Diploma Thesis

# MUSIC PERCEPTION WITH HEARING AIDS

The M.U.S.I.C. Test Battery

conducted at the
Signal Processing and Speech Communications Laboratory
Graz University of Technology, Austria

in cooperation with
Phonak AG
Stäfa, Switzerland

by
Martin Kirchberger, 0573153

Supervisors:
Dipl.-Ing. Dr.sc.ETH Harald Romsdorfer

Assessors/Examiners:
Univ.-Prof. Dipl.-Ing. Dr.techn. Gernot Kubin

Graz, September 2011

## ABSTRACT

This diploma thesis aims at making the music perception of hearing aid users measurable. To have the appropriate tool, the *M.U.S.I.C* test battery has been developed. The test battery provides the basis for the development of future strategies to improve music perception with hearing aids. The performance of new music processing algorithms can be measured and evaluated. The *Natural Instrument Recognition* test (*NIR*) is the first unit of the *M.U.S.I.C.* test battery. The NIR test contains eight familiar instruments from different instrument categories which are played back with purpose-composed melodies. The subject's task is to recognize the instrument played in the test stimuli. The second test unit (*MID*) measures a hearing impaired person's discrimination ability concerning four woodwind instruments. A new test methodology has been developed for the *MID* test. The subject compares a reference stimulus with four different test stimuli by turns. The subject has to identify the test stimulus whose instrument corresponds to the the instrument of the reference stimulus. The third test (*CCD*) focuses on the discrimination of sound color. Three piano chords are played back in succession. The harmony, lowest pitch and highest pitch of all three chords are identical. One chord differs from the other two chords in the arrangement of the harmony notes between the lowest and highest pitch. It is the subjects' task to detect this chord within the three stimuli. A pilot study with six normal hearing and seven hearing impaired persons was conducted to check the applicability of the *M.U.S.I.C.* test battery. According to the results, the *NIR* test is an appropriate music perception test for the profoundly impaired. *MID* and *CCD* are appropriate accross the whole range of hearing losses. Nevertheless, the results also indicate that the degree of hearing loss is not sufficient to classify impairments regarding music discrimination and music perception.

**Keywords: music perception, hearing aid, discrimination test**


## ZUSAMMENFASSUNG

Ziel dieser Diplomarbeit ist es, die Musikwahrnehmung schwerhöriger Hörgeräteträger messbar zu machen. Hierzu wird die *M.U.S.I.C.* Testbatterie entwickelt. Sie bildet die Grundlage für die Entwicklung von zukünftigen Strategien zur Verbesserung der Musikwahrnehmung mit Hörgeräten. Die Leistungsfähigkeit neuer Musikverarbeitungsalgorithmen kann bemessen und beurteilt werden. Den ersten Teil der *M.U.S.I.C.* Testbatterie bildet der Instrumentenerkennungstest (*NIR*). Acht allgemein bekannte Instrumente aus unterschiedlichen Instrumentenklassen werden mit speziell komponierten Melodien wiedergegeben. Der Proband hat die Aufgabe, das Instrument des Teststimulus zu erkennen. Die zweite Einheit (*MID*) prüft die Unterscheidungsfähigkeit der Hörgeschädigten bezüglich vier Holzblasinstrumenten. Hierzu wurde ein neues Testverfahren konzipiert. Der Proband vergleicht abwechselnd einen Referenzstimulus mit vier verschiedenen Teststimuli. Die Aufgabe des Probanden ist es, denjenigen Teststimulus zu detektieren, dessen Instrument dem Instrument des Referenzstimulus entspricht. Der dritte Test (*CCD*) untersucht die Diskriminationsfähigkeit verschiedener Klangfarben. Drei Klavierakkorde werden hintereinander wiedergegeben. Die Harmonie sowie die tiefste und höchste Note der drei Akkorde sind gleich. Ein Akkord unterscheidet sich von den beiden anderen in der Verteilung der Funktionstöne im dazwischenliegenden Notenbereich. Diesen Akkord gilt es zu erkennen. Um die Verwendbarkeit der *M.U.S.I.C.* Testbatterie zu überprüfen, wurde eine Pilostudie mit sechs normalhörenden und sieben schwerhörigen Probanden durchgeführt. Die Ergebnisse zeigen, dass sich der *NIR* Test für höchstgradig Schwerhörende eignet. Die beiden anderen Tests sind für alle Hörschädigunsgrade anwendbar. Nichtsdestotrotz indizieren die Resultate, dass der Grad der Hörschädigung nicht ausreicht, um die Beeinträchtigung der Wahrnehmung und Diskrimination von Musik zu klassifizieren.

**Schlüsselwörter: Musikwahrnehmung, Hörgerät, Diskriminationstest**

# Contents

# List of Tables

# List of Figures

# 1
# Introduction

*"Music is nothing if no one can hear it."*
*– Placido Domingo, Tenor and Conductor*

*Figure 1.1: Acoustic throne of King Goa VI (image from Oticon, Eriksholm museum).*

What has King Goa VI of Portugal to do with this diploma thesis? The king suffered from severe hearing loss. Unfortunately, his hearing disorder is nothing extraordinary, his hearing aid, however, is very special. The king's throne not only served as a chair but also as a hearing device. The armrests - shaped like the heads of roaring lions - were hollow inside. Sound was passed through and amplified by the resonator system inside the arm rests. Via an ear tube held by the king, the amplified sound could finally be conveyed to the royal ear. Visitors were required to kneel before the chair and to speak directly into the animal heads.

Fortunately, the technology has advanced since then and we are not required to kneel down when we converse. The state-of-the-art hearing aids resemble the size of a fingertip and disappear in the ear canal or behind the pinna. The most important reason to use hearing devices, however, has not changed since then. Hearing impaired people want to understand speech. An impaired auditory system affects speech intelligibility in several ways: the hearing threshold is elevated, the dynamic resolution and frequency resolution are decreased, the noise reduction and source localization are impacted.
Consequently, the research and development in the area of hearing aids has focused on developing strategies to enhance speech intelligibility. The subject 'music perception', however, has been neglected so far. Hearing disorders affect the enjoyment of music, the practice of an instrument and the motivation to join concerts and other social music events. People do not lose their love of music when they lose their hearing.

The objective of the research project *Music Perception with Hearing Aids* is to improve the music perception with hearing aids. First, a test battery is developed to assess and to learn how the hearing impaired can benefit from hearing aids. Second, the tests will be used to guide the development of signal processing concepts for music processing. Based on the test results, decisions can be made at last about which strategy to pursue. The diploma thesis marks the first step of the research project. Three discrimination tests are designed, implemented and put under test in a pilot study.

The remaining part of the diploma thesis is structured as follows:
In chapter 2, a glossary explains the technical terms that are used in the diploma thesis. Chapter 3 provides basic information about the technology of state-of-the-art hearing aids and the current music processing strategies. Inquiries about the music perception of the hearing impaired and existing music tests are also presented. Chapter 4 documents the building process of the M.U.S.I.C. test battery. In chapter 5, the setup of the pilot study is presented. The results of the pilot study are analyzed in chapter 6. A final discussion regarding the applicability of the M.U.S.I.C. test battery and the music perception with hearing aids follows in chapter 7. After a short summary in chapter 8, the subjects' audiograms and a manual for the M.U.S.I.C. test battery are attached in the appendices.

# 2
# Glossary

## Abbreviations

| | |
|---|---|
| **CCD** | Chord Color Discrimination |
| *gl.* | Glossary |
| **GNS** | German notation system |
| **HA** | Hearing aid |
| **HI** | Hearing impaired |
| **MID** | MIDI Instrument Discrimination |
| **M.U.S.I.C.** | Measuring the Understanding of Sounds, Instruments and Chords |
| **NH** | Normal hearing |
| **NIR** | Natural Instrument Recognition |

## Definitions

**audiogram** The audiogram is a standard plot to depict a person's hearing loss. Figure 2.1 shows an example of a profound hearing loss. The audiogram is set out with frequency (Hz) on the horizontal axis and the hearing loss (dB) on the vertical axis. The hearing loss is determined by the difference between the subject's measured hearing threshold and a standardized curve representing the normal hearing people. The hearing threshold of the left and right ear is illustrated by a blue line with crosses and a red line with circles respectively.

Apart from the hearing threshold, an audiogram can further plot the masking level and the uncomfortable level. Masking is needed in case of an asymmetric hearing loss to measure the hearing threshold of the more seriously damaged ear. Sound can be conducted via bone conduction to the auditory system of the healthier ear. To make sure that the subject cannot perceive the test signal with his better ear when the worse ear is actually tested, the better ear is exposed to noise.
The uncomfortable level indicates the frequency-dependent sound pressure threshold that is considered uncomfortable by the subject.

**bark scale** The bark scale represents the filter system of the human auditory system. Acoustic signals within the audible range are separated into 24 different sub-bands. The bark scale is displayed in figure 2.2.

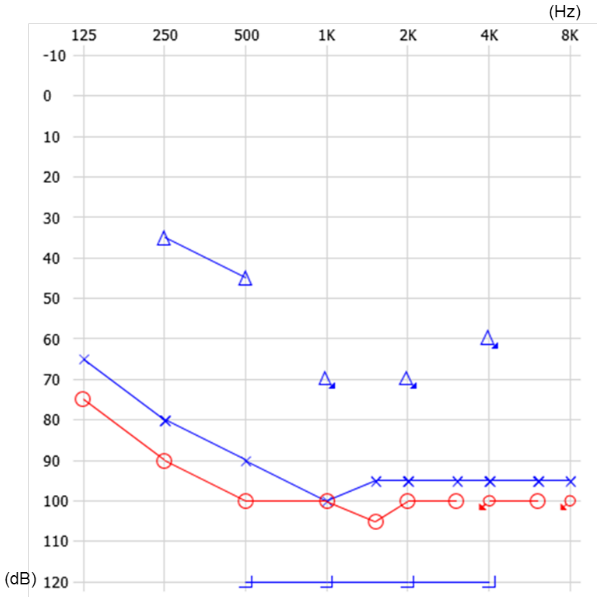


*Figure 2.1: Audiogram of a person with profound hearing loss. The measurement data in blue color (crosses, hooks, triangles) refers to the left ear, the data in red (circles) refers to the right ear. Crosses and circles represent the hearing threshold, hooks specify the uncomfortable level and triangles indicate the masking level.*

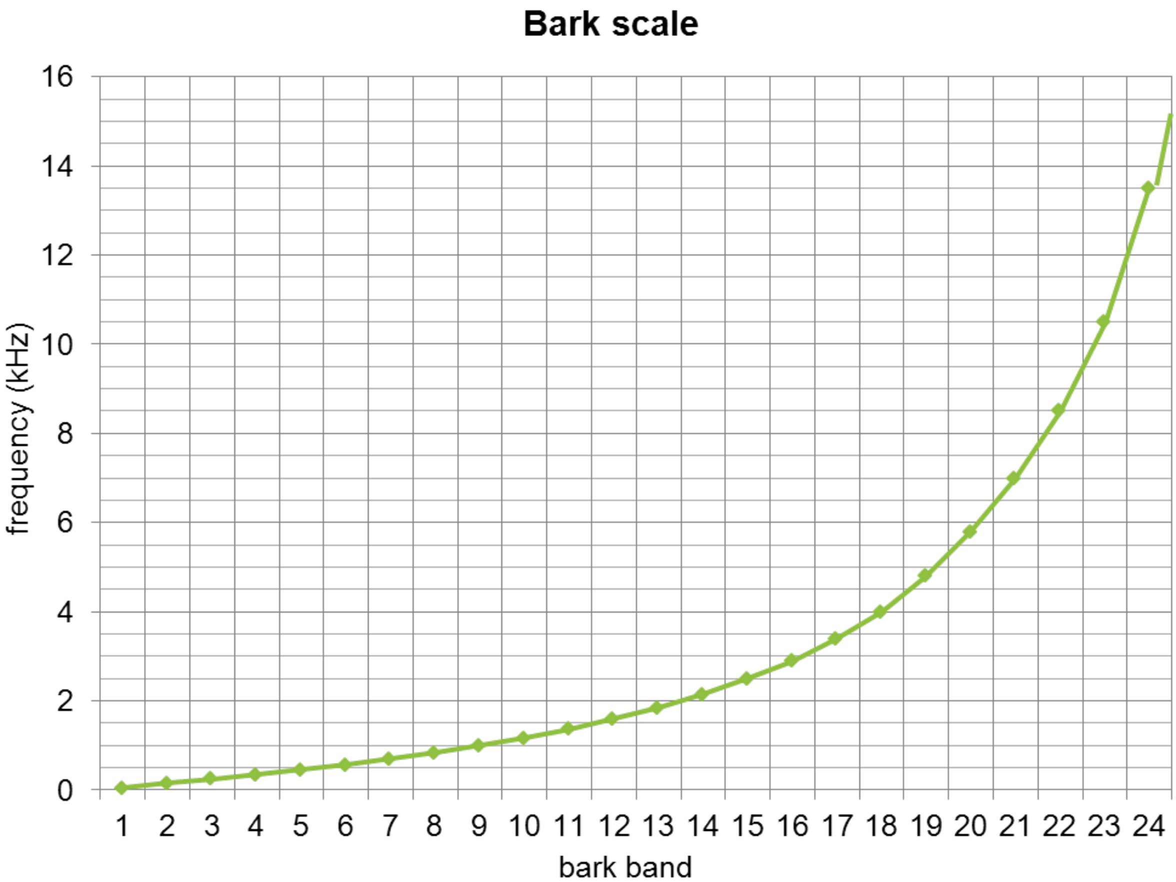


*Figure 2.2: The bark scale. The human auditory system filters the acoustic signals within the audible range into 24 sub-bands. The rhombi mark the center frequency of the sub-band. The vertical lines mark the lower and upper limit of the corresponding bark band.*

**chromatic / chromatic scale** The chromatic scale is a musical scale proceeding entirely by semitones. It contains all the notes that are available in traditional Western music and hence, does not have a definite key [1]. Figure 2.3 displays the chromatic scale. Without any limitation to universality, the starting key of the scale is *C4* (GNS: *c'*).

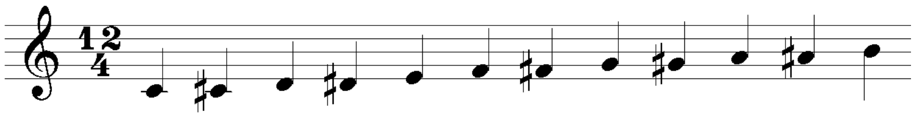

*Figure 2.3: The chromatic scale starting from* C4 *(GNS:* c'*)*

**compass** The compass is the range from the lowest pitch to the highest pitch of an instrument.

**degree of hearing loss** The degree of hearing loss is a measure to classify a person's hearing impairment. It equals the averaged hearing thresholds of the frequencies 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz. The national standards to categorize the degrees of hearing loss are not congruent. Within the framework of the diploma thesis, the degrees of hearing loss are classified as follows:

- less than 15 dB: normal hearing
- 15 dB to 35 dB: mild hearing loss
- 35, 1 dB to 55 dB: moderate hearing loss
- 55, 1 dB to 80 dB: severe hearing loss
- 80, 1 dB to 100 dB: profound hearing loss

Persons with a degree of hearing loss that is higher than 100 dB

**diatonic function** The diatonic function is the specific, recognized role of each of the seven notes and their chords in relation to the diatonic key. The function relates to the stability and degree of tension that the note creates [2].

**diatonic instrument** Within the framework of the diploma thesis, the technical term *diatonic instrument* is used as follows:
A diatonic instrument can play all semi-tones within its compass at discrete frequencies only. Diatonic instruments are the piano, the accordion or the harp for example. A cello is not a diatonic instrument, as it can vary the frequency of its pitches continuously. An alphorn is also not a diatonic instrument. The frequency of its pitches are fixed, but it cannot play all semi-tones within its compass.

**hearing aid** The classification of a hearing device to the technical umbrella term *hearing aid* is ambiguous in the literature. To avoid misunderstandings, all hearing devices whose output is sound or mechanic vibration are classified as hearing aids in this thesis:

- Behind-the-ear (BTE) hearing aid
- In-the-ear (ITE) hearing aid
- In-the-canal (ITC) hearing aid

- Completely-in-the-canal (CIC) hearing aid
- Bone-Anchored Hearing Aid

Cochlear implants and auditory brainstem implants output electronic stimuli are not types of hearing aids.

**instrument class** The technical term *instrument class* refers to a group of instruments that share common features. Depending on the classification system, the distinguishing features are different: the shape of the resonating body, the mode of excitation or the sound-producing material.
Hornbostel-Sachs [3] is an established musical instrument classification system. The four broad categories are defined by the sound-producing material: air column (aerophone), string (chordophone), membrane (membranophone) and the body of the instrument (idiophone). The next subcategory is defined by the mode of excitation, e.g. plucked or blown. All in all, over 300 basic categories are distinguished.

**instrument family** Instruments that share the same design but differ in size and tuning [3] belong to the same instrument family. Example: violine, viola, cello and contrabass.

**inversion** Using the example of the *C-major* chord, the principle of inversion is explained as follows:
The *C-major* chord contains three functions: the root *c*, the third *e* and the fifth *g*. The three functions do not have to be noted in the same octave but can be positioned in any register. Moreover, every function can be represented by more than one pitch. Three different examples of a *C-major* chord are displayed in figure 2.4.

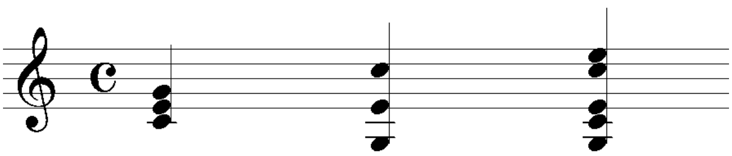

*Figure 2.4: Three examples of a* C-major *chord with the pitch classes* c, e *and* g.

According to the harmonics theory, it is only the choice of the bass note that defines the type of chord inversion. A chord with the root in the bass is called a root-position chord. If the third *e* or the fifth *g* is the deepest note, the chords are called first and second inversion. The corresponding examples are displayed in figure 2.5.

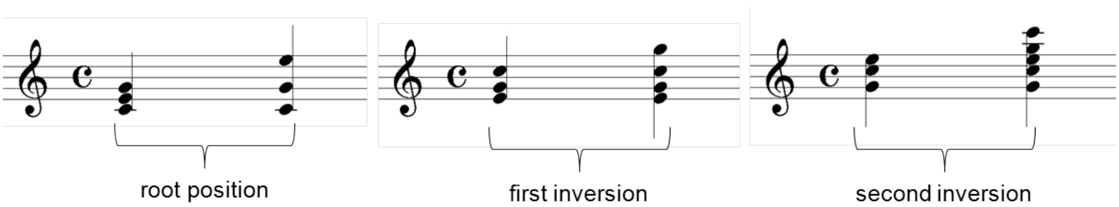


*Figure 2.5: Two examples of the C-major chord in root-position, as first inversion and as second inversion.*

**MIDI** MIDI (Musical Instrument Digital Interface) is an industry-standard protocol used for the communication of electronic instruments.

**MIDI instrument** In this thesis, the term *MIDI instrument* is used for synthesizers or electronic devices, that use audio sample libraries to generate sound.

**modern musical modes** The modern musical modes are a system of seven scales: Ionian, Dorian, Phrygian, Lydian, Mixolydian, Aeolian, and Locrian. Each of the scales is heptatonic and diatonic, i.e. it consists of seven notes that proceed in semitones and whole tones. The most common mode is the Ionian scale, better known as the major key. The other six modern musical modes are derived from the Ionian scale by simply choosing another of the scale's pitches as the root. All scales in the left half of figure 2.6 contain the same seven pitch classes. The first pitch, however, is different. As the first pitch is the tonal center of the scale and the other six pitches are perceived in relation to the root, the characteristic sound of each scale is different. It is the position of the semi-tone steps (marked by the brackets in figure 2.6) that defines the sound. Starting with the same pitch, the seven modes are displayed in the right half of figure 2.6.


*Figure 2.6: The seven modern musical modes. Starting with the Ionian scale, the other six scales are formed by choosing the next pitch of the scale as the new root (left half of the figure). The characteristic sound of each scales is defined by the position of the semi-tone steps which are marked by the brackets. In the right half of the picture, the seven modern musical scales are displayed with each scale starting with the same root* C4 *(GNS:* c'*).*

**musical aptitude** Musical aptitude is the the potential or capacity for musical achievement [4]. Synonyms are musical talent and musicality.

**music perception** Perception is the process whereby sensory stimulation is translated into organized experience [5]. Music perception is the auditory perception of musical stimuli and its qualities: pitch, timbre, harmony, rhythm [6].

**music understanding** *c.f. music perception*

**natural harmonic series** The natural harmonic series is composed of a fundamental pitch and its integer multiples. Without any limitation to universality, the scale is displayed for the key *C3* (GNS: *c*) in figure 2.7.

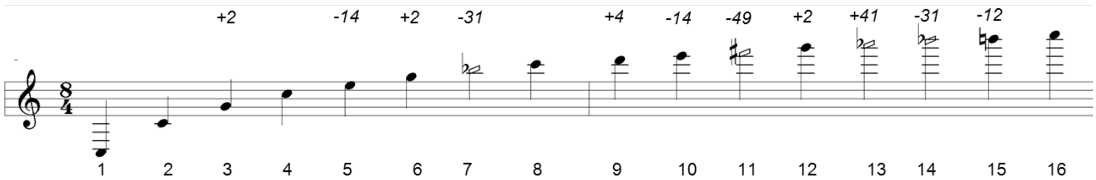


*Figure 2.7: The natural harmonic scale for the key* C3 *(GNS:* c*). The lower numbers indicate the number of the harmonic. The upper numbers become relevant in section 4.3.2. They indicate the deviation of the noted pitch and the frequency of the natural harmonic. The unit is cents.*

**notation system** In this thesis, the international Anglo-Saxon notation system is used to designate the notes. As the diploma thesis is written in Switzerland and Austria, the pitches are also specified according to the commonly used German notation system.
Regarding the Anglo-Saxon notation system, the registers of the pitches are consecutively numbered: *C0,C1,C2,C3,...* . The German system uses strokes to indicate the register: *"C,'C, C, c, c',c"...* . The Anglo-Saxon *C0* corresponds to the German *"C*.
Apart from *b* (GNS: *h*), both systems use the same labeling for their pitch classes. Figure 2.8 displays the Anglo-Saxon and German notation of the C-major scale.

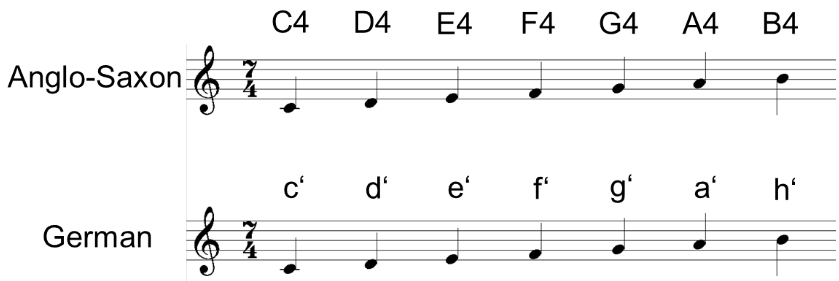


*Figure 2.8: The Anglo-Saxon and German notation of the C-major scale.*

**pitch** Pitch is defined by the American Standards Society as "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" [7]. It is the psychoacoustic correlate to the repetition rate of a sound [8]. According to [7], pitch perception is dependent on the intensity level of the stimulus. With the intensity increasing, the pitch perception decreases for frequencies below 2000 Hz and increases above 4000 Hz.

**pitch class** A pitch class is a group of pitches that are related to each other by intervals of integer multiples of an octave, e.g. ... ,C1, C2, C3, C4, ... (GNS: *... ,'C, C, c, c', ...*) [9].

**register** The register of an instrument is a set of pitches that are similar in timbre. A standard grand piano has a range of more than seven octaves. Within the framework of the diploma thesis, a register is defined as one octave of the piano range.

**sound color** Sound color is one aspect of timbre. Timbre includes the starting transient characteristics and the steady state properties in the sustain phase [10]. Sound color refers to the second aspect only.

