HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Electrical and Communications Engineering
Laboratory of Acoustics and Audio Signal Processing

Minna Ilmoniemi

Modification and Brain Recordings of Musical Instrument Tones

Master's Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.

Espoo, May 3, 2004

Supervisor: Professor Vesa Välimäki
Instructor: Docent Minna Huotilainen

HELSINKI UNIVERSITY OF TECHNOLOGY

ABSTRACT OF THE MASTER'S THESIS

Author:	Minna Ilmoniemi
Name of the thesis:	Modification and Brain Recordings of Musical Instrument Tones
Date:	May 3, 2004 **Number of pages:** 58
Department:	Electrical and Communications Engineering
Professorship:	S-89 Acoustics and Audio Signal Processing
Supervisor:	Prof. Vesa Välimäki
Instructor:	D.Sc. (Tech.) Minna Huotilainen

Modern brain recording techniques offer possibilities to study the functioning of the human brain. The mismatch negativity (MMN) response reflects a change detection from the frequently presented tone (i.e., a standard tone) to the infrequently presented tone (i.e., a deviant tone). Thus, the MMN is associated with the automatic functioning of the short-term auditory sensory memory and sound discrimination.

In this thesis the effect of the standard probability on the MMN response is studied. In order to perform a brain-research experiment in which dozens of deviants would differ from the standard with equal psychoacoustic steps, an appropriate set of sound stimuli was generated. A recorded cello tone was modified in several dimensions of timbre and the modifications were evaluated through a subjective listening test. Electroencephalographic (EEG) recordings were performed in order to study the MMN response.

This thesis discusses the timbre dimensions and the methods for timbre modifications. The subjective evaluation of timbre modifications is described, as well as the brain recordings. Within this thesis it was observed that it is possible to generate a set of sounds differing in timbre and having an equal psychoacoustic distance from the reference sound. The results also show that decreasing the probability of the standard decreases the amplitude of the MMN response, i.e., makes the automatic sound discrimination more difficult.

Keywords: acoustic signal processing, brain response, musical instrument tone, psychoacoustics, timbre

TEKNILLINEN KORKEAKOULU DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Minna Ilmoniemi
Työn nimi:	Soitinäänten muokkaaminen ja aivomittaukset
Päivämäärä:	3.5.2004 **Sivuja:** 58
Osasto:	Sähkö- ja tietoliikennetekniikka
Professuuri:	S-89 Akustiikka ja äänenkäsittelytekniikka
Työn valvoja:	Prof. Vesa Välimäki
Työn ohjaaja:	TkT Minna Huotilainen

Nykyaikaisten aivotutkimusmenetelmien avulla voidaan tutkia ihmisaivojen toimintaa. MMN-vaste (Mismatch Negativity) kuvaa eron havaitsemista usein toistuvan äänen (standardi) ja harvoin esiintyvän äänen (deviantti) välillä. MMN-vaste liitetäänkin lyhytkestoisen kuulomuistin toimintaan ja äänten erotteluun.

Tässä työssä on tutkittu standardin todennäköisyyden vaikutusta MMN-vasteeseen. Aivotutkimuskoetta varten tuotettiin sopiva ääniärsykejoukko, jossa kymmenet deviantit eroavat standardista yhtä suurin psykoakustisin eroin. Nauhoitettua sellon ääntä muokattiin useassa äänenvärin dimensiossa ja äänenvärieroja arvioitiin subjektiivisessa kuuntelukokeessa. MMN-vastetta tutkittiin aivosähkökäyrämittausten (EEG) avulla.

Tässä työssä käsitellään äänenvärin dimensioita ja tapoja muokata äänenväriä. Työssä esitetään äänenvärierojen subjektiivinen arviointi sekä aivomittaukset. Tässä työssä havaittiin, että on mahdollista tuottaa joukko äänenväriltään eroavia ääniä, joilla on yhtä suuri psykoakustinen etäisyys vertailuääneen nähden. Tulokset myös osoittavat, että standardin todennäköisyyden vähentäminen pienentää MMN-vasteen amplitudia eli vaikeuttaa automaattista äänten erottelua.

Avainsanat: aivovaste, psykoakustiikka, soitinääni, äänenkäsittely, äänenväri

Acknowledgements

This work has been carried out in the Laboratory of Acoustics and Audio Signal Processing at Helsinki University of Technology and in the Cognitive Brain Research Unit at University of Helsinki.

I would like to thank my supervisor Professor Vesa Välimäki and my instructor Docent Minna Huotilainen for providing me an interesting topic for my thesis. My gratitude goes to Vesa for his guidance during this work and to Minna for introducing me to the interesting world of brain research. Both of them have also provided helpful comments and suggestions during the writing process.

I wish to thank Dr. Hanna Järveläinen for her help in the listening test arrangements and Pasi Piiparinen for advising and helping me within the EEG recordings. I also want to thank Henri Penttinen and Matti Airas for their help in some practical things.

Finally, I would like to thank all my subjects that participated in the listening test and/or the brain recordings. I am also grateful to all those people who have somehow contributed to this work.

Espoo, May 3, 2004

Minna Ilmoniemi

Contents

List of Abbreviations

A1	Primary Auditory Cortex
CBRU	Cognitive Brain Research Unit
EEG	Electroencephalography
EOG	Electrooculogram
ERP	Event-Related Potential
FIR	Finite Impulse Response
HEOG	Horizontal Electrooculogram
LDR	Linear Derivation
LM	Left Mastoid, location of EEG electrode behind the ear
MDS	Multidimensional Scaling
MMN	Mismatch Negativity, ERP component
N100, N1	Negative component of the ERP 100 ms after the stimulus onset, related to sound perception
N400, N4	Negative component of the ERP 400 ms after the stimulus onset, related to discrepancy from an expected ending
P300, P3	Positive component of the ERP 300 ms after the stimulus onset, related to attention switching and selection of sounds
P600, P6	Positive component of the ERP 600 ms after the stimulus onset, related to discrepancy from an expected ending in music
RM	Right Mastoid, location of EEG electrode behind the ear
SNR	Signal-to-Noise Ratio
SOA	Stimulus Onset Asynchrony, time from stimulus onset to the onset of the next stimulus
VEOG	Vertical Electrooculogram

List of Symbols

a_k	Coefficients of the allpass filter
$A(z)$	Transfer function of the allpass filter
$c_{\mathrm{a}}(n)$	Curve for modifying attack time
c_{even}	Harmonic coefficient for even harmonics
c_{harm}	Harmonic coefficient, either c_{even} or c_{odd}
c_i	Brightness coefficient
c_m	Brightness coefficient, either c_i or c_{15}
c_{noise}	Noise coefficient
c_{odd}	Harmonic coefficient for odd harmonics
$D(z)$	Denominator of $A(z)$
f	Frequency
g	Scaling factor
G	Gain
$h_i(n)$	Harmonic components
$H(z)$	Transfer function of the inverse comb filter
$H_{\mathrm{fd}}(z)$	Transfer function of the fractional delay inverse comb filter
$H_{\mathrm{fdr}}(z)$	Transfer function of the fractional delay inverse comb filter with a resonator
$H_{\mathrm{r}}(z)$	Transfer function of the resonator
i	Index of the harmonic
L_{l}	Lower bound for a 95% confidence interval
L_{u}	Upper bound for a 95% confidence interval
M	Number of extracted harmonics
μ	Mean value of psychoacoustic distances
n	Discrete time index
N	Filter order

N_{s}	Sample size
R	Radius of the resonator's pole
$s(n)$	Processed sound signal
$s_{\mathrm{a}}(n)$	Sound signal processed in attack time dimension
$s_{\mathrm{b}}(n)$	Sound signal processed in brightness dimension
$s_{\mathrm{harm}}(n)$	Sound signal processed in harmonic dimension
$s_{\mathrm{noise}}(n)$	Extracted noise
$s_{\mathrm{noisy}}(n)$	Sound signal processed in noise dimension
$s_{\mathrm{ref}}(n)$	Reference sound
σ	Standard deviation
T	Sampling interval
θ	Angle of the resonator's pole
$x(n)$	Filter input
x_{a}	Attack time parameter
x_{b}	Brightness parameter
x_{harm}	Harmonic parameter
x_{noise}	Noise parameter
$y(n)$	Filter output
z^{-n}	Delay of n samples

Chapter 1

Introduction

The brain receives, stores, processes, and produces information. The brain consists of a number of lobes and areas that receive information from the different senses. With brain research methods the brain activity in different areas of the brain can be recorded. We can obtain information about the cognitive processes of humans by studying the relationship between the brain activity and perception.

