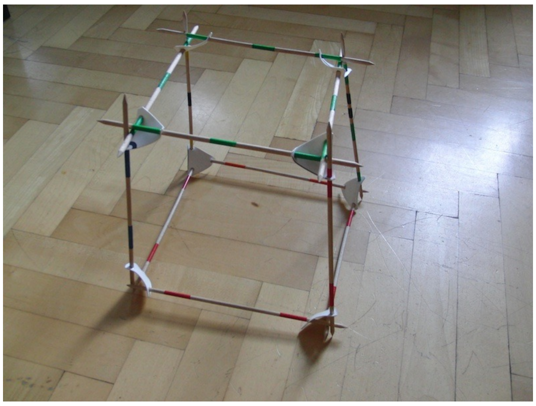
The final design is made up of a fixed platform, upon which rods can be mounted. Consequently, speakers can be mounted on the rods using DIN 3016 Type 1-8mm clamps. Every single speaker housing is mounted to a rod with a clamp on one side of the housing as seen in Fig.26(b). As the mounting hole in the speaker is placed concentric, the clamp can either be placed on the left or right side of the speaker enclosure. With help of the clamp the hight of each speaker can be adjusted as required. In this way it is possible to fulfill the requirement of modularity.

Each rod can be placed on the basis, which can have variable dimensions depending on the prototype being developed. The hole pattern on the base platform can be adapted to any shape for optimal speaker placement depending on the application. If no special hole pattern is required, a hole center to center distance of 2cm is recommended. Thus the different array shapes can vary from cubic to cylindric, even a spherical arrangement can be realized if required.

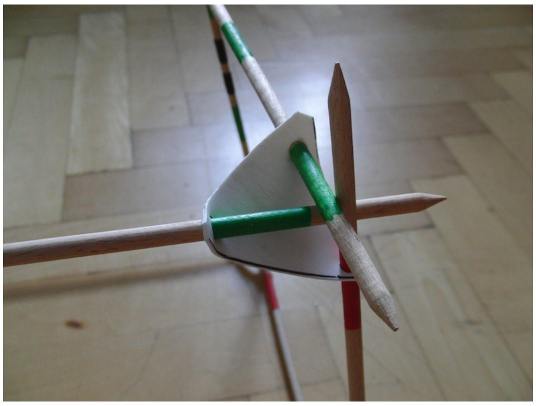
The rods used to mount the speakers are recommended to be of aluminum with an outer diameter of 6mm to ensure minimum space requirement and maximum flexibility. Because of the relatively small diameter the construction can get instable when a height of 30cm is exceeded. To avoid instability in this case a top plate has been designed to counter hold the rods. The top plate has to have the same hole pattern as the base plate to allow the rods to be fixed.

Eventually using rods of larger diameter will stabilize the construction further. This was not realized in the prototype made during the project because the clamps have to fit to the rods and we didn't want to waste already purchased clamps.

To assure the correct position of the speakers regarding to ergonomic aspects a crank angle for the plane formed by the speaker fronts had to be found. This angle should be around $20°$. Due to the used clamps this angle can be adjusted to not limit the possibilities of use scenarios.

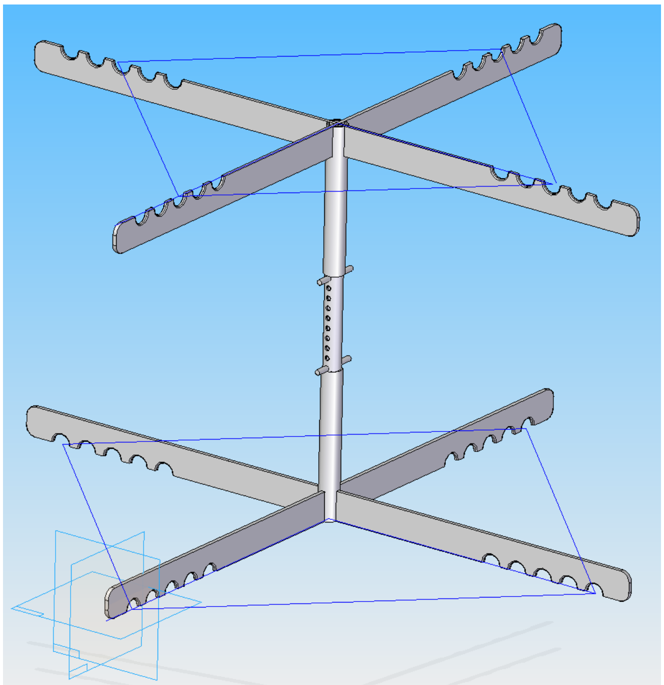


(a) Truncated pyramid



(b) Edge detail

Figure 27: Early prototype of the pyramid



(a) Microphone stand approach

(b) Mounting with rubber straps

Figure 28: Prototype based on microphone stand imitation and rubber strap mounting

# 4 Discussion

## 4.1 Application types

Two main application types have been identified, the one of sound reproduction and the one of interaction design. Within these fields, numerous designs can be developed and tested. Here, the examples are given on how someone could go forward in this direction.