**speech intelligibility** Speech intelligibility is the accuracy with which a listener can understand a spoken word or phrase.

**stereo panorama** Stereo panorama is the illusionary sound space between two loudspeakers.

**stereo position** Stereo position is the perceived spread of a signal in the stereo panorama.

**timbre** Timbre is defined as the quality of sound by which a listener can tell that two sounds of the same loudness and pitch are dissimilar. In a musical context, it designates the aspects of sound that allow an instrument to be identified and distinguished from others. Timbre breaks down into the transient characteristics in the settling phase and the sound color in the sustain phase [10].

# 3 Basics

## 3.1 Functionality of the hearing aids

In this section, the most important features of a digital hearing aid are introduced. The operating principle is explained using Phonak hearing aids as example.

### Gain model

State-of-the-art hearing aids are fully digital. A Phonak hearing aid samples the analog input signal with $f_s = 20480\,\text{Hz}$. Subsequently, a 128-bin fast Fourier transformation is applied. The resulting frequency resolution in the frequency domain is $\frac{20480\,\text{Hz}}{128} = 160\,\text{Hz}$. The frequency bins are divided into sub-bands that approximate the bark scale[gl.].
The hearing aid cannot transmit the full audible range. Due to a frequency resolution of 160 Hz, the first bark band is kept out. The upper limit is defined by the sampling frequency. According to the Nyquist criterion, only frequencies up to half the sampling frequency can be reconstructed. Hence, the theoretical upper limit is $\frac{f_s}{2} = 10240\,\text{Hz}$. The real upper limit is lower. The transfer function of the speaker limits the bandwidth to 8 kHz.
The filter bank of the hearing aids consists of 20 sub-bands. The assignment of the fft-bins to the bandpass filters follows the principle of the bark scale. Table 3.1 displays the mapping of the lower 64 fft-bins to the 20 bandpass filters.

The ranges of the sub-bands and the bark bands are not exactly congruent. Firstly, the frequency domain is discrete and the resolution is only $160\,Hz$. Secondly, in coordination with many other design aspects of a hearing aid, the current division turned out to perform better.

Hearing losses are very diverse. The hearing threshold and the uncomfortable level for example are frequency dependent. As the signal is split into 20 sub-bands, the gain characteristic can be adjusted for the 20 different frequency ranges individually. Figure 3.1 displays a typical gain characteristic of a sub-band.

The inherent noise level of the microphone is approximately 28 dB. Consequently, all signals with a lower intensity than 28 dB are not passed through. The signals which are more intense than 28 dB are amplified. After a certain threshold, that depends on the individual hearing loss, the gain of the input signals is reduced. Very intense signals are attenuated. To avoid that the output pressure level is harmful for the auditory system, a limiter is connected downstream.

| sub-band | # bins | fft bin | center frequency (Hz) | frequency range (Hz) |
|---|---|---|---|---|
| 1 | 1 | 2 | 160 | 160 |
| 2 | 1 | 3 | 320 | 320 |
| 3 | 1 | 4 | 480 | 480 |
| 4 | 1 | 5 | 640 | 640 |
| 5 | 1 | 6 | 800 | 800 |
| 6 | 1 | 7 | 960 | 960 |
| 7 | 1 | 8 | 1120 | 1120 |
| 8 | 1 | 9 | 1280 | 1280 |
| 9 | 2 | 10-11 | 1520 | 1440-1600 |
| 10 | 2 | 12-13 | 1840 | 1760-1920 |
| 11 | 2 | 14-15 | 2160 | 2080-2240 |
| 12 | 2 | 16-17 | 2480 | 2400-2560 |
| 13 | 3 | 18-20 | 2880 | 2720-3040 |
| 14 | 3 | 21-23 | 3360 | 3200-3520 |
| 15 | 4 | 24-27 | 3920 | 3680-4160 |
| 16 | 5 | 28-32 | 4640 | 4320-4960 |
| 17 | 6 | 33-38 | 5520 | 5120-5920 |
| 18 | 7 | 39-45 | 6560 | 6080-7040 |
| 19 | 9 | 46-54 | 8000 | 7200-8800 |
| 20 | 10 | 55-64 | 9520 | 8800-10240 |

*Table 3.1: The filter bank of a hearing aid. The frequency range of* 160 *Hz to* 10240 *Hz is divided into* 20 *filter banks following the bark scale. The frequency resolution of the Fourier transform applied is* 160 *Hz. The table shows how the fft-bins are subdivided into the sub-bands.*

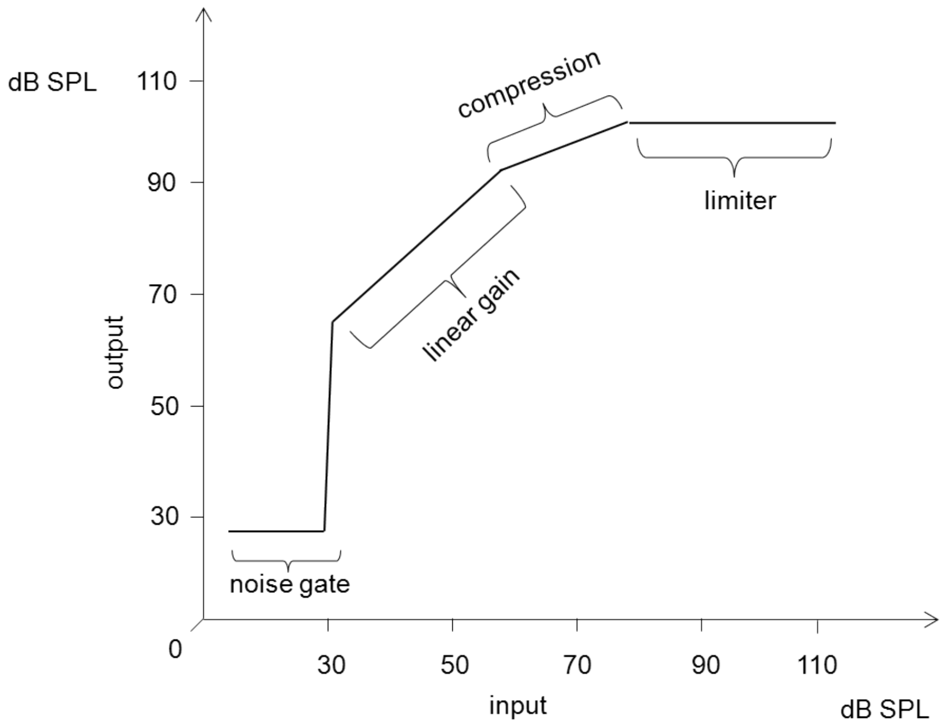


*Figure 3.1: A typical gain characteristic of a hearing aid. Inherent noise is suppressed by a noise gate, silent sounds are amplified and intense signals are compressed. Furthermore, a limiter determines the maximum output level.*

Hearing impaired persons not only suffer from a higher hearing threshold but also from a lower uncomfortable level. The reason for the reduced dynamics is the recruitment effect. Analogous to the hearing aid, the outer hair cells amplify silent signals and attenuate the loud signals. Damage or loss of these sensory cells impair the ability to compress the intensity of the audio signals.

### Adaptive Feedback Canceler, AFC

For all kinds of hearing aids, the distance between the sound output and the microphone is relatively small. If the output is not acoustically sealed, the signal is fed back to the microphone. Depending on the gain, the feedback loop can build up and an intolerably intense and disturbing tone is produced. One approach to avoid this effect is to block the acoustic path from the sound output to the microphones. The shells of the ITEs and ITCs as well as the earmolds of the BTEs do not seal the ear canal airtight. A minimal air interchange is necessary to drain the ear canal. Therefore, the amplified output signals are still transmitted to the microphone at a significant level.
With regard to BTEs, the less a person is hearing impaired, the more openly he is fitted to have the signals directly passed to the eardrum. An acoustic vent is drilled into the earmold or no earmolds are used at all. As a side effect, sound can also more easily be transmitted in the opposite direction from the speaker/ear canal to the hearing aid microphones.
Moreover, not only air but also the mechanic parts feed the vibration back to the microphone membrane. As a consequence, a further strategy is required to deal with feedback problems. The remedy is adaptive feedback canceling.

The concept of a feedback canceler is to estimate the transfer function of the feedback path. With the help of the estimated transfer function, the feedback signal can be approximated. By deducting the signal from the overall input, the feedback signal is canceled and the target signal is recovered.
The transfer function is not constant but varies subject to the setting. Putting on a hat or chewing gum for example changes the propagation paths considerably. With wearing a hat, the transfer function is different because the acoustic waves are reflected at its surface. While chewing gum, the jaws deforms the ear canal and impact the sealing of the BTE earmold or of the ITE and CIC shells respectively. As a consequence, the feedback canceler has to be adaptive in order to account for the temporal changes.

### Noise Canceler, NC

Another important topic for hearing aids is noise canceling. Hearing impaired people not only suffer from higher hearing threshold levels and limited dynamic resolution. Their capability to extract a target signal in a noisy environment is also weakened. By reducing the noise floor, the listening effort is reduced and a higher speech intelligibility is ensured.
The noise canceling algorithm in the hearing aids estimate the spectrum of the noise and subtract it from the input signal. The signal to noise ratio can be increased by up to 10 dB.

### Beamformer, BF

One skill of the human auditory system is the capability to locate the direction of a sound source. Moreover, a normal hearing person is also able to enhance the intelligibility of the signals that come from that direction. This skill is referred to as the cocktail party effect. To identify the direction, the auditory system analyzes the interaural time and level differences as well as the spectral shaping of the head and pinna. To enhance the intelligibility of a desired direction, it is assumed that a kind of cross-correlation is applied using the respective level and time differences as parameters.

Damages to the auditory system affect the localization capabilities. Hearing aids can compensate for this loss by means of beamforming techniques. The directionality is created by processing signals of multiple microphones. BTE and ITE models are equipped with two microphones and each device can form a cardioid beam. The newest technology in the hearing aid industry allows to create more sophisticated beam patterns by combining the signals of the hearing aids from both ears via wireless transmission.
By default, the beam is faced forward. The directionality, however, can also adaptively be adjusted to the auditory scene. When driving a car and looking ahead, the user wants to understand the passengers that are sitting on the front seat or behind.

Nevertheless, the applicability of an adaptive beamformer has a limit. If two target signals arrive from different directions, the beamformer cannot know which signal and direction the user wants to focus on. Nevertheless, new approaches to optimize the behavior of the beamformer in difficult auditory scenes are under development.

### Frequency Compression

The majority of hearing impaired people suffer from high frequency hearing loss. In very profound cases even the maximal amplification cannot make the high frequencies audible. For these types of hearing losses, an innovative concept can be applied: frequency compression. The high-pitch signals are mapped to lower frequencies that can be perceived by the hearing impaired person. As displayed in figure 3.2, two parameters can be adjusted: the frequency of the knee point and the compression ratio.

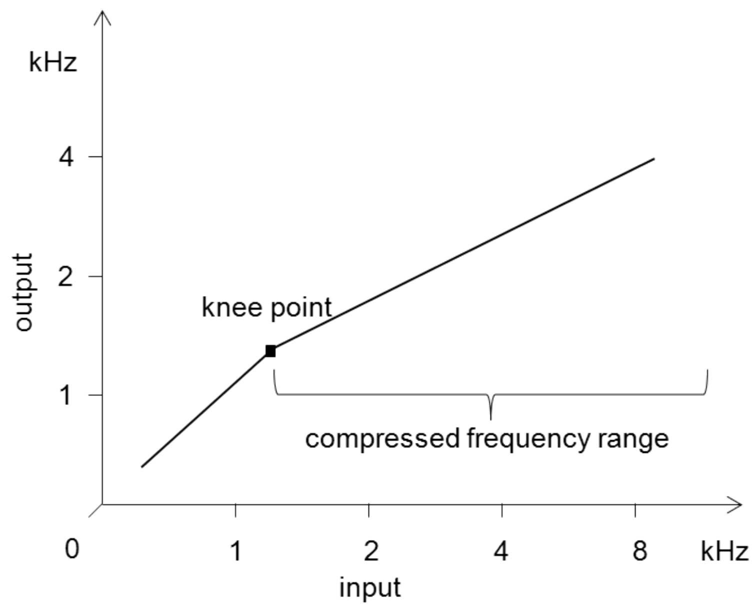


*Figure 3.2: Principle of frequency compression in hearing aids. Frequencies that are above a definable threshold are mapped to lower areas.*

The auditory system learns how to adapt to the frequency compressed signals. The lower the knee point frequency and the higher the compression rate, the more high-frequency information is conveyed to the auditory nerve. The speech intelligibility is increased but the sound of speech becomes less natural. An optimal trade-off between audibility and authenticity is desired.

#### Classifier

The functionality of a hearing device can be optimized by adjusting its features for different auditory scenes individually. A conversation in a restaurant for example requires another setting than watching a movie in a cinema.
The current version of a classifier distinguishes between four scenes:

1. clean speech
2. speech in noise
3. noise
4. music

The hearing aid not only adjusts its parameters optimally for the four scenes but also for any interstages. Every 0.8 seconds the auditory scene is analyzed and a probability distribution is calculated for the four classes. The results are low pass filtered and smoothed over time to avoid frequent and sudden changes between the scenes. The parameters of the classes are then weighted according to the probability distribution. Thereby, the music class is treated separately. The optimal adjustments for music are so different from noise and speech, that it is not reasonable to create interstages and mix their parameters.

## 3.2 Music processing

#### The singularity of music

As already mentioned, the music class is treated separately because its requirements are very different to those of the other categories. In [11], the singularity of music is contrasted with the characteristics of speech:

- The level of speech rarely exceeds 85 dB SPL. Music can easily reach levels higher than 120 dB SPL, be it classic or rock.
- The spectral characteristics of speech and its phonemes are well-established. Music, on the contrary, is very diverse and difficult to categorize from an audiological viewpoint.
- The crest factor of music (18 to 20 dB) is significantly higher than the crest factor of speech (12 dB).
- The transitions of silent to loud passages and vice versa are more abrupt in music.

These characteristics make it extremely difficult to process music in a hearing aid with sufficient quality. The project *music perception* focuses on digital signal processing and software solutions. Currently, the only software solution related to music processing is the music program.

#### The music program

When the music program is activated, the following adjustments are made:

- The noise canceler is switched off.
- The adaptation of the feedback canceler is slowed down.

- The beam-former is switched off.
- The gain-model is less compressive.
- The knee point for frequency compression is set to a higher value (cf. figure 3.2).
- An equalizer function is connected downstream to emphasize low and high frequencies.

Apart from the last item, the functionality of the music program is based on switching off or attenuating the effect of the features. Thereby, the authenticity of the signal should be enhanced.

For two reasons, the current status is not satisfactory. Firstly, the parameter adjustments rest upon the subjective evaluation of a handful of normal hearing persons. The applicability for hearing impaired people has not been examined thoroughly. Secondly, concepts to *actively* recreate the authenticity of music have not been considered so far.

## 3.3 Music perception

### Feedback of the hearing impaired

In order to improve the music perception, it is necessary to analyze the feedback and complaints from the hearing aid users. In [12], 221 hearing impaired persons between the age of 16 and 95 years were questioned about their listening habits. According to the results, one of three respondents is affected by his hearing disorder when practicing an instrument or listening to music. The specific problems are the distortion of sound, the inability to recognize melodies and the difficulty to distinguish instruments.
In [13], several analog and digital devices were put to test. According to the source, the analog devices still surpass the digital successors regarding sound quality. The latter models lack of precision and sonority. Musical passages are smeared, the presentation sounds distant and excitement or emotion are not sufficiently conveyed.

Above studies date back to the years 2006 and 2007. Consequently, the present[1] state-of-the-art hearing aids are not taken into account. In the latter study, the top of the line hearing aids of that time were used. In the first study, the people were randomly chosen. It has to be assumed that the hearing devices might not only be from previous launches but also from lower price and quality segments. The average life cycle of a hearing aid is six years. Thus, it can expected that the age of the used hearing aids was three years on average. Concerning music processing, the quality differences between the present models and the models from 2003/2004 are moderate. Analogously, the music processing of the top-of-the line models and the devices of the lower price segment is similar. Thus, it can be assumed that the results of both studies are still valid today[2].

To gain additional information about the performance of state-of-the-art hearing aids, further inquiries have been conducted. Three different hearing care professionals have been interviewed: Jochen Wied, Wied Hörforum (Ludwigsburg, Germany), Michael Beck, Beck Hörgeräte (Albstadt, Germany) and Christoph Schwob, Hörberatung Basel (Basel, Switzerland). They unanimously state that fitting the hearing aids for music is still a very difficult task. Especially audiophiles are not satisfied with the state-of-the-art devices. With regard to the music program of Phonak, the sound of the standard setup is considered too sharp in the high frequencies and too thin in the bass region. This misfit can easily be solved by the hearing care professional and his fitting software. The lack of separability of the instruments as well as the distorted sound, however, require new signal processing methods.

[1] 'present' refers the deadline of the diploma thesis: september 2011
[2] september 2011

### The relevance of music

In [14], the average music listening time for different age groups has been determined. For persons from 25 to 75 years, the average daily amount of music is higher than two and a half hours. The study was published in the year 2004. It must be expected that the music listening time has increased since then. The availability of music has risen because of the internet. Moreover, the diversity of playback devices and its market penetration has augmented.
To analyze the listening habits of hearing impaired people, the DataLogging tool of the Phonak hearing aids can be used. Amongst other things, this feature automatically tracks the amount of time that the music program has automatically or manually been activated. The measured share of the operating time is 2.7 % or 16 minutes on a daily basis respectively.
The difference to the measured average listening time of normal hearing people can be caused by three reasons:

- The classifier does not work properly.
- Hearing impaired people prefer to listen without hearing aids.
- Hearing impaired people listen less to music than normal hearing people.

As already mentioned, the classification is based on the probability distribution of the four classes, speech, speech&noise, noise and music. In ambiguous cases the first three classes are preferred to music. Speech intelligibility and noise reduction are more important than having the optimal adjustments for music. Consequently, it can be expected that the hearing aid was exposed to more music than the time measurement of the DataLogging tool indicates.
With regard to the latter two reasons, the paper of Edith Egloff[15][3] provides relevant information. According to her study, one in three interviewees prefers not to use his hearing aids when listening to music. Moreover, the majority of 70% does not enjoy listening to music as much as they did before being impaired.

To gain further information, the three hearing care professionals Beck, Wied and Schwob have been questioned about the relevance of music for their clients. Their responses are very different: Schwob indicates that approximately one third of his clients set a high value on music perception. Thereby, it has to be considered that Schwob is known for attaching great importance to fitting hearing aids for music. Consequently, the percentage of his audiophiles clients is certainly above average.
Beck and Wied state that most of their customers do not set great store by music. However, many hearing impaired people might not be aware of their loss of life quality. As music is not a basic need like food and shelter, people often resign to the fact that listening to music is not a pleasure any more. Consequently, the real relevance of music for hearing impaired people can be considered higher than they indicate.

## 3.4 Existing music tests

Two types of music tests have to be distinguished: music *aptitude* and music *perception* tests. The first type aims for measuring the musical talent of a subject. The latter type intends to assess the discrimination and enjoyment of musical signals.
As the M.U.S.I.C. test battery aims at researching the impact of hearing loss and hearing aid processing on the perception of music, only the second type is relevant.
The existing music perception tests can serve as reference for the design of the M.U.S.I.C. test

[3] The inquiries date back to the year 2003. As the music processing in hearing aid has not improved significantly since then, the statements can be regarded to be still valid today.

battery. With regard to the hearing impaired people, however, no music perception tests have been published so far. Concerning the tests for the normal hearing and the cochlear implantees, the Seashore test [16] and the Zurich Music Test Battery [17] are relevant as a reference for the music perception test design.

### The Seashore test

The Seashore test is designed for normal hearing people and can be carried out in five different degrees of difficulty. The test is subdivided into six subtests. Each of them is based on the discrimination of one of the following features:

- frequency
- loudness
- tone length
- timbre
- melody
- rhythm

The latter two subtests require the subject to memorize the melody line or the rhythm pattern. Consequently, the tests not only assess music perception but also musical aptitude. The latter four tests are discrimination tests. The fourth test serves as a model for two tests of the M.U.S.I.C. test battery. The NIR and MID tests are based on the idea to measure the discrimination of timbre.

### The Zurich Music Test Battery

The Zurich Music Test Battery is designed for cochlear implantees. It consists of five subtests:

- pitch ranking
- instrument recognition
- melody contour identification
- harmony discrimination
- music quality test

Pitch ranking is performed with interval distances of semitones. It is an easy version of the first subtest of the Seashore test. The second subtest assesses the recognition of instrument timbres. This concept is used in the design of the NIR test (cf. section 4.1). The other four subtests are not relevant for the diploma thesis: The melody contour identification test and the harmony discrimination test are too easy for hearing aid users. The musical quality test is not assessing a subject's discrimination abilities but the enjoyment of music. Musical enjoyment tests are part of the music perception research project but are not part of the diploma thesis. Within the framework of the diploma thesis, only discriminative tests are designed.

# 4
# The M.U.S.I.C. Test Battery

The M.U.S.I.C. test battery is a tool to research the impact of hearing loss and hearing aid processing on the perception of music. It is an instrument to benchmark music processing algorithms.

The two dimensions of music perception are the discrimination and the enjoyment of music. Music discrimination refers to the discrimination of musical components like pitch[gl.], timbre[gl.], harmony and rhythm. Music enjoyment considers the perceived quality and the emotional effect of music. The final version of the M.U.S.I.C. test battery is to include discrimination and enjoyment tests. Within the framework of the diploma thesis, only music discrimination tests are developed.

The three following discrimination tests have been developed:

- NIR - Natural Instruments Recognition[4]
- MID - Midi Instruments Discrimination
- CCD - Chord Color Discrimination

The NIR test is designed for severe to profound hearing losses. The MID test focuses on the mild and moderate hearing impaired. Both tests are based on the discrimination of timbre. The CCD test focuses on frequency resolution. Hence, not the transient settling time characteristics but only the sound color is on trial. The CCD test intends to be applicable for the whole range of hearing losses.

**Remark** The assignation of a test to a certain degree of hearing loss is only a first approach and does not claim to be correct. The classification of hearing losses into the categories mild, moderate, severe and profound is only based on the averaged hearing thresholds. Although this proceeding is well-established in the hearing aid community, the data is not sufficient to assess a persons hearing disorder completely. Frequency and temporal resolution are not accounted for.

[4] As the subjects not only have to discriminate but to identify the instruments, it is called Natural Instruments *Recognition*. Nevertheless, NIR is a discrimination test.

## 4.1 NIR - Natural Instruments Recognition

### 4.1.1 The instruments

**Considerations:** As the timbre perception is tested by recognition, it has to be made sure, that the instruments are familiar to the subjects. Furthermore, a higher musical education should not be necessary to distinguish the instruments. A normal hearing person without musical practice should be able to recognize the instruments. To meet this criterion, two requirements have to be satisfied:

1. The timbre characteristics of the instrument can be derived from its shape and the way it is excited. Not every person might be familiar with a vibraphone for example. Nevertheless, normal hearing persons can identify its sound easily as the characteristic short attack time of the timbre corresponds to its metallic look and the way it is played.

2. The instruments have to be chosen from different instrument categories[gl.]. Generally, the categorization of instruments can be based on different aspects: e.g. handling, excitation method or the vibrating medium. Considering the requirement to recognize the instrument, the important criteria are the shape of the instrument and its excitation method. With choosing the instruments for the NIR test, these aspects have to differ for every item so that a distinction is possible without musical expertise. Vibraphone & xylophone or trombone & trumpet for example cannot be part of the same set.

To avoid that pitch is a discrimination cue, the instrument selection has to satisfy a further requirement: The ranges[gl.] of all instruments need to have a common cut-set. With meeting this demand, it is possible to create a score that can be played by all instruments.

