ISSN 0280-6819

KAROLINSKA INSTITUTET

TECHNICAL AUDIOLOGY

Report TA134
Sept 1998

Continuous rating of sound quality

David Hedberg, Christopher Jansson

KAROLINSKA INSTITUTET
Department of Ear and Skin
Unit of Technical Audiology

REPORT

September 1998

Continuous rating of sound quality

David Hedberg, Christopher Jansson

This work is a master thesis by the two authors and was performed at the section of Technical Audiology, Karolinska institutet. The examination for the masters degree was done by the department of Vehicle Engineering at the Royal Institute of Technology in Stockholm.

The Marcus Wallenberg Laboratory at the Royal Institute of Technology has also published this report with the identification code

TRITA-FKT 1998:33
ISSN 1103-470X
ISRN KTH/FKT/EX--98/33--SE

Mailing address	*Visiting address*	*Phones*	*Fax*	*Org.nr:*
Teknisk Audiologi	Drottning Kristinas väg 33A	+46-8 411 66 60	+46-8 791 71 30	01 202100 2973 01
Box 70025	Kungl Tekniska Högskolan	+46-8 791 75 17	*E-Mail*	*Moms/VAT-No.:*
S-100 44 STOCKHOLM	Stockholm	+46-8 791 75 18	bjorn.hagerman@ood.ki.se	SE202100297301
SWEDEN		***Home page:*** http://www.ki.se/ees/taudiology		

Preface

This Master of Science thesis has been written at the Marcus Wallenberg Laboratory at the Royal Institute of Technology (KTH) in Stockholm. The project was carried out at the Department of Technical Audiology, Karolinska institutet, Stockholm.

We would like to thank our supervisor associate professor Björn Hagerman and the others at Technical Audiology: Åke Olofsson, Henrik Olsen, Jun Cheng, Alexander Häggström, Ann-Cathrine Lindblad and Tatiana Goriatcheva for assistance throughout the project.

Thank also to our examiner Dr Ulf Carlsson and everyone that participated in the listening tests.

Finally we would like to thank Ren Botanabe who has been proof-reader and Anders Sundström who has taken the photographs.

Abstract

The purpose of this project was to create a system which makes it possible to perform continuous estimations of psychoacoustic descriptors. Some minor listening tests with subjects were carried out in order to examine if the method works and is reliable.

The system was computer based and consisted of a sound card that could handle simultaneous playback and recording, one pair of headphones and one control box with a sliding potentiometer that could generate an input signal to the sound card.

Seven subjects participated in the test which consisted of two major parts; estimation of loudness and estimation of brightness. The test contained 18 different stimuli and lasted for about 45 minutes per subject.

In the analysis of the loudness test we calculated an envelope of the sound pressure level in dB(A) of each stimulus. This was compared with the received responses. In the analysis of the brightness test, to have something to compare with, we calculated the variation of the balance point of the spectrum of the stimulus over time. Reaction times, variations between subjects, reproducability and transfer functions were also studied. We found that it was easier to estimate loudness than brightness. The average time delay was about 0.5 s for loudness and about 0.8 s for brightness. The average uncertainty of the rating was about 0.7 scale units for loudness and about 0.8 scale units for brightness. From the results of the analysis of the loudness test we also tried to simulate the hearing process with a multi-step model. With this model we could simulate a fictive response to a stimulus.

CONTENTS

1 Introduction

In order to get an understanding of how humans perceive different sound properties, such as loudness, brightness and sharpness, listening tests are performed. In these tests subjects listen to short sound stimuli, typically of the length of one minute. The subjects then give a total impression of specific properties by making a mark on a scale. This method is well developed for overall estimations such as quality ratings of sound reproducing systems - loudspeakers, headphones, telephones etc. [3]. The method is also suitable for estimation of static sounds, such as vacuum cleaners [8]. However, most sounds in our environment fluctuate over time, and with this method it is difficult to be sure on the relation between the rating and the physical features of the sound. Therefore it is interesting to find a more appropriate estimation method. One possibility is to let subjects continuously estimate a sound property while listening to the stimulus. This creates a picture of how the judgment changes during the stimulus. For example if you continuously rate the annoyance of train noise you can get detailed information on which sounds that actually causes the annoyance (for example squeaking brakes).

The purpose of this project was to develop a system which make it possible to perform continuous estimation of psychoacoustic parameters. In order to examine the reliability of the method, i.e. if the subjects' responses are repeatable for identical stimuli, we then carried out some minor tests and analyzed the results.

We chose to concentrate on the estimation of loudness, which is the most basic psychoacoustic parameter, but we also studied the estimation of brightness. Only limited research has been done in the field of continuous estimation of psychoacoustic parameters, and this work is to be regarded as explorative.

2 Basic concepts

2.1 Anatomy

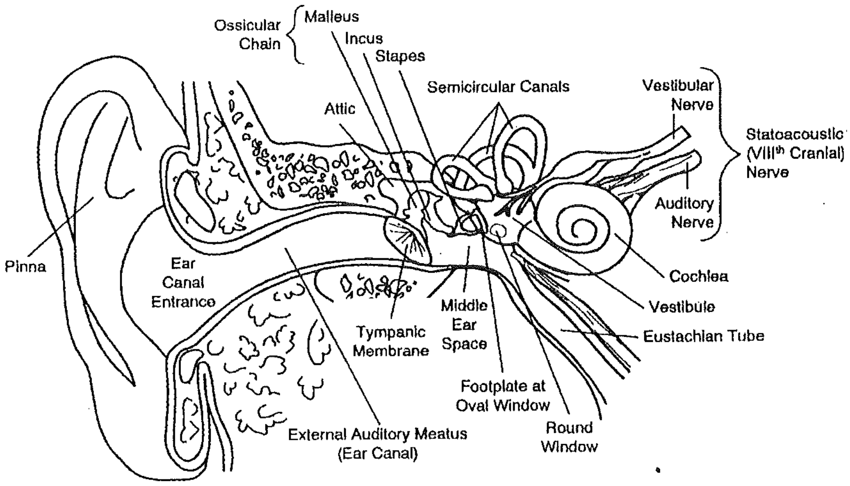


Figure 2-1. The major parts of the human ear [.

The human ear is generally divided into three main parts: the outer ear, the middle ear and the inner ear. The **outer ear** is made up of the *Pinna* and the ear-canal. The major function of the *Pinna* is to collect sound-waves like a hopper and lead them into the ear-canal. Through its asymmetric shape it gives us a sense of direction and enables us to determine the location of a sound source. The ear-canal ends with the eardrum (*Tympanic Membrane)* which separates the outer ear from the middle ear.

In the **middle ear,** the eardrum is mechanically connected via the hammer (*Malleus*), the anvil (*Incus*) and the stirrup (*Stapes*) to the oval window on the *Cochlea*, where the inner ear begins. The middle ear serves two purposes. First, the ear-bones mentioned above and the membranes belonging to them, creates a mechanical impedance transformation of airborne vibrations on the eardrum to liquid borne in the shell. Second the transmission of the ear bones can be modified by two small muscles to protect the inner ear from too strong signals. The *Tensor tympani muscle* pulls the hammer so that the eardrum stiffens and the *Stapedius muscle* turns the stirrup. The mass and the elasticity of the transmission work as a lowpass filter and limit the upper cut-off frequency of the hearing.

The **inner ear** is made up of the *Cochlea* and the balance organ. In the *Cochlea* you find the basilar membrane on which the hair cells are located. When the oval window vibrates, the basilar membrane moves, which stimulates the hair cells. The hair cells then fire of nerve signals which are sent to the brain where a hearing sensation appears.

2.2 Sound pressure and sound power

The definition of sound pressure is:

$$p(t) = p_{total}(t) - p_0 \tag{2-1}$$

where

p_{total} = total pressure

p_0 = static pressure ($\approx 10^5$ Pa)

The instantaneous value of sound pressure is of limited importance, it is more relevant to calculate the **r**oot **m**ean **s**quare-value (**rms**), $\tilde{p}$:

$$\tilde{p} = \sqrt{\frac{1}{T}\int_0^T p^2(t)dt} \tag{2-2}$$

since it gives information about the power of the signal. The mean value of the sound power is proportional to the squared **rms**-value of the pressure:

$$\overline{W} \propto \tilde{p}^2 \tag{2-3}$$

2.3 Levels and decibels

It is convenient to show sound quantities on a logarithmic scale. That way we get manageable numbers and the scale matches the hearing experience fairly well. The unit Bel is defined as the 10-logarithmic of the ratio between two acoustic powers. To get handy values it is practical to use a tenth of a Bel as a unit, deciBel (dB). The presentation in dB is referred to as level. A logarithmic scale has no zero value and therefore we must determine a reference point. For sound power the reference point is equal to 10^{-12} W. Thus sound power can be calculated using this formula:

$$L_w = 10\log\frac{\overline{W}}{W_{ref}} \tag{2-5}$$

Since $\tilde{p}^2 \propto \overline{W}$ the sound pressure level becomes:

$$L_p = 10\log\frac{\tilde{p}^2}{p^2{}_{ref}} \tag{2-6}$$

where

$p^2{}_{ref} = 2 \cdot 10^{-5}$ Pa

2.4 The hearing area

The hearing area spans, in frequency, from 20 Hz to 20 kHz. The hearing area's lower sound level limit is the threshold of hearing. This is strongly dependent on the frequency. The reference value for sound pressure mentioned in the previous section is determined from the threshold at 1000 Hz for people with normal hearing. The upper limit of the hearing area is the pain-threshold which is about 130 dB for all frequencies. Ordinary speech occur in the area showed by the 'speech banana' and non electric music is represented by the 'music pumpkin' (*figure 2-2*). The hearing is impaired with increasing age. This affects all frequencies, especially higher frequencies.

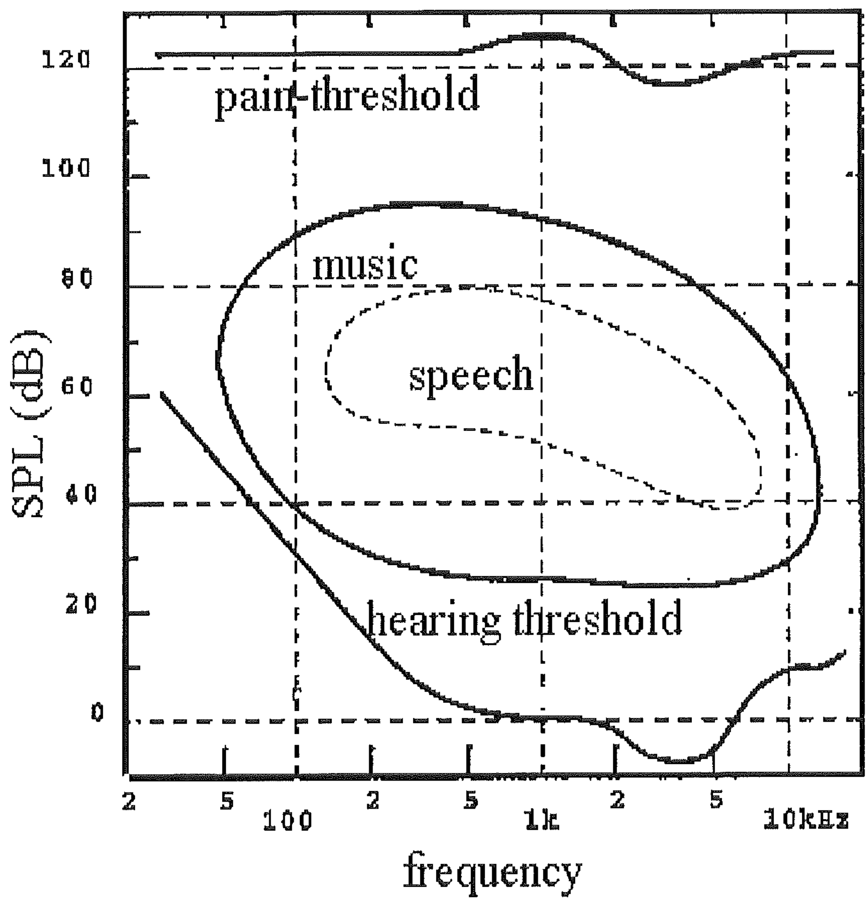


Figure 2-2. This diagram shows the areas that are used in non-electric music and speech, the 'music pumpkin' and the 'speech banana' [6].

2.5 Loudness

The human perception of how loud a sound is, its loudness, does not entirely correspond to the physical sound pressure level of that sound. Different frequencies with the same intensity are perceived as having different loudness. Loudness level is defined as the sound pressure level a 1 kHz sine tone should have to give the same perception as the sound you are judging. The unit of loudness level is *phon*. *Figure 2-3* shows curves which link tones with different frequencies but the same loudness together.