### 4.1.1 Type 1: playback device

When using the platform as a playback device for art installations or in a home environment where it is essential to have a specific shape the interaction aspect can be left aside. Then it is important to think about what happens when multiple speakers play back the same signal. Superposition and comb filtering effects will occur and affect the sound quality. This can be solved in two ways, either by using beamforming techniques, or by trying to provide as uniform a reproduction into the whole space using signal processing techniques. When using beamforming, the sound signal can be made louder in a certain direction, while other directions receive less sound energy. This could perhaps be beneficial, for listening in domestic situations, however this remains to be established in usability evaluation studies. A Bessel array approach could be a more interesting solution in order to increase the sound output of the device. As an example a constellation of 5 speakers with equal level, equal polarity and equally spaced leads to a strongly frequency dependent polar magnitude response (See Fig.29 top row). With the use of the Bessel array technique the signals of each speaker is weighted and also modified in sense of phase. The weighting factors are derived from the Bessel Function of first kind and order.

$$J_n(z) = \left(\frac{z}{2}\right)^n \sum_{k=0}^{\infty} \frac{(-z^2/4)^k}{k!(n+k)!} \tag{13}$$

As an example the calculation of the weights for a five-element array is done as follows [Kee90]. An argument value of $z = 1,5$ is found to be a good choice for for the five-element array. From the results of $J_{-2<=n<=2}(1,5)$ the weights can be approximated as $+0,5 : -1 : +1 : +1 : +0,5$. The values for $n$ beyond $\pm 3$ decrease very rapidly to very small values and are truncated. The resulting polar magnitude response is shown in Fig.29 in the bottom row. As you can see clearly the polar magnitude response is improved drastically. The results in the plots are simulated for a distance of the point of measurement to the array that is 20 times the width of the array.

### 4.1.2 Type 2: Output channel for an audio interface

When going towards interface design other aspects have to be investigated, as sounds need to be located at different positions on the array and be perceived clearly enough to support interaction. For interactive interfaces 3D audio algorithms can be used to present virtual sources in different positions on the surface or in front of it. At the same

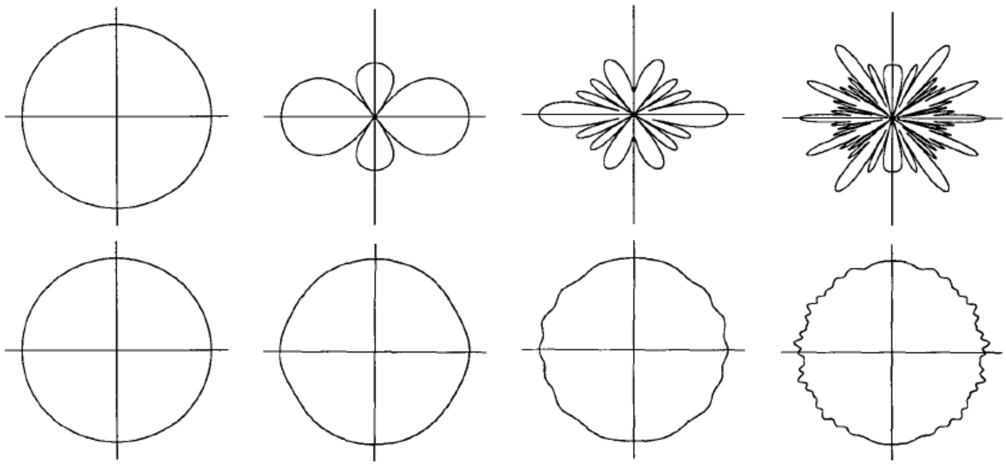


Figure 29: Polar magnitude response of a five source equal level, equal polarity and equally spaced line array for frequencies of 0,316Hz, 1Hz, 3.16Hz and 10Hz (top row) and polar magnitude response of an equally spaced five source Bessel array for the same frequencies (bottom row) [Kee90]

time to get full advantage of the array, it should be possible for the user to control it even when they are not directly facing the arrays. The performance of 3D audio algorithms on variable setups and user positioning that might be required by different products has however not been evaluated.