**Outcome:** In coordination with all the requirements, the following eight instruments are chosen:

- cello (stringed chordophone)
- guitar (plucked chordophone)
- piano (percussive chordophone)
- vibraphone (idiophone)
- human voice (voice)
- trombone (brass aerophone)
- accordion (squeezebox aerophone)
- saxophone (woodwind aerophone)

**Remark:** Originally, the clarinet was chosen as the woodwind aerophone instrument. Contrary to the saxophone, the visible material is wooden and not metallic. According to above requirements, the look of the clarinet is consistent with its sound. However, the saxophone was selected instead of the clarinet. No clarinet player was available to record the instrument. The requirement, that the instruments belong to different categories is imposed to make sure that the timbres of the instruments are different. It is not important whether the instrument is the best representative of its category. Therefore, the saxophone can be considered to be an equally good choice for the NIR test. Its timbre is distinct and also very well-known in our society.

### 4.1.2 The score

#### The type of stimuli

**Considerations:** To test the recognition of instruments, different types of stimuli could be played back: singular tones, chords, monophonic or polyphonic melodies. The saxophone, trombone, and the human voice are monophonic instruments. Nevertheless, polyphonic melodies or chords can also be created by overdubbing or by means of several musicians performing simultaneously. In brass bands and choirs, the polyphony of the latter two instruments is very common. Polyphonic saxophone music exists, but it is rather unusual.
The effort to record polyphonic music for all eight instruments is relatively high. To avoid parasitic cues, all parts have to be played very precisely. In the case of overdubbing, the effort is multiplied by the number of parts. If several musicians play simultaneously, the effort might be even higher as all instrumentalist have to play faultlessly at the same time. It is already difficult for one musician to play his part accurately enough to satisfy the quality requirements for a timbre discrimination test. Moreover, the attack of polyphonic audio material is smeared. The discrimination is rather based on the sound and does not include the temporal resolution adequately.
Consequently, the polyphonic options are not suitable. A decision between the two alternatives - singular tones and monophonic melodies - has to be made. With using singular tones as test stimuli, the subjects might become bored very easily.

**Outcome:** Monophonic melodies are used as test stimuli. To make sure that the melodies don't favor any instrument, they have to meet two important requirements:

- The melodies are unknown.
  A well-known melody could unconsciously favor the instruments by which it was played in the original.
- The melodies are playable by every instrument.
  A fast vibrato for example is easy to be played with most of the instruments in the set, but very hard to be play with the trombone.

#### The scale of the melody

**Considerations:** All the instruments in our set are chromatic$^{gl.}$, i.e. they can play every semi-tones within their range. Consequently, any modes that are common in our Western music culture lend themselves to the composition of the melodies: the chromatic scale, the modern musical modes$^{gl.}$ or the natural harmonic series$^{gl.}$.

Chromatic melodies are not pleasant as test stimuli. Being annoyed or emotionally moved by the stimuli might influence the concentration and bias the results. Therefore, it is important that the subjects are neutral to the stimuli. Modern musical modes as well as natural harmonic series lend themselves to compose neutral melodies[5]. The choice of instruments decides about the applicability of the scales. Natural instruments cannot play other pitches than those of their harmonic scale. The current set of the NIR test does not contain any natural harmonic instrument. The instruments of the set can play both, the natural harmonic series and the modern musical modes. However, if a later version of this test should include natural harmonic

[5] Every modern musical mode or natural harmonic series is a subset of the chromatic scale. However, the expression *chromatic melodies* refers to a melody that proceeds in semi-tone steps.

instruments (e.g. alphorn), only the natural harmonic scale of the instrument is suitable. Nevertheless, with adding a natural instrument to the set, the problem of deviating pitches has to be considered. The frequencies of the tone that a natural instrument produces does not match the frequency that a diatonic instrument$^{gl}$ produces playing the same tone. The diatonic instruments are equal-tempered. The Western musical notation system is based on the equal temperament. The octave is composed of 12 semi-tone steps whose interval is defined by the same frequency ratio $1 : \sqrt[12]{2}$. The frequency ratios of the notes of the natural harmonic series, however, are the quotient of the respective harmonic numbers. Consequently, as the ratio must be rational, it differs from the irrational ratios of the equal temperament. The deviation of the harmonic pitches from the equal tempered pitches is indicated by the numbers in the upper section of figure 2.7 on page 17. The corresponding unit is cents. 100 cents equal the interval of one well-tempered semi-tone, 1 cent equals the ratio $1 : \sqrt[1200]{2}$. The note heads of harmonic number $7, 11, 13$ and 14 in figure 2.7 are not filled as their frequencies deviate significantly from equal temperament.
The pitch deviation could be used as a discrimination cue if natural harmonic and equal-tempered instruments are used. Moreover, if more than one natural instrument is included, they need to have the same tune. Otherwise, the common cut-set of the score is minimized. A Swiss alphorn in *Fis* and a natural horn in *F* for example do not have one tone in common[6]. Consequently, with introducing natural harmonic instruments to the set, the effect of the pitch deviation must be observed. With choosing the right key of a natural harmonic series and of the modern musical mode, however, the pitch deviation is not too critical. Thus, adding a natural harmonic instrument to the set might be an option for future research.

To specify the register$^{gl.}$ of the scale, the common cut-set has to be determined. Figure 4.1 displays the compasses$^{gl.}$ of the eight instruments.

Depending on which type of saxophone or singer is used, the common cut-set of the eight instruments ranges from *F2* (GNS: *F*) to *A4* (GNS: *a'*).

**Outcome:** A modern musical scale and a natural harmonic scale are chosen for the composition of the melodies. The compass from *F2* (GNS: *F*) to *A4* (GNS: *a'*) can be used to compose the melodies.

### The melodies

**Considerations:** It is reasonable to compose several melodies in order to prevent learning effects. With having eight different melodies, each instrument can be randomly played back with a different melody. Thereby, it has to be ensured, that the recognition cues of all melodies are equal. The following rules for the design of the melodies have been established:

1. Each melody consists of two bars.
2. The same five pitches are used for every melody.
3. Each melody consists of seven notes, thus two pitches are duplicated (see rule 2). In the following, the duplicated pitches are called double-pitch, the non-duplicated pitches are called single-pitch.
4. The pitch of two succeeding notes are always different.
5. Only eighth, quarter and dotted quarter notes are used.

[6] Nevertheless, both instruments have several *notes* in common, e.g. the harmonic 13 of the *F* scale and the harmonic 12 of the *Fis* scale. The frequencies of those notes, however, differ significantly.

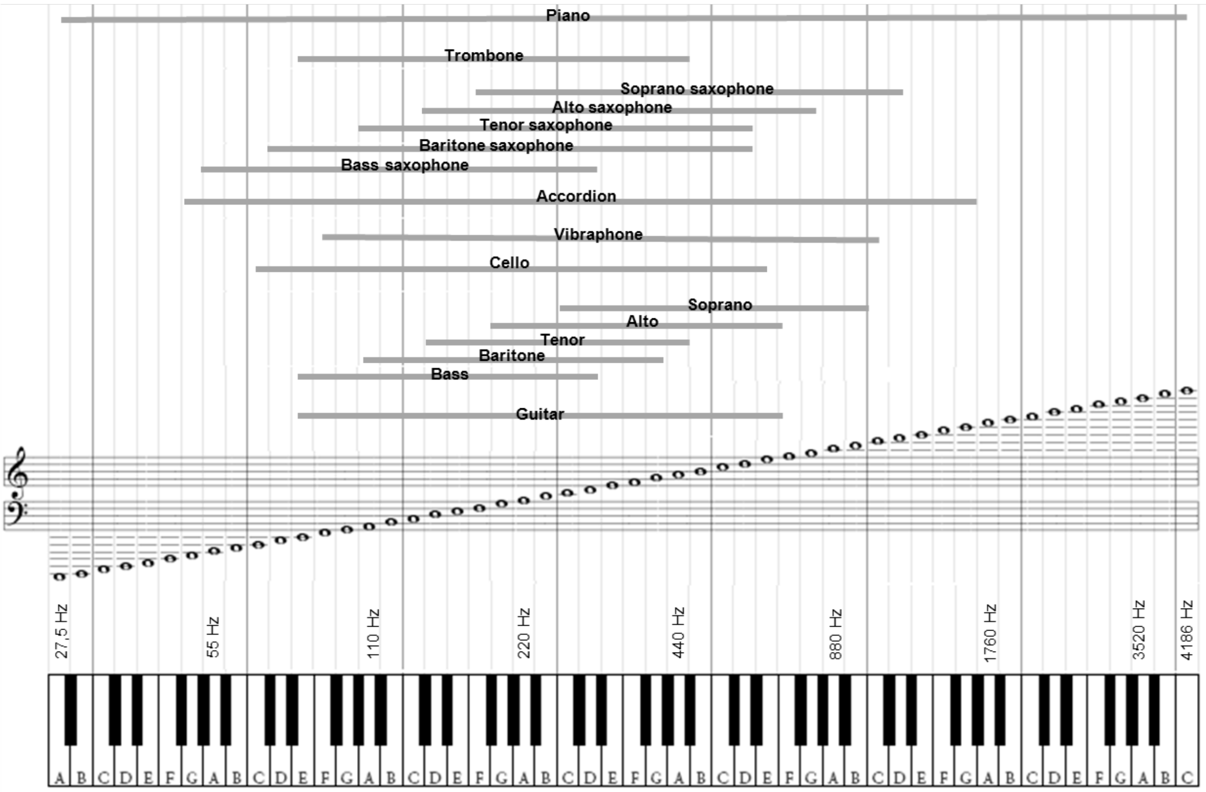


*Figure 4.1: The range of the instruments that are used in the NIR test.*

6. The length of the first note is not less than a quarter note.
7. The last note of the melody is always a dotted quarter note.
8. Exactly two eighth rests are embedded in the melodies.
9. The note prior to each eighth rest is an eighth.
10. One of those eighth notes of rule 5 is a double-pitch note, the other one is a single-pitch note.
11. The length of the second double-pitch note of rule 10 is at least a quarter.
12. One accentuation is embedded in each melody. It is applied to the second double-pitch pair that is not used in rule 7. The accentuated note must not be at the beginning or the end of the melody and has to be at least one quarter long.

**Outcome:** In coordination with all the rules, eight different melodies have been developed for each of the two tonality systems, the modern musical mode and the natural harmonic series. With regard to the latter system, the fourth till eighth harmonic of the *Fis1* (GNS: *'Fis*) key were selected (cf. figure 4.2).*Fis1* (GNS: *'Fis*) is the key of the Swiss alphorn. Consequently, the alphorn and any other natural instruments with the same key could be added to the NIR set in a later version. Concerning the modern musical mode, the first five notes of the *A3* (GNS: *a*) Ionian / major key were chosen (cf. figure 4.3). The key of the scale was determined such that the selected pitches of both scales are in the same register[gl.]. The melodies of both systems are displayed in figure 4.4.

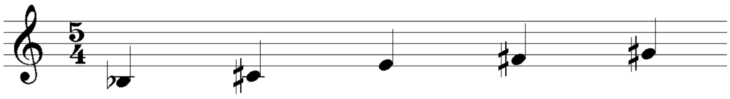

*Figure 4.2: NIR test: The fourth till eighth overtone from the* Fis1 *(GNS:* 'Fis*) harmonic series.*

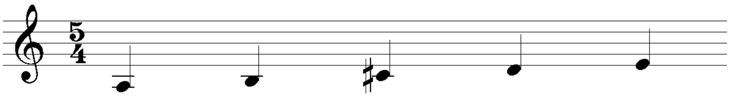

*Figure 4.3: NIR test: The first five pitches from the* A3 *(GNS:* a*) major scale.*

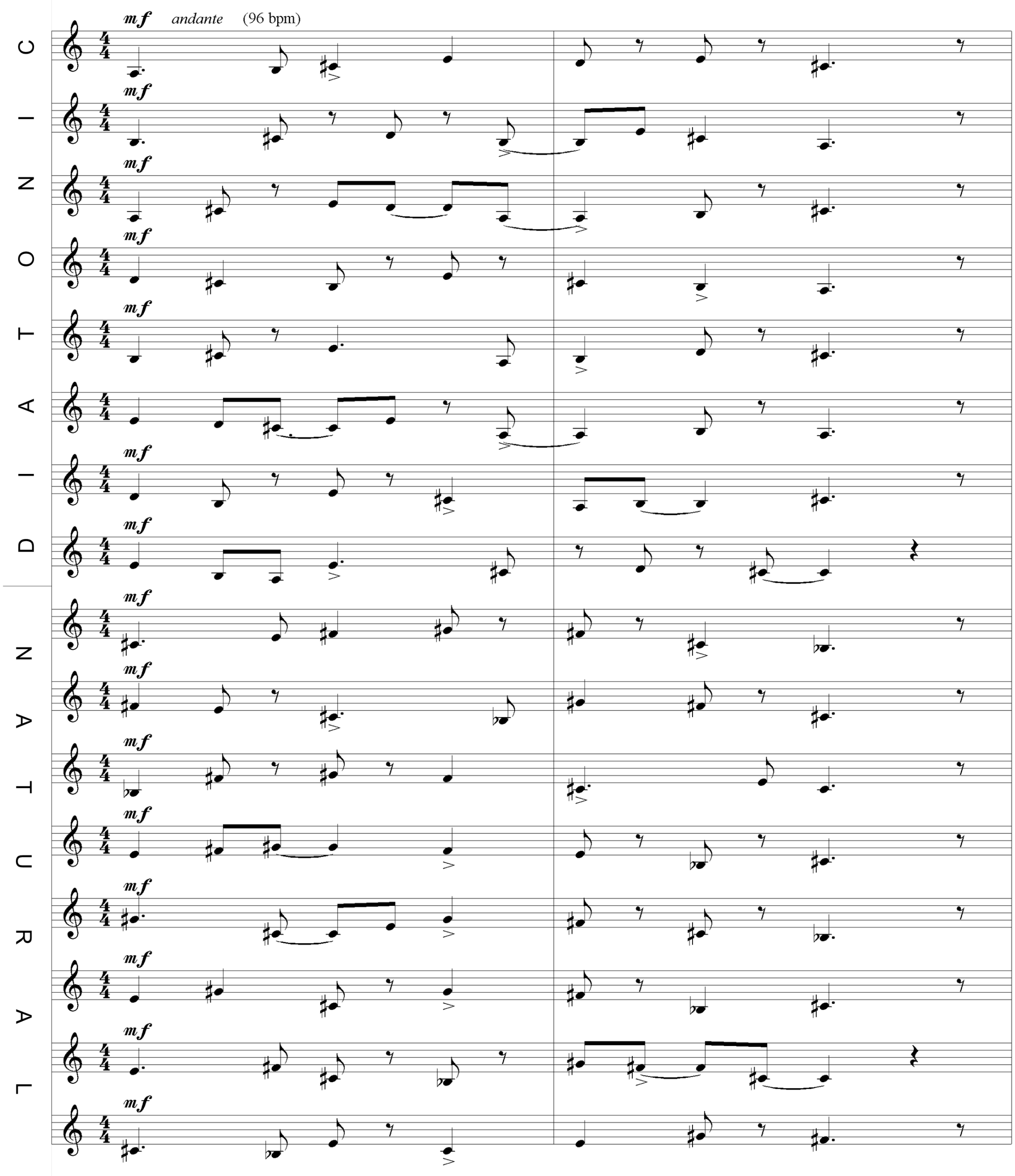


*Figure 4.4: The score for the NIR test: eight diatonic melodies and eight melodies of the natural harmonic series melodies.*

### 4.1.3 Generating the audio material

#### Discussion: MIDI vs. live

To generate the audio material, either natural instruments or MIDI instruments can be used. MIDI software stations either work with sample banks of instrument recordings or with electronically synthesized sounds. The latter option is not suitable due to the poor sound quality. Using sample banks of instrument recordings is an alternative if the quality is sufficient. One criterion for the quality of the sample bank is the extent of its recordings. Low-grade work stations may only have recordings for a subset of pitches and generate the missing pitches by means of pitch shifting. Moreover, the effect that the sound color of an instrument changes with volume might not be taken into consideration. In many cases, one single recording is mapped to all different volume levels by simple amplification. This method as well as the pitch shifting reduce the authenticity of the instruments' timbres.
As a further criterion, the neutrality of the instruments' timbre has to be checked. Even for high standard sample banks, it is common to highlight the authenticity of a sample by emphasizing the parasitic noises. The scratching sound of a cello or the sliding noises of a guitarist for example are desired to a certain degree. Although these effects are part of a natural instrument, the parasitic noises must be avoided. Their discriminative cue is too strong, so that the actual task to discriminate the instruments based on their timbre would be undermined. Regarding the aspect of authenticity, it has to be considered that most of the music we listen to is not live but playback. The playback material is digitally mastered and any parasitic noises are removed. Consequently, we are used to the edited sounds and consider them authentic without their natural noises.

The sample bank that was available (*Sample Tank 2 L* [18]) does not meet the demands for all the instruments on the short list. A more sophisticated sample station might be an alternative. However, even the most distinguished state-of-the-art sample bank does not compensate for another quality feature: the transitions. An instrument vibrates with a characteristic pattern when a tone is produced. This vibration pattern influences the settling time of the subsequent tone. Moreover, the play of a musician varies according to the succession of tones. The timbre of a note depends on its musical context.

As a consequence, the melodies of the NIR test have been recorded.

#### Recording

The piano, accordion and guitar were played by the author. To record the other instruments, professional musicians were recruited within the Phonak company, the Zurich opera orchestra and the Zurich University of the Arts. The recording of the voice, the trombone and the cello are especially difficult tasks. In contrast to the other instruments, these instruments do not produce discrete pitches but continuous pitches. Consequently, the performers instrumental skills have to be very sophisticated in order to hit the defined pitch exactly.

As the recording locations varied, the miking had to be very close to the acoustic source to emphasize the direct signal and to avoid any spatial cues. By this means the recorded signal is dry but unfortunately, the parasitic noises like the sliding of the guitar or the scratching from the cello bow are more prominent. Those unwanted parts of the signal are generally high frequency components. The high frequency components are more dampened by air than the broadband target signal. The closer the miking is to the instrument, hence, the more prominent are the parasitic noises. Therefore, the instrumentalists were required to play very accurately.

**Remark:** From a conceptual point of view, the generation of the audio material is a small step. Regarding the work effort, it is immense. The reason therefor is the fidelity that is required to avoid parasitic cues like ambient noise, playing errors, differing spatiality or stereo position$^{gl.}$ and any pitch inaccuracies.

### 4.1.4 Audio editing

**Stereo panorama** A stereo signal is preferred to a mono signal. Firstly, its sound quality is better. Secondly, we are more accustomed to listen to stereo recordings than to mono recordings nowadays.
It is reasonable to play back the stimuli via two loudspeakers arranged according to the stereo triangle (cf. figure 5.1). This setup is a standard in the audio engineering community. The playback via earphones or headphones is not an alternative. With being tested while wearing a hearing aid, the earphones do not fit into the auditory canal anymore. Headphones would cover the hearing aids and occlude them acoustically. Consequently, the test stimuli must be played back via loudspeakers. Thereby, the stereo position of the instruments might differ. The perceived location and width of the instrument within the loudspeakers could serve as discrimination cue. To avoid that the position the signals in the stereo panorama is different, the following two steps have been applied to all recordings:

1. All files are mixed down to mono.
2. The mono files are filtered with the same stereo reverb.

By this means, the resulting test stimuli are still stereo but their position is identical.

**Reverb** A compromise has to be made concerning the strength of the reverb. On the one hand, it should be minimal to avoid excessive smearing of the transient effects during the settling phase. On the other hand, it must be strong enough to cover the parasitic noises from the instrument. By means of subjective evaluation, it is ensured that both criteria are met.

**Loudness** As already mentioned, the NIR test is a timbre discrimination test. Timbre is any aspect of sound, that is not intensity or pitch. Consequently, any discriminative cue related to intensity must be avoided. With applying the Zwicker model [23], the loudness of the stimuli can be equalized for normal hearing people. Hearing impairments, however, affect the perception of loudness. A hearing impaired person might perceive loudness differences between test stimuli that are perceived as equally loud by normal hearing people.

As a countermeasure, level roving is applied. Before being played back, the signals are amplified by a random factor. With regard to the range of the amplification factor, a trade-off has to be found. On the one hand, a high range reduces the probability for a subject to use loudness as a discriminative cue. On the other hand, it must be considered that the perception of sound color is not totally invariant to the intensity. With the intensity increasing, the human auditory system emphasizes the bass and high frequency components relatively to the mid frequencies. Moreover, the instruments change their timbre subject to the intensity. Furthermore, the hearing aids signal processing and the sound depends on the input volume. The signals might not be audible if the amplification factor is too low and the subjects are tested without wearing a hearing aid.
Having conducted preliminary tests, a level roving range from $[-1.5\,dB, 1.5\,dB]$ has proved to be sufficient for the NIR test. The range is comparable to the perceived level differences caused by a subjects' head movements during the test.

### 4.1.5 Methodology and implementation

The methodology of the test is simple and straight-forward. The subjects listen to a stimulus and have to recognize the instrument.

#### Test units

To counteract a possible learning effect, an introduction round and a training round is carried out prior to the test round. During the introduction round, each instrument is played back once and the respective instrument is being marked simultaneously. This step ensures that the subject can recall the less common instruments like the vibraphone for example.
The training round and the test round are similar. The only difference is that a feedback to the response is only provided in the training.

#### Playlist

**Length:** During the test, the 64 combinations of the eight instruments and the eight diatonic melodies of the modern musical modes are played back. The eight melodies of the natural harmonic series could also have been chosen instead of the diatonic scale. With testing both series, however, the duration of the test would be too long.
The decision to test all 64 combinations has a certain reason. As described in section 4.1.3, a large number of design principles is applied to ensure that the amount of timbre cues within each melody is equal. With testing every combination, it can be checked in retrospect whether this criterion is actually satisfied.

**Order:** As preliminary studies showed, playing back the different instruments with the same melody causes a bias. The subject gets accustomed to the melody, the full attention can be drawn to the discrimination of timbre and the discrimination capability increases with every consecutive stimulus. This effect can be avoided by implementing that the melody and the instrument of two succeeding test stimuli are always different.

Apart from the rules that the melody and instrument of consecutive stimuli are always different, a further constraint regarding the order of the stimuli is implemented. The 64 stimuli are subdivided into eight consecutive blocks that contain eight stimuli each. Within these blocks every instrument and every melody is presented once. The presentation order within the block is random. This design allows to retrace the learning effect. Block by block, the hit rates can be accumulated for every subject and the performances can be compared over time[7].
Another bias has been taken care of. The subjects tend to avoid choosing the same instrument in a row although they might have perceived its timbre correctly. Therefor, it is programmed that the melodies and instruments of adjacent blocks, e.g. stimulus 8 from block 5 and stimulus 1 from block 6 are always different.

A counter-argument for the block-design has to be mentioned: The instrument of the eighth stimuli of each block can be deduced from the previous seven stimuli. Nevertheless, this possible bias implies that

- the logic regarding the subdivision into blocks is detected by the subject.
- the subject can remember the previous instruments.

[7] Had the files been completely randomized, the comparability of the instruments' learning curves would be limited. One example: The piano might be played at number 2, 6, 25, 27, 33, 44, 52 and 57. The guitars positions could be 49, 51, 53, 55, 58, 60, 62 and 64. The learning effect of the guitar might be very different to that of the piano just because of the positioning.

- the subject recognizes the vast majority of the preceding samples correctly. In this case, the subject is not likely to benefit from having detected the block pattern. The subject's decision rather relies on the perception and the discriminative capacity.

Consequently, a bias related to the block division is not to be expected.

### Layout

A snapshot of the NIR test is displayed in figure 4.5.