An important ability of the brain considering processing of auditory stimuli is the discrimination of sounds. It is vital to humans to distinguish ”normal” sounds from sounds that are warning about danger. The processing of auditory information is complex and the discrimination of the different features of the sounds, e.g., frequency, pitch, timbre, and location, needs the cooperation of different levels and areas of the brain. An objective measure of the automatic (i.e., unconscious) sound discrimination of the brain is the mismatch negativity (MMN) response that is associated with the short-term auditory sensory memory. The MMN response is elicited by an infrequently-presented tone, or deviant, that differs from the frequently-presented tone, or standard. By examining the MMN response the sound recognition and discrimination can be studied. The difference between the deviant and standard tones can be in one or several physical parameters, e.g., in the frequency or duration. In addition to these also a change in tone timbre will elicit the MMN response.

Musical timbre denotes the ”tone quality” or ”tone color” of a sound. Timbre is an interesting feature of a sound because it allows a listener to discriminate sounds that have the same pitch, duration, and loudness. Timbre is a multidimensional attribute to which many features of sounds, including the spectral and temporal patterns, have been found to contribute. The perceptual differences between sounds differing in timbre are not easy to evaluate, especially when different timbre dimensions are considered.

Within this thesis both timbre perception and brain activity are studied. The main goal of this work was to gather information about the functioning of the short-term auditory memory. In order to reach this goal also another target was set, i.e., the generation of appropriate test stimuli. Thus, this work can be divided into two tasks. First, to generate a sound set, or a sound matrix, that consists of the natural-sounding tones differing with equal steps in timbre space. Second, to study the effect of the standard probability on the MMN response by using the sounds of the sound matrix as test stimuli in the brain recordings.

Typically pure sine waves have been used as test stimuli in the MMN studies. In the brain recordings performed within this thesis, natural-sounding stimuli were required in order to obtain more accurate results considering the human cognition in everyday life. The number of stimuli was required to be relatively large in order to ensure that the standard tone would occur more often than each deviant tone in all test cases. Also an equal psychoacoustic distance, or the perceived difference, between the standard and each deviant was required in order to make sure that the brain responses would derive from the differences in probabilities in which different sounds occur and not from the differences in psychoacoustic distances within the sounds. These requirements were the motivation for timbre modifications. Since timbre is a multidimensional attribute of the sound, the required sound set can be obtained by modifying different features of timbre.

The outline of this thesis is as follows. Chapter 2 gives an overview of the theory behind the topic of this thesis. Also the previous research results described in the literature are discussed. In Chapter 3 the modifications of timbre and the generation of test stimuli are described. In Chapter 4 the description of the listening test arranged in order to evaluate timbre modifications is given and test results are presented. The electroencephalographic (EEG) recordings performed in order to study the brain responses evoked by the stimuli are described and the results are presented in Chapter 5. Chapter 6 concludes the thesis by summarizing the goals and results of this work.

Chapter 2

Background Theories

2.1 Overview of the Human Auditory System

The human auditory system can be divided into two parts: the peripheral auditory system (ears) and the auditory nervous system (auditory nerves and the brain) [26]. The auditory system detects sounds, transforms them from air pressure changes into electrical signals, and processes these electrical signals so that the different features of sounds, e.g., pitch, loudness, timbre, and location, are compared and perceived. The human auditory system is not equally sensitive to all frequencies. The audibility range of humans is about 20 Hz–20 kHz from which the hearing system is the most sensitive (i.e., the threshold for hearing is lowest) at frequencies between 2 and 4 kHz [5]. The frequencies in the range of 2–4 kHz are also the most important ones for understanding speech.

The ear is divided into three sections: the outer ear, the middle ear, and the inner ear. The outer ear, which consists of the pinna and the ear canal, collects the incoming sound and amplifies the frequencies in the range of 2–5 kHz. The middle ear begins with the tympanic membrane, or eardrum, located at the end of the ear canal. The middle ear matches the impedance between the air in the outer ear and the liquid in the inner ear. The three small bones of the middle ear, called ossicles (malleus, incus, and stapes), are attached to the eardrum. The oscillating eardrum transforms the air pressure variations into mechanical vibrations of ossicles. The ossicles transmit the vibrations to the inner ear via the oval window. The main structure of the inner ear is the cochlea, which is filled with liquid. As a result of vibrations of the ossicles, the liquid of the cochlea begins to vibrate, too. This results in vibration of the basilar membrane that is located in the cochlea. The place of the peak amplitude of the traveling wave in

the basilar membrane depends on the frequency of the sound: the higher the frequency the shorter the distance from the oval window to the peak amplitude. This is how the initial frequency analysis takes place in the cochlea. The organ of Corti, which is located on the basilar membrane, contains receptors called inner and outer hair cells. The hair cells are organized tonotopically, i.e., according to the frequency. The hair cells near the oval window are activated by high-frequency sounds while the hair cells in the far end of the basilar membrane are activated by low-frequency sounds. Also the adjacent hair cells are activated by the frequencies that are near to each other [31]. The vibrations of the basilar membrane cause the hair cells to bend in consequence of which the hair cells transform the vibrations into the electrical impulses and transmit these impulses to the brain via the auditory nerve. In this way the information about the location and amplitude of the vibration in the basilar membrane is transmitted to the higher levels of analysis.

From the cochlea the sound information is transmitted to the auditory cortex via auditory pathways. In the brain the electrical signals are transmitted by the firing of the neurons, which means that if the potential change over the cell membrane in a presynaptic neuron (i.e., the sending neuron) reaches the threshold value, the neuron sends an action potential to propagate along the neuron's axon without attenuating. When the action potential reaches the end of the axon it interacts with a postsynaptic neuron (i.e., the receiving neuron). The auditory nerve transmits the information signal from the cochlea to the cochlear nucleus. From the cochlear nucleus, most of the nerve fibers cross to the superior olivary nucleus located in the opposite side of the brain stem. From there, the fibers go to the inferior colliculus of the midbrain and the medial geniculate nucleus of the thalamus. From the thalamus, which is regarded as a "gateway to cortex", the fibers go to the primary auditory cortex (A1) located in the temporal lobe.

The auditory cortex with columnar arrangement is also tonotopically organized. The auditory pathways are bilateral, which means that they exist on both the left and right sides of the brain. The auditory pathways cross over the two cerebral hemispheres so that most part of the information from the right ear is transmitted to the left temporal lobe and vice versa. Although a significant part of the crossing takes place in the brain stem, there are connections between the both sides of the brain at each level of the auditory pathways. The two auditory cortices are connected with each other through the corpus callosum that is the most important communication highway between the two cerebral hemispheres. In addition to the ascending neural pathways from periphery to cortex, the auditory system also includes descending tracts from cortex to periphery.