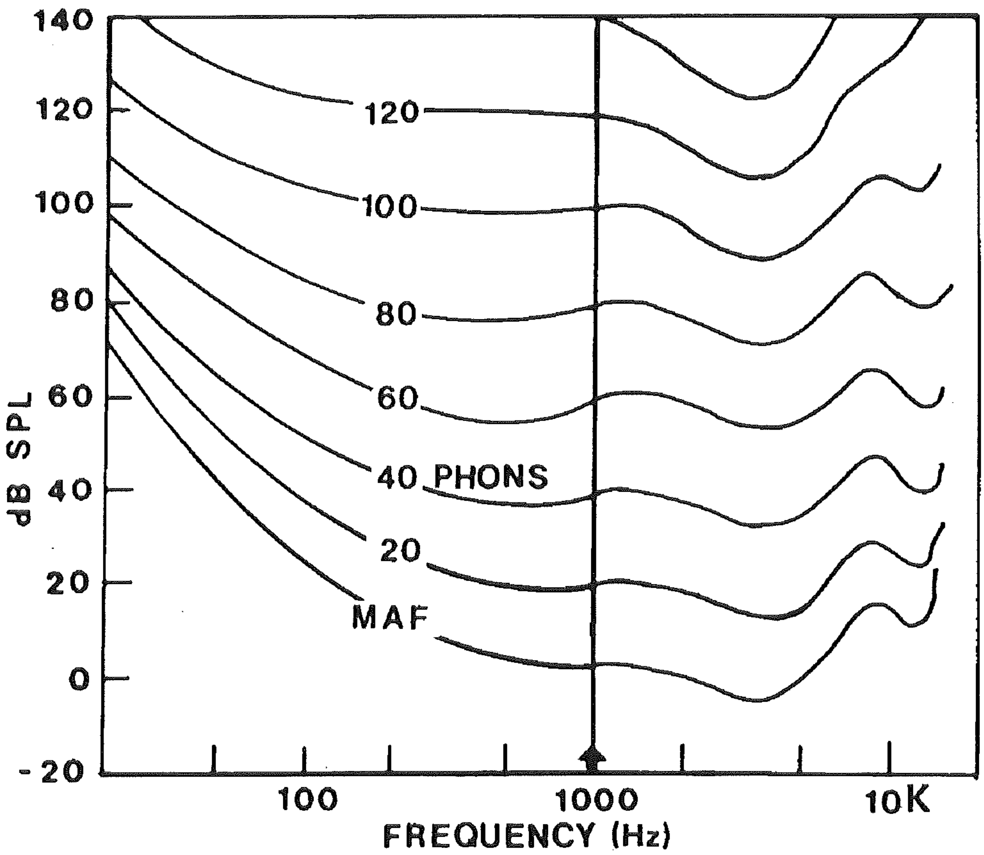


Figure 2-3. Equal loudness level contours (phon-curves) for pure tones. The lowest curve shows the threshold of hearing for normal-hearing subjects in a free sound field [4].

The unit of loudness is *sone*. One *sone* is 40 *phons* and when a sound is perceived twice as strong, the value is two *sones*. As a rule of thumb each doubling in *sones* corresponds to a 9 *phons* increase.

2.6 Weighting curves

In order to simulate human hearing, standardized frequency weighting curves have been developed from the *phon*-curves. They are used to get more adequate values of the annoyance *(figure 2-4)*. The A-weighting is derived from the 40 *phon*-curve, the B-weighting from the 70 *phon*-curve and the C-weighting from the 90 *phon*-curve. The newer D-weighting was developed for measuring jet-noise. Originally the A-weighting was supposed to be used for low sound pressure values, the B-weighting for medium values and the C- weighting for high values but today the A- weighting is commonly used regardless to the level.

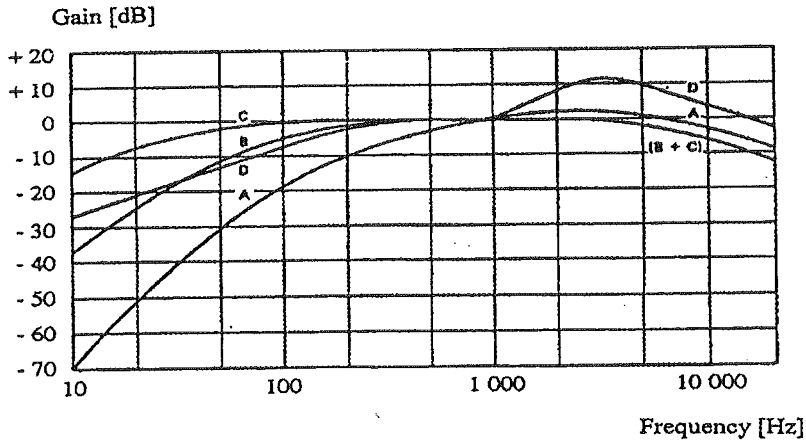


Figure 2-4. IEC-standardized weighting curves [1].

2.7 Brightness

Brightness is a subjective quantity of how bright a sound is perceived. It is closely dependent on which frequency components the sound is containing. Naturally, the more power of the higher frequency components the brighter the sound is perceived. For pure tones a physiologic scale has been developed for pitch. The unit is *mel* (not to be confused with musical pitch). The reference point is 1000 Hz which corresponds to 1000 *mel*. The principle is that a tone with a pitch twice as high has the double amount of *mel*. At low frequencies *mel* and Hz are nearly proportional, above say 1000 Hz the *mel*-scale becomes almost logarithmic [6]. That fact is connected to the width of the critical bands i.e. the width of the filters of the inner ear. Therefore it can be practical to show the frequency in *bark*. This frequency scale is namely directly related to the critical bands. The relation between frequency, f, in Hz and bark is defined by this formula [9]:

$$bark = 13 \times \arctan(0.00076 \times f) + 3.5 \times \arctan((\frac{f}{7500})^2) \qquad (2\text{-}7)$$

3 The test equipment

To be able to register how a subject perceives a sound we needed a measuring system able to handle simultaneous playback and recording. *Figure 3-1* shows the equipment used.

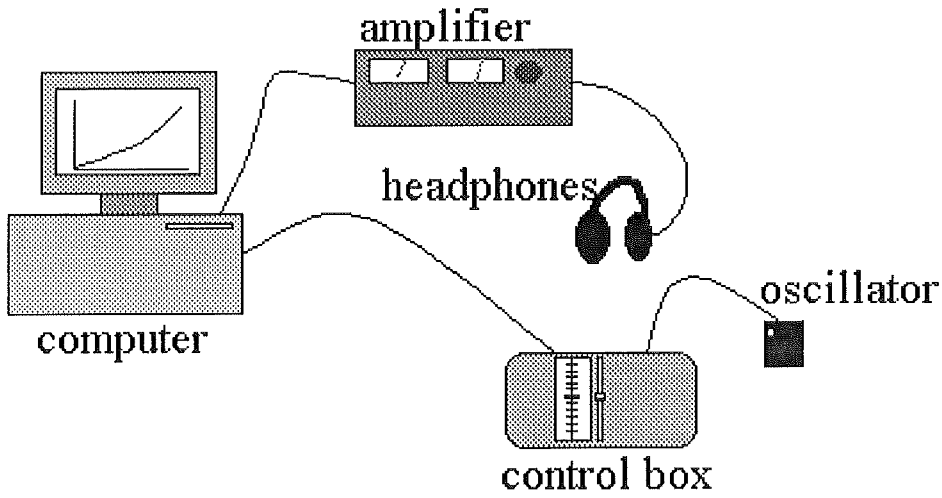


Figure 3-1. The test equipment.

The playback procedure was performed in the following way: the sound file, stored on the hard-drive, was played back by the sound card, amplified and presented to the subject via the earphones.

The recording part was done by using an oscillator to generate a square wave voltage. This voltage could be changed with a sliding potentiometer. The signal was then received by the sound card and stored on the hard-drive in a response file.

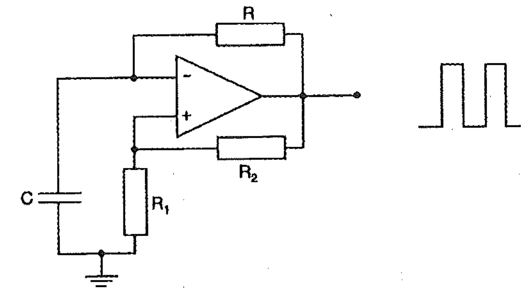
The computer:
The computer used was a *Dell* PC with a Pentium 166 MHz processor, 2.5 Gb hard-drive and 64 Mb RAM.

The software:
Since the number of sound- and response- files was large, we needed a program which could handle all the files. Therefore, a Visual Basic program we called *Test Manager* was developed which could take care of that (see chapter 4.4).

The sound card:
The sound card used was a *Turtle beach, FIJI* which has full duplex, i.e. it can do simultaneous play-back and recording. Since the card is AC-coupled the input signal needed to be an alternating voltage. Therefore we built a square wave oscillator.

The oscillator:

The oscillator circuit was constructed out of the following circuit diagram:



$$f_0 \approx \frac{1}{2\,RC\,ln\,(\frac{2\,R_1}{R_2} + 1)}$$

Figure 3-2. A diagram of the oscillator circuit [2].

The source was a 9V-battery. To keep the output voltage on a constant level a zener-diode was soldered on. The circuit was placed inside a small plastic box on which a switch, a light-diode and a RCA-contact was mounted.

The control box:

A linear sliding potentiometer from 0 to 10000 ohm with a lever travel of ten centimeters was mounted into the lid of an inclining box (see photo in appendix 4) with the following dimensions:

front height: 47 mm

rear height: 84 mm

lid: $308 \cdot 167$ mm

The output amplifier

To get sound pressure levels high enough we needed to amplify the output signal. The amplifier used was a *Yamaha Natural Sound B2.*

The headphones:

The headphones, *AKG K240*, enclose the whole ear (supra aural). Therefore they are easy to calibrate with an ear simulator, *Brüel &Kjaer 4153* (see *chapter 5.4* for details about the calibration). The ear simulator was also used to estimate the frequency response of the headphones *(see figure 3-3)*.

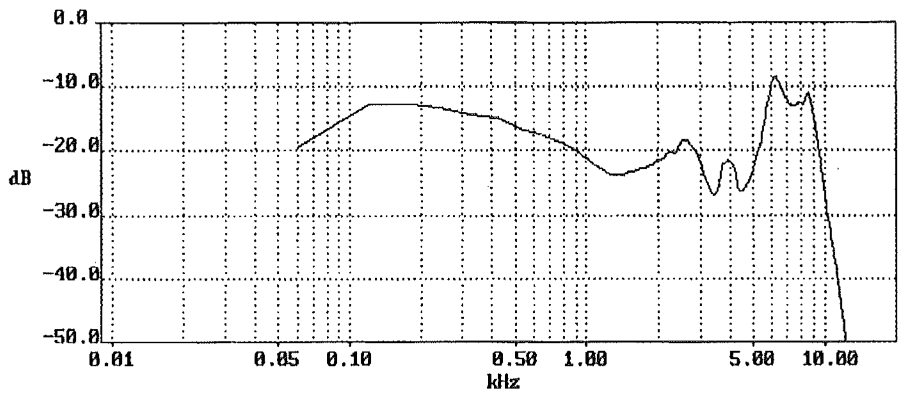


Figure 3-3. The frequency response of the headphones measured in ear simulator Brüel &Kjaer 4153.

4 The listening test

The test was divided into two parts, estimation of loudness and brightness, respectively. In the loudness test we used 16 different stimuli, 11 artificial noises, three pieces of music and two sounds from everyday life. In the brightness test we used six stimuli, four artificial noises (of which two were used in the loudness test) together with the everyday sounds.

4.1 The subjects

The subject group consisted of six males and one female in the age of 24-55 years. All participants declared themselves to have normal hearing.

4.2 The scales

The scales were chosen to have a range between 0 and 10 which has been shown to be the most appropriate graduation for listening test [3]. The scale mark 1 corresponds to *very quiet* and *very dull* respectively. A 9 represents *very loud* and *very bright (see figure 4-1 and 4-2).*

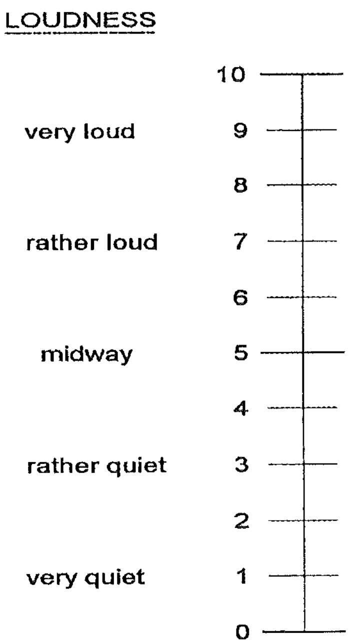


Figure 4-1. The scale used in the loudness test.

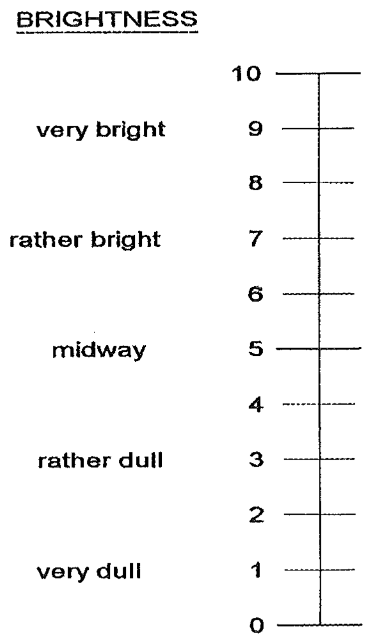


Figure 4-2. The scale used in the brightness test.

4.3 The test procedure

The test was divided into two parts, and started with loudness rating. After the subject had entered the sound isolated test room, he had a chance to practice on a few stimuli in order to familiarize himself with the control box and the scale. The actual test consisted of three rounds. Each round consisted of the 16 stimuli. They had a length of 24 seconds each, and were separated by a few seconds pause. After each round, the order of the stimuli was randomized and the next round took place. The total time for the three rounds was approximately 30 minutes. The brightness test was performed in a similar way but this time with 6 different sounds and with a total time of about 12 minutes.

4.4 Playback and recording

To take care of the test procedure we used our Visual Basic program, *Test Manager* (see code in appendix 2). The program consists of two forms and works as following: In the first form you choose the directory containing your sound files and add the files you like to use in the test to the *play list*. Then you enter the subject's initials and go to the next form.

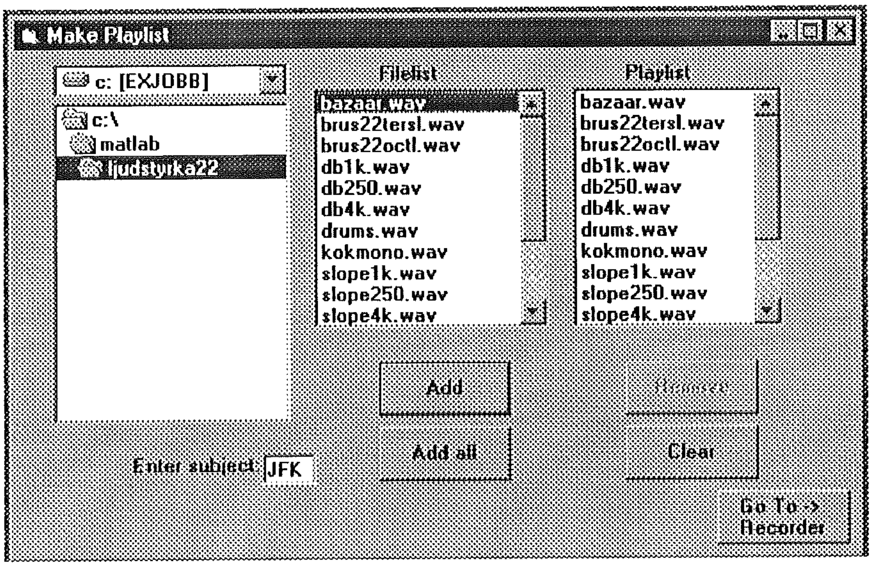


Figure 4-3. In the first form of the "Test Manager" you can pick the stimulus you like to have in your test.