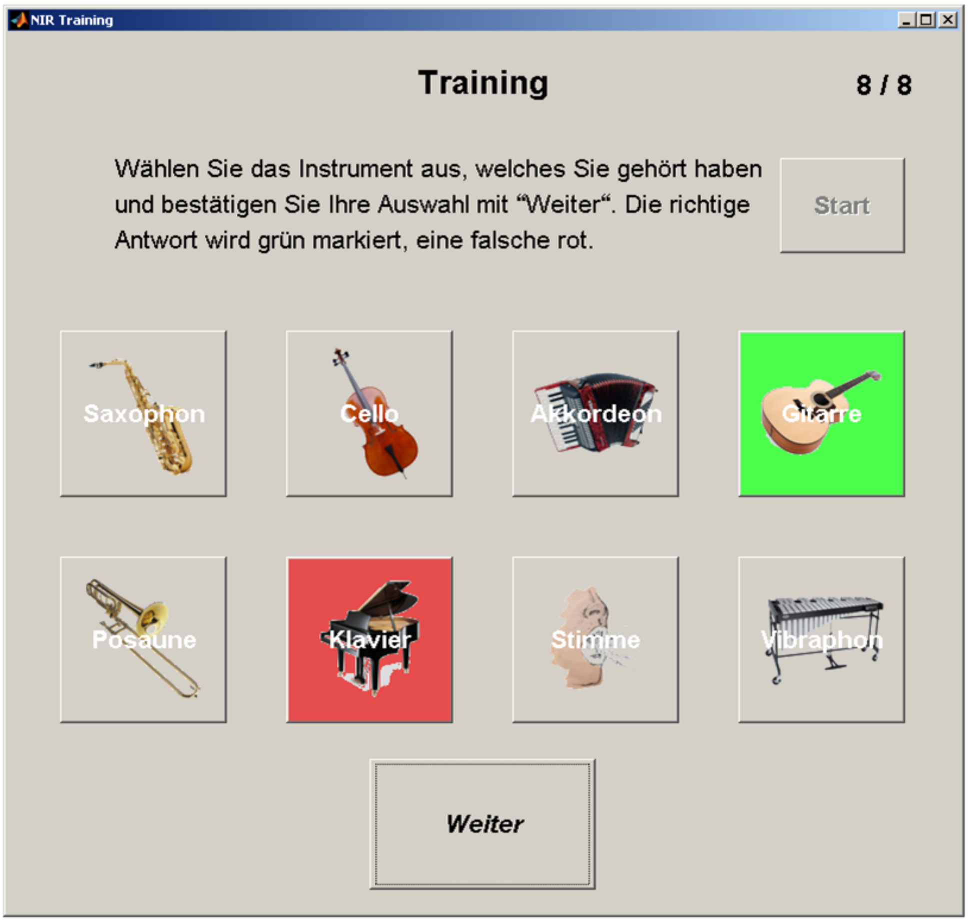


*Figure 4.5: Snap-shot of the NIR test, training round.*

The buttons in figure 4.5 not only contain the name but also a drawing of the instruments. The sizes of the illustrations are identical although the sizes of the instruments differ. The feedback of a subject of a preliminary study indicated that different sizes of the pictures can cause a bias. Louder and fuller sounds are rather assigned to the instrument with the larger illustration. To avoid that bias, the size of all pictures is identical.

### Procedure

A test stimulus is played back once. The subject has to choose one instrument and confirm the selection. As mentioned above, the subjects are provided with a feedback in the training round. The correct response button is highlighted in green, an incorrect response is marked in red. In the test round, no further feedback is given in order not to influence the subjects. One complete run including the introduction round and the training round takes 20 minutes on average.

## 4.2 MID - Midi Instruments Discrimination

### 4.2.1 The instruments

**Considerations:** The MID test intends to be sensible for mild and moderate hearing losses. Analogous to the NIR test, it is a timbre discrimination test. Compared to the instruments of the NIR test, the timbres of the MID instruments have to be more similar to meet the adequate degree of difficulty. In contrast to the NIR test, the instruments are not chosen from different categories but from the same.

To generate the test stimuli, the instruments' ranges are required to have a common subset. Otherwise the instruments cannot play the same score. The following four instrument sets meet the requirement:

1. saxophone, oboe, bassoon, clarinet (category: woodwind)
2. trumpet, trombone, tuba, French horn, cornet (category: brass wind)
3. violin, viola, violoncello, contrabass (category: string bowed)
4. guitar, harp, harpsichord, zither (category: plucked string)

Before making a choice, the qualities of both audio generating methods, live recording and MIDI programming, must be reevaluated. Their advantages and disadvantages have been discussed earlier in section 4.1.3. With having made the experience of generating the audio material for the NIR test, new arguments have to be considered. The recording and audio editing for the NIR test has turned out to be a highly time-consuming task. The aspect time efficiency is a determining factor. Although MIDI programming cannot achieve the authenticity of live recordings, this method is chosen to generate the test stimuli of the MID test. Therefor, the work station *Sample Tank 2* has been used. The quality of the instruments' sounds has been inspected by means of subjective evaluation. The quality of most instruments of the categories brass wind (2) and string bowed (3) is not satisfactory. The quality of the plucked string instruments (4) is excellent, however, the sounds are not suitable for the MID test. Parasitic noises like the sliding of the guitar are very prominent. As discussed in section 4.1.3 this property is counterproductive for a timbre discrimination test. The sound quality of the woodwind instruments is fair and appropriate for the MID test. Nevertheless, several pitches of the oboe stand out because a clicking sound is very prominent. As mentioned above, this noise is nonconstructive to test timbre. With avoiding these singular pitches, however, the instruments of the woodwind category are qualified for the MID test.

**Outcome:** The following four woodwind instruments are used in the MID test:

- saxophone
- oboe
- clarinet
- bassoon

The audio files are generated with the MIDI work station *Sample Tank* 2 *L* [18].

### 4.2.2 Concept and test methodology

Contrary to the NIR test, the MID test cannot assess the perception of timbre by means of *recognizing* the instruments. The instruments do not meet the necessary requirements (cf. page 28). They belong to the same category and their timbre is similar. Consequently, a considerably high musical expertise is needed to assign the stimuli correctly to the instruments.
The discrimination of the instruments' timbres can be tested, however, by means of comparison. In the following, the applicability of several comparison methods for the MID test are discussed.

#### Paired comparison

A pair of stimuli is played back and the subject is to decide whether both stimuli belong to the same instrument or not. The success probability by guessing is 50 %.

To modify the forced choice task into an unforced choice task, the option *don't know* could be added. The reliability of the results is increased with this modification [19]. However, it must be considered that the subjects react differently to this response option. Ambitious characters tend to avoid it. They rather accept the chance of being wrong than not being able to accomplish the task perfectly. The subjective threshold to decide for a definite response or to choose *don't know* is different for each person. Thus, the results are distorted as not only the subject's perception but also the personality determines the measurement results.

The test duration of the paired comparison method is one criterion that determines the applicability for the MID test. To estimate the duration of the test, the number of stimuli presentations serves as a good reference.
The four woodwind instruments can form six different pairs of two different (heterogeneous) instruments. Each of the heterogeneous pairs should be tested at least six times to obtain a resolution that is fine enough to rank the subjects' differences in timbre discrimination. The same amount of homogeneous pairs has to be added so that both response options, *the instruments are equal* and *the instruments are different* are pari passu.
Concerning the playback of the pairs, two manners have to be discussed. Either the audio files are played back a defined number of times or the subject can listen to the files as often as desired. In the latter case, a bias would be imposed. Ambitious subjects might listen to the samples more often and increase the probability to perceive differences. Although their perceptive capabilities may not differ from another person with a comparable hearing disorder, their higher commitment might lead to better results. Consequently, the first option with a defined number of playback is preferred because no personality bias is introduced. With playing back the pairs only once, a temporal lack of concentration is likely to distort the results. A second playback of the files is considered to be enough to avoid the concentration bias. More repetitions would unnecessarily prolong the test. Nevertheless, the pairs do not stringently have to be played back twice. If the subject is concentrated and ready to make a decision after the first presentation, a response can be given without listening to the second presentation. In the worst case, however, the number of presented stimuli adds up to $6(\text{pairs}) \cdot 2(\text{hom/het}) \cdot 2(\text{stimuli}) \cdot 6(\text{repetitions}) \cdot 2(\text{trials}) = 288$ stimuli.

Carrying out the test, each subject provides 72 responses. One half of the responses belongs to the heterogeneous stimuli pairs and the other half belongs to the homogeneous pairs. To evaluate the subjects' performance, only the heterogeneous stimuli pairs are relevant[8]. Consequently, the

[8] The homogeneous pairs do not contribute to the hit rate. Still, they can be used to scan whether the subjects' responses are reliable or not. If a homogeneous pair is played back and the subject does not choose the response options *instruments are different*, it can be assumed that the subject guesses.

evaluation of the subjects' performance is based on 36 responses.

### Three-alternative forced choice (3AFC), Version I

An alternative to the paired comparison method is the 3AFC method: Three samples are presented, two of them belong to the same instrument, one sample belongs to a different instrument. The order of the differing instrument is randomized. It can be the first, second or third stimulus. The subject is to detect, which of the three instruments is different. The guess probability equals 33 %.

To modify the forced-choice into an unforced-choice task, the response option *don't know* is added. The subjects do not have to guess if they cannot detect the different instrument. Thus, the reliability of the responses is increased, however, a new bias is introduced. As mentioned above, the subjects' threshold to use the *don't know* response option or to try guessing is different. The results do not exclusively depend on the person's perception but on the attitude as well.

The test duration of the three-alternative methods is one criterion that determines the applicability for the MID test. To estimate the duration of the test, the number of stimuli presentations serves as a good reference.
Six different heterogeneous combinations can be composed. These six combinations result in 12 different stimuli triples as either the first or the second instrument of the combination pairs can be doubled. Nevertheless, both versions of a triple are equivalent regarding the subjects' requirements to discriminate the instruments' timbre. Therefore, only six triples are distinguished. To gain an adequate resolution of the subjects' discrimination performance, each triple should be tested at least six times. In order to avoid measurement errors due to a temporal lack of concentration, it is reasonable to allow one playback repetition for each trial. Totally, the maximal number of stimuli presentations sums up to $3(\text{triple}) \cdot 6(\text{combinations}) \cdot 6(\text{repetitions}) \cdot 2(\text{trials}) = 216$ stimuli.

Each subject provides six responses for each of the six stimuli triples; 36 responses altogether.

### Three-alternative forced choice (3AFC), Version II

A modification of the 3AFC method can be created by adding homogeneous triples[9] to the stimuli set and a fourth response option *All instruments are the same*. If the test is set up as a forced-choice task, the guess probability is reduced to 25 %. If the response option *don't know* is added, the reliability of the responses is further increased.
With adding the homogeneous triples to the test set, the duration of the test is prolonged. One in four triples must be homogeneous so that all response options, *A is different*, *B is different*, *C is different* and *All instruments are the same*, are considered equal[10]. With adding the homogeneous triples to the test stimuli set, the maximal number of presented stimuli rises by $216 \cdot (1 + \frac{1}{3}) = 288$.

[9] In the following, the term *homogeneous triples* is used for a triple that contains three identical stimuli.

[10] A lower proportion of homogeneous triples might also be sufficient. The important aspect is not that the distribution of the triples is balanced, but that the subjects consider it to be balanced. Otherwise, the decision might be biased in hardship cases.
A small deviation of the ratio in favor of the heterogeneous triples might not be noted by the subjects. It is the objective of the test to have a degree of difficulty so that timbre differences can not always be perceived by the subjects. Consequently, the option *All instruments are the same* might be chosen more often than there are actually homogeneous triples in the test material.

Carrying out the test, each subject provides 48 responses: six results for each of the six heterogeneous[11] stimuli triples and one third of that amount for the homogeneous triples. To evaluate the discrimination performance, only the heterogeneous stimuli pairs are relevant[12]. Consequently, the evaluation of the subjects' performance is based on 36 results.

### The 2n+1 comparison method

As a further alternative to the paired comparison and the three-alternative methods, a new test method has been specifically created for the MID test: *2n+1 comparison.*
In each trial, $n+1$ different audio samples are presented: the reference instrument and $n$ test stimuli. Each of the $n$ test stimuli holds one different instrument of the instrument set in random order. The subject is to detect which of the instruments of the $n$ test stimuli corresponds to the reference instrument. The test stimuli and the reference are played back alternately:

*reference - test stimuli 1 - reference - test stimuli 2 - reference - ... - test stimuli n - reference*

The line-up starts and ends with the reference. By this means, each test stimuli can be compared twice with the reference instruments, once with the preceding and once with the succeeding one. Therefore, it is not necessary to provide the opportunity to repeat a trial. Given the fact that the $n$ test stimuli are embedded in $n+1$ reference files, the methodology is called *2n+1-comparison.*

The success probability by guessing is 25 %. The reliability of the responses could be further increased by adding the *don't know* response button. To evaluate the advantage and disadvantage of this supplementary response option, please refer to page 38.

To obtain sufficient results to evaluate the subjects' performance, each instrument should be put under test six times at least. The woodwind set contains 4 instruments. According to the line-up, $4 \cdot 2 + 1 = 9$ stimuli are played back in each trial. Altogether 9(stimuli) $\cdot$ 6(repetitions) $\cdot$ 4(instruments) = 216 stimuli are presented. Each subject provides 24 responses.

### Summary

To give an overview regarding the qualities of the discussed test methods, table 4.1 and table 4.2 compare the relevant attributes quantitatively and qualitatively.

| | # presentations | guess probability | # responses |
|---|---|---|---|
| paired comparison | 144 – 288 | 50 % | 36 |
| three-alternative, version I | 108 – 216 | 33 % | 36 |
| three-alternative, version II | 144 – 288 | 25 % | 36 |
| 2n+1 method | 216 | 25 % | 24 |

*Table 4.1:* Quantitative *comparison of different methods regarding duration (# presentation), reliability (guess probability) and resolution (# responses) of the MID test.*

[11] In the following, the term *heterogeneous triples* is used for a triple of two identical stimuli and one different stimulus.

[12] The homogeneous triples of the three-alternative method do not contribute to the hit rate. Still, they can be used to scan whether the subjects' responses are reliable or not. If a homogeneous triple is played back and the subject does not choose one of the response options *don't know* or *All instruments are the same*, it can be assumed that the subject guesses.

| | # presentations | guess probability | # responses |
|---|---|---|---|
| paired comparison | o | - | + |
| three-alternative, version I | + | o | + |
| three-alternative, version II | o | + | + |
| 2n+1 method | o | + | o |

*Table 4.2:* Qualitative *comparison of different methods regarding duration (# presentation), reliability (guess probability) and resolution # responses) for the MID test. The three grades are: (+) first, (0) second, and (-) third place.*

Before deciding on which method to apply, the following two aspects have to be taken into consideration:

- Not only the amount but also the information content of the responses must be evaluated. Concerning the first three methods, the subjects discriminate only between two different stimuli. With regard to the *2n+1* method, a test stimuli is compared to four instruments within each trial. Consequently, the information content of the latter responses is considered to be higher.
- A subject approximately needs one and a half hour to pass the M.U.S.I.C. test battery including the NIR, MID and CCD test. To avoid boredom and exhaustion, it can be helpful to use different methodologies for each subtest of the battery. The three-alternative method is already applied in the CCD test (cf. section 4.3.4).

According to the qualitative comparison, both three alternative methods rank first. After considering the supplementary arguments above, the $2n + 1$ method is used in the MID test.

### 4.2.3 Creating the test material

Analogous to the considerations in section 4.1.3, monophonic melodies are a reasonable choice as test stimuli for the MID test. Certain requirements have to be met: The melodies must be unknown so that the subjects do not unconsciously favor the instrument that played the melody in the original. Moreover, the tempo and rhythm of the melody has to be ordinary for all the instruments. To ensure that the discrimination cues of the melodies are similar, all the requirements that are enumerated on page 30 have to be met.
Due to the specific line-up of the reference (cf. page 40) and the test stimuli, one further constraint is imposed:

- The last note's pitch of the preceding melody and the first note's pitch of the succeeding melody must be different.

As the instrument samples are played back in a row without interruption, the first note of the succeeding melody follows directly the last note of the preceding melody. If the pitches of those two notes are the same, the timbre and sounds can directly be compared. The discrimination of two instruments is easier if they play the same pitch. The subject can fully concentrate on the perception of timbre and is not distracted by different pitches. Consequently, a bias would be introduced. Therefore, it has to be made sure, that the last note's pitch of the preceding melody is always different to the first note's pitch of the succeeding melody.
To avoid a learning bias, several melodies are composed. The subject is less likely to recognize possible parasitic noises of the instruments if several melodies and test files are available for each instrument. As a consequence, the discrimination is not likely to be based on the detected abnormalities but on the perception of timbre. To render the randomization of the playback order possible with keeping the constraint from above, the following criteria have to be met:

- The first pitch and last pitch of a melody are identical.
- Every melody of the set starts with a different pitch.

As one trial of the MID test contains five different audio samples, a five pitch scale is sufficient. The first five functions of a major scale have been used[13]. Two sets of melodies have been composed. By this means, a randomization of the melodies is possible not only within but also between the melody sets. As mentioned in section 4.2.1, a MIDI software has been used to generate the audio files. Due to sound anomalies of certain pitches as described in section 4.2.2, the *A3* (GNS: *a*) major diatonic scale from the NIR test could not be carried over. Instead, the *B3* (GNS: *h*) major diatonic scale was used. The two melody sets are depicted in figure 4.6.

**Remark:** The fact, that the scales and pitches from the NIR and MID test are not identical is not a problem. The decisive aspect is that the scale is part of the subset of the instruments ranges. The criterion is met by the *B3* (GNS: *h*) major diatonic scale.

### 4.2.4 Implementation

#### Test units

Before the actual test starts, the subject have to get familiarized with the handling of the test. Besides, the learning curve should ideally have reached its saturation so that no training effect has to be accounted for in the test round. For both reasons a training round is connected ahead to the test round.

#### Playback

**Length:** Within the test round, each of the four woodwind instruments, saxophone, oboe, bassoon and clarinet, is presented six times as the reference. Totally, $4 \cdot 6 = 24$ references are on trial. As one trial includes 9 stimuli, $4 \cdot 6 \cdot 9 = 216$ stimuli are presented.

**Order:** The 24 trials are divided into six successive blocks. Each block contains all four different instruments as reference stimuli. The order of presentation within each block is randomized. The melodies of the set *M* (cf. figure 4.6) are randomly mapped to the reference and the four test stimuli before each trial. With having a second melody set *N*, it is also randomized if a melody of the *M* set is replaced by its corresponding substitute of the *N* set, e.g. *N4* replaces *M4*.

**Remark:** The division into blocks seems to have a disadvantage: the fourth reference instrument of one block can be derived from the previous three trials. Analogous to the discussion in section 4.1.5, however, the pattern is not likely to be discovered. The previous instrument would have to be detected correctly and in that case, the subject rather relies on the perception than on the assumed pattern. Besides, knowing the reference instrument does not help to accomplish the task. The subject does not have to *name* the reference instrument; the subject has to *detect* the reference instrument within the four test stimuli. The order of the test stimuli, however, is randomized. Consequently, the subdivision into the six blocks does not introduce a bias. As explained in section 4.1.5, the partition into blocks is necessary to keep track of the learning effect.

[13] Any other modern musical mode or the natural harmonic scale are adequate as well.

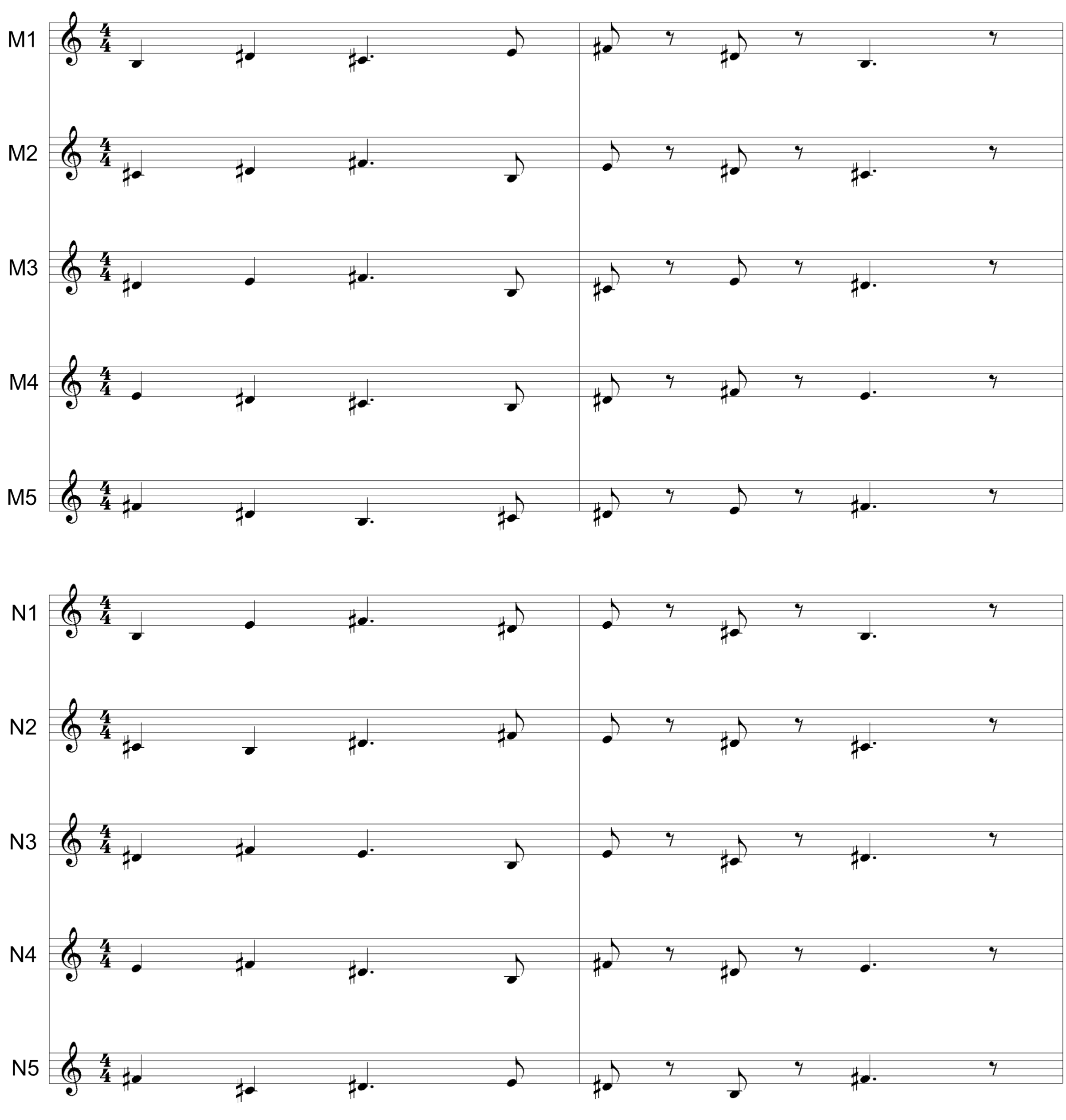


*Figure 4.6: Score of the MID-test: Two sets of melodies (M1-M5) and (N1-N5) are depicted. Every melody of a set starts with a different pitch. The pitch of the first note and the last note of a melody is identical.*

### Procedure

Analogous to the NIR test, the subject is provided with feedback in the training round. The correct response button is highlighted in green, an incorrect response is marked in red. During the test round, no further feedback is given. One complete run of the test takes 25 minutes on average.

The layout of the MID-test is depicted in figure 4.7.