As mentioned in section 2.3 current spatialization algorithms suffer from the following problems. Precise localization of virtual sources only works in the sweet spot area. Depending on the algorithm this area varies in size, but is a persistent problem in interactive interface design because user movement is important. High cost and space requirement because of the demand for a large number of speakers and corresponding amplifiers are obstacles in the way of any interaction designer. These problems can be overcome for the purpose of interaction design and product development by using the developed small speaker platform. Using single speakers as independent positions for virtual sources can eliminate problems that occur when using 3D Audio algorithms. At the same time the platform provides the possibility to directly evaluate the algorithms used for single speaker reproduction.

The creation of an interactive interface for such systems poses a lot of questions that were outside the scope of this project. For example it needs to be found what kind of mappings work best for presenting the interface objects or tasks and how someone can input information and get feedback accordingly. A very promising direction for input is gesture and speech interaction.

In order to give the user power to control and manipulate an 3D auditory system with virtual non-physical objects, gesture recognition is essential because one cant physically touch virtual objects. The ability to track a person's movement and determine what gestures they ma be performing can be achieved through various tools. Besides pointing and manipulating some global gestures like *Volume Up/Down*, *Mute* as well as others that are more application specific may be necessary. Lately there is a large amount of research going on regarding video/image based tracking systems. The latest models work very well and even found a way into our homes with the "Kinect" system for Microsoft's Xbox. The IEM Cube is equipped with a Vicon Motion Tracking System [Vic] that could be used for experiments with gesture interaction. User control can also be realized by speech input. Speech recognition can be used to activate predefined procedures or allow a refinement of commands when used as text input method. The possibilities of input techniques are wide ranging. In future interaction projects different interaction modalities will be evaluated to see which ones perform better when used with different speaker array constellations.

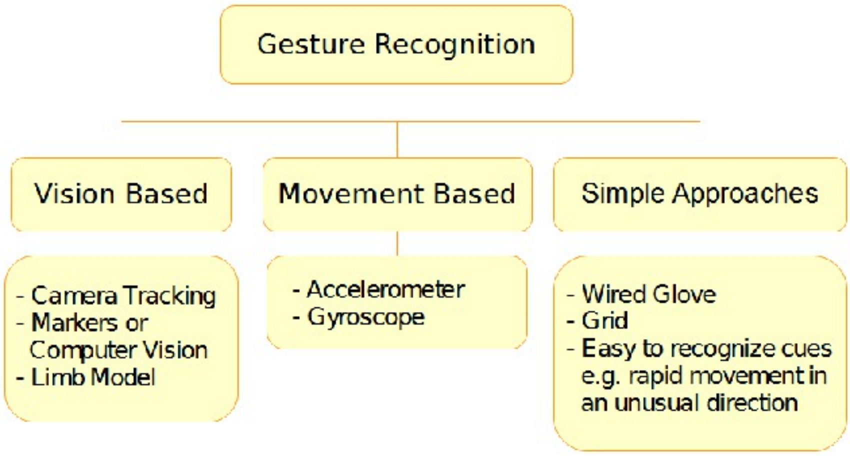



Figure 30: Gesture recognition techniques

### 4.1.3 A Design Proposal

As an initial design for an interactive surface a pyramid was created. The pyramid offers the possibility to play back sound at different directions depending on where the user is located and thus allows for a device that can be controlled from different views and from a distance, in this way integrating better to the everyday life of the user. To achieve this task however an extension of the Bessel Array technique int two dimensions would be required. Using sound spatialization, the pyramid can be made less dense as sounds can appear between the speakers. The exact setup however needs to be further examined in future work. In the final design the mounting platform is realized as an 29mm thick medium-density fiberboard with an outer dimension of 50cm featuring a 24 x 24 hole matrix with a hole to hole center distance of 20mm (See Fig.31).