In the second form you can choose the response round and which descriptor you like to test. The response files automatically gets the names according to this convention: *"Initials-Descriptor round-stimuli name"*. When you have the right settings you can start the test by clicking the 'play / record'-button. Between the response rounds you can change the play order of the stimuli by clicking the button 'randomize'.

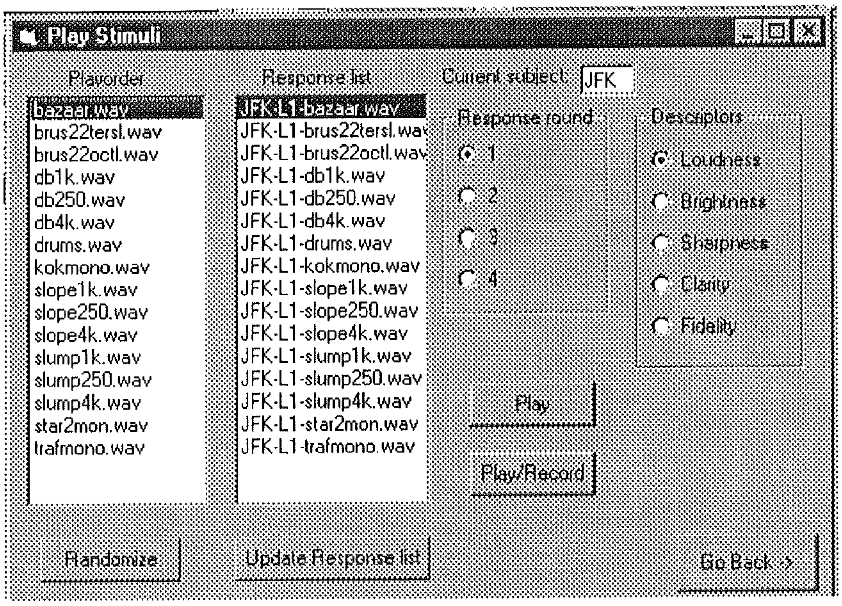


Figure 4-4. The second form takes care of the 'play/recording'-part.

5 Description of the stimuli and the response files.

All sound files were stored as Wave-files, a Microsoft standard which offers three sample frequencies, two bit formats and mono or stereo. The sample frequency 22050 Hz, 16-bits and mono was used for all files. The sample frequency 22050 Hz leads to the fact that the highest frequency component the sound files can contain, is about 11 kHz (half the sampling frequency). 16-bits gives 2^{16}=65536 discrete levels and thus the dynamic are approximately 96 dB ($20\log|2^{16}| \approx 96$ dB) which is enough in this listening test. Furthermore the sound files were presented in mono because the test does not demand the spaciousness of stereo sound. These three facts all together reduces the quantity of data to a quarter of what would have been the case if we had used full CD-quality *(see table 1)*.

Quality of Recording	Amount of Hard Drive Used
11 kHz, 8 bit, Mono	661K/Minute
11 kHz, 8 bit, Stereo	1.3 Meg/Minute
11 kHz, 16 bit, Mono	1.3 Meg/Minute
11 kHz, 16 bit, Stereo	2.6 Meg/Minute
22 kHz, 8 bit, Mono	1.3 Meg/Minute
22 kHz, 8 bit, Stereo	2.6 Meg/Minute
22 kHz, 16 bit, Mono	2.6 Meg/Minute
22 kHz, 16 bit, Stereo	5.3 Meg/Minute
44.1 kHz, 8 bit, Mono	2.6 Meg/Minute
44.1 kHz, 8 bit, Stereo	5.3 Meg/Minute
44.1 kHz, 16 bit, Mono	5.3 Meg/Minute
44.1 kHz, 16 bit, Stereo	10.5 Meg/Minute

Table 1. Various recording formats and the amount of data they are causing.

5.1 Music and everyday sounds

There were three music stimuli, one Christmas carol performed by a church choir, one drum piece and one piece with drums, clarinet and Hammond organ. The two stimuli from everyday life were one section of traffic noise and one sequence from a kitchen. The five stimuli in table 2 were all used in the loudness test, the everyday sounds were also used in the brightness test.

Name	Type	Source
star2mon	choir	**Live recording** S:t Klara Motett Choir, Stockholm 1997 **piece:** *"Det strålar en stjärna"*
drums	congas, bongos etc.	**CD:** *"Bazaar Musik"* **track:** 16 **time:** 0:00-0:24
bazaar	clarinet, organ, drums	**CD:** *"Bazaar Musik"* **track:** 17 **time:** 0:00-0:24
kokmono	variant kitchen sounds	**CD:** Brüel & Kjaer *"Sound tailor fits all"* **track:** 10 **time:** 0:20-0:44
trafmono	traffic noise	**CD:** Widex *"Real-life Environment Sound examples"* **track:** 4 **time:** 1:08-1:32

Table 2. The music- and everyday sound used in the tests.

5.2 Artificial noises

All together we used 13 different noise stimuli which were generated in Matlab by modifying white noise (see code in appendix 1). They can be separated into two groups. The sounds in the first group, (nine stimuli), were both filtered and amplitude modulated whilst the sounds in the other group, (four stimuli), were only filtered. The creation of the first group was made in the following way: white noise was filtered to three octave bands *(figure 5-1)* with the center frequencies 250, 1000 and 4000 Hz. We modulated each of the three octave bands by multiplying each of them with three different types of amplitude functions. One was exponential, the second linear and the third changed instantaneously from one level to another *(figure 5-2, 5-4 and 5-6, respectively)*. The levels, the inclinations and the interval times were randomly generated. This resulted in three types of sound files *(figure 5-3,5-5 and 5-7)*.

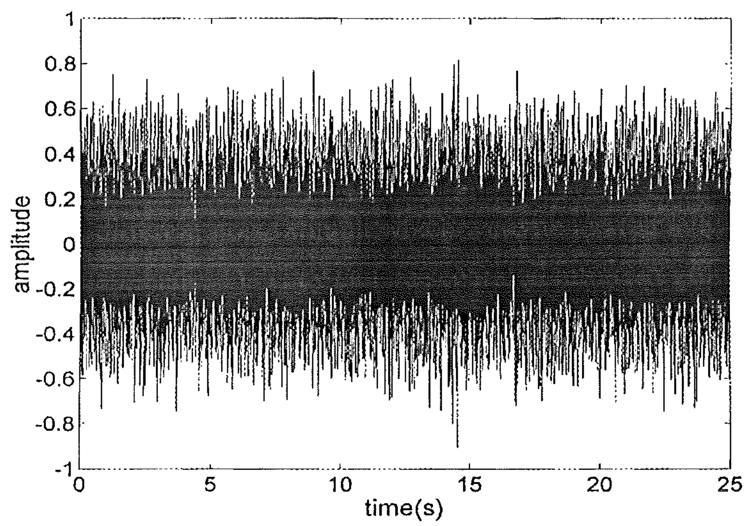


Figure 5-1 The time signal of octave filtered noise.

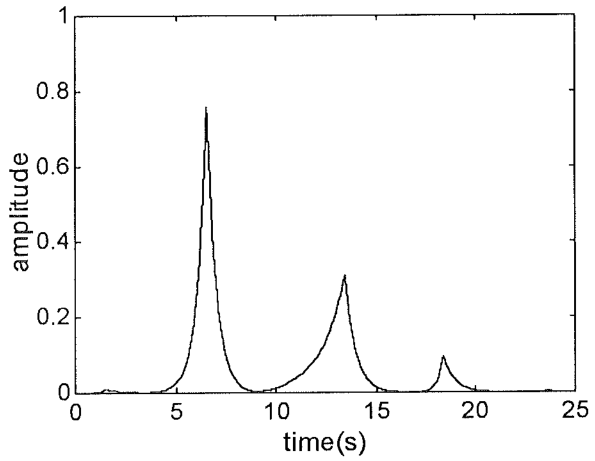


Figure 5-2. The first type of amplitude function (exponential).

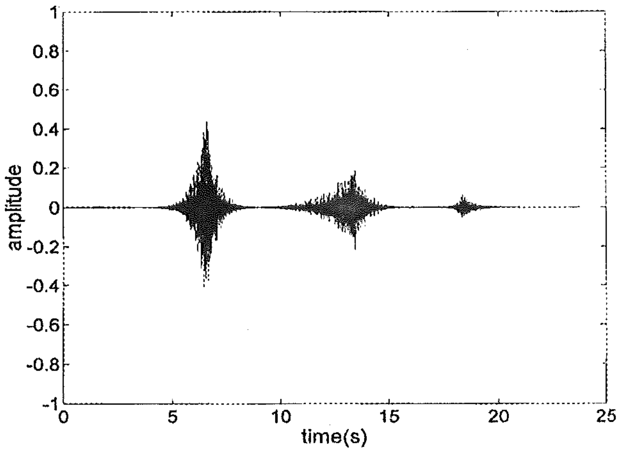


Figure 5-3. The resulting amplitude modulated noise.

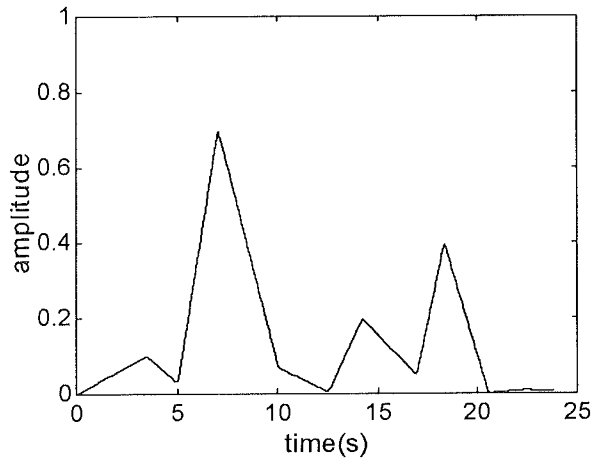


Figure 5-4. The second type of amplitude function (linear).

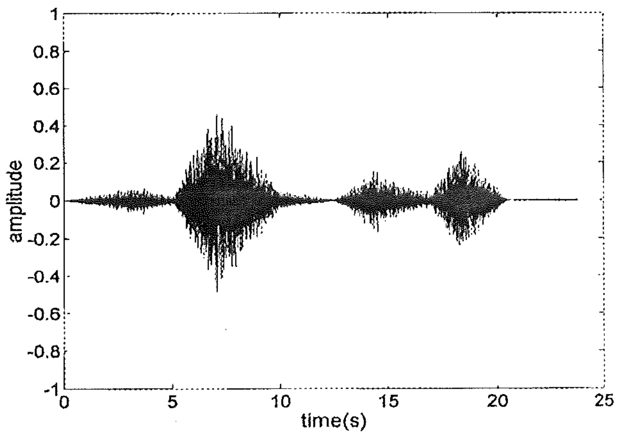


Figure 5-5. The resulting linear amplitude modulated noise.

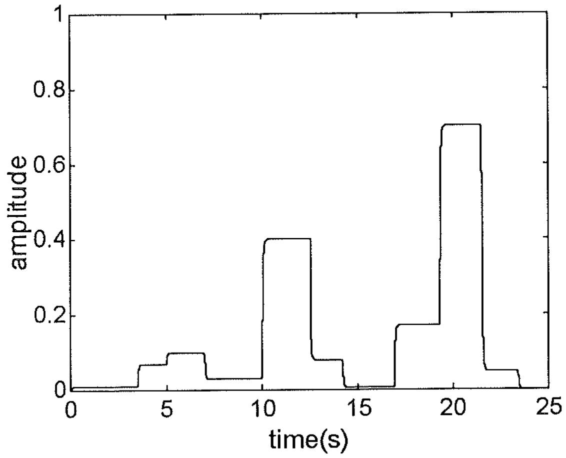


Figure 5-6. The third type of amplitude function

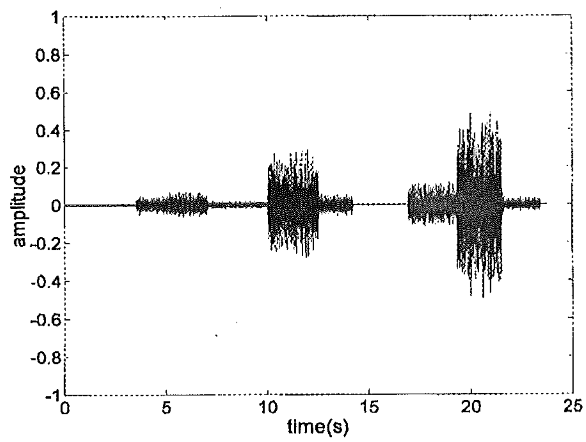


Figure 5-7. The resulting stepwise amplitude modulated noise.

After these steps we had nine artificial noise stimuli which were only used in the loudness test *(table 3)*.

name	**type of change**	**bandwidth**
db250	exponential	177<f<344
db1k	exponential	707<f<1414
db4k	exponential	2828<f<5656
slope250	linear	177<f<344
slope1k	linear	707<f<1414
slope4k	linear	2828<f<5656
slump250	stepwise	177<f<344
slump1k	stepwise	707<f<1414
slump4k	stepwise	2828<f<5656

Table 3 Nine artificial noises used in the loudness test.

When creating the second group, white noise was filtered with four different kinds of filters, lowpass-, highpass-, octave- and third-octave filters. The limiting- and the center frequencies were randomly selected. After each interval time, which was also randomly selected, the cut-off and the center frequencies changed *(see figure 5-8)*.