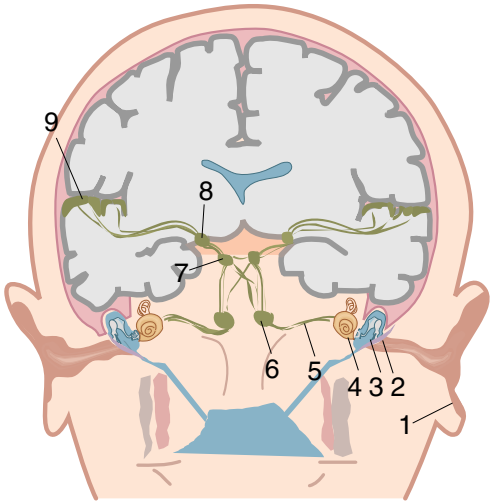

Figure 2.1: The structure of the human auditory system: (1) pinna and ear canal, (2) eardrum, (3) ossicles, (4) cochlea, (5) auditory nerve, (6) cochlear nucleus, (7) inferior colliculus, (8) thalamus, and (9) auditory cortex [31].

The structure of the human auditory system is illustrated in Figure 2.1.

The first information of the sound arrives at the auditory cortex about 15 ms after the onset of the sound. However, it takes about 100 ms from the onset before the largest responses to the sounds are evoked. From the primary auditory cortex the signals travel to other auditory areas in the temporal lobe: the secondary auditory cortex and the auditory association cortex. From these cortices the signals are furthermore transmitted to other areas in the temporal, frontal, and parietal lobes in order to process more complex features of the sounds. All these areas and their cooperation are important for the humans so that they are able to discriminate accurately the different features of the sounds, e.g., frequency, duration, intensity, and timbre. The cooperation of all areas is also needed for the discrimination of the sounds of different instruments and the voices of different speakers.

2.2 Musical Timbre

A musical tone is often described by four attributes: pitch, loudness, duration, and timbre. Musical timbre is defined by the American Standards Association [2] as ”that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar”. In this

definition the duration of the sound is not taken into account but in general it is thought that in timbre judgment the sounds need to have equal duration as well. A revised definition for timbre is suggested by Pratt et al. [25]: ”Timbre is that attribute of auditory sensation whereby a listener can judge that two sounds are dissimilar using any criteria other than pitch, loudness, or duration”. So if two sounds have the same pitch, loudness, and duration, their timbre allows a listener to discriminate these sounds.

Timbre is a multidimensional attribute on which both the spectral content and temporal patterns have an effect. The dimensions of timbre have been studied in listening tests utilizing the multidimensional scaling method (MDS) [14] in interpreting the results. In multidimensional scaling, perceptual data consisting of similarity judgments are gathered. The pairwise similarity judgments are regarded as psychoacoustic distances from which the multidimensional geometric map of the stimuli is constructed. The map describes the psychoacoustic distances of the stimuli in a multidimensional geometric space where the dimensions are the different features of timbre. A short distance between two sounds in the timbre space corresponds to a high perceived similarity whereas a long distance corresponds to a high perceived dissimilarity.

In many similarity scaling experiments the relative amplitudes of lower and higher harmonics and the variations in amplitude envelopes have been found to contribute to the perception of timbre and similarity judgments of sounds [6, 12, 34]. Grey [6] studied timbral similarities using the MDS. He proposed a three-dimensional scaling solution for timbre features. The dimensions of timbre that he proposed are the spectral energy distribution, the presence of synchronicity in the transients of the higher harmonics, and the presence of low-amplitude, high-frequency energy in the initial attack segment. Also Wessel [34] found that the spectral energy distribution and the nature of the onset transient are strongly related to the timbre. The spectral energy distribution is highly correlated to the spectral centroid (i.e., the mean frequency of the spectrum) and the brightness of a sound. In addition to the sound energy and the balance between the low and high components in the spectrum, the width of the spectrum has also been found to have an effect on timbre [7].

Iverson et al. [12] proposed in their studies that the main attributes contributing to similarity judgments are the centroid frequency and the variations in amplitude envelopes. They also proposed that the attributes contributing to the timbre perception are present throughout the sounds, not only on onsets or remainders (i.e., sounds with the onsets removed). However, the attack is very important for the timbre perception and especially for the sound source identification. In the case of musical instruments the timbre is often determined more by the onset than by the remainder. It is known

that many wind instruments cannot be identified if the onset is removed from the sound [13]. Also the automatic musical instrument classification in terms of timbre has been studied [1]. In this study the features of timbre considered were the inharmonicity, the spectral centroid, and the energy contained in the first partial. It was found that the string instruments were the most misclassified instrument family whereas the classification of wind instruments was quite accurate.

In this thesis, timbre modifications were performed in four dimensions. These are described in Chapter 3.

2.3 Electroencephalography

The neurons, or nerve cells, of the brain constitute a complex network in which information is transmitted as electro-chemical signals to different brain areas. The information is transmitted within voltage changes (around 100 mV) called the action potentials. However, these changes do not add up when recorded from the scalp surface because they last less than 10 ms. The voltage changes associated with synaptic potentials, i.e., the voltage changes that do not reach the threshold value, occurring at the neuronal contact points, or synapses, are smaller (less than 10 mV) but they last hundreds of milliseconds and so they can be recorded outside the head. It has been proposed that the voltage changes recorded from the scalp are mainly caused by the postsynaptic potentials in the pyramidal cells of the cortex [31]. The voltage change caused by an individual synapse is very small. Therefore thousands of synapses have to be activated synchronically in order to be able to record the activation from the scalp.

The electroencephalography (EEG) is a method with which the electric voltage changes caused by the brain activity can be recorded. The EEG is recorded from the scalp by using electrodes that pick up the potentials associated with the brain's electrical activity. The electrodes are attached to the scalp with a conducting layer of paste. Typically the EEG is recorded simultaneously from several locations in order to obtain information about the distribution of the brain activity across the scalp. The voltage changes in the electrodes are recorded towards a reference electrode that is attached e.g. to the mastoid or the tip of the nose. The frequency range of the EEG is about 1–50 Hz and the amplitude of the voltage changes is in the range of 5–300 μV [11]. The EEG signal recorded between two electrodes is amplified, digitized, and stored for further processing.

The EEG is a noninvasive and relatively cheap brain recording method. It has a time resolution of about 1 ms and a spatial resolution of 1–2 cm. The problem of the

EEG method is the spontaneous and background brain activity that may obscure the impact of the test stimulus. The spontaneous brain activity, which is not related to the test stimulus, is 2 to 100 times stronger than the responses evoked by the test stimulus itself. Therefore the same stimulus is presented many times so that the response to the stimulus can be observed after averaging the evoked signals.

2.4 Brain Responses

The brain activity can be studied with brain responses evoked by the stimuli. In this section the event-related brain activity is discussed. First the event-related potentials (ERPs) and their components are described and then the mismatch negativity (MMN) component is more accurately studied.

2.4.1 Event-Related Potential

From the ongoing EEG the segments following each stimulus are extracted and averaged. This method produces the waveforms called event-related potentials (ERPs), which are time-locked to sensory events and describe voltage fluctuations caused by the sound stimuli. The ERP waveforms plot the voltage change as a function of time. Since the time resolution of the EEG is excellent, the ERPs provide accurate information about the time course of the brain activity [4, 24].