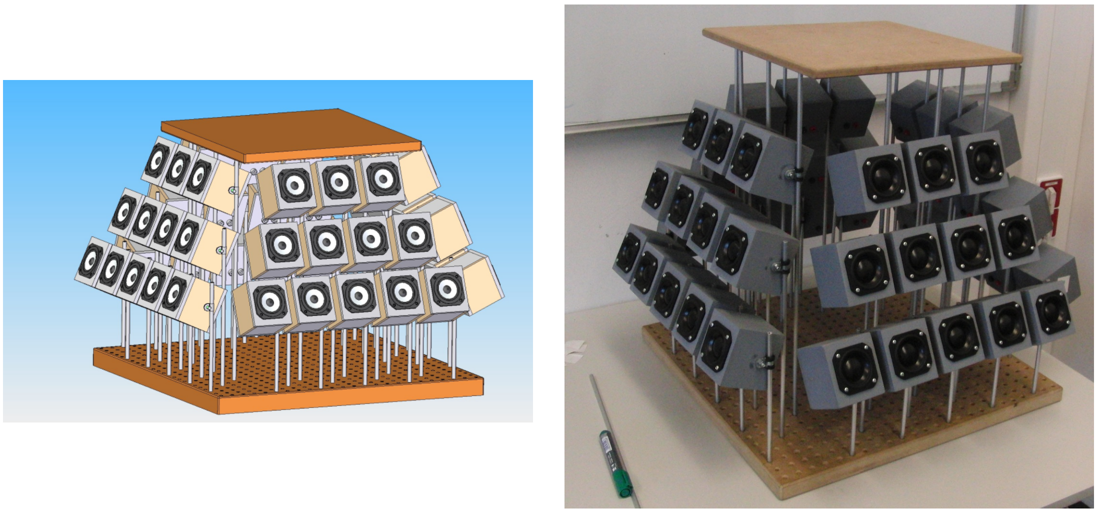


Figure 31: CAD design and finalized version of the 48 speaker pyramid approach

## 4.2 Conclusions and outlook

*Amplifiers and their optimal operational range*
To gain the best audio quality regarding THD+N ratings the amplifiers should be operated at supply voltages as high as possible. The range for the supply voltage is from 10 to 30V. However, this results in a higher noise floor. Due to the tonal noise occurring above 15V the ideal supply voltage is below that value. Also the input signal Level should not be higher than $0dBU$ respectively $774,5mVRMS$ to keep the level of harmonic distortion below 1%. There has been a few faults in the PCB layout, so some connections have to be corrected to allow the amplifier to work. In the final PCB layout files this errors have been corrected.

*PC Interface*
To improve the connection between personal computer and amplifiers an interface with firewire or USB connection could be implemented in a future project. To achieve this the use of digital input class D amplifiers is recommended.

*Speakers and how they interact with the amplifiers*
Any 4 Ohm Speakers can be used in combination with this amplifier board. An adaption of the output reconstruction filter is necessary if one wants to use different speaker impedance values. The resulting frequency response with the Peerless speakers is shown in chapter 3.6.1 Fig.25.

*Mounting platform*
The mounting platform could be improved by using more stable rods, to avoid horizontal movements of the speakers when using longer rod lengths. The top plate could then be omitted. The mounting of the speakers on the rods could be improved to be changeable more easily. Mechanical components were a big problem during this project because the availability is not as good as desired and building parts from scratch in small quantities is very expensive.

The final positioning of speakers on the platform is highly dependent on the application. Which shapes and playback techniques will work best has to be investigated by future projects. For this project the decision was made to work with multiple small speakers to get a wide range of possible applications and shapes.

Using a speaker system as proposed in this project allows multi user applications, the user is free to do any movement and the system can be adapted to the desired size and form. A modular form is sought for this project, since a large number of unknown parameters require extensive experimentation. Making it possible do quickly change the placement of speakers to test different arrangements for different applications is thus necessary.

This "development platform" enables researchers to investigate and experiment with new forms of speaker arrangements in future projects at the IEM.

# References


[ADT01] V. R. Algazi, R. O. Duda, and D. M. Thompson, "THE CIPIC HRTF DATABASE," *IEEEWorkshop on Applications of Signal Processing to Audio and Acoustics*, pp. 99–102, 2001.

[Aro08] B. Arons, "A Review of The Cocktail Party Effect," *Lecture Notes Conversational Computer Systems, MIT Media Lab*, 2008.

[BMB09] D. Beer, S. Mauer, and S. Brix, "Flat panel loudspeaker consisting of an array of miniature transducers," *Audio Engineering Society, Convention Paper 7685*, 2009.

[Bux07] W. Buxton, *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann, 2007.

[BVV93] A. J. Berkhout, D. D. Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[CF95] R. Conor and D. Furlong, "Effects of Headphone Placement on Headphone Equalisation for Binaural Reproduction," *AES Convetion Paper*, vol. 98, 1995.

[CL91] M. Cohen and L. F. Ludwig, "Multidimensional audio window management," *International Journal of Man-Machine Studies*, vol. 34, no. 3, pp. 319 – 336, 1991.