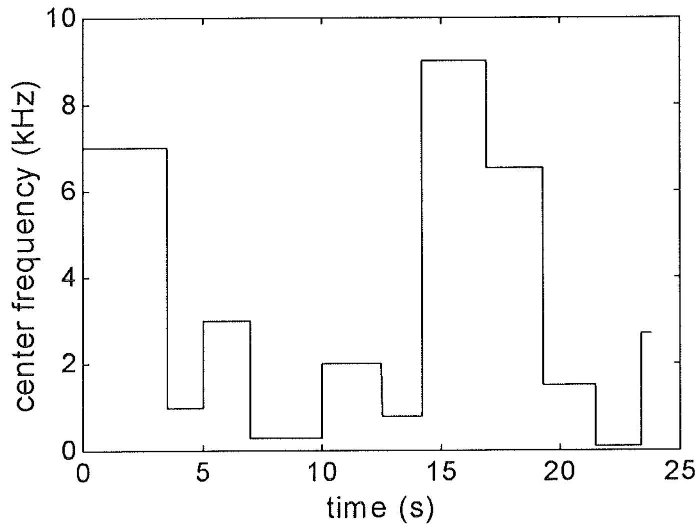


Figure 5-8. An example on how the center frequency changes with time.

These four stimuli were used in the brightness test. 'brus22oct' and 'brus22ters' were also used in the loudness test.

name	filter type
brus22ters	third-octave
brus22oct	octave
brus22low	lowpass
brus22high	highpass

Table 4. Four different types of filtered noise used in the brightness test.

In total eighteen different stimuli was used in the tests. Table 5 shows all the stimuli and in which tests they occurred.

name	loudness	brightness
slope250	x	
slope1k	x	
slope4k	x	
db250	x	
db1k	x	
db4k	x	
slump250	x	
slump1k	x	
slump4k	x	
star2mon	x	
drums	x	
bazaar	x	
kokmono	x	x
trafmono	x	x
brus22ters	x	x
brus22oct	x	x
brus22low		x
brus22high		x

Table 5. All 16 stimuli used in the tests.

5.3 Calibration

To be able to present desirable levels to the subject, we had to calibrate the output system. We therefore made a 1 kHz sine tone in Matlab with the amplitude 1 which is the maximum value for a Wave-file. The headphones were then placed on an ear simulator, *Brüel & Kjaer 4153,* with a microphone inside (calibrated with a pistonphone) that was connected to a sound level meter. Since we only had a dynamic range of 96 dB, we chose the interval 14 to 110 dB SPL and for **rms**-values 11 to 107 dB SPL. We played the tone and adjusted the output amplifier so that the output level became 107 dB SPL. It was then possible to calculate the sound pressure levels of all the stimuli.

5.4 Calculation of the sound pressure levels of the stimuli

In order to get a physical value of the strength of the sounds, they have been processed in several steps. Eventually we get a curve for each stimulus showing its dB(A)-value as a function of time. This curve is supposed to correspond fairly well to the loudness judgment. First, the sound file was A-filtered by transforming the time signal to the frequency domain. It was then multiplied with the A-weighting curve and transformed back to the time domain *(figure 5-9 to 5-13)*.

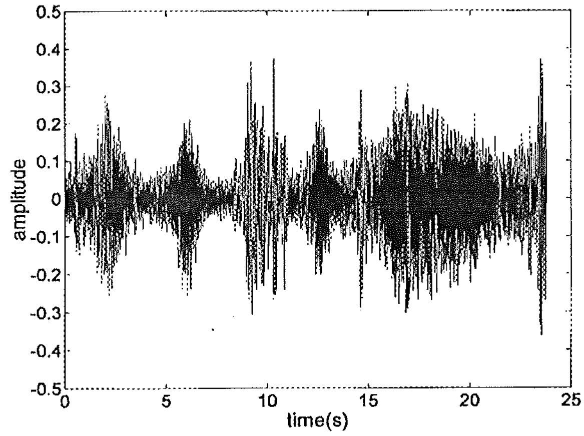


Figure 5-9. Traffic noise.

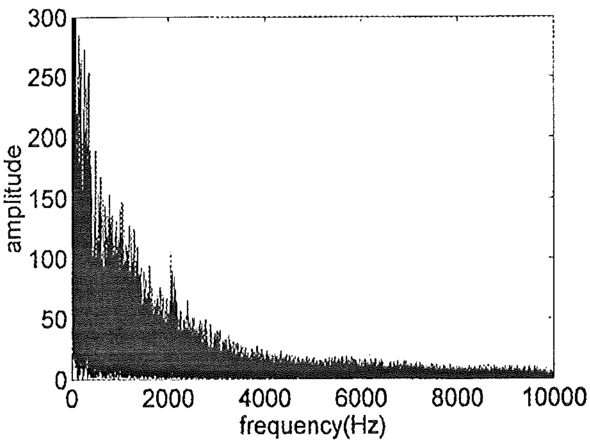


Figure 5-10. Spectrum of traffic noise.

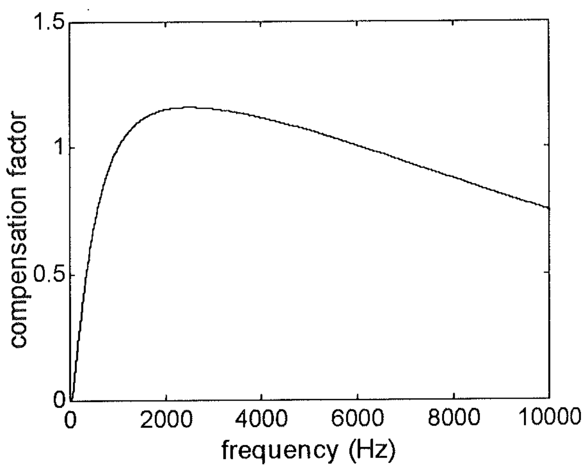


Figure 5-11. A-weighting curve.

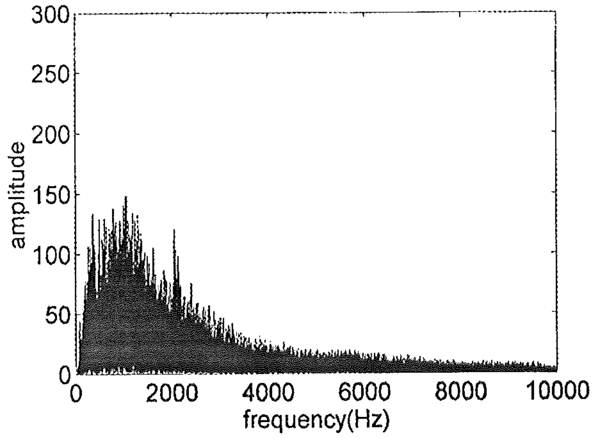


Figure 5-12. Spectrum after A-weighting.

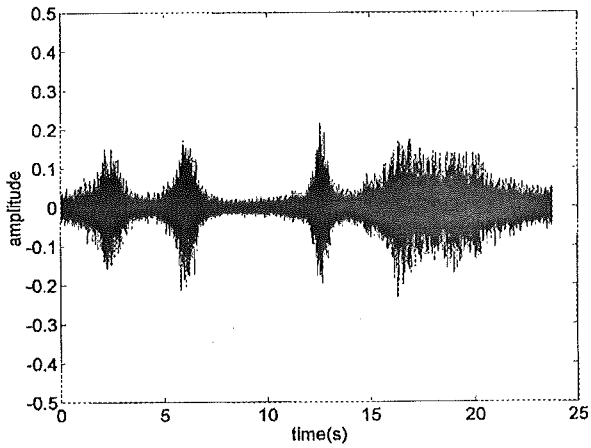


Figure 5-13. A-weighted traffic noise.

In order to get information of the power of the signal the **rms**-value was calculated. This was done in three steps:

* **Squaring** *(figure 5-14).*

* **Lowpass filtering**, which is equal to calculating the mean value over a certain time corresponding to the time constant of the filter.

* **Square rooting** *(figure 5-15).*

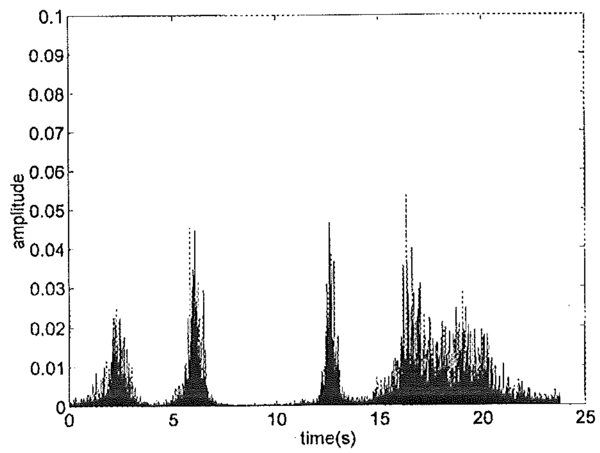


Figure 5-14. Squared traffic noise.

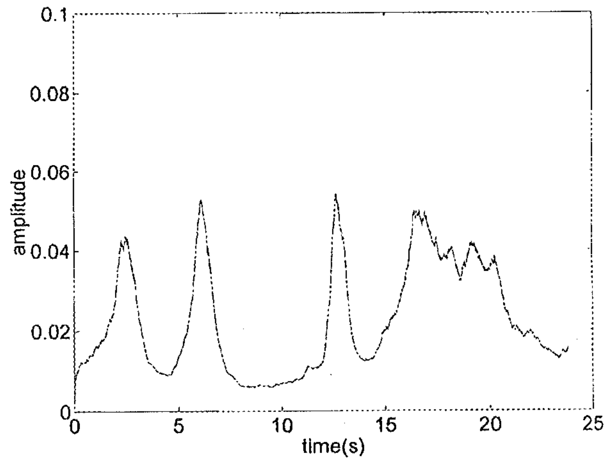


Figure 5-15. The envelope after lowpass filtering and square rooting.

By finally calculating the logarithm of the obtained **rms**-value, we get the sound pressure level in dB(A). Figures 5-16 to 19 show four types of envelopes in SPL: one everyday sound (the traffic noise) and the three types of amplitude modulated noise.

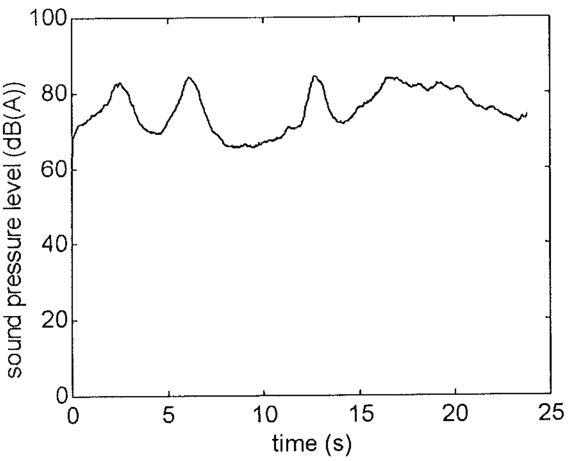


Figure 5-16. The envelope of the traffic noise in dB(A).

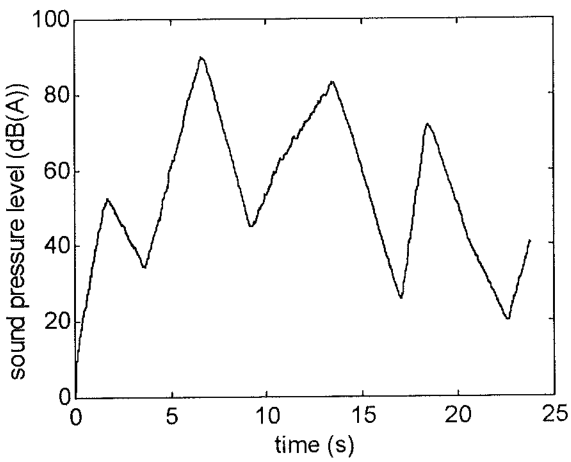


Figure 5-17. The envelope of an exponential amplitude modulated noise.

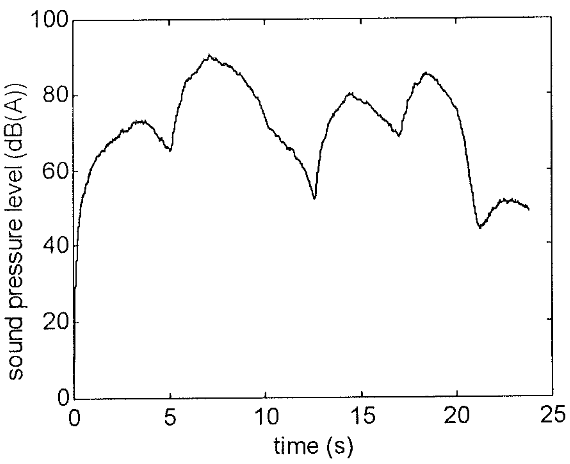


Figure 5-18. The envelope of a linearly amplitude modulated noise

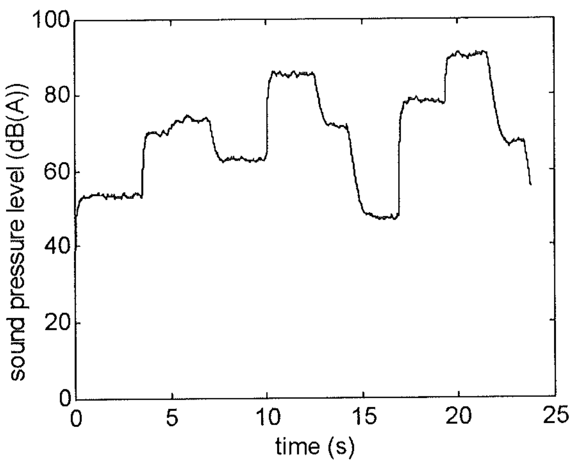


Figure 5-19. The envelope of a stepwise amplitude modulated noise

In this state it was all right to resample the envelopes in order to reduce the quantity of data. That was done by picking out every hundred value of the envelope and save those in a Matlab data file (Mat-file). Thus the envelopes were represented with 220.5 samples per second.