In the human auditory system the largest responses are not elicited until 100 ms after the sound stimulus [11, 31]. The N100, or N1 component of the ERP, appears as negative on electrodes on frontal and central parts of the scalp and occurs approximately 100 ms after the stimulus onset. The N1 response is correlated with the physical properties of the sound, e.g., frequency, and is large when the sounds are repeated with long intervals. The N1 is related to the perception of the sound and is the main response to sounds in adults. If repeated sounds are sometimes replaced with a different sound, the MMN (see Section 2.4.2) response is elicited about 100–200 ms after the sound onset. In addition to the MMN, the positive P300, or P3, component can be observed 300 ms after the onset of the tone in the response to the deviant tone. This component is thought to be correlated with the change of the attention from the task (e.g., reading a book or watching a movie) to the deviant tone. N400, or N4, is a negative response elicited 400 ms after the presentation of something incongruent, for example a word that is not suitable to the verbal continuum in a semantic meaning. P600, or P6, response is a positive musical counterpart of the N400. It is elicited 600 ms after

the sound onset if the sound violates the musical context. An example of the brain response with N1, MMN, and P3 components of the ERP is illustrated in Figure 2.2. Following the convention in brain research the amplitude scale in Figure 2.2 is reversed so that the negativity increases upwards and the positivity increases downwards.

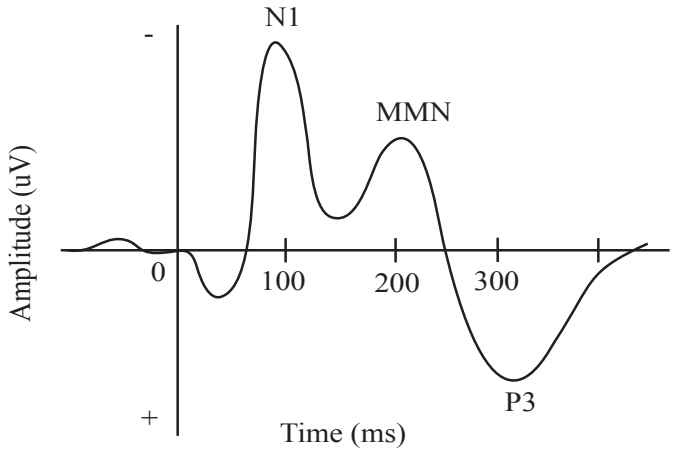

Figure 2.2: Event-related potential with N1, MMN, and P3 components.

2.4.2 Mismatch Negativity

Mismatch negativity (MMN) is a negative component of the ERP that is elicited by an infrequently presented tone (deviant) differing from the frequently presented tone (standard) [19, 20, 22]. The deviant tone can differ from the standard tone in physical parameters like frequency, duration, or intensity. The MMN is also generated by a change in the spectral component of tone timbre or in the temporal parameters of the stimulation [30, 32]. The main generators of the MMN component are located in or near the primary auditory cortex. The MMN appears with maximal amplitude at frontal and central scalp locations usually 100–200 ms after the onset of the deviant stimulus.

The MMN is not elicited by the first stimulus in the test sequence or if the deviant is presented alone without the intervening standard stimuli. Therefore it is thought that the MMN reflects a change detection when a memory trace representing the standard stimulus is violated by the deviant. The first standard stimuli in the sequence develop a memory trace that represents the features of the standard. If the deviant is presented while the memory trace is still active, the MMN is elicited. It is estimated that the duration of these auditory sensory memory traces is approximately 10–12 s [19, 35]. It means that after that time interval the memory trace representing the standard no

longer exists and the MMN cannot be elicited by the deviant because there is nothing with which to compare the deviant. The MMN can be elicited in the absence of attention, which describes the brain's ability to automatically detect changes in auditory stimulation. In fact, if the attention is directed to the sound stimuli the MMN can be overlapped by other responses that are correlated with conscious change detection. The MMN is currently the only objective measure of the accuracy of central auditory processing in the human brain.

In addition to the short-duration memory traces, also permanent auditory memory traces can be studied by using the MMN method. Typically these traces represent the phonemes of one's mother tongue. When using vowels as stimuli it has been found that a larger MMN was elicited in those subjects to whom the deviant vowel was familiar (i.e., the vowel exists in their mother tongue) than in the subjects to whom the deviant vowel was unfamiliar [21]. The standard vowel used in this experiment was familiar to both subject groups.

The reliability and quality of the MMN can be described with signal-to-noise ratio (SNR) that is the size of the response divided by the standard deviation of the noise [3, 27]. The noise consists of the measurement noise and the noise caused by the spontaneous brain activity. The amount of the responses to be averaged, the amplitude of the response, and the amplitude of the background brain activity have an influence on the SNR. The SNR of the MMN can be low and therefore it is important to optimize the SNR in order to obtain reliable results. The SNR can be improved by developing better recording procedures or by involving different signal processing strategies for analysis.

Chapter 3

Generation of Test Stimuli

3.1 Introduction

In order to generate a brain-research experiment in which dozens of deviants would differ from the standard with equal steps, a multidimensional sound matrix was generated. The aim was to generate a $3 \times 3 \times 3 \times 3$ (i.e., four dimensional) sound matrix consisting of sounds that have the same pitch, duration, and loudness but differ from each other in terms of musical timbre in one to four dimensions. Each timbre has an equal perceived distance from the reference sound, which is the middle sound of the matrix. Since in the brain recordings the number of stimuli was required to be relatively large and the psychoacoustic distance between each deviant and standard was required to be equal, it was a natural choice to modify timbre of the sounds because of its multidimensionality. For example, by modifying the frequency, only two suitable deviants, i.e., one sound with lower and another with higher frequency, would have been obtained, which is not enough for this experiment.

The sounds of the sound matrix are generated with Matlab by processing a natural recorded sound of the cello. A bowed cello tone was selected due to its harmonicity, attack segment properties, and amplitude envelope. The musical timbre of the sounds is altered in four dimensions that are carefully selected according to psychoacoustical research. In previous studies, the timbres of musical instruments have been characterized [6, 12, 34]. Within the sounds of the sound matrix the most important features of timbre found in those studies are modified. Thus, the timbre dimensions of the sound matrix are the ratio of even and odd harmonic components, brightness, the attack time, and the amount of noise. The original cello tone is taken from the McGill University Master Samples library [23]. This sound has the fundamental frequency of 524 Hz

(C5). The sampling rate used is 44100 Hz. Since the duration of the sounds of the matrix was required to be 200 ms (8820 samples), the cello tone is truncated after 200 ms from the onset. The amplitude in the final 20 ms is attenuated linearly. The magnitude response and the waveform of the original truncated sound signal are shown in Figures 3.1 and 3.2, respectively.