[dVSV94] D. de Vries, E. W. Start, and V. G. Valstar, "The Wave Field Synthesis Concept Applied to Sound Reinforcement: Restrictions and Solutions," *AES*, vol. 96th Conve, 1994.

[Far00] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," *Journal of the Acoustical Society of America*, 2000.

[GM99] S. Goose and C. Möller, "A 3D Audio Only Interactive Web Browser : Using Spatialization to Convey Hypermedia Document Structure," *Proceedings of the seventh ACM international conference on Multimedia*, pp. 363–371, 1999.

[HA05] J. Honda and J. Adams, "Class d audio amplifier basics, application note an-1071 class d audio amplifier basics," 2005.

[Ins07] T. Instruments, *Datasheet of TPA3122D2*, 2007.

[Kee90] B. Keele, "Effective Performance of Bessel Arrays," *Journal of the Acoustical Society of America*, vol. 38, 1990.

[LBdlH11] T. Lossius, P. Baltazar, and T. de la Hogue, "Dbap - distance-based amplitude panning," 2011.

[LBPE05] J. J. López, S. Bleda, B. Pueo, and J. Escolano, "A Sub-band approach to Wave-Field Synthesis Rendering," *Audio Engineering Society Convention Paper 6403*, vol. 118, 2005.

[Max07] Maxim Aplication Note, "Class D Amplifiers : Fundamentals of Operation and Recent Developments," 2007. [Online]. Available: http://www.maxim-ic.com/app-notes/index.mvp/id/3977

[MB04] G. Marentakis and S. A. Brewster, "A Study on Gestural Interaction with a 3D Audio Display," *Lecture Notes in Computer Science*, pp. 180–191, 2004.

[MB06] ——, "Effects of Feedback , Mobility and Index of Difficulty on Deictic Spatial Audio Target Acquisition in the Horizontal Plane," *Conference on Human Factors in Computing Systems*, pp. 359–368, 2006.

[Mor05] S. S. Moreno. (2005) Class d audio amplifiers - theory and design. [Online]. Available: http://sound.westhost.com/articles/pwm.htm

[Nie98] K. Nielsen, "Audio power amplifier techniques with energy efficient power conversion," PhD Thesis, Technical University of Denmark, 1998.

[Noi10] M. Noisternig, "Beschallungstechnik - lecture notes," 2010.

[Pul01] V. Pulkki, *Spatial sound generation and perception by amplitude panning techniques*. Helsinki University of Technology, 2001.

[Shn93] B. Shneiderman, *Sparks of Innovation in Human-Computer Interaction*. Ablex Publishing Corporation, 5 1993. [Online]. Available: http://amazon.com/o/ASIN/1567500781/

[SKR03] S. Spors, A. Kuntz, and R. Rabenstein, "An approach to listening room compensation with wave field synthesis," in *AES 24th International Conference on Multichannel Audio*, 2003, pp. 49–52.

[Sma72] R. H. Small, "Closed-box loudspeaker systems," *Journal of the Acoustical Society of America*, vol. 20, no. 10, pp. 798–808, 1972.

[SRA08] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *124th AES Convention*, 2008, pp. 17–20.

[SRDV97] E. Start, M. Roovers, and D. De Vries, "In situ measurement on a wave field synthesis system for sound enhancement," *PREPRINTS-AUDIO ENGINEERING SOCIETY*, 1997.

[The77] G. Theile, Gunther; Plenge, "Localization of lateral phantom sources," *Journal of the Acoustical Society of America*, vol. 25, no. 4, pp. 196–200, 1977.

[Vic] Vicon. Vicon motion systems. [Online]. Available: http://www.vicon.com

[VVTB95] E. Verheijen, P. Van Tol, and M. Boone, "Evaluation of loudspeaker arrays for wave field synthesis in audio reproduction," *PREPRINTS-AUDIO ENGINEERING SOCIETY*, 1995.

[WAKW93] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman, "Localization using nonindividualized head-related transfer function," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[ZZ93] M. Zollner and E. Zwicker, *Elektroakustik*. Springer-Verlag, 1993, vol. 174.