5.5 A physical representation of brightness.

The brightness of a sound depends on the relation between the power of the high and the low frequency components. The more power of the higher components the brighter the sound is perceived. One simple thing you can do to describe this relation is to calculate the balance point of the spectrum of the signal:

$$f_{bp} = \frac{\sum_{i=1}^{n} A_i \cdot f_i}{\sum_{i=1}^{n} A_i} \qquad (5\text{-}1)$$

where

f_{bp} = *the balance point of the spectrum.*

A_i = *the amplitude of the spectral component i.*

f_i = *the frequency of the spectral component i.*

To get an envelope corresponding to how bright a sound is we divided the signal into small sections (0.2 s/each) and calculated the **balance point** of each section. This resulted in a jagged curve which could be smoothened by a lowpass filter *(figure 5-20)*.

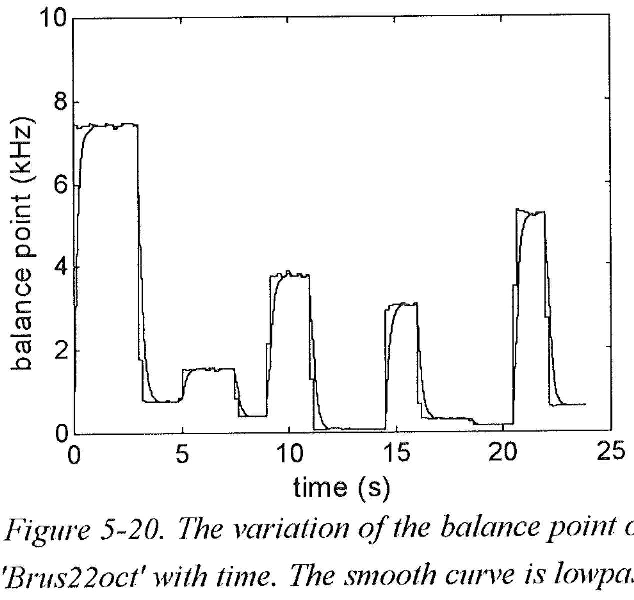


Figure 5-20. The variation of the balance point of 'Brus22oct' with time. The smooth curve is lowpass filtered.

5.6 The response files

In order to get a curve corresponding to the ratings the subject had done the response files, also sampled with 22050 Hz, had to be processed. First, the absolute value of the sampled signal was calculated, the result was then lowpass filtered. The received envelope was resampled in the same way as the sound signal and finally standardized so that it spanned between zero and ten.

6 Analysis methods and results

6.1 Analysis and results of the loudness test

We have divided the analysis of the sound and response files of the loudness test into the following sections:

* **Reaction times** - how the reaction times differ between the subjects and how the reaction times are affected by different large and fast level changes.
* **Variation between subjects** - how much the subjects differ from each other in time delay and level.
* **Reproducibility** - how much each subject differs between the different rounds.
* **The sensitivity curve of the ear** - plotting of the perceived loudness as a function of the sound pressure level.
* **The transfer function-** calculation of the transfer function and the coherence function to the system SPL in dB(A) → rated loudness.
* **Simple model of the hearing process**

All the analyses were performed in Matlab using self made programs (see code in appendix 1). Since the sound and response files were digitized it was convenient to do the analysis in Matlab, which has many tools for signal processing. The analysis was, like the test, divided into two parts; estimation of loudness and brightness respectively.

6.1.1 Plotting the envelopes

It is desirable to be able to compare the response envelope with the sound envelope in the same plot. We found that the highest sound level occurring in our tests was approximately 100 dB(A) and we know that the responses spanned between 0 and 10. Therefore we used two y-scales when we plotted the sound and response together. We did not include the first second in the plot because the subject needed a certain time to get started. *Figure 6-1* shows an example of a plot of a sound envelope and a corresponding response envelope.

Every subject has three response curves to each stimulus and we have calculated the average curve of those three. This we will call *average curve*. We have also calculated an average curve of all subjects to each stimulus. This we will call *mean curve over subjects*.

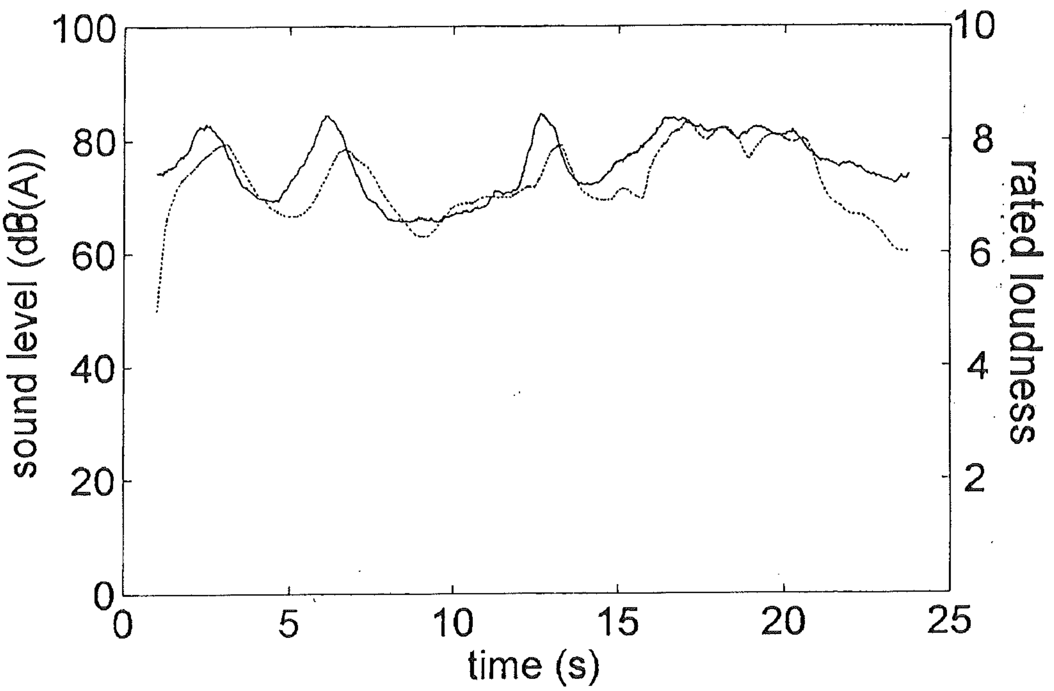


Figure 6-1. The sound envelope of the traffic noise (heavy line) and the response of one subject (dashed line).

6.1.2 Cross correlation technique

You can get a value of the average time delay between two signals by using cross correlation technique [7]. The covariance function is obtained if the mean values of the signals are subtracted before cross correlation is performed. By calculating this and studying the time for its maximum value, we get information about the time delay between the two signals. *(figure 6-2 and 6-3).*

$$r_{xy}(\tau) = \int_{-\infty}^{\infty} (v_1(t) - v_{1m}) \cdot (v_2(t+\tau) - v_{2m}) dt \qquad (6\text{-}1)$$

where

$r_{xy}(\tau) =$ *the covariance function*

$\tau =$ *lags (s)*

$v_1(t) =$ *the first time signal*

$v_{1m} =$ *the mean value of the first signal*

$v_2(t) =$ *the second time signal*

$v_{2m} =$ *the mean value of the second signal*

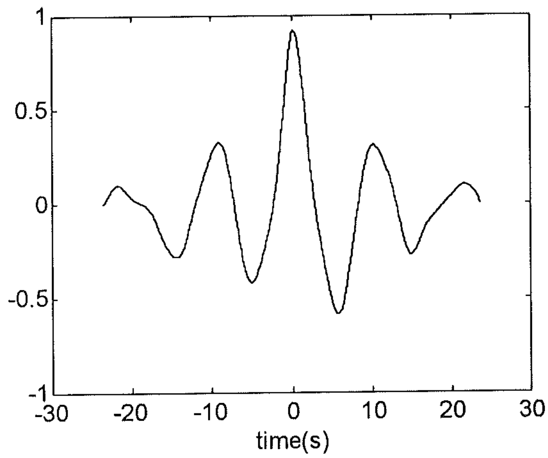


Figure 6-2. The covariance function between two signals gives information of the time shift between them.

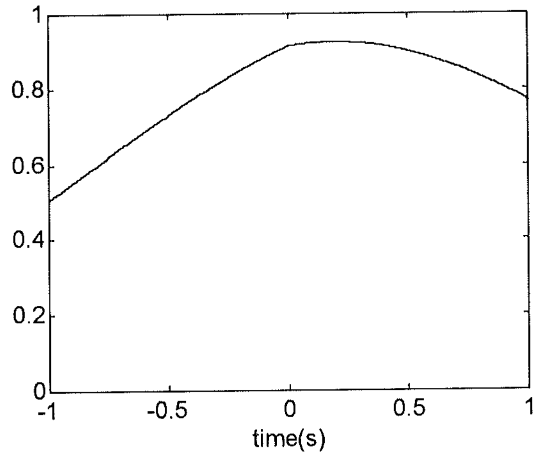


Figure 6-3. The same covariance function magnified.

6.1.3 Reaction times

By using the correlation technique between response and sound we get the average delay between those signals. This time-shift can be regarded as a measurement of the subject's average reaction time to that stimulus. We have studied all subjects' reaction times for five of the stimuli *(table 6)*.

subject nr	db1k	slope1k	slump1k	trafmono	drums	mean	st. dev.
1	0.31	0.41	0.63	0.46	0.59	0.48	0.12
2	0.15	0.62	0.49	0.55	0.40	0.44	0.16
3	0.02**	0.46	0.33	0.41	0.37	0.39	0.05
4	0.15	0.22	0.63	0.39	0.44	0.37	0.17
5	0.18	0.32	0.53	0.55	-6.98* / 0.75	0.47	0.20
6	0.29	0.57	0.44	0.54	-7.65* / 0**	0.46	0.11
7	0.59	0.55	0.54	0.57	-7.51* / 0.34	0.52	0.09
mean	0.24	0.45	0.51	0.50	0.48	0.45	
std	0.18	0.14	0.11	0.07	0.16		

Table 6. The reaction times in seconds of all subjects to five of the stimuli.

* This value is the maximum value of the covariance function. Since it is not realistic, we took the time for the first peak after zero.

** This reaction time is too fast and is not included in the mean value and the standard deviation calculation.

Note that the average reaction time of 'db1k' is much better than for the other stimuli. However the standard deviation is high and it is not certain that the difference is significant.

The average time delay does not give any information about the reaction times for each event. By an event we mean when the sound level changes instantaneously (see figure 5-19) or when the sound level alternates between inclination and declination (see figure 5-17). First we have studied instantaneous level leaps as in the three stimuli 'slump250', 'slump1k' and 'slump4k' to see how the sizes of the leaps affect reaction times. We also studied continuous level changes as in the 'dB'-stimulus to see how the gradient of the slopes affect the reaction times. The reaction times and the size of the changes were calculated from the figure by marking the events with the cursor *(figure 6-4)*. By subtracting the time for a peak in the sound curve from the corresponding peak in the response curve, we got the reaction time for that peak and by subtracting the amplitude of the following valley from the amplitude of the peak we got the level change of that event.

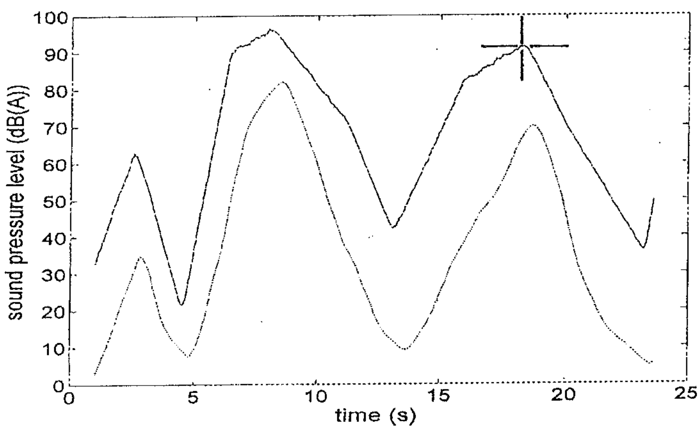


Figure 6-4. The reaction times was estimated by marking on the curve with a cursor.

By using this method we calculated the reaction times for each stimulus. After doing that for all subjects' *average curves* to the 'db'-sounds and the 'step'-sounds respectively, we could plot the reaction times for the 'db'-sounds towards their corresponding inclination in dB/s *(figure 6-5)* and the reaction times for the 'step'-sounds towards their corresponding level change in ΔdB *(figure 6-6)*.

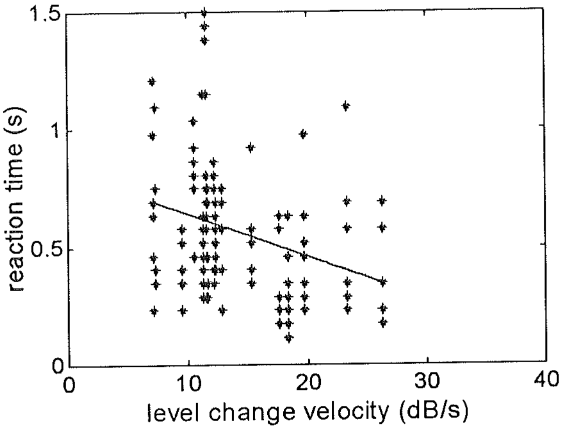


Figure 6-5. The reaction times of all subjects for the events of the three 'db'-sounds plotted towards their corresponding slopes and an adapted linear regression to all dots.

The equation of the line in figure 6-5 is:
y=0.82-0.018*x

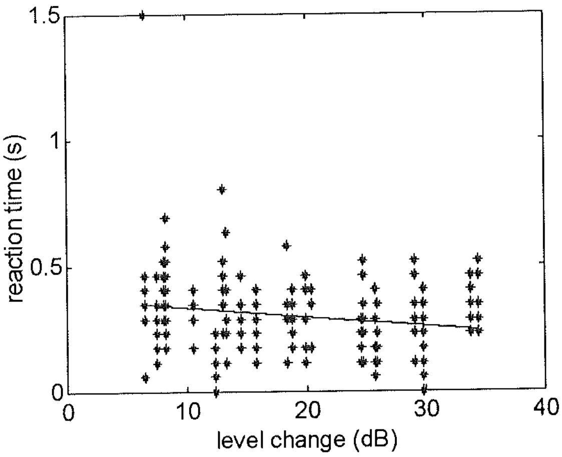


Figure 6-6. The reaction times of all subjects for the events of the three 'step'-sounds plotted towards their corresponding leaps and an adapted line to all dots.