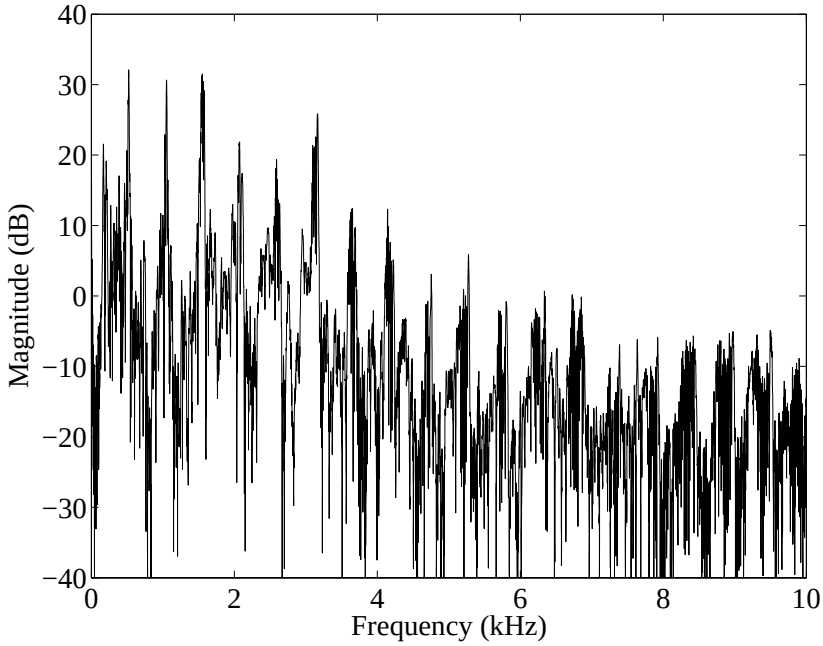

Figure 3.1: The magnitude response of the original sound signal.

3.2 Separation of Harmonic Components

The harmonic components of the original sound need to be separated when the sound is processed in dimensions of harmonics, brightness, and attack time. Also in noise dimension all harmonics need to be extracted. The harmonic components are separated by filtering the original truncated sound with a fractional delay inverse comb filter with a resonator that picks up single harmonics. So the filter passes one partial and attenuates the others from the signal. The method for extracting the harmonic components has been accepted for publication in an international conference [33].

The transfer function of the Nth order FIR inverse comb filter is given by [28]

$$H(z) = g(1 - z^{-N}) \ , \tag{3.1}$$

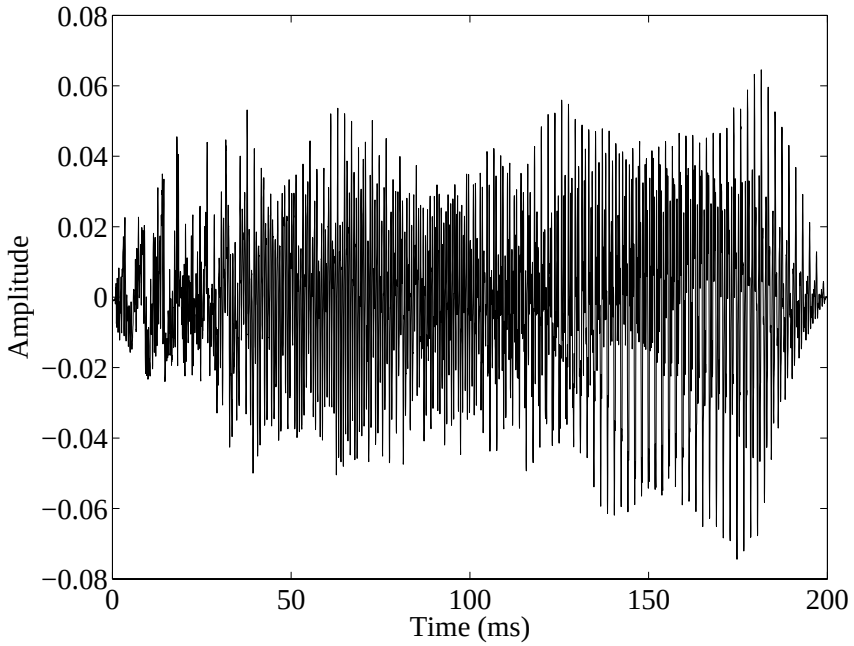

Figure 3.2: The waveform of the original sound signal.

where g is the scaling factor and z^{-N} is the delay line. In the magnitude response of the inverse comb filter there are notches at the integer multiples of the fundamental frequency. A Thiran allpass filter (see [15]) with an order of 84 and a fractional delay of $0.1611T$ where T is the sampling interval was used to implement the delay line of the inverse comb filter and to make the notch frequencies more accurate so that they would be exactly at the frequencies of the harmonic components. The transfer function of the Nth order allpass filter is expressed as [15]

$$A(z) = \frac{z^{-N}D(z^{-1})}{D(z)} = \frac{a_N + a_{N-1}z^{-1} + \ldots + a_1 z^{-(N-1)} + z^{-N}}{1 + a_1 z^{-1} + \ldots + a_{N-1}z^{-(N-1)} + a_N z^{-N}} \ , \tag{3.2}$$

where the terms a_k are the coefficients of the allpass filter and the numerator polynomial is a mirrored version of the denominator $D(z)$. When Equations (3.1) and (3.2) are combined and the scaling factor has a value of 1/2 in order to set the gain between the notches to unity, the transfer function of the fractional delay inverse comb filter becomes

$$H_{\mathrm{fd}}(z) = \frac{1 - A(z)}{2} \ . \tag{3.3}$$

The implementation of this filter that cancels all harmonics is shown in Figure 3.3.

A resonator is cascaded with the fractional delay inverse comb filter in order to pick

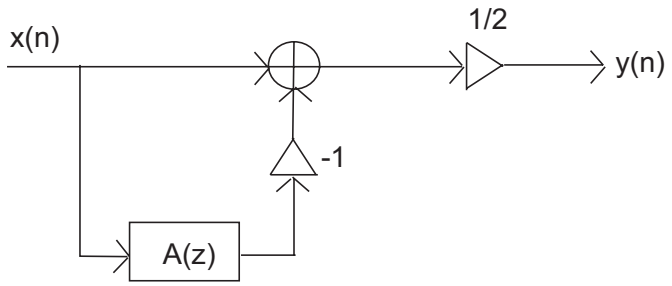

Figure 3.3: The implementation of the fractional delay inverse comb filter.

up single harmonic components. The transfer function of the resonator is given by [28]

$$H_{\mathrm{r}}(z) = \frac{1}{(1 - Re^{j\theta}z^{-1})(1 - Re^{-j\theta}z^{-1})} = \frac{1}{1 - 2R\cos(\theta)z^{-1} + R^2z^{-2}} \ , \qquad (3.4)$$

where R is the radius and θ is the angle of the poles of the resonator in the z-plane. R has a value of 1. Combining Equations (3.3) and (3.4) gives the final transfer function of the fractional delay inverse comb filter with a resonator

$$H_{\mathrm{fdr}}(z) = \frac{GH_{\mathrm{r}}(z)[1 - A(z)]}{2} = \frac{GH_{\mathrm{r}}(z)[D(z) - z^{-N}D(z^{-1})]}{2D(z)} \ , \qquad (3.5)$$

where the gain G is given by

$$G = \frac{1}{\max\left|\frac{H_{\mathrm{r}}(z)[1-A(z)]}{2}\right|} \ . \qquad (3.6)$$

The magnitude responses of both the inverse comb filter canceling all harmonic components and the inverse comb filter with the resonator picking up the 5th partial (f = 2620 Hz) are shown in Figure 3.4. The pole zero plots of these filters are shown in Figure 3.5. The poles and zeros of the filter evoke peaks and notches, respectively, in the magnitude response of the filter [17]. From Figures 3.4 and 3.5 it is evident that the poles on the unit circle at the same points with zeros cancel the effect of the zeros. So the notch in the magnitude response of the inverse comb filter, which attenuates the harmonic component, is replaced by the peak that picks up the required (in this case 5th) partial. In addition to the unit circle the poles and zeros exist also inside the unit circle. These poles and zeros are evoked by the allpass filter and resonator.