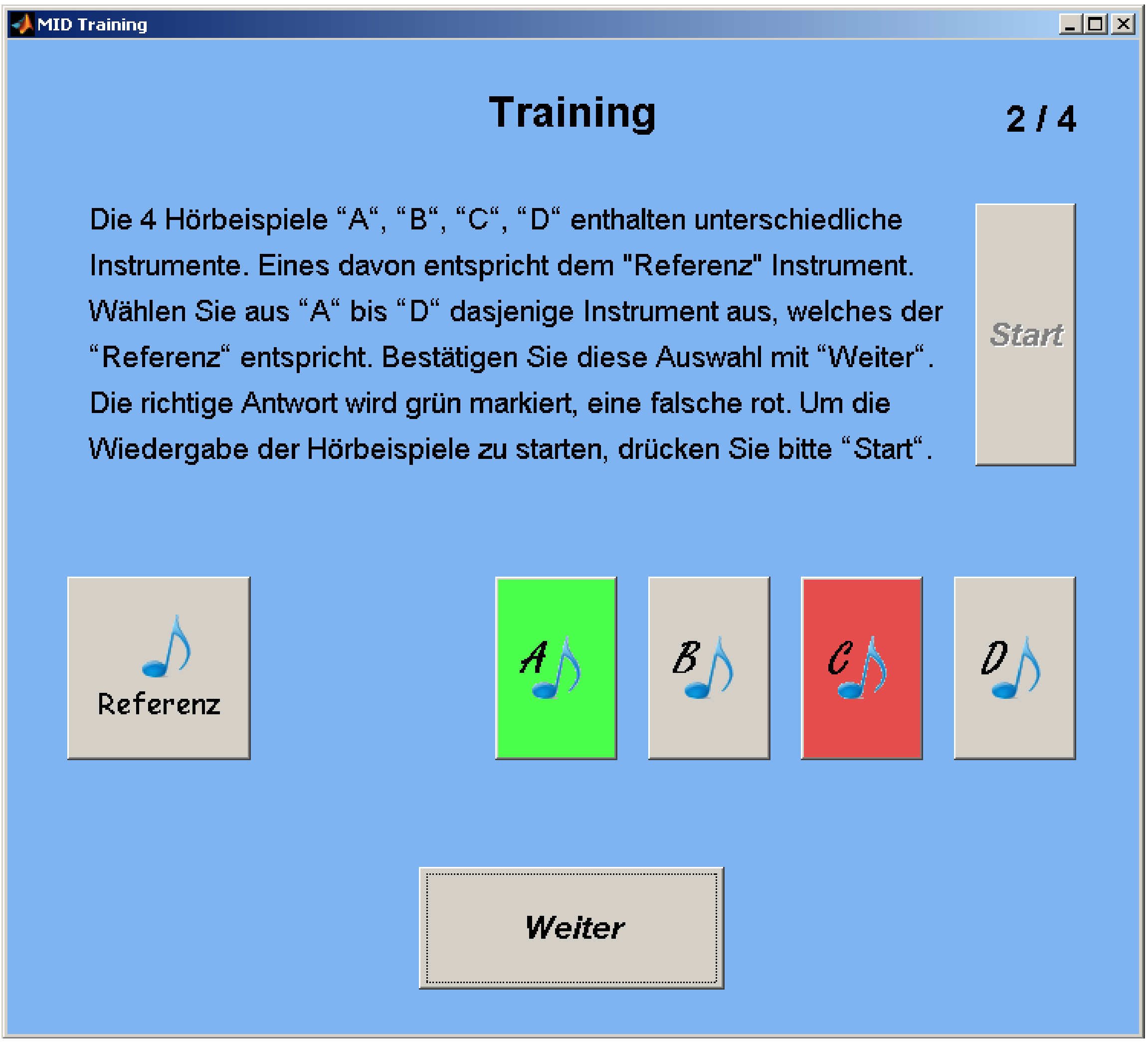


*Figure 4.7: MID-test: snapshot of the training round.*

## 4.3 CCD - Chord Color Discrimination

### 4.3.1 Instruments

The CCD test focuses on the discrimination of sound color. It intends to test a subject's frequency resolution of a musical stimulus. To create audio samples that differ in sound color but not in other aspects of timbre, the same instrument is used for all test stimuli. As the sound color cannot only be varied by intonation but also by modifying the arrangement of a harmony, the design choices are more flexible if a polyphonic instrument is used. Monophonic instruments can also produce polyphonic sounds if multiple instruments are used. Nevertheless, to develop ideas and concepts, it is practical to work with a polyphonic instrument. No further musicians are needed and the polyphony can instantly be monitored. The most common polyphonic instruments are the guitar and the piano. Whereas the guitar's polyphony is limited to six notes according to the number of strings, the piano offers 88 keys.

Consequently, the piano is chosen as the instrument of the CCD test.

### 4.3.2 The score

The subjects have to discriminate piano chords in the CCD test. As the characteristics of the transient attacks are similar, the discrimination is based on the different sound colors. Regarding the design of the chord, a trial and error approach has been pursued. The following paragraph presents the ideas that resulted in the final design of the chords.

By conducting preliminary studies with hearing impaired persons, it became apparent that the discrimination of

- chords of different keys, e.g. C-minor and D-minor
- chords of the same key but different harmonies, e.g. C-minor and C-major

is too easy even for profoundly impaired persons. The stimuli have to be more similar. One option is to create different chords of the same key whose harmony is the same. An obvious approach to create those accords is the musical method called inversion[gl.].
After having conducted preliminary tests with inverted chords, it turned out that the concept of inversion is not suitable for a sound color discrimination test. The crucial aspect of sound color perception is the frequency resolution. The changing bass tone of the inverted chords, however, is a stronger discrimination cue than the sound color.

#### Design aspect: Fixing the compass of the chords

To design appropriate chords for the CCD test, further restrictions have to be introduced. Not only the key and the harmony but also the bass tone of the chords and for similar reasons, the highest tone of the chords must be equal. A changing highest tone is a strong discriminative cue even though it is less prominent than a changing bass note. A single tone is not only composed of its fundamental frequency but also of its harmonics. The frequencies that the highest note of a chord contain are therefore overlapped by the overtones of the deeper notes. Consequently, adding or leaving the highest pitch is perceptually not as crucial as modifying the bass tone. Nevertheless, the effect is more prominent than the sound color itself. Thus, a changing highest pitch has to be avoided in order not to introduce additional discrimination cues.
As a consequence of above considerations, all chords not only have the same root but also their highest pitch is identical.

#### Design aspect: The diatonic functions of the root and the highest note

Regarding the functions of the root and the highest note, the following aspects have to be considered:

- To have the full harmony, each chord has to contain one example of the root, the third and the fifth. To conserve maximal flexibility to create different chords, it is reasonable to choose different functions for the bass tone and the highest pitch.
- To design a wide variety of different chords, the interval between the bass tone and the highest note must be sufficiently large. Thereby, the pitches should not exceed the most commonly used registers of the piano.

The criteria are optimally met by the following constellation: the root is in the bass and the highest pitch is the fifth that is positioned two octaves higher. Without loss of generality, the interval is illustrated in figure 4.8 using the example of the root *C4* (GNS: *c'*).

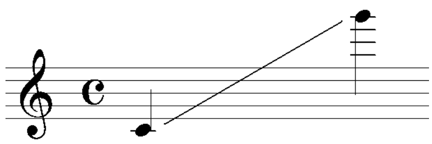

*Figure 4.8: The interval of the chords that are used in the CCD test, exemplarily displayed with the root* C4 *(GNS:* c'*).*

#### Design aspect: Symmetry

As already mentioned, the objective of the test is to measure the perception of sound color. Additional discrimination cues have to be avoided. It is desired that the different tones of each chord melt into one overall sound. The balance of high and low frequency components should be consistent for each chord. This requirement can be best met, if the chords are symmetric.

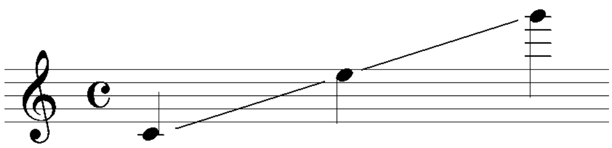

*Figure 4.9: The* C-major *chord with the third* e *representing the point of symmetry.*

As depicted in figure 4.9, the third *E5* (GNS: *e"*) is the point of symmetry of the *C-major* chord with the key *C4* (GNS: *c'*). In the following, it is calculated how many different symmetric chords can be designed that meet all above requirements.

#### Amount of different chords

The above design aspects limit the amount of different chords.

As depicted in figure 4.10, the six other notes of the *C-major* chord that lay within the compass can be grouped into three symmetric pairs. If the pitch *E5* (GNS: *e"*) is noted, the three functions are complete and any of the $2^3 = 8$ combinations of the symmetric pairs represents a *C-major* chord. If the pitch *E5* (GNS: *e"*) is not noted, the only pair that contains the pitch class *e* has to be noted. The residual two pairs can be added, so that $2^2 = 4$ further

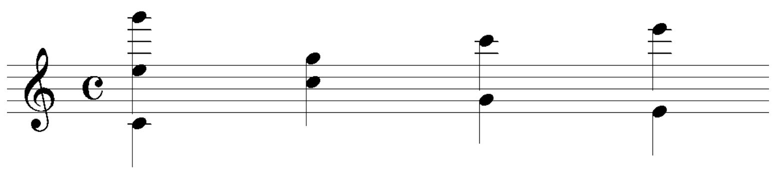

*Figure 4.10: The* C-major *chord and the three symmetric pairs.*

combinations are possible. Summing up, 12 different chords meet the requirements for the MID test. By means of subjective evaluation, it is confirmed that all 12 chords meet the criteria of a balanced sound.

### Evaluation of the applicability of the chord pairs

As explained later in section 4.3.4, the discrimination is tested with the three-alternative method. Three stimuli are played back in a row whereof one is different. Out of 12 chords, $12 \cdot 11/2 = 66$ different chord pairs can be put together that constitute the heterogeneous stimuli triples. Having thoroughly listened to all 66 combinations, an interesting observation can be made: Although the sound of each individual chord is balanced, the direct comparison causes singular notes to stick out perceptually. These combinations are inadequate as the discrimination of sound color is undermined by the discrimination of the singular pitches. After having disqualified the unsuitable combinations, ten combinations are left.

### Different degrees of difficulty

As already mentioned, the CCD test intends to be applicable for the whole range of hearing loss from mild to profound. Thereby, it has to be determined how many degrees of difficulty should be distinguished. According to the wide range of different hearing disorders, the more nuances there are, the more precise a measurement could be taken. On the other hand, a high variety in the measurement tools complicates the standardization of the test and the comparability of its results. With considering both aspects, three different degrees of difficulty are distinguished in the CCD test.

### The final selection of the chords

Each degree of difficulty is represented by one chord pair. Within the ten chord pairs that are left, three chord pairs with different degree of difficulty could be determined. The chosen pairs are displayed in figure 4.11[14].

The chords have been designed and acoustically evaluated for the key *B3* (GNS: *h*). The pitches of the chords range from *B3* (GNS: *h*) to *Fis6* (GNS: *fis"'*). This range is common in piano music. Chords up to one octave lower and one octave higher are still within the ordinarily used range. However, it is impossible to test chords at every semi-tone within this range. Three registers[gl.] are chosen to represent the common range:

- the lower boundary *C3* (GNS: *c*)
- the middle register *B3* (GNS: *h*)
- the upper boundary *B4* (GNS: *b'*)

[14] To be consistent with the previous explanations and demonstration material in figure 2.5, 4.8, 4.9 and 4.10 the key of the chords in figure 4.11 is also *C4* (GNS: *c'*).

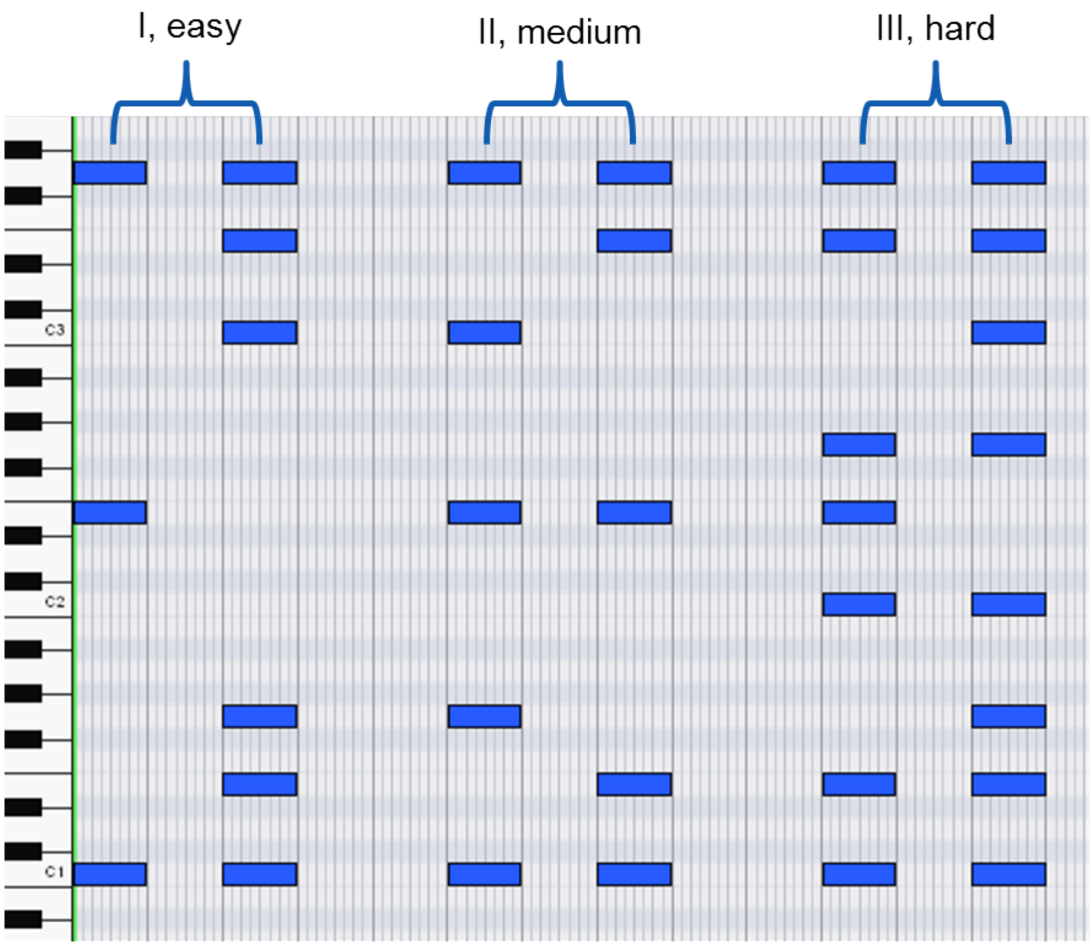


*Figure 4.11: The three chord pairs of the CCD test. The marked fields represent the piano key and pitch that is part of the chord.*

The three positions are combined with three degrees of difficulty. Altogether, nine chord pairs are put to test.

### 4.3.3 Generating the audio material

#### Advantage of MIDI

The audio samples were generated with the MIDI software tool *SampleTank* 2 *L* [18] for several reasons. Firstly, the piano sound quality is excellent and authentic. Secondly, the disadvantage of MIDI not to account for the transition of consecutive tones is not relevant for the CCD test. As chords are played back, the notes are attacked simultaneously.
Concerning the loudness balance, MIDI programing is advantageous: As displayed in figure 4.11, the chords contain up to eight notes. To ensure comparability within and between the chords, it is important that the chords have the same loudness. Regarding the sound balance, each note of the chords should be equally loud. The loudness of the chords can be digitally adjusted after recording. The balance of the pitches, however, cannot be modified. For an instrumentalist, it is an extremely difficult task to play every pitch with the same loudness. MIDI, however, allows to define the loudness of the pitches. The balance of the chords is guaranteed.

#### Loudness considerations

The loudness of each pitch is identical, the loudness of the six chords, however, is different. One reason is that the number of notes that every chords contains is not identical. The number of notes range from three to eight (cf. figure 4.11). In addition, not only the amount but also the position of the notes determines the perceived loudness. Loudness summation is subject to the frequency of the pitches. By equalizing the loudness according to the Zwicker model [20], this additional discrimination cue can be removed for normal hearing people. Nevertheless, a hearing impaired person might perceive a loudness difference between the chords, especially if his hearing loss is one-sided, e.g. high-frequency hearing loss. Because of the symmetric structure and the balanced spectrum of all chords, the differences in loudness perception between the chords, however, should not be significant.

### 4.3.4 Methodology

The methodologies that have been introduced in section 4.2.2 are also applicable for the CCD test:

1. paired comparison
2. three-alternative method, version I
3. three-alternative method, version II
4. *2n+1* method

The latter option should not be applied in the CCD test. This method is already used in the MID test. In order not to bore the subjects, it should be avoided to apply the same test method twice.

#### Applicability of the test methods

To decide which method is most suitable, the qualities are compared. The three decisive aspects are duration (# presentations), reliability (guess probability) and resolution (# responses). The data related to the guess probability can be adopted from table 4.1. The amount of presentations and the amount of responses are determined as follows:

- paired comparison: The nine chord pairs are presented six times. The same amount of homogeneous pairs is added. Each trial can be repeated once if the subject has temporarily been inattentive. Altogether, between 216[15] and 432[16] stimuli can be presented. Only the heterogeneous pairs are accounted for in the hit statistics. Therefore, the amount of results is $9 \cdot 6 = 54$.
- three-alternative, version I: The nine different chord pairs are presented as stimuli triples. The stimuli triples are played back six times and each trial can be repeated once. Altogether, between 162[17] and 324[18] stimuli can be presented. The amount of results is $9 \cdot 6 = 54$.
- three-alternative, version II: The supplementary homogeneous triples account for one third of the heterogeneous triples. The amount of presentations ranges from $162 \cdot (1 + \frac{1}{3}) = 216$ to $324 \cdot (1 + \frac{1}{3}) = 432$. Only the heterogeneous triples contribute to the hit rate statistics. Thus, the amount of results is $9 \cdot 6 = 54$.

[15] 9(chord types) · 2(heterogeneous & homogeneous) · 2(stimuli) · 6(presentations) = 216
[16] 216 · 2(repetitions) = 432
[17] 9(chord types) · 3(triples) · 6(repetitions) = 162
[18] 162 · 2(repetitions) = 324

The data is summarized in table 4.3. Both versions of the three-alternative methods are supe-

| | # presentations | guess probability | # responses |
|---|---|---|---|
| paired comparison | 216 – 432 | 50 % | 54 |
| three-alternative, version I | 162 – 324 | 33 % | 54 |
| three-alternative, version II | 216 – 432 | 25 % | 54 |

*Table 4.3: Comparison of the applicability of different methods for the CCD test regarding duration (# presentation), reliability (guess probability) and resolution (# responses).*

rior to the paired comparison method concerning the length (# presentation) and the reliability (# guess probability) of the test. Version I is expected to last shorter while the reliability of version II is higher. As time efficiency is considered more important than the gain in reliability, version I of the three-alternative method is used in the CCD test.

### 4.3.5 Implementation

#### Test units

Before the actual test starts, the subjects have to get familiarized with the handling of the test. Besides, the learning curve should be flat within the test round. Preceding training can absorb a possible training effect. For both reasons a training round is connected ahead to the test round.

#### Playback

**Length:** One run consists of 54 trials, each of the nine triples is played back six times.

**Order:** The presentation order is pseudo-randomized. The 54 triples are separated into six consecutive blocks. Each of the blocks contains the nine different triples. Within the blocks, the presentation order is randomized. This design allows to retrace the learning effect. Block by block, the hit rates are accumulated and the performance can be compared over time.
As mentioned earlier, the challenge to detect the different stimulus of a triple is expected to be independent from which stimulus is contained twice in the triple. In other words, the discrimination of the chord triple $A$, $A$, $B$ and the chord triple $B$, $B$, $A$ is considered equally difficult. Nevertheless, to foreclose a bias, both versions are employed in the test to the same degree. The versions alternate from block to block.

#### Procedure

The subject is provided with feedback in the training round. The correct response button is highlighted in green, an incorrect response is marked in red. During the test round, no further feedback is given. One complete run of the test takes 28 minutes on average.

A snapshot of the CCD test is displayed in figure 4.12.

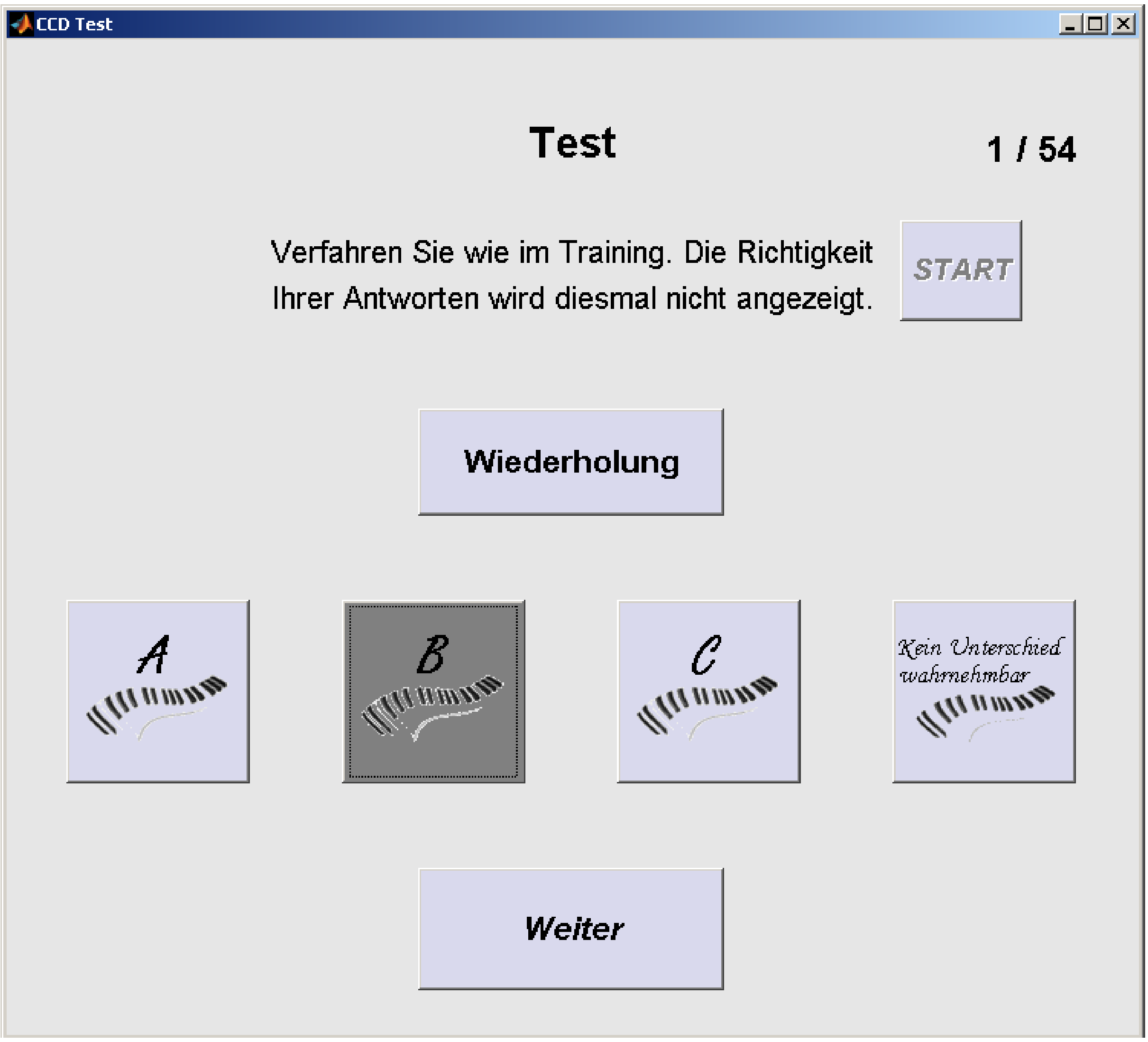


*Figure 4.12: Snapshot of the CCD test.*

# 5 Pilot Study with M.U.S.I.C.

## 5.1 The subjects

The pilot study was conducted with 13 subjects. Six of them are normal hearing and do not wear hearing aids. They serve as the control group. The other seven subjects have mild to profound hearing loss and use hearing aids. Table 5.1 specifies the participants and their qualities. The participants are ordered according to their degree of hearing loss.

| subject | age | hearing loss (dB) | usage of HA | type of HA | musical training |
|---|---|---|---|---|---|
| 1 | 25 | normal, -1 | no | - | yes |
| 2 | 37 | normal, -1 | no | - | no |
| 3 | 31 | normal, 0 | no | - | no |
| 4 | 36 | normal, 4 | no | - | yes |
| 5 | 31 | normal, 5 | no | - | no |
| 6 | 40 | normal, 10 | no | - | yes |
| 7 | 56 | mild, 19 | often | Phonak Audéo smart V | no |
| 8 | 60 | mild, 23 | occasionally | Phonak Audéo S Yes IX | no |
| 9 | 38 | moderate, 38 | always | Phonak Audéo S Yes IX | no |
| 10 | 61 | moderate, 38 | often | Audéo smart IX | no |
| 11 | 27 | moderate, 48 | always | Audéo S Yes IX | no |
| 12 | 48 | profound, 98 | always | Widex Quattro Q-32 | yes |
| 13 | 31 | profound, 99 | always | Phonak Naída S 9 SP | no |

*Table 5.1: Subjects of the pilot study and their qualities*

## 5.2 Test setup

Figure 5.1 displays a sketch of the test setup. The loudspeaker and the position of the subject are arranged as an equilateral triangle. This setup is a standard in the audio engineering society [21]. The measurement took place in the acoustic research laboratory of the Phonak AG. The reverberation times RT60 for several frequencies are listed in table 5.2.