The equation of the line in figure 6-6 is:
y=0.37-0.004*x

You can see that the reaction times are shorter for larger changes. The relation is not as clear in the second plot as in the first. That can depend on the fact that it is easy to hear an instantaneous leap even though it is small.

6.1.4 Variation between subjects

It can be interesting to compare the subjects' response curves with each other. That way we get information on how easy the stimulus was to follow. In figure 6-7 we see the sound envelope to the stimulus 'db1k'. We see in figure 6-8 that that stimulus seems to be an 'easy' stimulus since all subjects' *average curves* have the same shape. In figure 6-9 and 6-10 respectively we see the sound and response envelopes to 'drums', that seems more 'difficult' since the curve shapes differs quite much from each other.

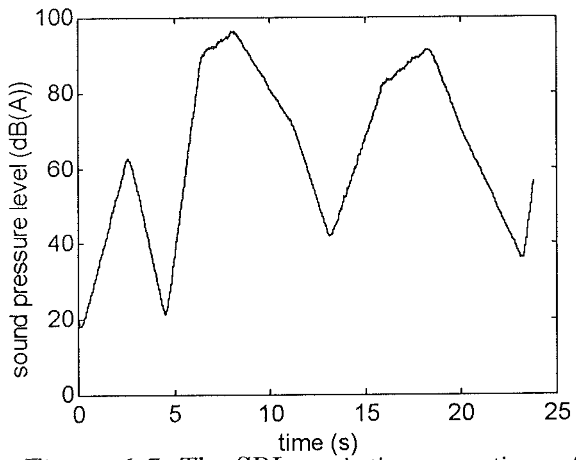


Figure 6-7. The SPL-variation over time of 'db1k'.

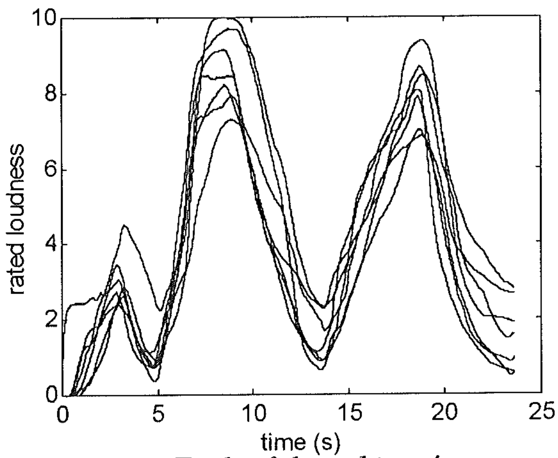


Figure 6-8. Each of the subjects' average response curve to 'db1k'

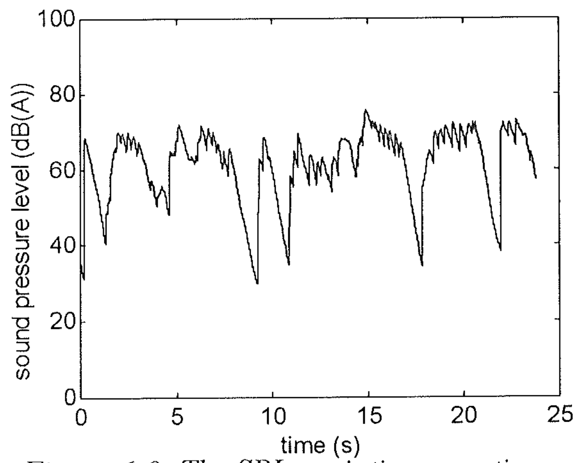


Figure 6-9. The SPL-variation over time of 'drums'.

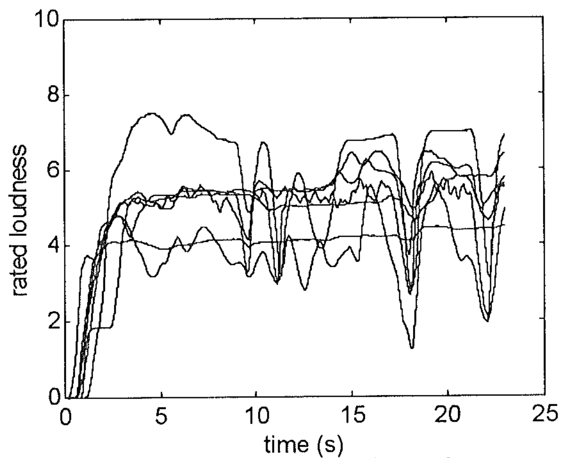


Figure 6-10. Each of the subjects' average response curve to 'drums'.

Now we will try to find a measure of how easy each stimulus was to follow. In section 6.1.3 we calculated the standard deviations of the reaction times which gives information on the deviation in time between the subjects (see table 7 column 2). It is now desirable to calculate the deviation in rating between the subjects. That was performed in the following way: First, to remove constant level differences, we subtracted the mean value of the amplitude of each subject's *average curve* (see end of 6.1.1 for definition). From the resulting curve, we then calculated the standard deviations of the amplitude between all subjects' *average curves* for each discrete time. This was done for each of the five stimuli in table 8. This resulted in a series of values for each stimulus for which we calculated the mean value (see table 8 column 2) with the following formula:

$$\bar{s} = \sqrt{\frac{\sum_{i=1}^{n} s_i^2}{n}} \tag{6-2}$$

where

$\bar{s}$ = *average standard deviation*

s = *standard deviation of each discrete time value.*

n = *the number discrete time values.*

The deviations in time and level respectively do not give complete information on how easy the stimulus was to follow since it is more likely to get larger deviations between the subjects if the variation of the response level of the *mean curve over subjects* (see 6.1.1) is large. To get values of how large these variations are, we calculated the standard deviations of each *mean curve over subjects* (see column 3 in table 7 and table 8). This we will call *standard deviation of response level*. We then normalized the average standard deviations by dividing them with their corresponding *standard deviation of response level*. This results in one quote per table, quote 1 and quote 2 (see column 4 in table 7 and table 8). Finally we multiplied these two quotes with each other and got an *index* of how easy each stimulus was to follow or actually how equally it was perceived by the subjects. The lower the index the easier the stimulus was to follow (see column 5 in table 8).

stimuli	st. dev of reaction times (s)	st.dev of response level (scale units)	quote 1
slump1k	0.11	1.72	0.064
db1k	0.18	2.50	0.072
slope1k	0.14	1.48	0.095
trafmono	0.07	0.81	0.086
drums	0.16	0.50	0.32

Table 7. Five selected stimuli from the loudness test. The standard deviations of the reaction times.

stimuli	mean standard deviation in rating (scale units).	st.dev of response level (scale units)	quote 2	index $*10^{-2}$
slump1k	0.66	1.72	0.384	2.5
db1k	0.74	2.50	0.296	2.1
slope1k	0.53	1.48	0.358	3.4
trafmono	0.59	0.81	0.728	6.3
drums	0.68	0.50	1.36	44

Table 8. Five selected stimuli from the loudness test. The standard deviation of the ratings.

We can see on the index values that the noise stimuli were almost equally easy to follow. The traffic noise was rather difficult and the drums wer very difficult to follow. This agrees with what the subjects claimed after the test.

6.1.5 Reproducibility

A stimulus should be perceived the same way regardless of how many times you hear it. Therefore it is interesting to see how much a subject's response differed between different rounds, both in time delays and levels. Figure 6-11 shows one subjects three response curves to 'db1k'.

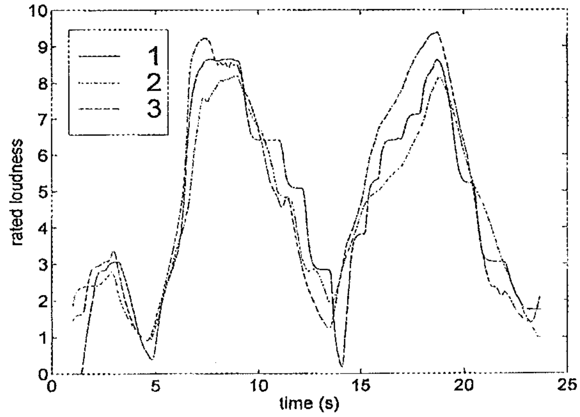


Figure 6-11. One subject's three response rounds for the stimulus 'db1k'.

First we calculated the average time shift between two rounds with the correlation technique described in section 6.1.2. After having compensated for that time shift, we estimated the mean value of the absolute value of the differences in each sample between the two rounds. We compared rounds 1 with 2 and 2 with 3 for five of the stimuli. The examination of the stimulus 'db1k' resulted in the following table:

db1k				
subject	**Round 1&2**		**Round 2&3**	
	time shift (s)	**level deviation (scale units)**	**time shift (s)**	**level deviation (scale units)**
1	-0.67	2.27	0	0.66
2	0	0.46	0	0.57
3	-0.20	0.90	0	0.91
4	-0.12	0.75	0	0.80
5	0	0.68	-0.11	0.69
6	0.02	1.01	0	0.63
7	0	1.02	0	0.80
mean	**-0.14**	**1.01**	**-0.02**	**0.72**

Table 9. The subjects' time-shift and level deviation between the rounds for the stimuli 'db1k'.

We see that the differences between round 2 and 3 are smaller than between round 1 and 2, both in level deviation and time shift. This was also the case for the four other sounds we studied (see appendix 3). The cause of that can be that the subjects get more secure on their rating the more times they listen to a stimulus. It indicates that a certain amount of training for the subject is needed to obtain reproducible results.

6.1.6 The sensitivity curve of the ear

To study if there is a linear relation between sound pressure level and perceived loudness, we plotted every rated loudness sample towards its belonging SPL-value. The average time delay between the sound and the response was calculated with the correlation technique. After having compensated for that time delay we plotted rated loudness towards SPL. This resulted in a set of dots to which we could adapt a polynomial of the third order by using the least square method. In figure 6-12 we see the set of dots and the adapted polynomial. Note that the dots are so close together that they form lines.

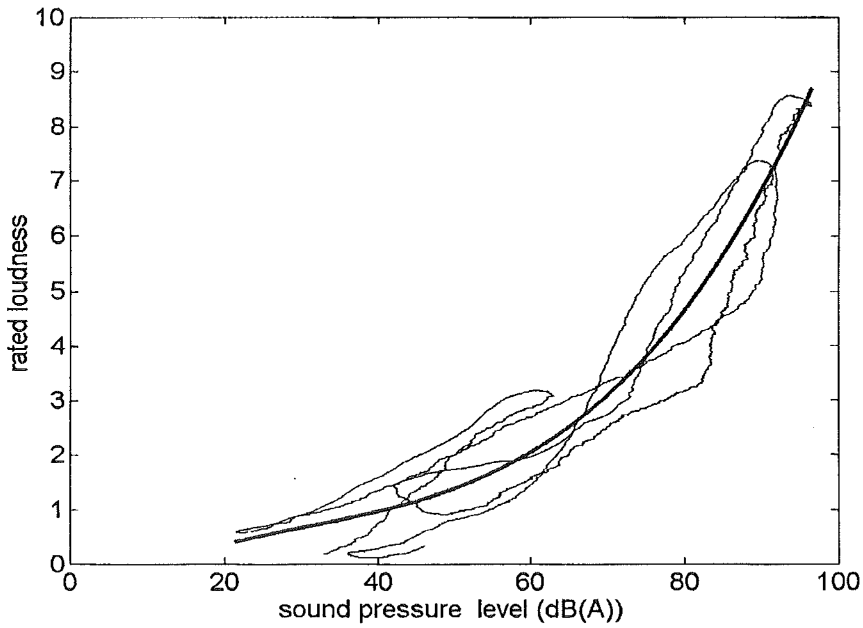


Figure 6-12. One subject's response plotted towards the sound pressure level of 'db1k'. A third degree order polynomial adapted to the set of dots (heavy line).

This gives information about how sensitive a person is to different loud levels. If we plot all 16 *mean curves over subjects* (see 6.1.1) towards the corresponding SPL-value we get a new set of dots to which we can adapt a third degree polynomial that should correspond to an average hearing (see figure 6-13 and further in section 6.2.1).

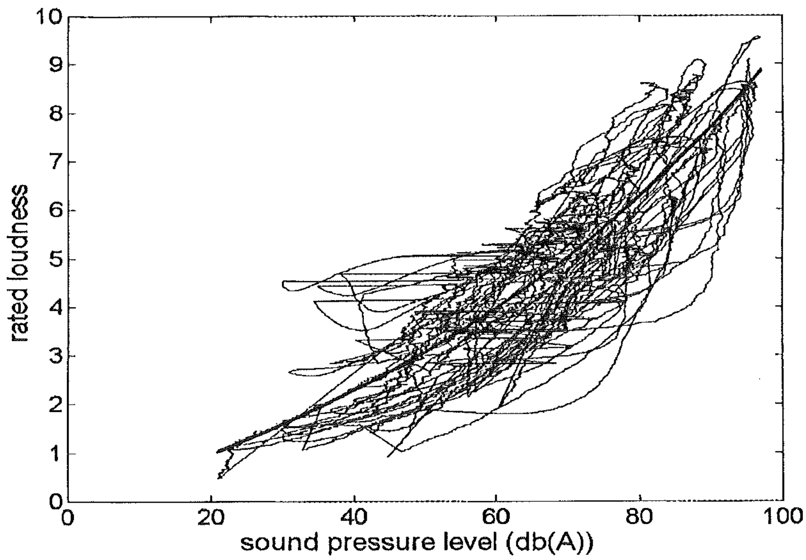


Figure 6-13. An adapted polynomial (heavy line) to the 16 average response curves plotted towards their corresponding SPL-values.