The 30 lowest harmonic frequencies are extracted from the original sound by filtering the original signal with the transfer function of Equation (3.5). It was noticed by listening to the filtered signals that filtering the signal only once is insufficient, i.e., the fundamental frequency can still be perceived although only one higher partial should

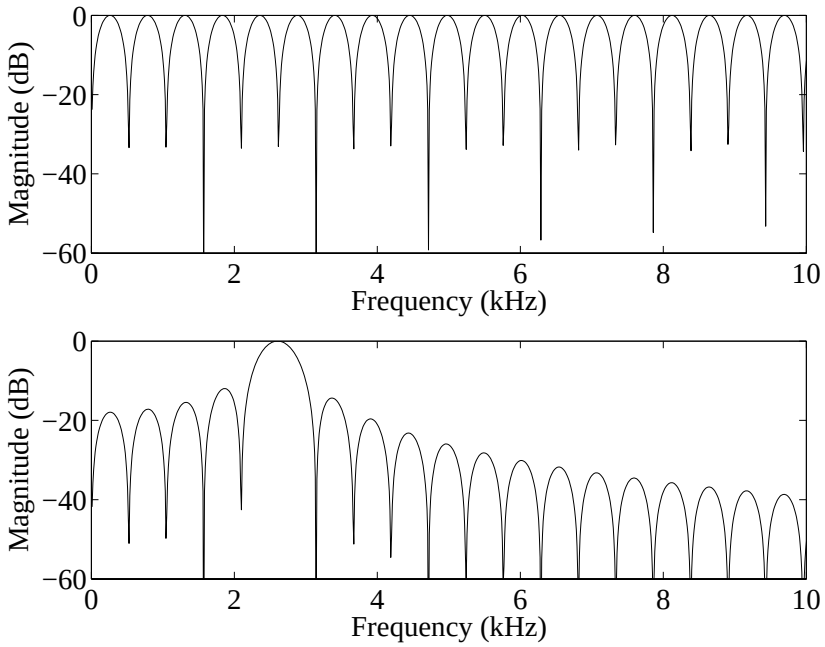

Figure 3.4: The magnitude response of the inverse comb filter (top) canceling all harmonics and (bottom) with the resonator picking up the 5th partial.

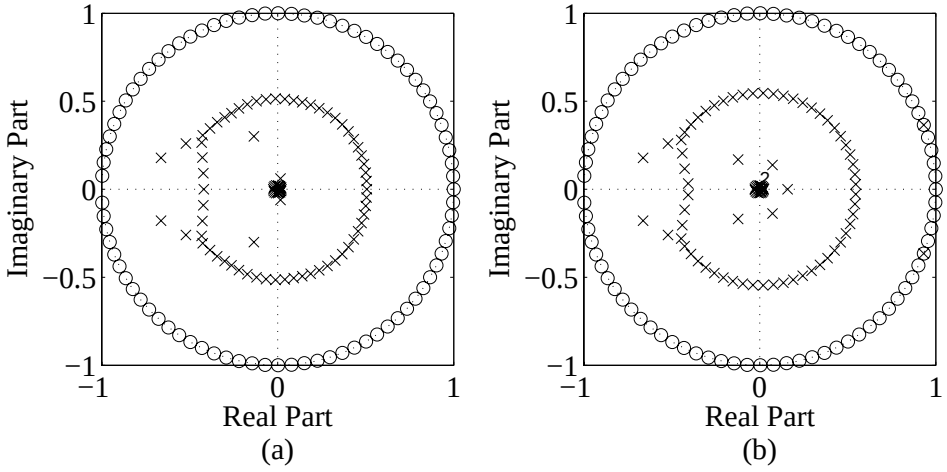

Figure 3.5: The pole zero plot of the inverse comb filter (a) canceling all harmonics and (b) with the resonator picking up the 5th partial.

exist in the signal. That is why the signal is filtered twice in each case. In Figure 3.6 it is shown that after filtering there is only one partial left, in this case it is the 5th partial.

The reference sound $s_{\mathrm{ref}}(n)$ (i.e., the middle sound of the matrix), whose magnitude response is shown in Figure 3.7, is generated by summing up all the separated

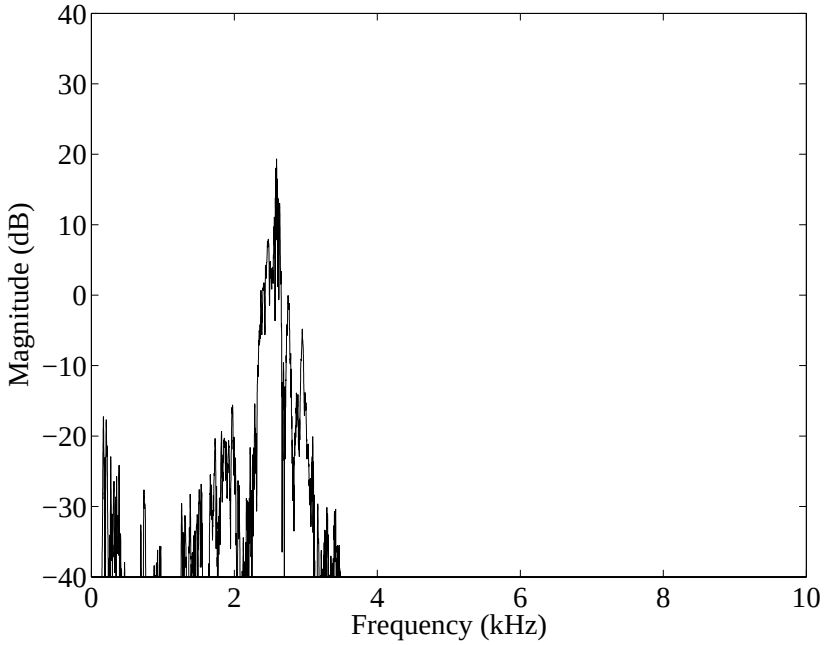

Figure 3.6: The magnitude response of the 5th partial of the original signal.

harmonics. This implementation is defined as

$$s_{\text{ref}}(n) = \sum_{i=1}^{M} h_i(n) \ , \tag{3.7}$$

where n is discrete time index, terms $h_i(n)$ denote different harmonics and M is the number of separated harmonics. In this work M has a value of 30.

3.3 Modification of Even and Odd Harmonic Components

The first dimension of the sound matrix is the ratio of even and odd harmonic components. This dimension is implemented by processing the separated harmonics of the original sound. In the one end of the ratio there are only even harmonics in the signal (Figure 3.8). Since the first harmonic is suppressed this signal sounds like a tone played an octave higher. In the other end of the ratio only odd harmonics exist in the signal (Figure 3.9). Between these the ratio is specified so that even harmonics are amplified and odd harmonics are attenuated, or vice versa, according to timbre parameter and harmonic coefficients. The coefficients for even and odd harmonics c_{even} and c_{odd},