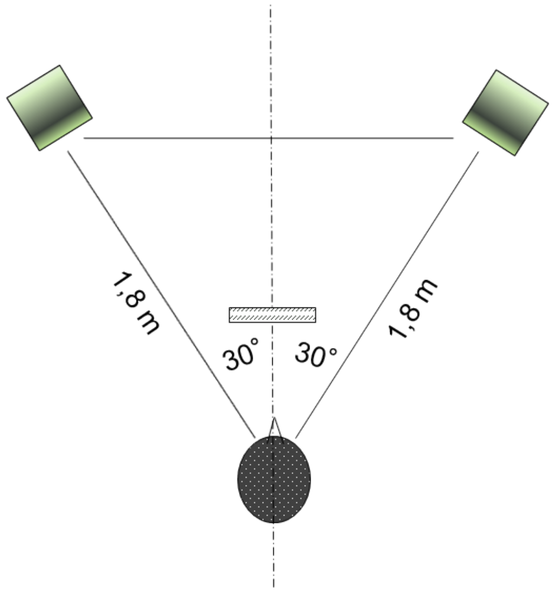


*Figure 5.1: The setup of the M.U.S.I.C. pilot study. The stimuli are played back by two loudspeakers that form an equilateral triangle with the position of the subject as third corner. The side length is 1, 8 m. The rectangle in front of the subject's head marks the position of the touchscreen.*

| | 125 Hz | 250 Hz | 500 Hz | 1000 Hz | 2000 Hz | 4000 Hz | 8000 Hz |
|---|---|---|---|---|---|---|---|
| RT60 | 0.71 | 0.16 | 0.13 | 0.13 | 0.12 | 0.12 | 0.1 |

*Table 5.2: Reverberation time RT60 (in sec) of the acoustic research laboratory where the pilot study was conducted.*

Apart from the measured value at 125 Hz, the reverberation times of the room are low. The reliability of the results is increased because the influence of the room acoustics on the results is negligible.

A touchscreen is installed in front of the subject. It guides through the M.U.S.I.C. test and serves as an answering tool. Regarding the position of the touchscreen, it must be ensured that the acoustic propagation path from the speakers to both ears of the subject is not blocked.
The overall playback volume is set to 75 dB. This level has been chosen for two reasons. Firstly, it is within the range that people are normally exposed to when listening to music. Secondly, the level is high enough so that mild and moderate impaired persons can also be tested without wearing their hearing aids.

The M.U.S.I.C. test is implemented in Matlab code. The program and installation descriptions are specified in the appendix A. Besides the written instructions as part of the Matlab GUI, the participants are also orally briefed to make sure that the assignment is understood. The order of the tests is fixed:

1. NIR
2. MID
3. CCD

Whenever necessary, the subjects are allowed to take a break. One complete test run takes 85 minutes on average.

# 6 Results

## 6.1 Performance with and without hearing aids

In order to evaluate the effect of hearing aids on music perception, the hearing impaired subjects were asked to take the test with and without their hearing aids. The profoundly hearing impaired, however, could not be tested without their hearing devices because the playback volume was lower than their hearing threshold.
Four of the mild and moderate hearing impaired subjects agreed to repeat the M.U.S.I.C test without wearing hearing aids. Subject 7 and 10 took the CCD test twice, subject 9 and 11 took all three tests twice. The results of the subjects are displayed in table 6.1.

| subject | NIR (with/without HA) | MID (with/without HA) | CCD (with/without HA) |
|---|---|---|---|
| 7 | | | 0.85 / 0.80 |
| 8 | 1.00 / 1.00 | 0.92 / 1.00 | 0.93 / 0.93 |
| 10 | | | 0.85 / 1.00 |
| 11 | 1.00 / 1.00 | 1.00 / 1.00 | 0.93 / 0.94 |

*Table 6.1: Comparison of the NIR, MID and CCD hit rates of four hearing impaired persons for two scenarios:* Wearing *hearing aids and* not wearing *hearing aids. The labeling of the subjects refers to table 5.1.*

Regarding the results of the NIR test, the subjects 8 and 11 answered all trials correctly with and without their hearing aids. Concerning the MID test, subject 8 performed perfectly with and without hearing aids. Subject 11 improved the result of 92 % to 100 %. Both subjects stated their discrimination is easier without using the hearing aids. Regarding the CCD test, subject 7 performed better with the hearing aids. The hit rates of subject 8 with and without hearing aids are identical. Subject 10 and 11 improved their performance without hearing aids.
Overall, the subjects' performance without hearing aids is better than with hearing aids.

**Remark:** If not noted differently in the following, the test results of the hearing impaired always refer to the measurements with hearing aids.

## 6.2 Training effect

The playback order of the stimuli of the three M.U.S.I.C. tests is subdivided into blocks. Each block contains the same set of test stimuli and the playback order within the blocks is random. The subdivision into blocks is implemented to observe a possible training effect.

Regarding the results of the normal hearing subjects in the NIR and MID test, it is not possible to observe a training effect. The results of all blocks are almost perfect. Therefore, a learning bias cannot be derived from the measurement values. As a consequence, the learning curves for both tests are displayed only for the hearing impaired in figure 6.1 and figure 6.2. For the CCD test, the learning curves of both groups are shown in figure 6.3 and figure 6.4.

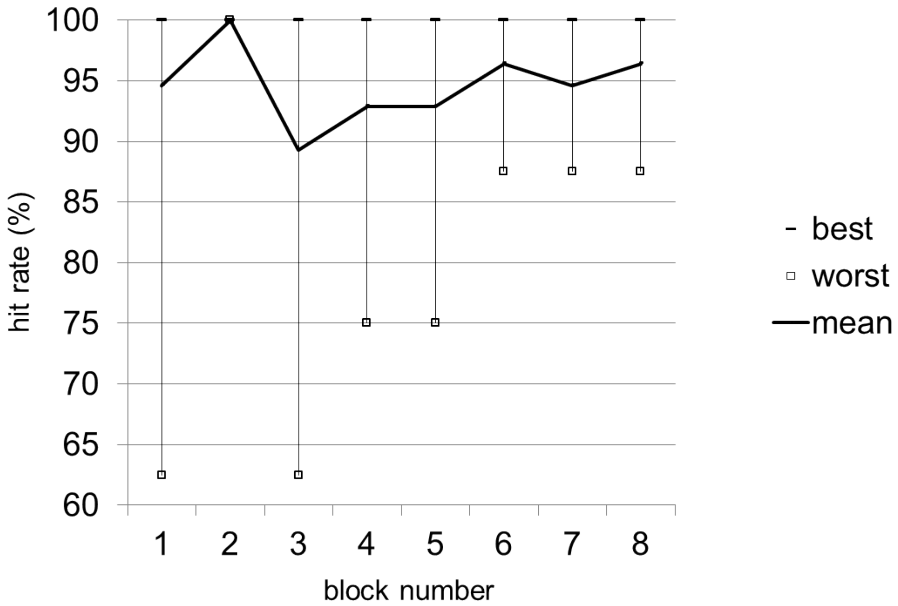


*Figure 6.1: NIR test: the learning curve of the hearing impaired subjects. The 64 test stimuli are divided into eight consecutive blocks. For each block, the hit rates of the hearing impaired subjects are accumulated. The best, mean and worst performance of the subjects are displayed for each block.*

As displayed in figure 6.1, the mean learning curve of the NIR test does not have a clear-cut trend. The worst performance, however, has improved from 63 % in the first block to 88 % in the last block. With regard to the MID test (cf. figure 6.2), the mean results increase from 71 % in the first half to 88 % in the second half. The worst results of each block also improve from the first three blocks to the last three blocks. Concerning the learning curves of the CCD test in figure 6.3 and figure 6.4, no clear trend can be observed neither for the normal hearing nor for the hearing impaired subjects.

## 6.3 Hit rates of NIR, MID and CCD

Table 6.2 displays the hit rates of all three tests and the rates that the *don't know* button (dk-rate) or the repetition button (rep-rate) have been selected in the CCD test. Furthermore, the corrected hit rate of the CCD test (CCD chr) is added. The corrected hit rate refers to the percentage of trials that a subject responded correctly based on the perception. It is approximated by deducting an estimation of the guessed hits from the measured test results. Before carrying out the CCD test, the subjects were explicitly instructed to choose the answer option *don't know* in case of an ambiguous perception. If a response is wrong, it can be assumed that

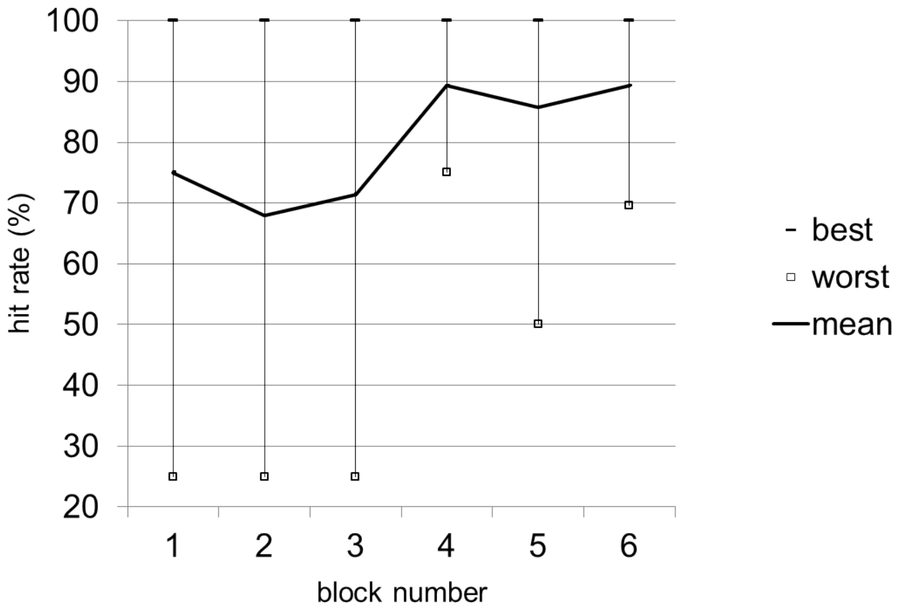


*Figure 6.2: MID test: the learning curve of the hearing impaired subjects. The 24 test stimuli are divided into six consecutive blocks. For each block, the hit rates of the hearing impaired subjects are accumulated. The best, mean and worst performance of the subjects are displayed for each block.*

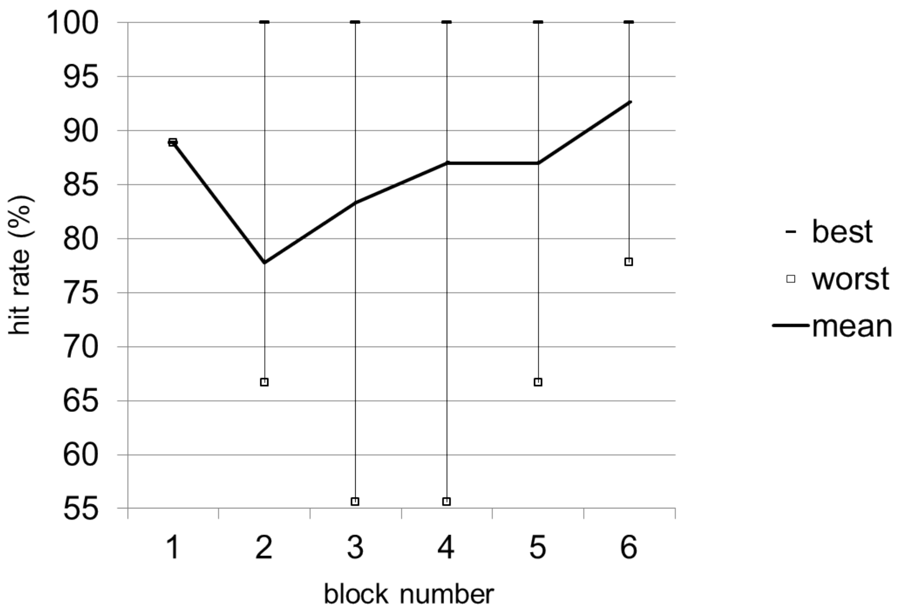


*Figure 6.3: CCD test: the learning curve of the normal hearing subjects. The 54 test stimuli are divided into six consecutive blocks. For each block, the hit rates of the hearing impaired subjects are accumulated. The best, mean and worst performance of the subjects are displayed for each block.*

the subject tried to guess. To approximate the corrected hit rate, the guessed hits are deducted from the results. With assuming that every error is caused by guessing, the number of guessed hits is estimated as follows:
The guess probability of hit to error equals 1 : 2. According to that proportion, it can be estimated that half the amount of guessed hits equal the errors. The error rate is determined as the difference of the sum of the measured hit rate and the dk-rate to 100 %.

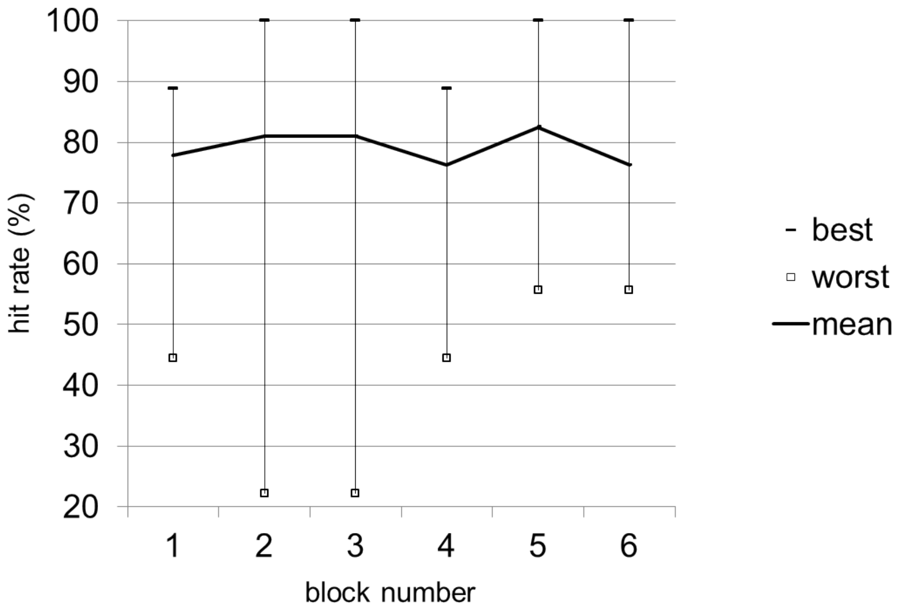


*Figure 6.4: CCD test: the learning curve of the hearing impaired subjects. The 54 test stimuli are divided into six consecutive blocks. For each block, the hit rates of the hearing impaired subjects are accumulated. The best, mean and worst performance of the subjects are displayed for each block.*

| subject | NIR hitrate | MID hitrate | CCD hitrate | CCD dk-rate | CCD rep-rate | CCD chr |
|---|---|---|---|---|---|---|
| 1 | 100 % | 100 % | 96 % | 4 % | 26 % | 96 % |
| 2 | 100 % | 92 % | 74 % | 6 % | 78 % | 64 % |
| 3 | 100 % | 100 % | 89 % | 4 % | 54 % | 85 % |
| 4 | 100 % | 100 % | 85 % | 4 % | 17 % | 80 % |
| 5 | 92 % | 83 % | 80 % | 7 % | 65 % | 73 % |
| 6 | 100 % | 100 % | 93 % | 2 % | 17 % | 90 % |
| 7 | 100 % | 100 % | 85 % | 11 % | 19 % | 83 % |
| 8 | 100 % | 92 % | 93 % | 4 % | 30 % | 91 % |
| 9 | 91 % | 46 % | 78 % | 11 % | 94 % | 72 % |
| 10 | 100 % | 88 % | 85 % | 7 % | 26 % | 81 % |
| 11 | 100 % | 100 % | 93 % | 6 % | 43 % | 92 % |
| 12 | 80 % | 58 % | 41 % | 44 % | 76 % | 33 % |
| 13 | 92 % | 75 % | 80 % | 17 % | 63 % | 78 % |

*Table 6.2: The results of the pilot study. The labeling of the subjects corresponds to table 5.1. The* hit rates *of NIR, MID and CCD refer to the* measured *results.* CCD dk-rate *is an abbreviation and stands for the percentage of trials that the subject chose the* don't know *button.* CCD rep-rate *is an abbreviation for the percentage of trials that the subject made use of the* repetition *button.* CCD chr *is the corrected hit rate of the CCD test.*

$$\begin{aligned}\text{CCD chr} &= \text{meas. hit rate} - \tfrac{1}{2} \cdot \text{error rate} = \\ &= \text{meas. hit rate} - \tfrac{1}{2} \cdot (100\% - \text{meas. hit rate} - \text{dk rate})\end{aligned}$$

**Remark:** Unless noted otherwise below, the term *hit rate* refers to the corrected hit rate and not to the measured hit rate.

**NIR test** As displayed in table 6.2, five of the six normal hearing subjects assigned all 64 stimuli of the NIR test correctly. The remaining normal hearing subject achieved a hit rate of 92 %. Regarding the hearing impaired subjects, four of seven were able to recognize all instruments. The remaining three subjects answered 91 %, 92 % and 80 % of the trials correctly.

**MID test** Regarding the normal hearing subjects, four of six answered all 24 trials correctly. The remaining two normal hearing subjects achieved a hit rate of 92 % and 83 %.
With regard to the hearing impaired subjects, the test results range between 46 % and 100 %. All subjects stated they had to concentrate harder to accomplish the MID test than the NIR test.

**CCD test** The normal hearing persons' results range between 74 % and 96 %, the results of the hearing impaired range between 41 % and 93 %. A perfect score has not been achieved.

## 6.4 Confusion matrix for the NIR and MID test

The confusion matrices in figure 6.5 and figure 6.6 reveal which instruments have been played back and which instruments have been selected as the response.

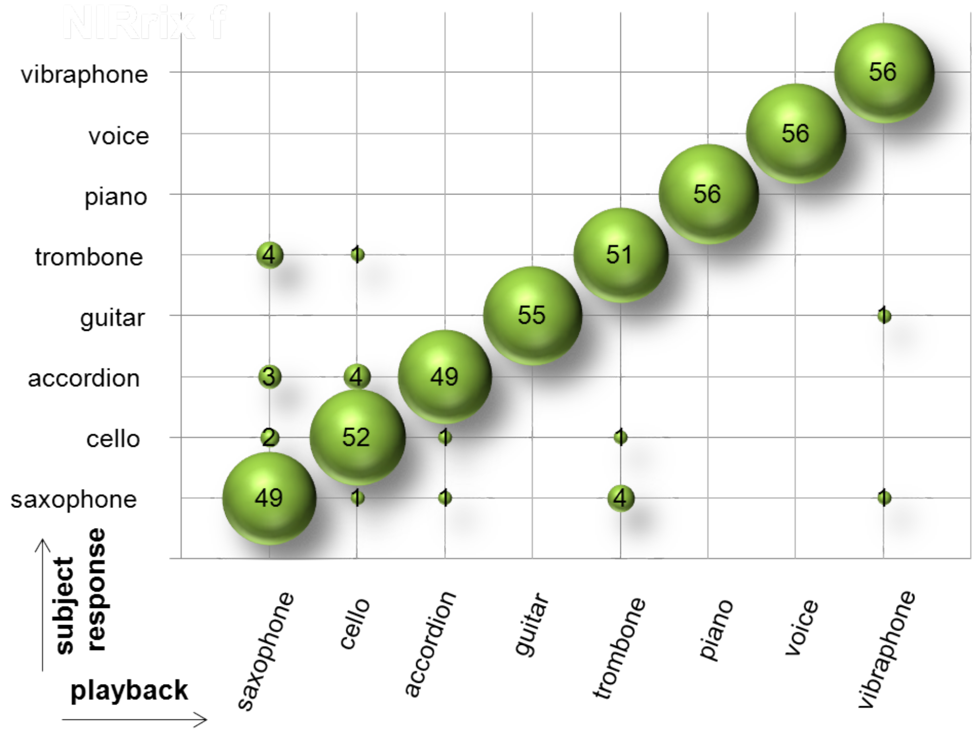


*Figure 6.5: NIR test: Confusion matrix of the profoundly hearing impaired subjects. The x-axis refers to the playback, the y-axis refers to the subject response. The bubbles and the numbers represent the number of subject responses.*

According to figure 6.5, the pairs *saxophone/trombone*, *accordion/cello* and *saxophone/accordion* were most often confused in the NIR test. *Piano* and *voice* were never mistaken. The subjects' feedback after the test confirmed that the latter instrument is easily discriminated.

Regarding the CCD test, errors were committed on all possible combinations of the played back instruments and the response instruments. The corresponding confusion matrix is displayed

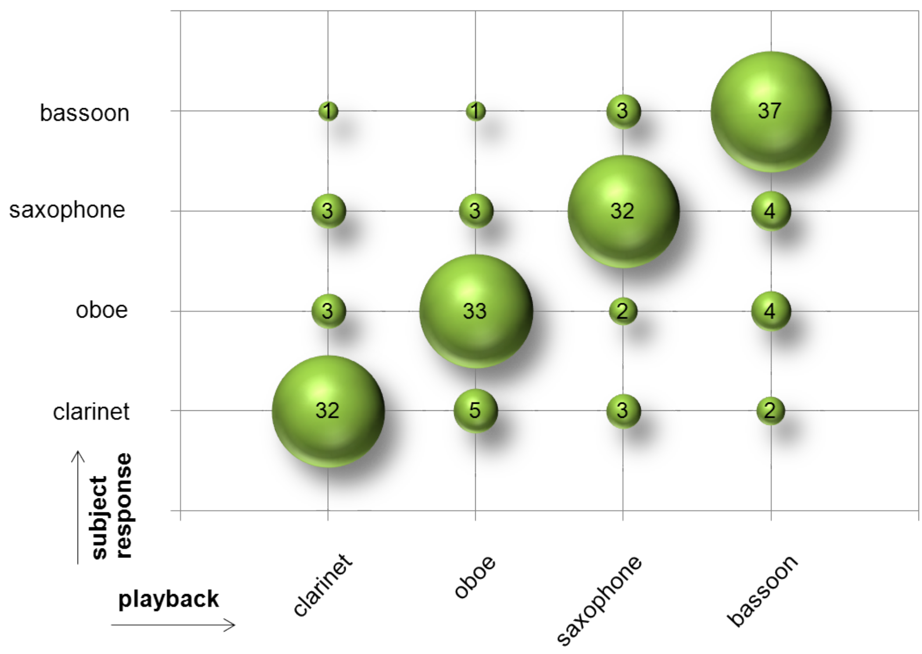


*Figure 6.6: MID test: Confusion matrix of the hearing impaired subjects. The x-axis refers to the playback, the y-axis refers to the subject response. The bubbles and the numbers represent the number of subject responses.*

in figure 6.6.