The equation of this polynomial is

$$y=-0.18+0.058*x-9.6*10^{-5}*x^{2}+4.8*10^{-6}*x^{3}$$

6.1.7 The transfer function

The relationship between the SPL of a stimulus and its perceived loudness can be treated as a transfer system. Therefore we have estimated the corresponding transfer function and the coherence of that system. In figure 6.14 we show an example of a transfer function and we can see that it has a lowpass character. We also estimated the coherence function of the system *(figure 6.15)*and found that it was fairly good up to one Hz which means that the system is not quite linear above that frequency.

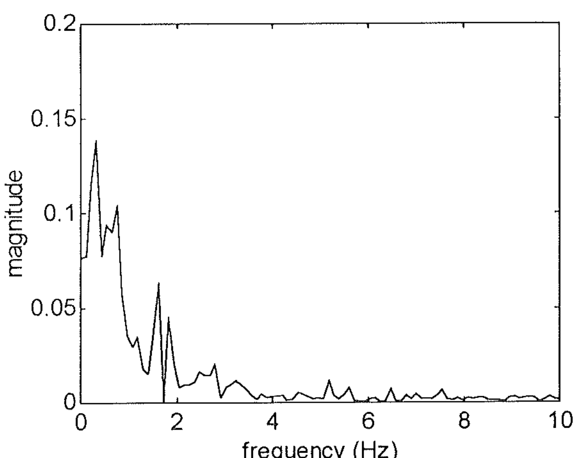


Figure 6-14. A transfer function between a sound envelope and a response.

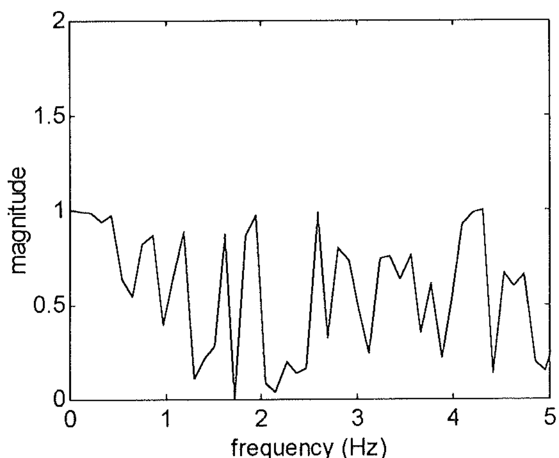


Figure 6-15. The coherence function between the same signals.

6.2 Application of the results from the loudness test

One application of the results from the analysis above is to try to estimate a simple model of the hearing process. This would enable you to pick any sound you like and simulate how an average person's response of that sound would look like. The path of incoming sound pressure variations from a subject's ear to the response curve he gives is complex, thus our model contains several steps, namely:

1 A-weighting
2 RMS-calculation
3 Logarithm-calculation
4 Level compensation
5 Lowpass filtering

The first three steps have already been done.

6.2.1 Level compensation

In the section *Sensitivity curve of the ear* we examined the relationship between SPL in dB(A) and loudness and found that it is not linear. In this model we use the coefficients for the polynomial that was estimated in section 6.1.6 to change the envelope according to the following:

$$newenv = a_0 + a_1 \cdot soundenv + a_2 \cdot soundenv^2 + a_3 \cdot soundenv^3 \qquad (6.3)$$

where

newenv=the compensated sound envelope

soundenv=the envelope in dB(A)

$a_0 = -0.18$

$a_1 = 0.058$

$a_2 = -9.6 \cdot 10^{-5}$

$a_3 = 4.8 \cdot 10^{-6}$

Figure 6-16 shows the envelope we used to compensate the envelope and figure 6-17 shows the result of that compensation.

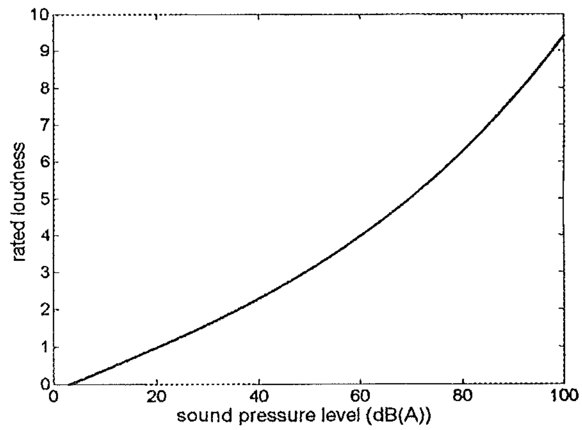


Figure 6-16. The polynomial we used to compensate the sound envelope.

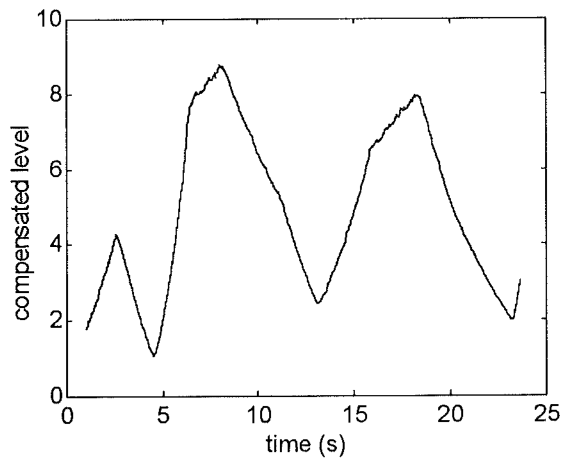


Figure 6-17. The envelope of 'db1k' after level compensation.

6.2.2 Lowpass filtering

Since humans are not able to respond to variations faster than a few Hz, we function as lowpass filters. We saw in chapter 6.1.7 that the transfer function has a lowpass character. By using the *mean curves over subjects* (6.1.1) we could estimate a transfer function for each stimulus *(figure 6-18* shows all 16 transfer functions) and then estimate an average transfer function of all 16 stimuli *(figure 6-19)*.

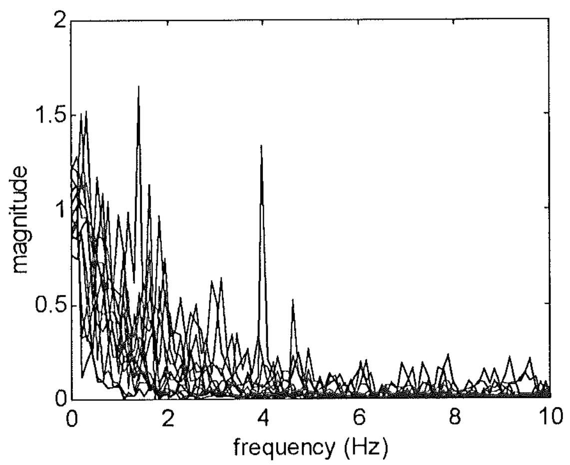


Figure 6-18. All 16 transfer functions.

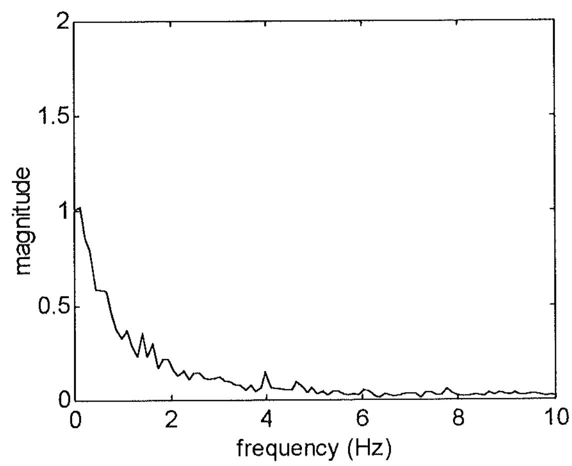


Figure 6-19. Average transfer function.

From this we adjusted a Butterworth lowpass filter of first order, with a cut-off frequency of 0.35 Hz *(figure 6-20)*. Using this lowpass filter on the sound envelope we get something we can call a simulated response *(figure 6-21)*. By performing the five steps mentioned in the beginning of chapter 6.2 we are able to pick a sound and simulate how the average response of that sound would look like.

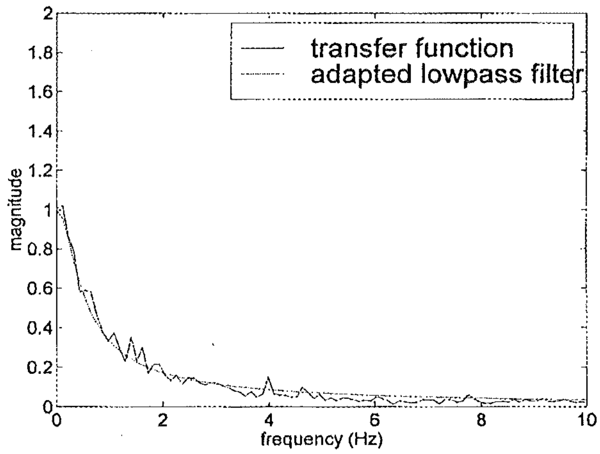


Figure 6-20. The frequency response for the adapted low pass filter together with the average transfer function.

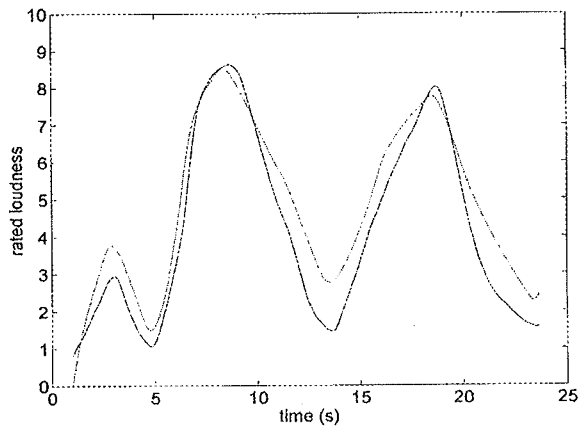


Figure 6-21. A simulated response (dotted curve) plotted together with the average response of 'db1k'.

We see in figure 6-21 that the simulated response to the stimulus 'db1k' closely matches the *mean curve over subjects*.

6.3 Analysis and results of the brightness test

In the analysis of the brightness test we have studied the following aspects:

* **Reaction times**
* **Variation between subjects**
* **Reproducability**
* **The relationship between the center point of the spectrum and perceived brightness**

6.3.1 Reaction times

In table 10 we have investigated the reaction times for five of the stimuli in the brightness test. As in the loudness test, the reaction times were calculated with correlation technique.

subject nr	brus22oct	brus22ters	brus22low	brus22high	kokmono	mean	std
1	0.93	0.89	1.22	0.92	0.55	0.90	0.24
2	0.70	0.86	0.77	0.78	0.37	0.70	0.19
3	0.57	0.49	0.46	0.62	0.54	0.54	0.06
4	0.77	0.68	0.76	0.72	0**	0.73	0.04
5	0.69	0.67	0.77	0.78	0.31	0.64	0.19
6	0.74	0.83	0.82	0.80	0.60	0.76	0.10
7	1.01	0.94	0.87	0.78	0.22	0.76	0.32
mean	0.77	0.77	0.81	0.77	0.37	0.70	
std	0.15	0.16	0.20	0.09	0.21		

Table 10. The reaction times of five of the stimuli used in the brightness test.

** This reaction time is too fast and is not included in the mean value and the standard deviation calculation.

We see in the table that the reaction times are generally higher than in the loudness test. This indicate that humans do not have the same intuitive perception of brightness and must think more to decide their rating. Furthermore we see that the reaction times of the four noise stimuli used in the test are almost the same. One of the everyday sound used in the brightness test, 'kokmono', had a little bit shorter mean reaction time.

6.3.2 Variation between subjects

The variation between subjects in the brightness test was calculated the same way as in the loudness test. In figure 6-22 and 6-24 we see how the balance point (see section 5.5) of the spectrum varies with time for the stimulus 'brus22oct' and 'kokmono' respectively. In figure 6-23 and 6-25. the corresponding responses are shown.

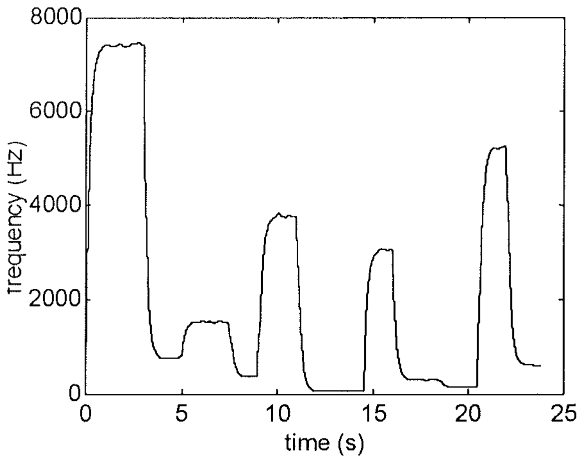


Figure 6-22. The variation of the balance point of the spectrum with time for 'brus22oct'.

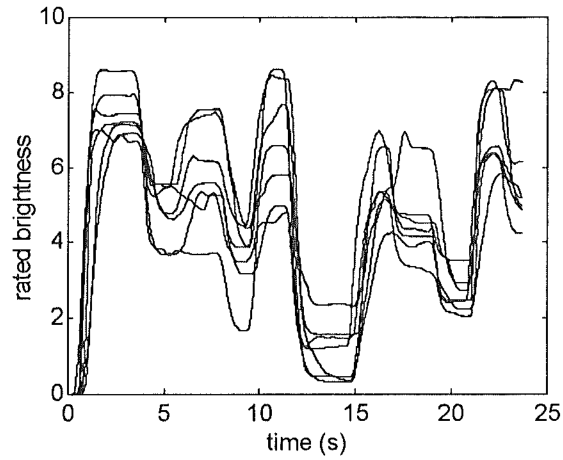


Figure 6-23. All subjects average response envelopes to 'Brus22oct'.