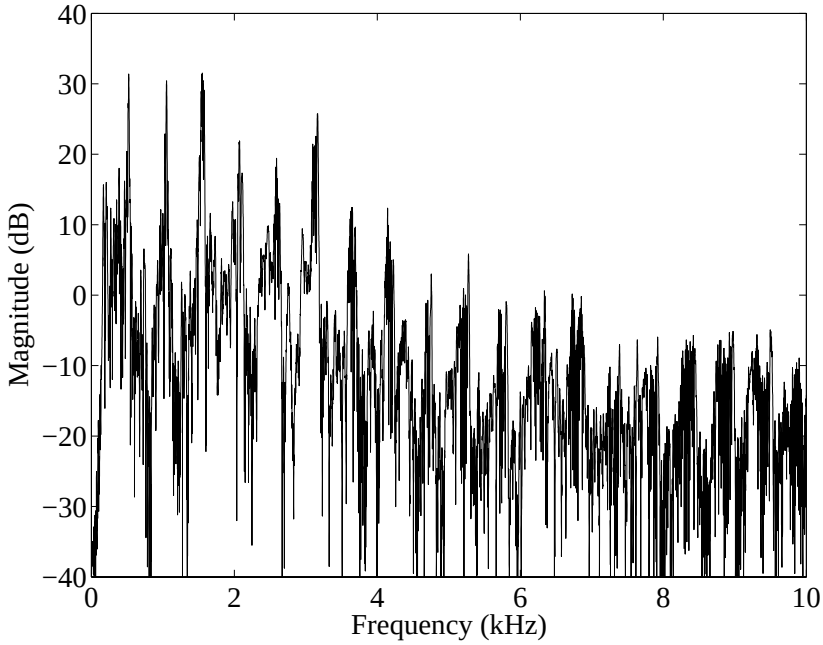

Figure 3.7: The magnitude response of the reference sound.

respectively, are defined as

$$c_{\text{even}} = 1 - x_{\text{harm}} \ , \tag{3.8}$$

$$c_{\text{odd}} = x_{\text{harm}} + 1 \ , \tag{3.9}$$

where x_{harm} is the timbre parameter for harmonic dimension ($x_{\text{harm}} \in [-1,1]$). If harmonic parameter x_{harm} is smaller than 0 even harmonics are amplified and odd harmonics attenuated and vice versa in the case of x_{harm} being greater than 0. In the reference sound the harmonic parameter is 0, which means that the ratio of even and odd harmonics is not modified.

When processing the sounds of this dimension the harmonic parameter and harmonic coefficients are specified first. The even and odd harmonics are multiplied by c_{even} and c_{odd}, respectively, and the processed harmonics are summed up. This implementation is expressed as

$$s_{\text{harm}}(n) = \underbrace{\sum_{i=1,3,\ldots}^{M-1} c_{\text{odd}} h_i(n)}_{\text{odd}} + \underbrace{\sum_{i=2,4,\ldots}^{M} c_{\text{even}} h_i(n)}_{\text{even}} \ , \tag{3.10}$$

where $s_{\text{harm}}(n)$ denotes the sound processed in harmonic dimension.

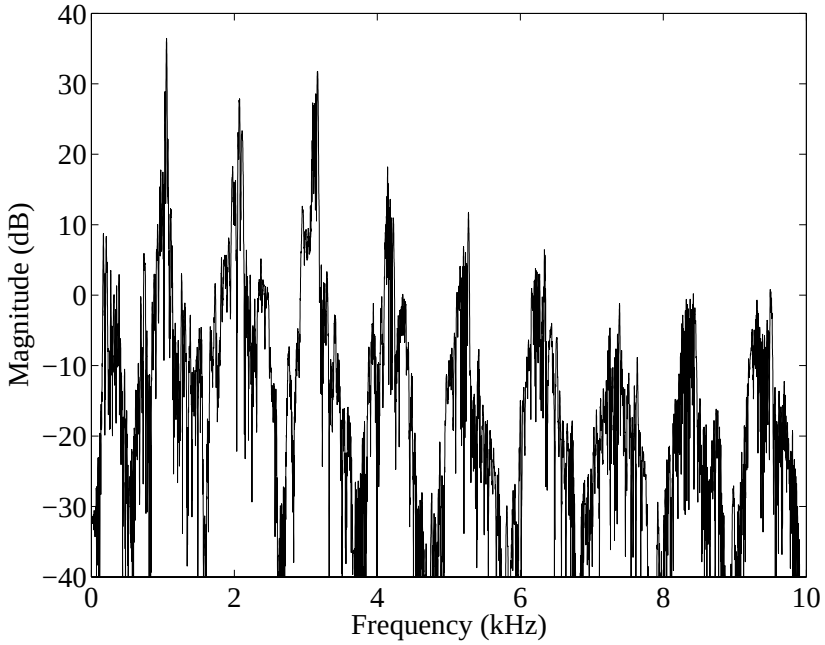

Figure 3.8: The magnitude response of the sound consisting of even harmonics.

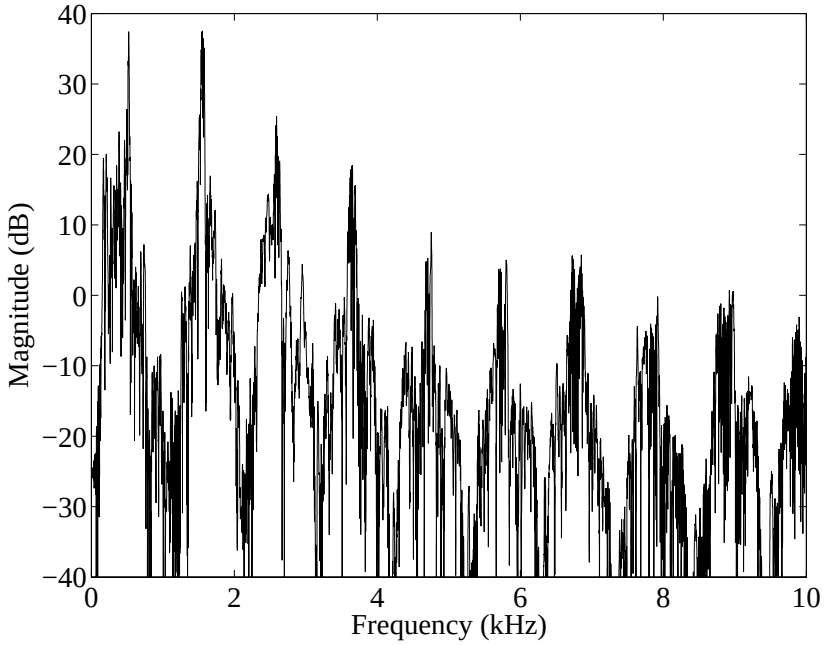

Figure 3.9: The magnitude response of the sound consisting of odd harmonics.

3.4 Modification of Brightness

The second dimension of the sound matrix is the brightness of the sound. Also this dimension is implemented by processing the separated harmonics of the original signal. The brightness of the sound is related to the centroid frequency of the sound spectrum so that the higher the centroid frequency the brighter the sound. So the modification of the brightness is implemented by altering the centroid frequency. The harmonics are either amplified or attenuated exponentially depending on whether the centroid frequency is required to be higher or lower, respectively.

The brightness coefficients c_i are expressed as

$$c_i = x_{\mathrm{b}}^{i-1} \quad , \tag{3.11}$$

where x_{b} is the timbre parameter for brightness dimension ($x_{\mathrm{b}} \geq 0$) and i tells which harmonic is considered (i = 1,2,...,30). If the brightness parameter is smaller than 1 the sound is required to be darker and so the centroid frequency has to be lower. In this case each harmonic is multiplied by corresponding brightness coefficient. If the brightness parameter is greater than 1 the sound is required to be brighter. In this case the harmonics are processed as in previous case but only up to the 15th harmonic. For the rest of the harmonics the brightness coefficient is the same as for the 15th harmonic. The value of the brightness parameter in the reference sound is 1. Then also all coefficients have a value of 1 and harmonics are not modified at all. In all cases the processed harmonics are summed up after multiplications. These implementations are expressed as

$$s_{\mathrm{b}}(n) = \begin{cases} \sum_{i=1}^{M} c_i h_i(n) & \text{for } x_{\mathrm{b}} \leq 1 \quad , \\ \sum_{i=1}^{15} c_i h_i(n) + \sum_{i=16}^{M} c_{15} h_i(n) & \text{for } x_{\mathrm{b}} > 1 \quad , \end{cases} \tag{3.12}$$

where $s_{\mathrm{b}}(n)$ is the sound processed in brightness dimension.