# A Drafts / Schematics / Parts

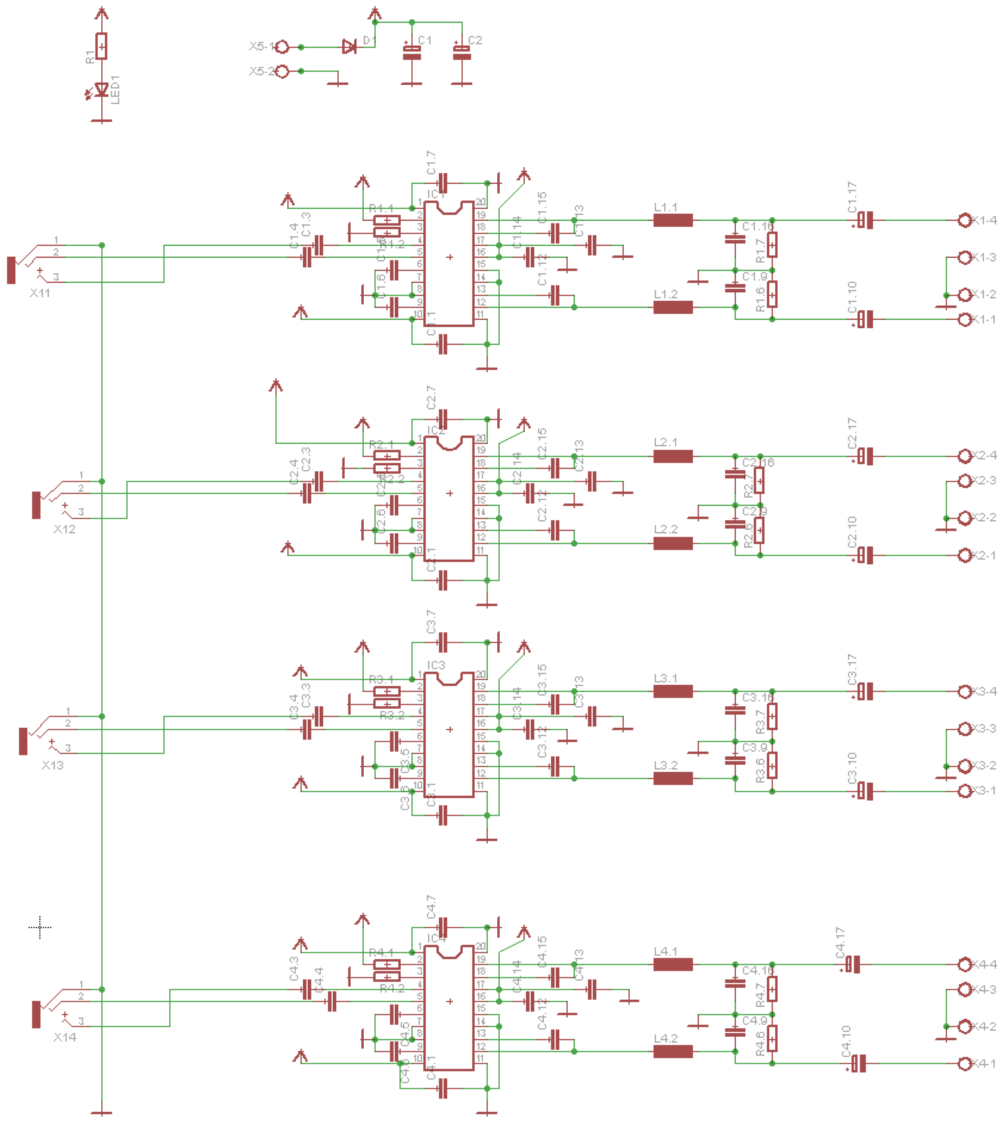


Figure 32: Board schematic

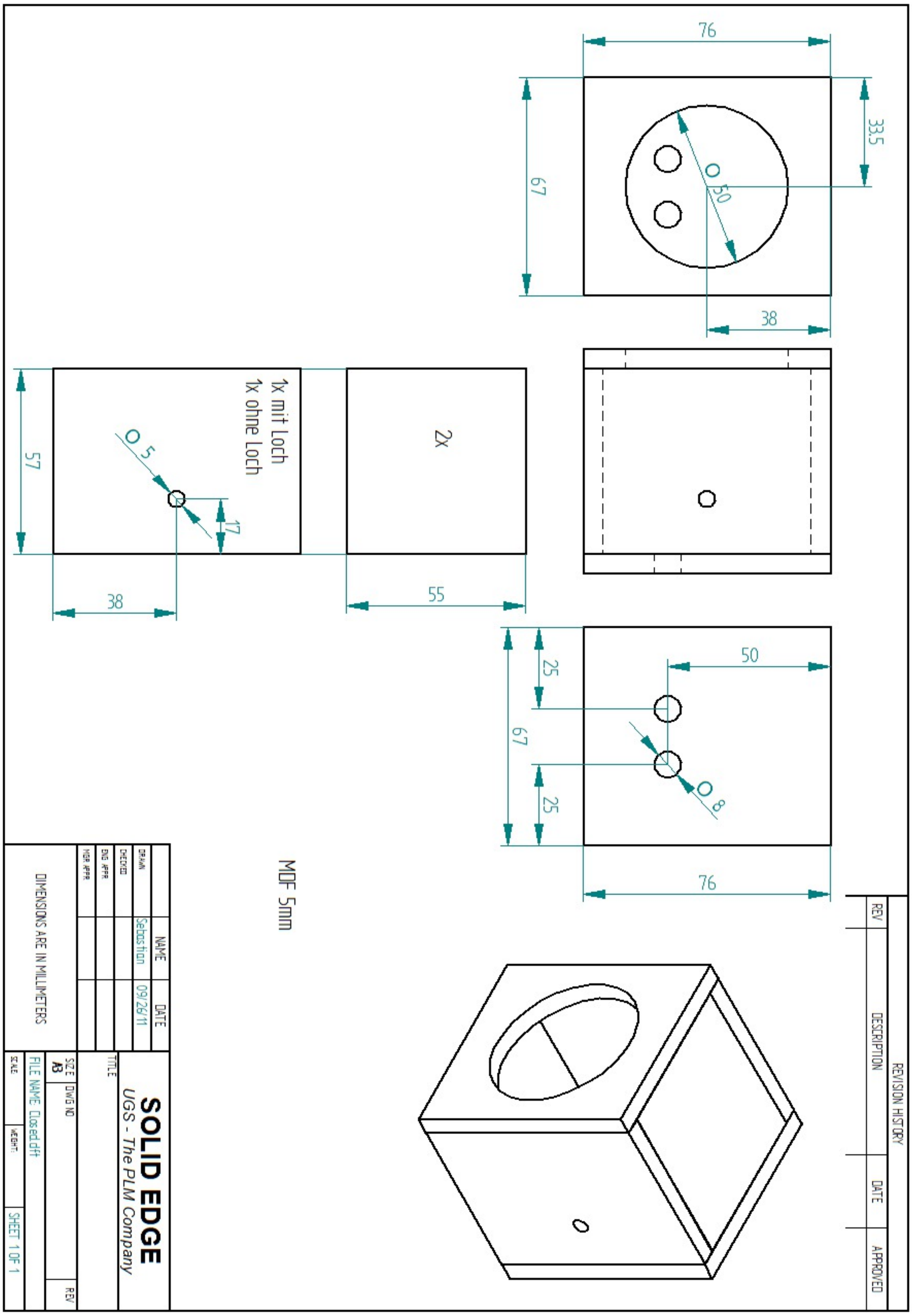



Figure 33: Draft of speaker enclosures for the closed box design

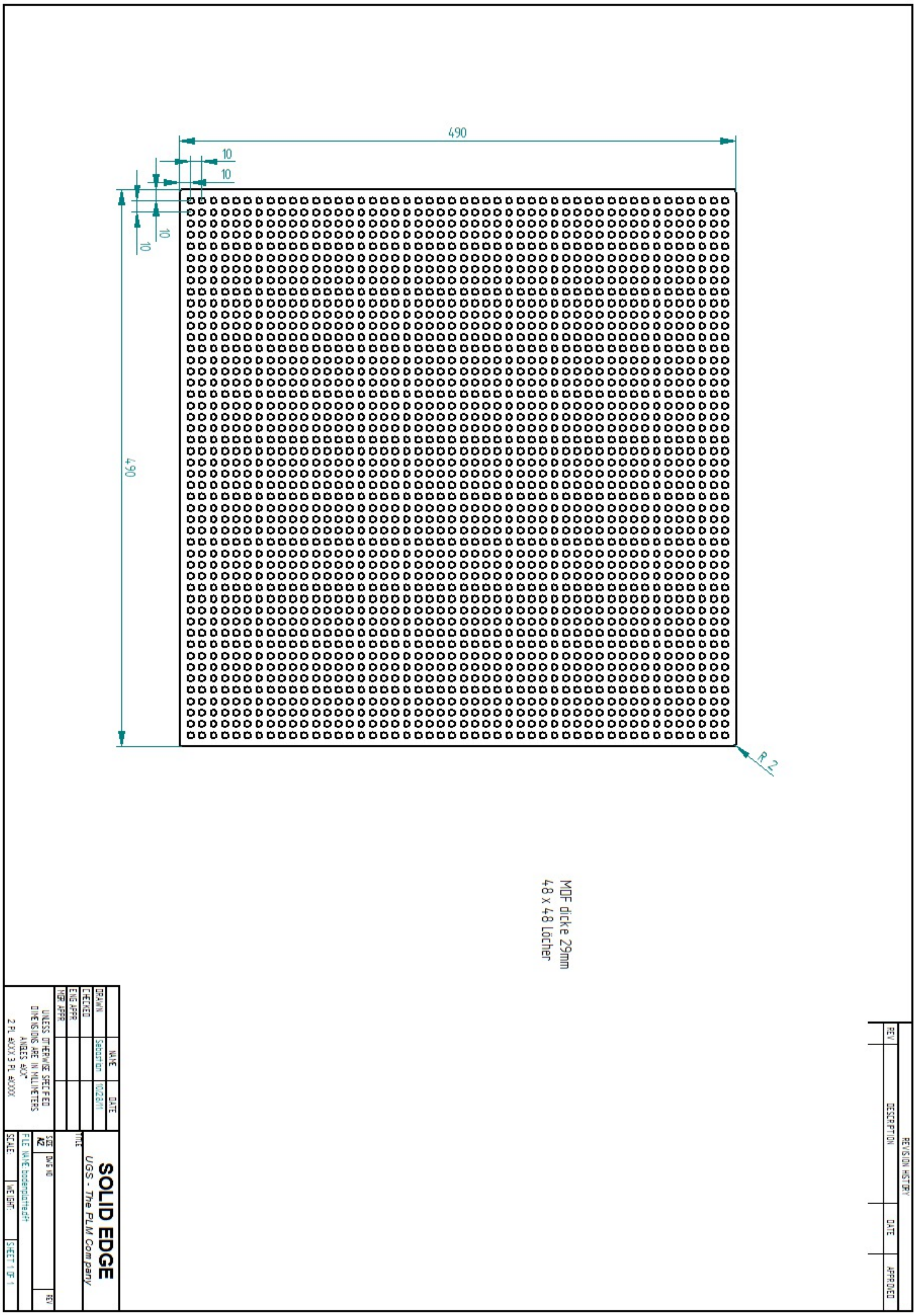



Figure 34: Draft of the ground plane

| Pcs | Value | Description | Distributor | Order. Nr. | Name |
|---|---|---|---|---|---|
| 12 | $0.1\mu F$ | SMD Capacitor<br>C-EU 0603 | Farnell | 1692286 | Cx.1, Cx.7, Cx.14 |
| 8 | $0.22\mu F$ | SMD Capacitor<br>C-EU 0603 | Farnell | 1828891 | Cx.12, Cx.15 |