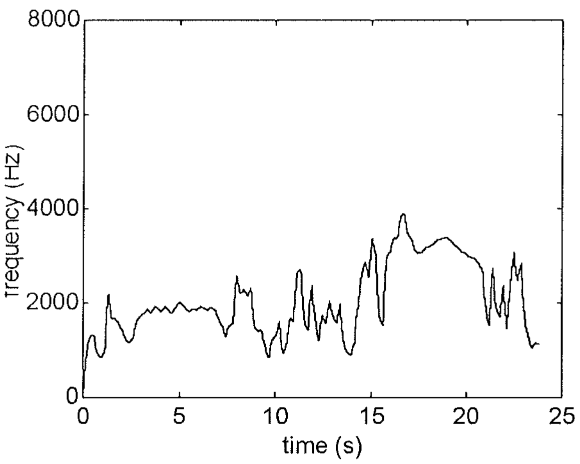


Figure 6-24. The variation of the balance point of the spectrum with time for 'kokmono'.

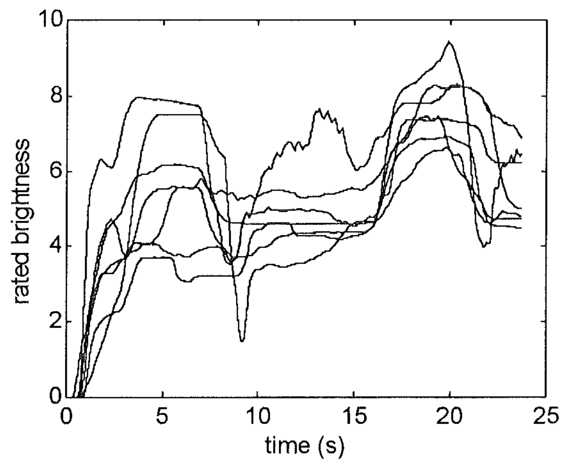


Figure 6-25. All subjects average response envelopes to 'kokmono'.

If we compare the response curves with the variation of the balance point of the spectrum, we see that they significantly differ (the low values of the balance point are not rated that low), that indicates that linear frequency might not be the most appropriate to use (see further section 6.3.4).

The same way as with the loudness test we calculated an index to get information about how easy the stimuli were to follow (table 11 and 12).

stimuli	st. dev of reaction times (s)	st. dev of response level (scale units)	quote 1
brus22oct	0.15	1.85	0.081
brus22ters	0.16	1.78	0.090
brus22low	0.22	1.22	0.180
brus22high	0.09	1.36	0.066
trafmono	0.21	0.47	0.447

Table 11. Five selected stimuli from the brightness test. The standard deviations of the reaction times.

stimuli	mean standard deviation in rating (scale units).	st. dev of response level (scale units)	quote 2	index $*10^{-2}$
brus22oct	0.88	1.85	0.476	3.9
brus22ters	1.04	1.78	0.584	5.3
brus22low	0.64	1.22	0.525	9.5
brus22high	0.94	1.36	0.691	4.6
trafmono	0.56	0.47	1.19	53

Table 12. Five selected stimuli from the brightness test. The standard deviations of the ratings.

We can see on the index values that three of the filtered noises were about equally easy to follow. They can be compared with the noise stimuli in the loudness test. The lowpass filtered noise was a little bit more difficult and the traffic noise was very difficult, much more difficult than in the loudness test.

6.3.3 Reproducibility

It is also interesting to study how much a subject's answers differed between the rounds in the brightness test and compare that with the results in the loudness test. Figure 6-26 shows one subject's three response curves to the stimulus 'brus22oct'.

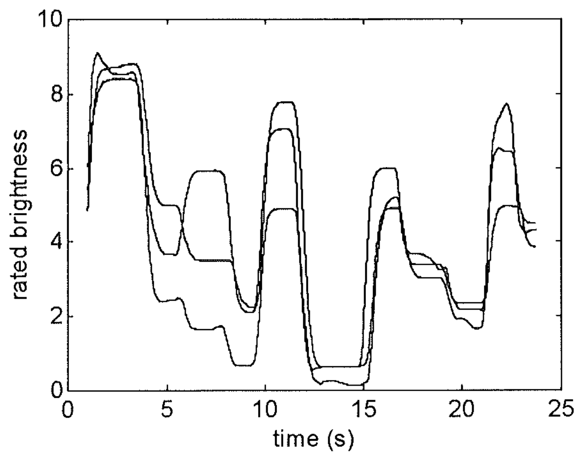


Figure 6-26. One subject's three brightness response curves to 'brus22oct'.

The reproducibility was studied in the same way as in the loudness test for five of the stimuli. For 'brus22ters' it resulted in this table:

brus22ters				
subject	**Round 1&2**		**Round 2&3**	
	time shift (s)	**level deviation**	**time shift (s)**	**level deviation**
1	-0.15	1.09	0.37	1.03
2	-0.01	1.10	0	0.83
3	-0.027	1.00	0	0.95
4	0.06	0.80	-0.11	0.75
5	0	0.71	0	0.81
6	0	0.86	0.38	0.29
7	0	1.08	-0.09	1.17
mean	**-0.02**	**0.95**	**0.09**	**0.83**

Table 13. The subjects' time-shift and level deviation between the rounds for the stimuli 'brus22ters'.

In this case we see that the differences between round 2 and 3 are a little bit smaller than between round 1 and 2 in level deviation but not in time. The differences was not as clear as in the loudness test though. This was also the case for the other four stimuli we studied (see appendix 3).

6.3.4 The relationship between the balance point of the spectrum and perceived brightness

To find out what kind of relation there is between the balance point of the spectrum and perceived brightness, we compensated for the time delay and plotted the rated brightness *(mean curve over subjects)* towards the logarithm of the balance point for all six stimuli used in the brightness test. As before this results in a set of dots to which we adapted a third order polynomial *(figure 6-27)*.

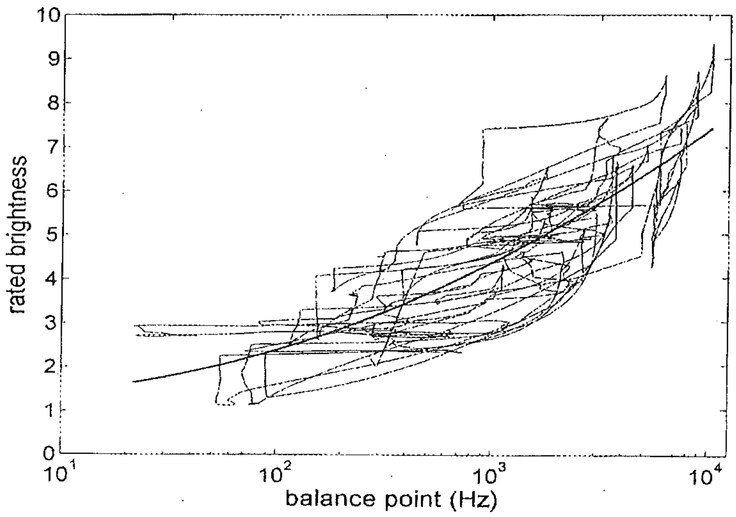


Figure 6-27. An adapted polynomial to the 6 average response curves plotted towards their corresponding balance point in Hz..

The equation of this polynomial is:

$$y = a_1 + a_2 \cdot \log f + a_3 \cdot (\log f)^2 + a_4 \cdot (\log f)^3$$

where

$a_1 = 1.52$

$a_2 = -0.61$

$a_3 = 0.51$

$a_4 = 0.0006$

We see that the curve flattens against lower frequencies and therefore we transformed the center point frequency to *bark* using the formula defined in section 2.7. As you can see in figure 6-28 this curve is more linear and therefore gives a better a description a of how we perceive brightness.

Note that

5 bark $\Leftrightarrow$ 530 Hz

10 bark $\Leftrightarrow$ 1260 Hz

15 bark $\Leftrightarrow$ 2710 Hz

20 bark $\Leftrightarrow$ 6420 Hz

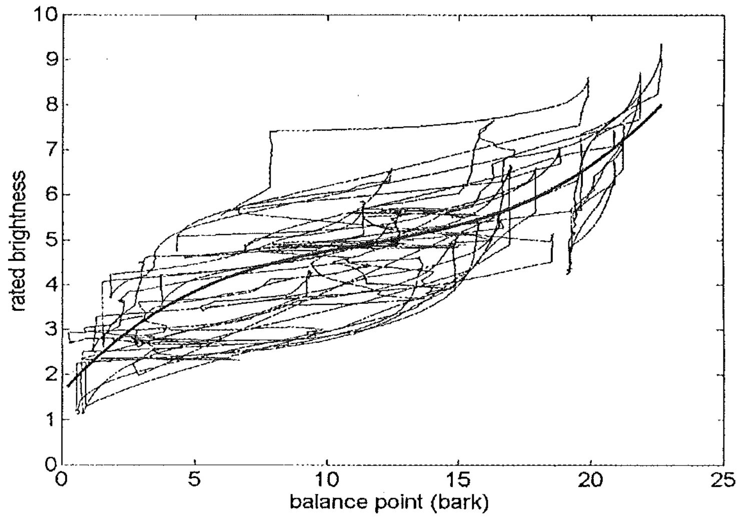


Figure 6-28. The same as figure 6-27. but with the balance point in bark.

The equation of this polynomial is:

$$y = a_1 + a_2 \cdot b + a_3 \cdot b^2 + a_4 \cdot b^3$$

where

$a_1 = 1.61$

$a_2 = 0.65$

$a_3 = -0.048$

$a_4 = 0.0014$

Figure 6-29 and 6-30 show that the sound envelope becomes more like the response envelope when using bark on the y-axis.

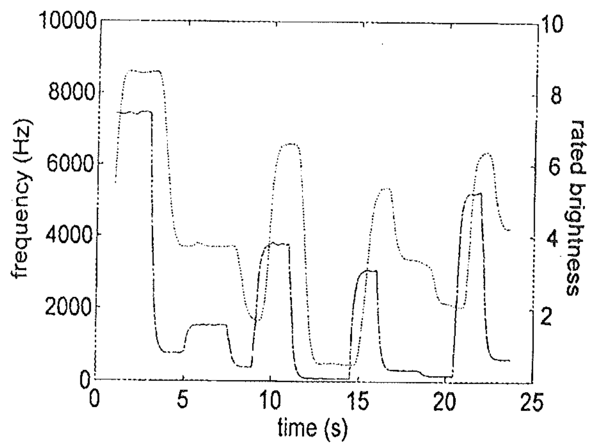


Figure 6-29. The sound envelope of 'Brus22oct' with the balance point in Hz and one subject's response plotted together.

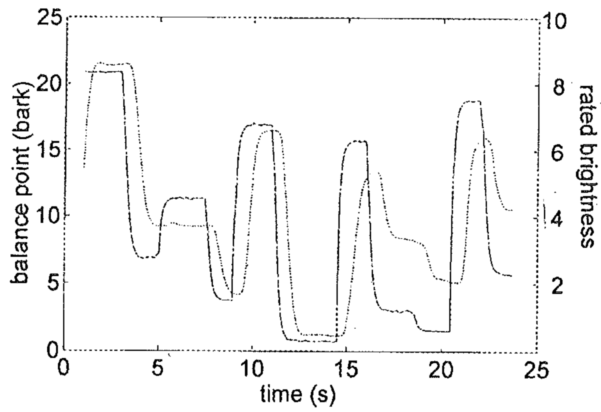


Figure 6-30. The sound envelope of 'Brus22oct' with the balance point in bark and one subject's response plotted together.

7 Discussion

7.1 Comments of the method

In this project we have found that our system for continuous estimation works well for loudness and brightness although some of the stimuli seemed very hard to follow. The general opinion among the subjects was that the brightness test was more difficult than the loudness test. This was also confirmed by the analysis. We saw in the results that the subjects got successively more secure in their responses, therefore you might want to see the first round as training and make conclusions from the two other rounds.

7.2 Suggestions for future research

To get more general results it is desirable to have a larger number of participants in the tests. You can also do a more thorough comparison between different types of stimuli as music sounds and noises. The next step could be to examine other descriptors such as sharpness, fullness, clarity and annoyance. Another research you can do is to compare continuous estimation with the classic listening test described in the introduction. A study of the relationship between overall loudness and instantaneous loudness, has been done in [5].

7.3 Applications

This method can be useful in the field of audiology, for example to help adjusting hearing aids and evaluate how well it works for the patient. It may also be used in order to get an increased understanding of the annoyance of fluctuating sounds, such as industrial and traffic noise.

7.4 Sources of errors

* The frequency response of the headphones was not linear, which means that the output level may be different at different frequencies.
* The correlation technique only gives information of the average time shift between two signals. When compensating for that to get the signals more alike, it may not be the best compensation for different intervals. Therefore it would be better to calculate a time shift for a shorter interval and use it for compensation of that interval.
* The values calculated from the marks done with the cursor may not be totally accurate since the marks were subjectively picked.

8 Conclusions

The described method of continuous sound quality rating worked well. The average time delay was about 0.5 s for loudness and about 0.8 s for brightness. The average uncertainty of the rating was approximately 0.7 scale units for loudness and approximately 0.8 scale units for brightness. These values were obtained from round two and three. Round one was regarded as training.

References

1. Bodén, H. et. al. 1998. *Ljud och vibrationer.* MWL, KTH, Stockholm, Sweden.

2. ELFA™. Catalog of products.

3. Gabrielsson et. al. 1990. *Perceived sound quality of reproductions with different frequency responses and sound levels.* Department of Technical Audiology, Karolinska Institutet, Stockholm.

4. Gelfand, S.A. 1981. *Hearing - An Introduction to Psychological and Physiological Acoustics.*

5. Kuwano, S. and Namba, S. *Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness.* Journal of Acoustical society of Japan.

6. Liljencrants, J. 1993. *Elektroakustik.* Institutionen för talöverföring och musikakustik, KTH, Stockholm, Sweden.

7. Svärdström, A. 1993. *Tillämpad signalanalys.*

8. Söderholm, M. *Listening test as a tool in sound quality work.* Master of science thesis at MWL, KTH, Stockholm.

9. Zwicker, E. & Terhardt, E. 1974. *Analytical expressions for critical-band rate and critical bandwith as a function of frequency.* Journal of Acoustical Society of America, Vol. 68, No. 5, Nov 1980.