## 6.5 CCD test

### 6.5.1 The nine chord pairs

The CCD test includes nine different chord pairs. Table 6.3 and table 6.4 display the hit rates of the normal hearing and the hearing impaired subjects broken down into the nine chords. The frequency to use the repetition button or the *don't know* button are displayed in table 6.5, table 6.6, table 6.7 and table 6.8.

| Type / position | *c* | *h* | *h1* |
|---|---|---|---|
| I, easy | 96 | 92 | 83 |
| II, medium | 96 | 88 | 83 |
| III, hard | 65 | 89 | 40 |

*Table 6.3: CCD test: Normal hearing subjects'* corrected *hit rate of the nine different chords. 'Type' refers to the chord structure (cf. figure 4.11), 'position' refers to the root of the chord. The German notation system is used to specify the pitch of the root.*

### 6.5.2 The degree of difficulty of the chord pairs

Originally, the repetition button was implemented to avoid a concentration bias. However, it has to be considered that it was not only used when a subject was temporarily inattentive but also when the subject could not detect the different chord for the first time. Consequently, the amount of times that the repetition button was pressed can be regarded as another indicator for the difficulty of the test. Nonetheless, it has to be considered that the personality of a subject

| Type / position | *c* | *h* | *h1* |
|---|---|---|---|
| I, easy | 86 | 85 | 88 |
| II, medium | 89 | 90 | 74 |
| III, hard | 51 | 86 | 33 |

*Table 6.4: CCD test: Hearing impaired subjects'* corrected *hit rate of the nine different chords. '*Type*' refers to the chord structure (cf. figure 4.11), 'position' refers to the root of the chord. The German notation system is used to specify the pitch of the root.*

| Type / position | *c* | *h* | *h1* |
|---|---|---|---|
| I, easy | 25 % | 19 % | 47 % |
| II, medium | 33 % | 56 % | 28 % |
| III, hard | 64 % | 53 % | 58 % |

*Table 6.5: CCD test: The proportion of trials that the normal hearing subjects' repeated broken down by the nine different chords.* Type *refers to the chord structure (cf. figure 4.11),* position *refers to the root of the chord. The German notation system is used to specify the pitch of the root.*

| Type / position | *c* | *h* | *h1* |
|---|---|---|---|
| I, easy | 38 % | 26 % | 45 % |
| II, medium | 41 % | 48 % | 52 % |
| III, hard | 67 % | 52 % | 81 % |

*Table 6.6: CCD test: The proportion of trials that the hearing impaired subjects' repeated broken down by the nine different chords.* Type *refers to the chord structure (cf. figure 4.11),* position *refers to the root of the chord. The German notation system is used to specify the pitch of the root.*

| Type / position | *c* | *h* | *h1* |
|---|---|---|---|
| I, easy | 0 % | 10 % | 0 % |
| II, medium | 0 % | 0 % | 0 % |
| III, hard | 14 % | 3 % | 22 % |

*Table 6.7: CCD test: The proportion of trials that the normal hearing subjects choose the response button* don't know *broken down by the nine different chords.* Type *refers to the chord structure (cf. figure 4.11),* position *refers to the root of the chord. The German notation system is used to specify the pitch of the root.*

is another factor that determines the repetition rate. An accurate subject is likely to use the repetition button more often than an impatient subject. To check the correlation between the repetition rate and the hit rate, both parameters are set against each other in figure 6.7. The correlation factor is $-0.64$.

In the following analysis, the hit rate and the repetition rate are weighted equally to evaluate the chord pairs' degree of difficulty. In figure 6.8 and figure 6.9 both parameters are set against each other for the nine chords. The first figure relates to the normal hearing subjects, the latter relates to the hearing impaired subjects. In table 6.9, the ranking is further subdivided into the three registers 'low', 'middle' and 'high'.

According to the evaluation, the chord pairs of type III are most difficult to discriminate in all three positions. The normal hearing discriminate chord type II better in the high position than chord type I. The hearing impaired subjects' discrimination of chord type II is better than of chord type I in the low position. This observation is confirmed by a second study. The chord pairs $II_{low}$ and $I_{low}$ were further tested with the hearing impaired subjects 12 and 13. Both

| Type / position | *c* | *h* | *h1* |
|---|---|---|---|
| I, easy | 7 % | 12 % | 5 % |
| II, medium | 0 % | 2 % | 12 % |
| III, hard | 38 % | 7 % | 45 % |

*Table 6.8: CCD test: The proportion of trials that the hearing impaired subjects choose the response button* don't know *broken down by the nine different chords.* Type *refers to the chord structure (cf. figure 4.11),* position *refers to the root of the chord. The German notation system is used to specify the pitch of the root.*

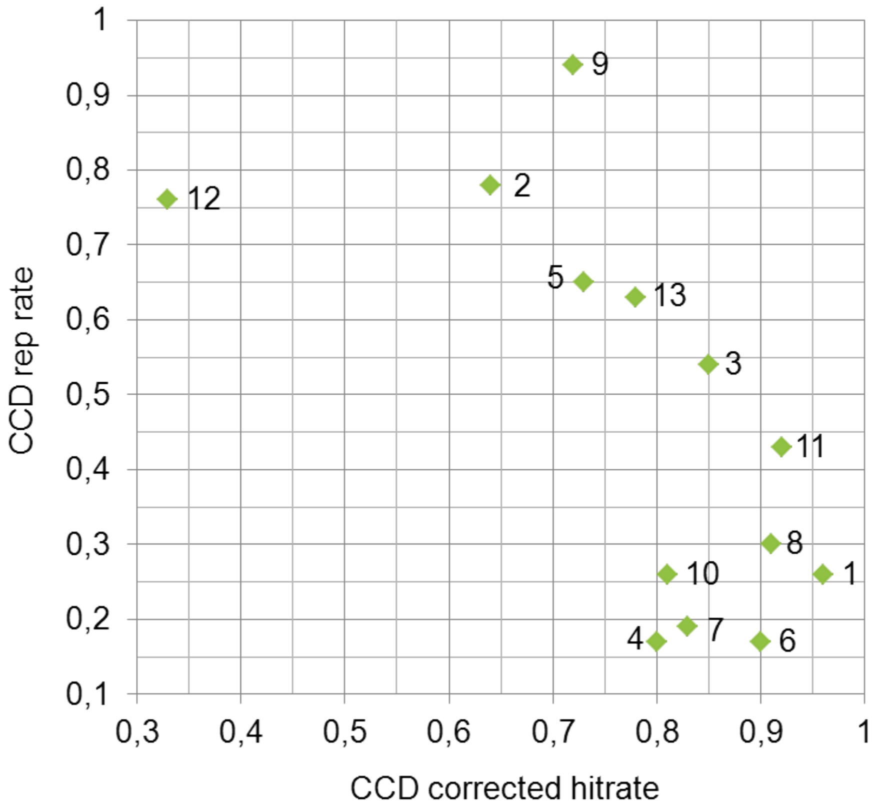


*Figure 6.7: CCD test: The corrected hit rate set against the repetition rate. The subjects are indicated by the numbers of table 5.1.*

subjects ranked the degree of difficulty according to the pilot study and hence contrary to the normal hearing subjects.

With ignoring the register of the chords, another subdivision can be made regarding the degree of difficulty. The subdivision is indicated by the circles in figure 6.8 and figure 6.9. Table 6.10 lists the respective chords.

As displayed in table 6.10, it is most difficult to discriminate the chord pair $\mathrm{III}_{\mathrm{high}}$. The distribution of the other chords indicate that not only the type but also the register determines the chord pairs' degree of difficulty. The coherence of both factors, however, cannot be extracted from the data.

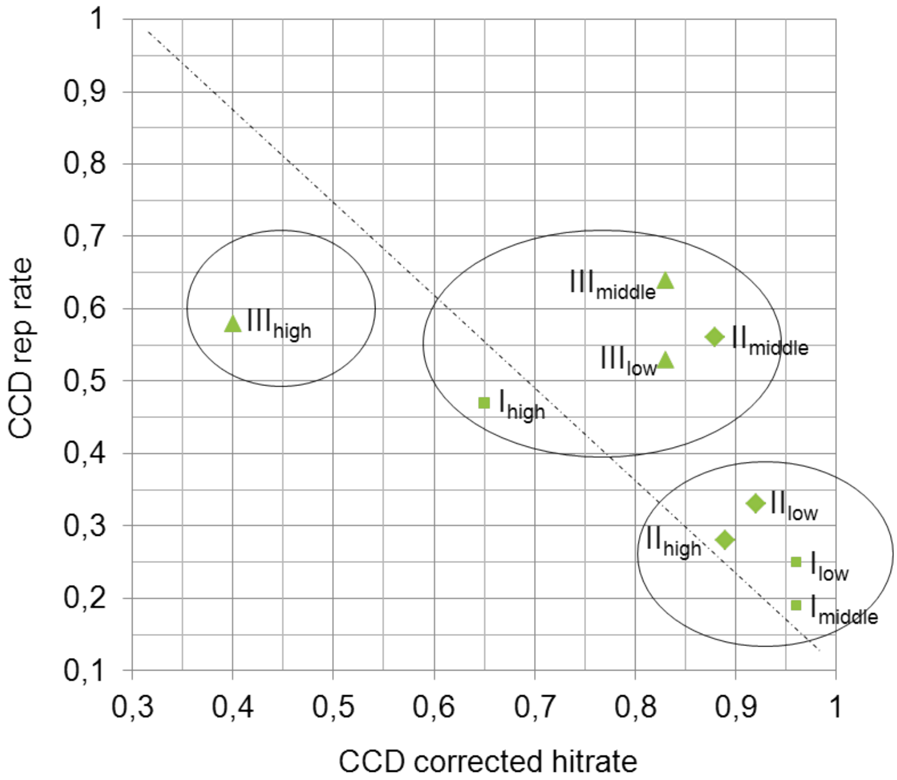


*Figure 6.8: CCD test: The hit rates and the repetition rates of the normal hearing subjects are set against for each of the nine chords. 'I','II','III' refers to the type of chord (cf. figure 4.11), 'low', 'middle' and 'high' refers to the register (cf. figure 4.3.2).*

| | **NH** | | | **HI** | | |
|---|---|---|---|---|---|---|
| Difficulty / Position | low | middle | high | low | middle | high |
| difficult | III | III | III | III | III | III |
| medium | II | II | I | I | II | II |
| easy | I | I | II | II | I | I |

*Table 6.9: CCD test: The chord types 'I', 'II' and 'III' (cf. figure 4.3.2 aligned to the degrees of difficulty according to the results of the pilot study.*

| Degree of difficulty | normal hearing | hearing impaired |
|---|---|---|
| difficult | $III_{high}$ | $III_{high}$ |
| moderate | $I_{high}$, $II_{middle}$, $III_{low}$, $III_{middle}$ | $III_{low}$ |
| easy | $I_{low}$, $I_{middle}$, $II_{low}$, $II_{high}$ | $I_{low}$, $I_{middle}$, $I_{high}$, $II_{low}$, $II_{middle}$, $II_{high}$, $III_{middle}$ |

*Table 6.10: CCD test: The nine chord pairs aligned to the degrees of difficulty according to the results of the pilot study.*

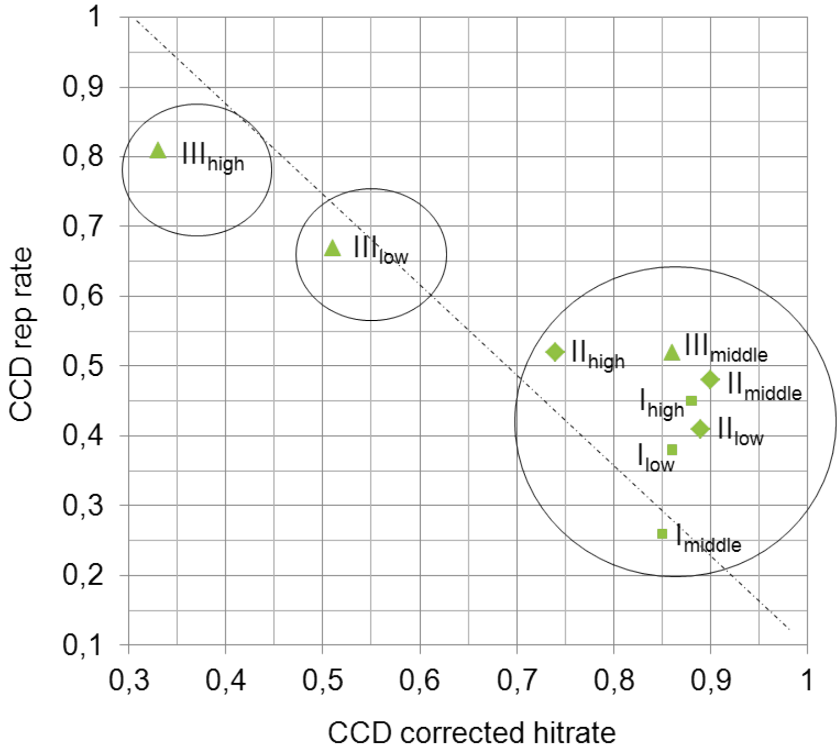


*Figure 6.9: CCD test: The hit rates and the repetition rates of the hearing impaired subjects are set against for each of the 9 chords. 'I','II','III' refers to the type of chord (cf. figure 4.11), 'low', 'middle' and 'high' refers to the register (cf. figure 4.3.2).*

# 7
# Discussion and Outlook

The pilot study pursues two different objectives. First, gaining insights concerning the music perception of the hearing impaired and second, checking the applicability of the test.

With regard to the first objective of the pilot study, no comparative measurement is available to evaluate the results. Characteristic values that indicate any impairment related to music perception have not been developed so far. The degree of hearing loss serves as a first reference.

With regard to the second goal, the applicability of the M.U.S.I.C. test battery can be checked by evaluating its degree of difficulty. The tests are appropriate if the measurement of the subjects' performances is sensitive. Figure 7.1 pictures a showcase function of the degree of difficulty and the hit rates for a psychometric test [22] in general. The nearer a subject's results are to the inflection point, the more sensitive is the measurement.

## 7.1 Music perception of the hearing impaired

**Influence of the degree of hearing loss** The correlation coefficients of the degree of hearing loss and the test results of the three tests NIR, MID and CCD are $-0,77$, $-0,46$ and $-0,68$ respectively. According to the coefficients, the correlation regarding the NIR and CCD test is rather strong. Nevertheless, especially in the case of the MID test, the degree of hearing loss is not a sufficient indicator to evaluate or predict a subject's performance in the music perception tests. The degree of hearing loss is the averaged hearing threshold of a discrete number of frequencies. Frequency resolution, temporal resolution and the cognitive part of a hearing loss are not accounted for. It must be assumed, that impairments concerning music perception and music discrimination are more complex than the degree of hearing loss could indicate. Analogous to the *Speech Intelligibility Index (SII)*[23], a *Music Perception Index* should be developed.

**Music perception with and without hearing aids** Four hearing impaired subjects took the test twice, once with their hearing aids and once without. On average, the subjects performed better without hearing aids. It has to mentioned, however, that the subjects took the test without hearing aids *after* having passed the test with their hearing aids. Nonetheless, a learning effect can be ruled out. The amplified signals of the hearing aid are different to the original signals. Passing the test with hearing aids is not likely to improve the test results without hearing aids.

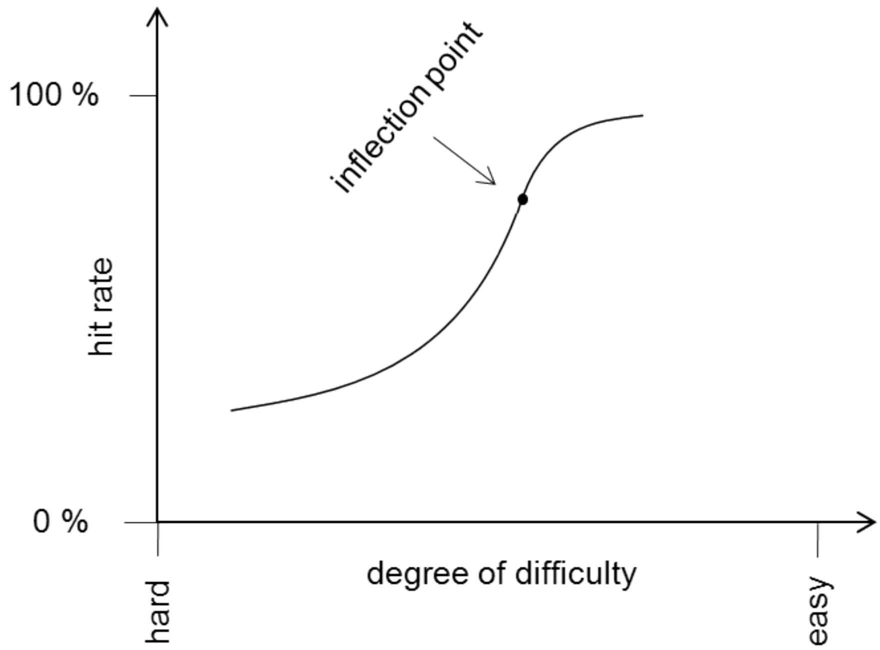


*Figure 7.1: Showcase psychometric function for the M.U.S.I.C. test battery. If a subject's performance is near the inflection point, the sensitivity of the test is optimal.*

A plausible explanation for the better results without hearing devices might be the comb-filter effect. The hearing aids of the four respective subjects are open-fitted. Consequently, the natural sound is passed through the ear canal and interferes with the output of the loudspeaker. Depending on the gain, comb-filter effects can affect the timbre and make it harder to discriminate. This theory might explain why the subjects performed better without their hearing aids. However, it does not account for the fact that some hearing impaired subjects performed even better than normal hearing people. One explanation might be above conclusion. The degree of hearing loss does not exclusively define the impairment concerning music perception. Moreover, impairments might even help to discriminate. Certain sounds may create distortions or artefacts in an impaired auditory system that facilitate the differentiation.

## 7.2 Applicability of the M.U.S.I.C. tests

### NIR

**Status quo** The NIR test is adequate for severe and profound hearing losses only. Although the degree of hearing loss might not be the exclusive indicator of a hearing loss related to music perception, it can be assumed that the NIR test is too easy for mild and moderate hearing losses.

Apart from the voice, the instrument set is appropriate. The timbres of all instruments are well-known and equally difficult to recognize. The timbre of the voice, however, is too different from the other instruments and can be recognized easily.

**Outlook** In order to enlarge the target group, two methods have potential to lower the degree of difficulty:

- Shortening the length of the melodies

- Adding noise

Regarding the first approach, it is not possible to cut the existing melodies. New melodies have to be created to ensure that the amount of recognition cues within each melody is identical. With regard to the second approach, the type of noise to be used should reflect a realistic situation. The white transmission noise of a radio or the harmonic noise of an orchestra playing in the background could be adequate.

Regarding the instrument set, the voice has to be substituted. The current recordings of the voice contain the syllable '*la*'. Recording the melodies with a humming voice might make it more difficult to discriminate the voice from the other instruments. If this option is not successful, the voice will be left out or a substitute instrument has to be found.

### MID

**Status quo** The subject's degree of hearing loss is not an appropriate reference to predict whether the MID test is applicable or not. Based on the results of the pilot study, subjects with all degrees of hearing loss can be sensitive to the measurement.

The choice of instruments is appropriate. The errors are evenly spread across all combinations of the played back instrument and the falsely chosen instrument.
Concerning the test methodology *2n+1 comparison*, the subjects' feedback diverges. The worst performers are satisfied with the test method whereas the better subjects consider it very tiresome. The automatic playback of the audio samples takes approximately 50 seconds. During this period of time, the subjects do not participate actively in the test but remain silent and listen.

**Outlook** In the next version of the CCD test, the methodology will be modified correspondingly. The following approaches have potential:

- The playback can be aborted by the subject. Thus, the subject does not have to listen to all the consecutive stimuli if the choice of the right test stimuli is already clear.
- The files are not automatically played back in a row. The subject has to activate the playback by pressing the respective button.
- Another of the methodologies discussed in section 4.2.2 could be chosen instead of the *2n+1 comparison* method.

### CCD

**Status quo** The CCD test is appropriate for all kinds of hearing losses. The chord pairs under test have different degrees of difficulty. The ranking of the chord pairs' difficulty depends on the type of chords and on the register. The discrimination of normal hearing and hearing impaired is not consistent. Hearing impaired persons are likely to use different cues for discrimination than normal hearing people. Moreover, the hearing aid signal processing might emphasize the discrimination cues depending on the chord.
The methodology of the CCD test is adequate. The subjects are comfortable with the handling and consider the test as pleasant and interesting.

**Outlook** In future research, the coherence of the chord type and the register will be analyzed. The objective target is to determine three chord pairs that can clearly be assigned to three different degrees of difficulty. Moreover, the differences of the normal hearing and hearing impaired persons' perception will further be researched.

### Enlargement of the M.U.S.I.C. test battery

The three music tests focus on the discrimination of timbre and sound color. A further discrimination test is planned to assess the temporal resolution. Apart from discrimination, music enjoyment will be tested. The tests focus on the perception of sound quality and emotion.

# 8 Summary

The diploma thesis marks the initial step of a long-term commitment to improve music perception with hearing aids. The long-term commitment includes three stages:

1. The music perception of the hearing impaired is assessed both with and without wearing hearing aids. The results are compared and benchmarked with the performances of normal hearing people. To measure the music perception, a M.U.S.I.C. test battery is developed. It includes discrimination tests and tests for music enjoyment.
2. Based on the insights and discoveries in step 1, strategies to improve music processing are researched. The approaches focus signal processing.
3. The strategies developed in step 2 are evaluated by means of the M.U.S.I.C. test battery. The optimization of the strategies requires an iteration of step 2 and step 3.

Within the framework of the diploma thesis, three discrimination tests have been developed: Natural Instrument Recognition (NIR), MIDI Instrument Discrimination (MID) and Chord Color Discrimination (CCD).

Regarding the NIR test, the subjects have to recognize the instruments being played back. Eight instruments from different categories are put under test: saxophone, cello, accordion, guitar, trombone, piano, voice and vibraphone. Two melody sets are composed for the NIR test. One set is based on the first five notes of the *A3* (GNS: *a*) Ionian scale. The other set uses the fourth till eighth harmonic of the *Fis1* (GNS: *'Fis*) natural harmonic series. Every melody of a set has the same amount of discrimination cues.

With regard to the MID test, the subjects have to discriminate four instruments of the woodwind category: saxophone, clarinet, oboe and bassoon. To present the stimuli and to retrieve the subjects' responses, the new test method *2n+1* has been developed. The crucial aspect of the *2n+1* method is the playback order of the stimuli. The reference stimulus and the test stimuli are alternately played back so that every test stimuli is preceded and succeeded by a test stimuli. The playback order determines the composition of the melodies that are used in the test. The first and last pitch of a melody is identical. Each melody of a set starts with a different pitch.

The CCD test focuses on the discrimination of sound color. The subjects have to distinguish piano chords. The chords are designed for three degrees of difficulty and are tested in three different registers. Per trial, three chords are played back in a row. The subjects have to detect which of the three chords is different from the other two chords.

A pilot study with 13 subjects was conducted. Six subjects were normal hearing, the other seven subjects were hearing impaired and used hearing aids. Based on the subjects' performance, the applicability of the M.U.S.I.C. test battery for hearing aid users could be confirmed. Nevertheless, it has to be considered that the degree of hearing loss is not sufficient to evaluate or predict a subject's appropriateness to a test. Besides the degree of hearing loss, other aspects of hearing loss like temporal or frequency resolution may also be determining.

A selection of mild and moderate impaired subjects took the test twice, at first with their hearing aids and then without their hearing aids. The performances without the hearing aids were better.

Based on the subject's feedback, optimization approaches for the existing M.U.S.I.C test battery are discussed. The applicability of the NIR test will be enlarged and the test method of the MID test will be modified. Furthermore, the M.U.S.I.C. test battery will be enlarged. A further discrimination test to assess the temporal resolution will be added and music enjoyment tests will be developed.

# A
# M.U.S.I.C. GUI

Please notice that the M.U.S.I.C. test battery has been designed for German speaking subjects and supervisors. Therefore, the language of the Matlab GUI is German.

## Installation

To install the M.U.S.I.C. test battery, please insert the accompanying CD into the drive. Copy the folder *MUSIC* into the directory of your choice. Add the subfolder *programs* to your Matlab path.