| 8 | $0.68\mu F$ | Foil Capacitor<br>C-EU 50-025X075 | Neuhold | - | Cx.9, Cx.16 |
| 16 | $1\mu F$ | SMD Capacitor<br>C-EU 0603 | Farnell | 1759408 | Cx.3, Cx.4, Cx.5, Cx.6 |
| 1 | - | Diode<br>1N4446 | Neuhold | - | D1 |
| 8 | $4.7k\Omega$ | R-EU 0204/5<br>Resistor | Neuhold | - | Rx.6, Rx.7 |
| 4 | $10\mu F$ | SMD Capacitor<br>C-EU 1210 | Farnell | 1759368 | Cx.13 |
| 8 | $10k\Omega$ | SMD Resistor<br>R-EU 0603 | Farnell | 1469749 | Rx.1, Rx.2 |
| 9 | - | Wago Clamp<br>W237-102 | Neuhold | - | X1, X2, X3, X4 ,X5 |
| 8 | $22\mu H$ | Inductor<br>BS11 | Farnell | 1749100 | Lx.1, Lx.2 |
| 8 | $470\mu F$ | Electrolytic Capacitor<br>CPOL-EUE5-105 | Farnell | 1692339 | Cx.10, Cx.17 |
| 1 | $1400\Omega$ | R-EU 0204/7<br>Resistor | Neuhold | - | R1 |
| 2 | $1800\mu F$ | Electrolytic Capacitor<br>CPOL-EUE75-16 | Farnell | 1848398 | C1, C2 |
| 4 | - | Headphone Jack<br>LUMBERG-1503 04 | Farnell | 1216981 | X11, X12, X13, X14 |
| 1 | red | LED 3MM | Neuhold | - | LED1 |
| 4 | - | Socket DIL20S | Neuhold | - | IC1, IC2, IC3, IC4 |
| 4 | - | TPA 3122 | Farnell | 1755372 | IC1, IC2, IC3, IC4 |

Table 6: Parts list for electronic assembly of 1 PCB