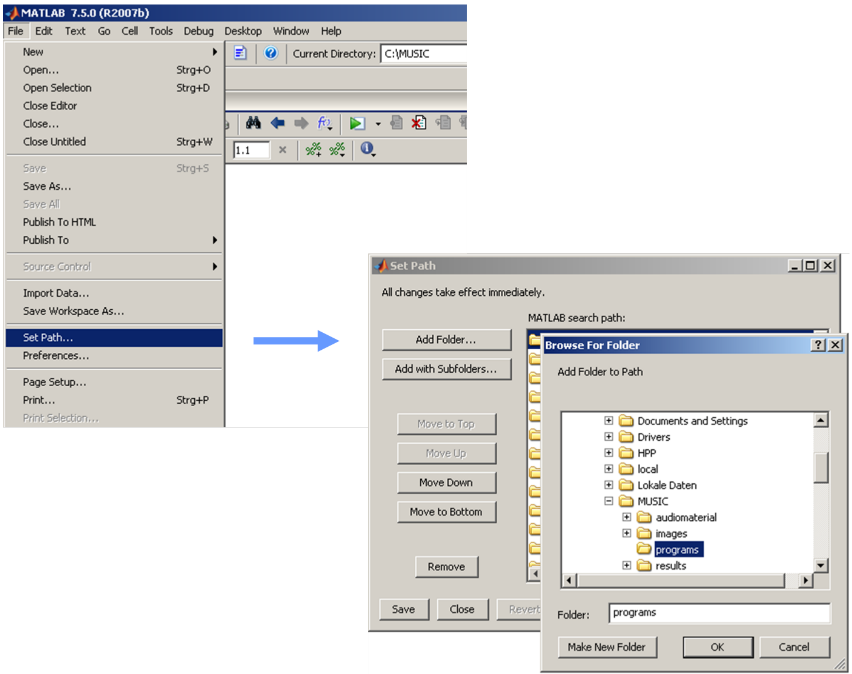


*Figure A.1: GUI: Adding the subfolder* programs *to the Matlab path.*

## Initialization

Run the program *MUSIC.m* that is located in the *MUSIC* folder. The initialization window *Initialisierung M.U.S.I.C* pops up.

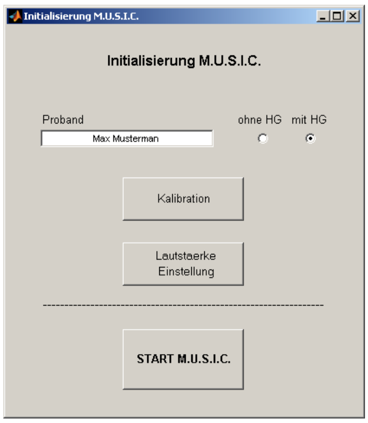


*Figure A.2: GUI: Initialization window*

Enter the subject's name and specify whether the subject is tested with or without hearing aids.

### Calibration

Open the calibration menu by clicking the *Kalibration* button of the initialization window.

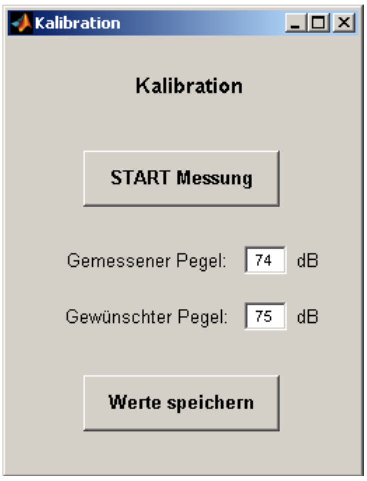


*Figure A.3: GUI: Calibration window.*

Start the measurement (*START Messung*), enter the measured value (*Gemessener Pegel*) and define the desired sound level (*Gewünschter Pegel*). The measurement has to be executed at the position of the subject's head. Save the settings (*Werte speichern*).

### Level adjustment

Within the framework of the pilot study, the output level was kept constant for all subjects. Other studies might prefer to test the subjects at their individual, most comfortable level. In order to adjust the level accordingly, press the button *Lautstaerke Einstellung* of the initialization window to open the respective menu *Lautstaerke-Einstellung*.

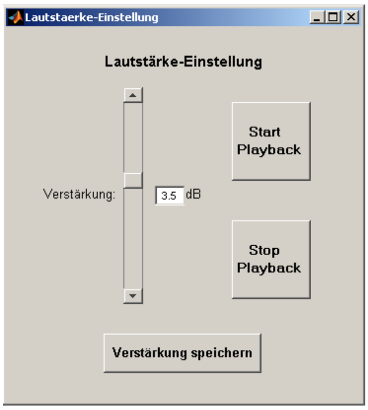


*Figure A.4: GUI: Adjusting the output level.*

Start the playback of the audio file and adjust its level either by moving the slider or manually entering the desired gain. Save the settings with *Verstärkung speichern*.

## Starting M.U.S.I.C

Press the button *START M.U.S.I.C.* of the initialization window to get to the welcome screen. Click anywhere on the welcome screen and the main menu *M.U.S.I.C.* pops up.

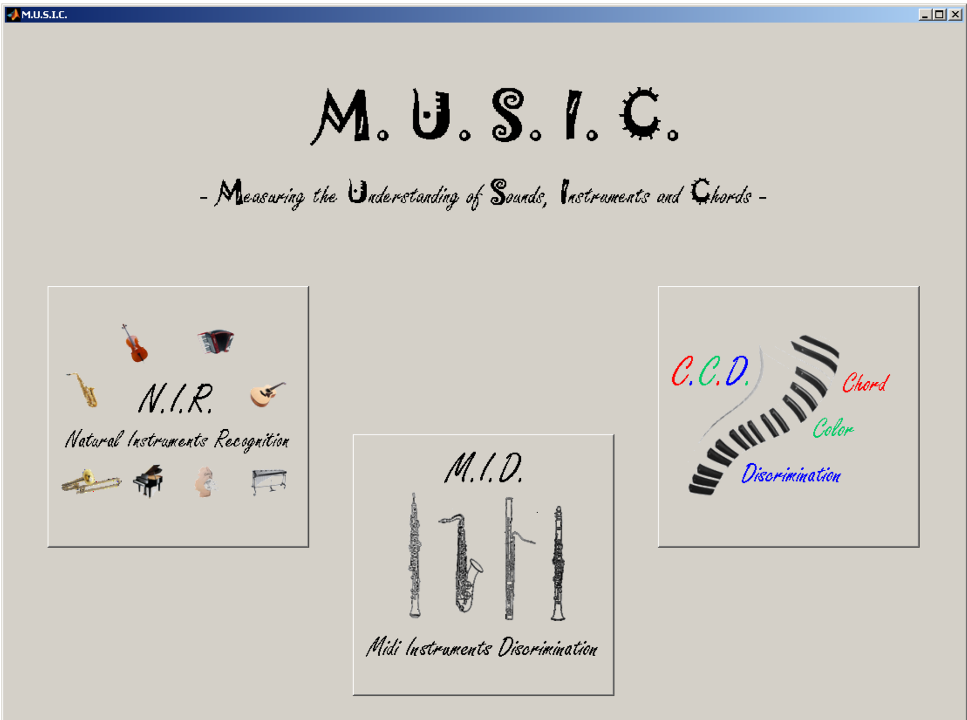


*Figure A.5: GUI: The main menu*

Start the test by pressing the corresponding button. The subsequent instructions guide through the test. All results are automatically saved.

# B Audiograms

In the following, the audiograms^gl.^ of the 13 participants of the pilot study are displayed. To protect personality rights, the subjects are indicated by numbers. The subjects are ordered according to their degree of hearing loss.

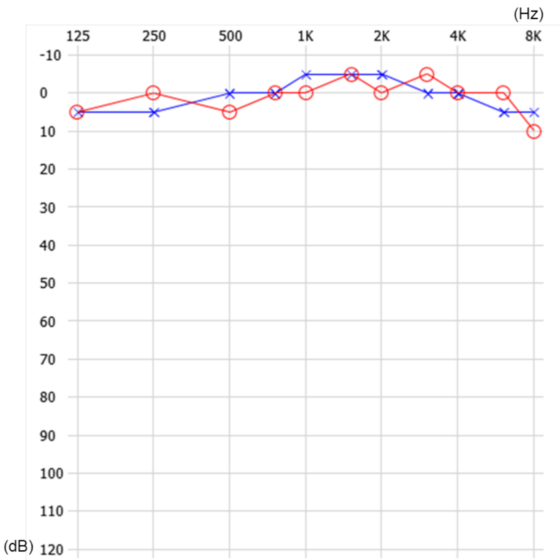


*Figure B.1: Audiogram of subject 1.*

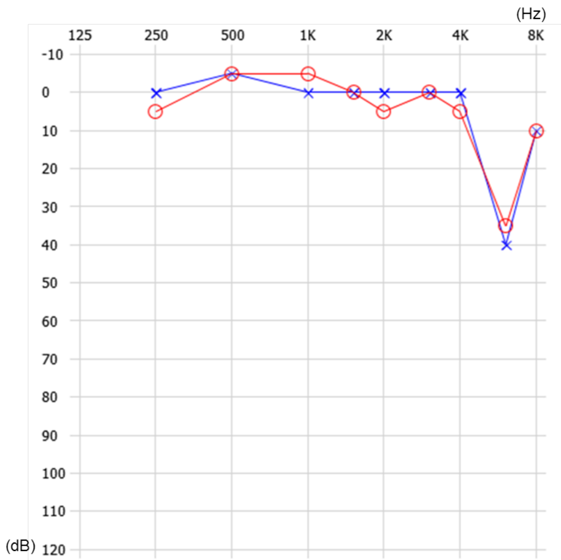


*Figure B.2: Audiogram of subject 2.*

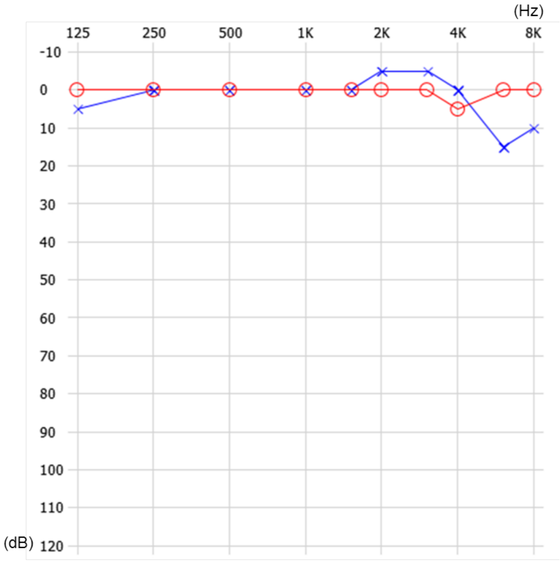


*Figure B.3: Audiogram of subject 3.*

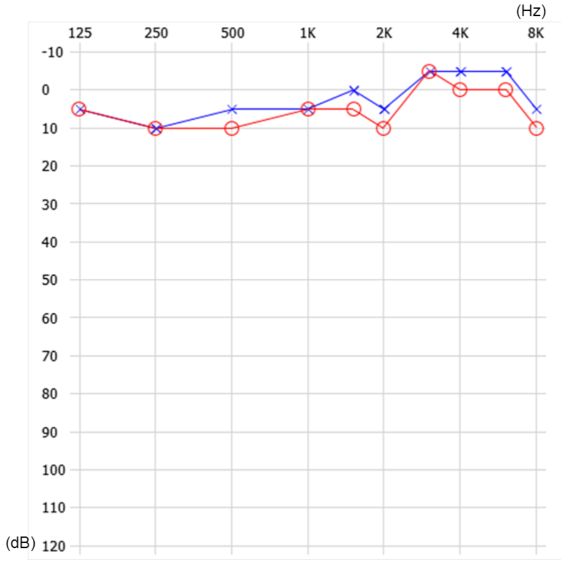


*Figure B.4: Audiogram of subject 4.*

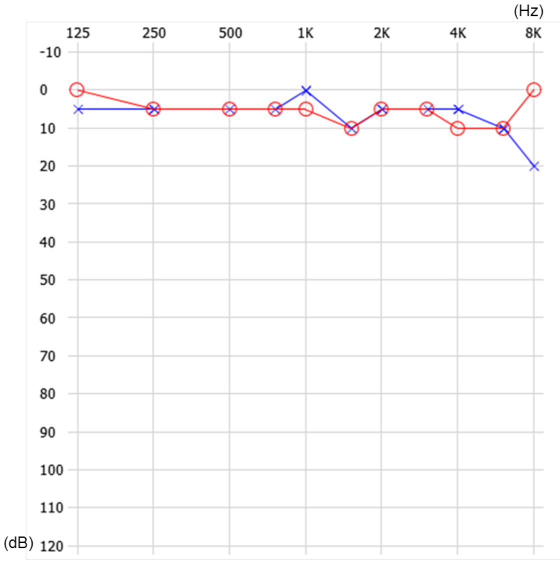


*Figure B.5: Audiogram of subject 5.*

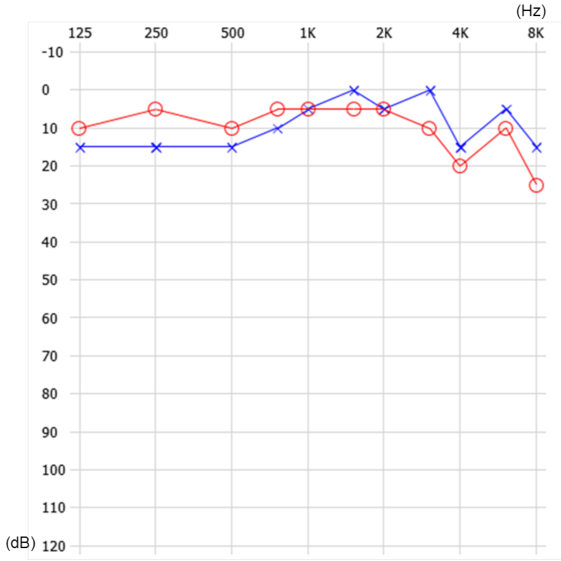


*Figure B.6: Audiogram of subject 6.*

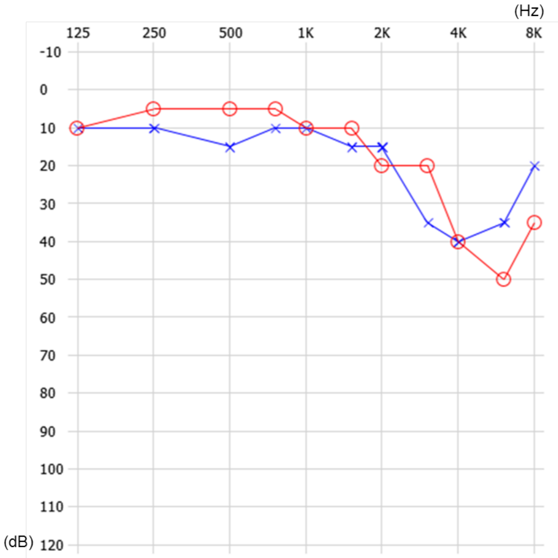


Figure B.7: Audiogram of subject 7.

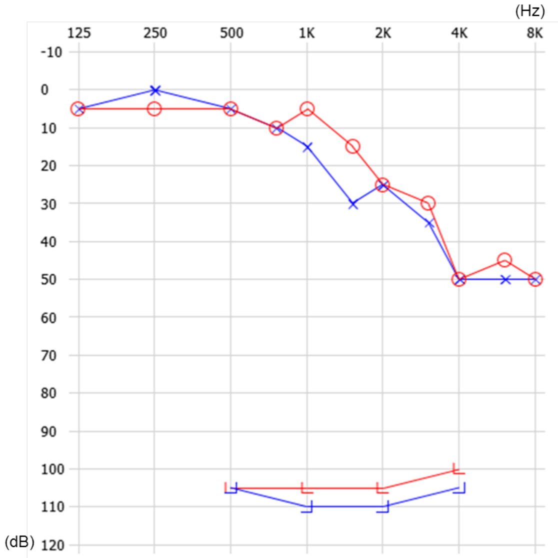


Figure B.8: Audiogram of subject 8.

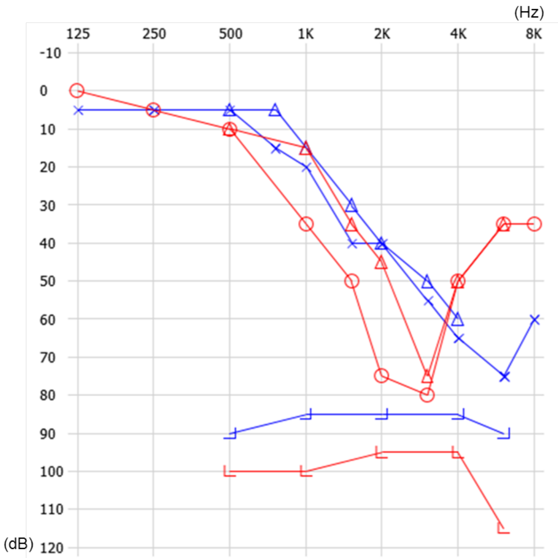


Figure B.9: Audiogram of subject 9.

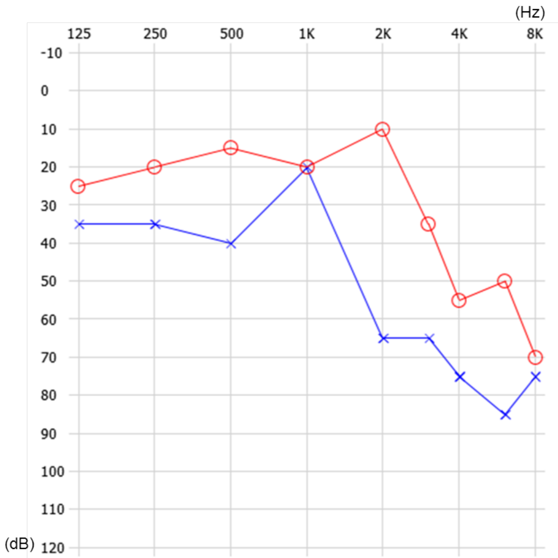


Figure B.10: Audiogram of subject 10.

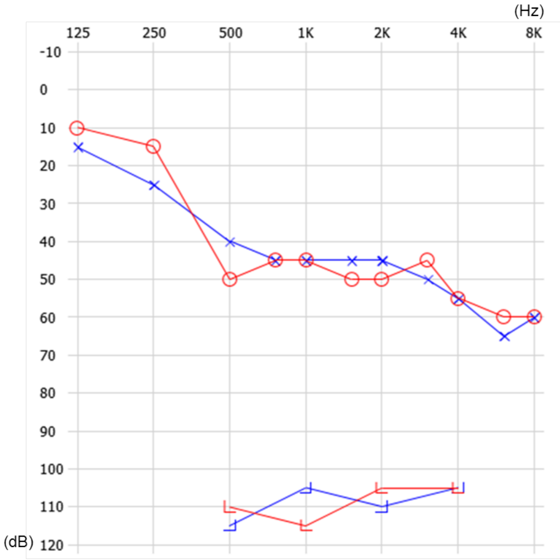


*Figure B.11: Audiogram of subject* 11.

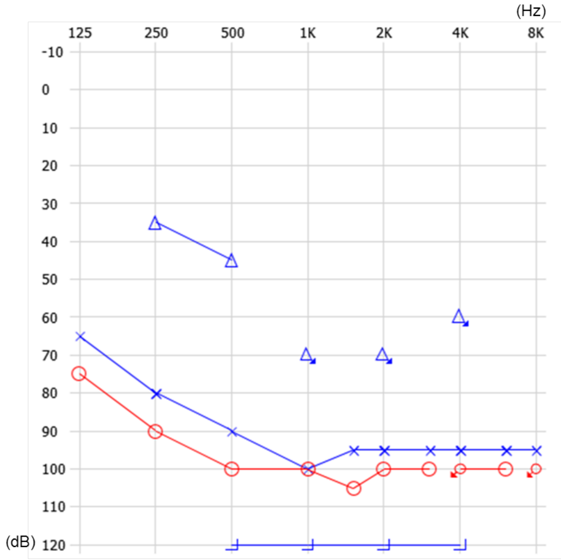


*Figure B.12: Audiogram of subject* 12.

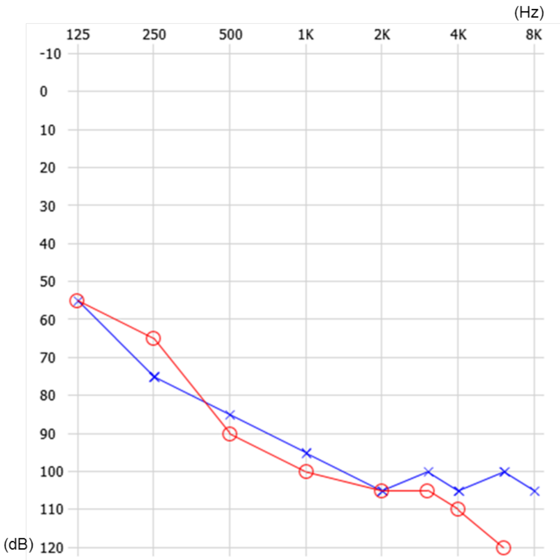


*Figure B.13: Audiogram of subject* 13.

# Bibliography


[1] Cook, Perry R., "Music, cognition, and computerized sound: An introduction to psychoacoustics," Cambridge, Mass., 2001.

[2] Piston, Walter, "Harmony," New York, 1987.

[3] von Hornbostel, Erich Moritz Sachs Curt, "Systematik der Musikinstrumente. Ein Versuch." pp. 553–590, 1914.

[4] Gordon, Edwin, "Primary Measures of Music Audiation," Chicago, 1979.

[5] Lindsay, Peter H. and Norman, Donald A., "Human information processing: An introduction to psychology," New York, 1977.

[6] Jones, Mari Riess and Fay, Richard R. and Popper, Arthur N., "Music Perception: Springer Handbook in Auditory Research (SHAR)," s.l., 2010.

[7] Moore, Brian C. J., "An introduction to the psychology of hearing," Bingley, 2008.

[8] Winter, Philipp G., "The Development and Pilot Testing of a Music Quality Rating Test Battery for New Zealand and Australian MED-EL Cochlear Implant Recipients," Ph.D. dissertation, University of Canterbury, 2010.

[9] Purwins, Hendrik, "Profiles of Pitch Classes Circularity of Relative Pitch and Key - Experiments, Models, Computatinal Music Analysis, and Perspectives," Ph.D. dissertation, Technical University of Berlin, Berlin, 2005.

[10] Pollard, Howard F., "Tonal portrait of a pipe organ," *The Journal of the Acoustical Society of America*, vol. volume 106, no. 1, pp. 360–370, 1999.

[11] Chasin, Marshall, "Music and hearing aids," *Hearing Journal*, vol. 56, pp. 36–41, 2003.

[12] Neukomm, Ernst, "Musikhören bei Schwerhörigkeit mit und ohne Hörgerät." [Online]. Available: http://www.pro-audito.ch/fileadmin/daten/Neukommpas_musikstudie.pdf

[13] Sarchett, Wayne, "Hearing Aid Issues for the Hearing Impaired Audiophile," 2006. [Online]. Available: http://www.hometheaterhifi.com/volume_13_2/hearing-aids-6-2006-part-1.html

[14] Bersch-Burauel, Antje, "Entwicklung von Musikpräferenzen im Erwachsenenalter: Eine explorative Untersuchung," Paderborn, 02.03.2004.

[15] Egloff, Edith, "Musik und Hörbehinderung," Bern, Juni 2003.

[16] Cary, H., "Are you a Musician? Professor Seashore's Specific Psychological Test for Specific Musical Abilites," 1923.

[17] Büchler, Michael, "The Zurich Music Test Battery," 16.6.2008.

[18] IK Multimedia, "Sample Tank 2: Sample Workstation," 2010.

[19] Kaernbach, Christian, "Adaptive threshold estimation with unforced-choice tasks," *Perception & Psychophysics*, vol. 63, no. 8, pp. 1377–1388, 2001.

[20] Zwicker, Eberhard and Fastl, Hugo, "Psychoacoustics: Facts and models," Berlin, 1990.

[21] Dickreiter, Michael, "Handbuch der Tonstudiotechnik," München, 1997.

[22] Klein, Stanley A., "Measuring, estimating, and understanding the psychometric function: A commentary," Berkeley, pp. 1421–1455, 2001.

[23] American National Standards Institute, "ANSI S1.1 - USA Standard Acoustical Terminology," New York, 1960.