The coefficients for modifying the centroid frequency and brightness are shown in Figure 3.10. The brightness parameter has values of 0.8 and 1.2 in Figures 3.10(a) and 3.10(b), respectively. The magnitude responses of the sounds modified in brightness dimension are shown in Figure 3.11 (x_{b} = 0.8) and Figure 3.12 (x_{b} = 1.2).

3.5 Modification of Attack Time

The third dimension of the sound matrix is the attack time of the sound. Like previous dimensions this dimension is also implemented by processing separated harmonics of the original sound.

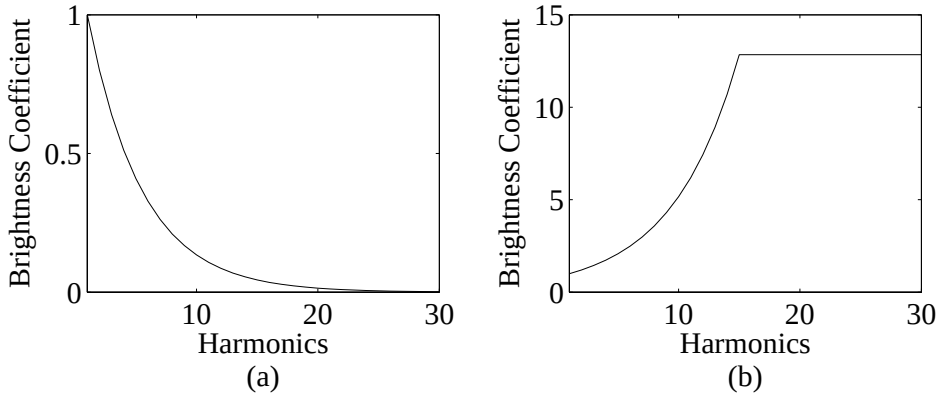

Figure 3.10: The coefficients for modifying the brightness, (a) $x_{\mathrm{b}} = 0.8$ and (b) $x_{\mathrm{b}} = 1.2$.

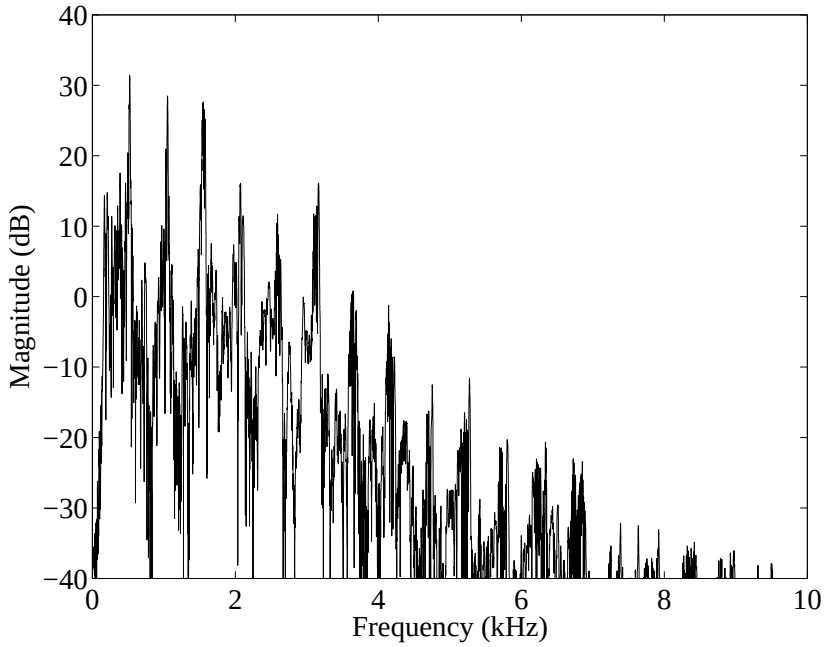

Figure 3.11: The magnitude response of the sound with the brightness parameter of 0.8.

Since the largest MMN responses are elicited usually 100–200 ms after the stimulus onset, it was decided that the attack time of the sound would be at most 80 ms. The timbre parameter for attack time dimension x_{a} can have values in the range of $[-0.08, 0.08]$. The negative value of the attack time parameter means that the attack time is required to be longer. If the parameter is positive the attack time is required to be shorter. In the reference sound the attack time parameter is 0, which means that the attack time is not modified at all. The curves for modifying the attack time are shown

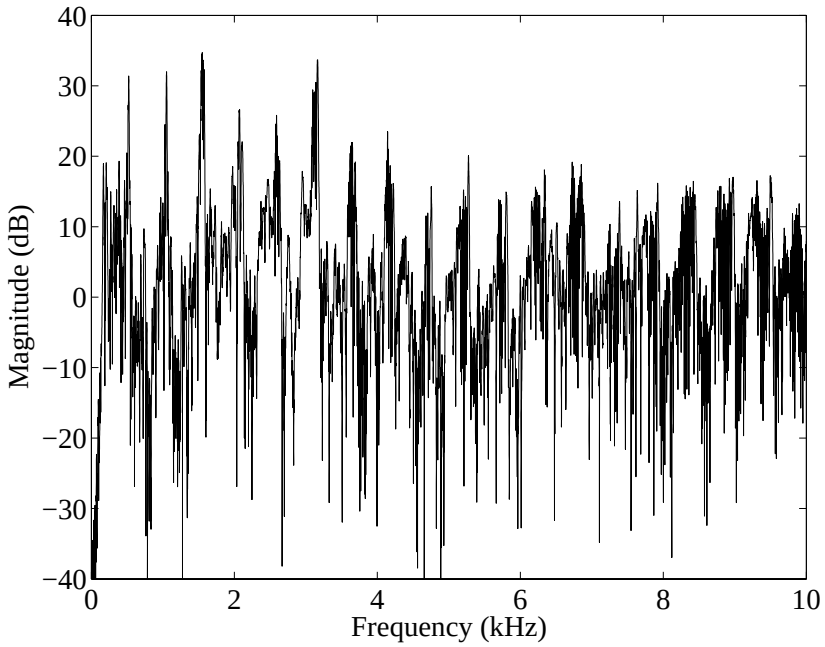

Figure 3.12: The magnitude response of the sound with the brightness parameter of 1.2.

in Figure 3.13. The attack time parameter has a value of -0.08 in Figure 3.13(a) and 0.08 in Figure 3.13(b). Each harmonic is operated with one of these curves depending on whether the attack time is required to be longer or shorter. Like in previous timbre dimensions the processed harmonics are summed up. The implementation is defined as

$$s_{\mathrm{a}}(n) = \sum_{i=1}^{M} c_{\mathrm{a}}(n) h_i(n) \quad , \tag{3.13}$$

where $s_{\mathrm{a}}(n)$ is the sound processed in attack time dimension and $c_{\mathrm{a}}(n)$ is the curve for modifying the attack time.

The reference sound and the sounds which are modified in the attack time dimension are shown in Figure 3.14. When the attack time is modified as described the spectrum of the sound is modified too. The change in timbre perception may result from the change both in the waveform and the spectrum of the sound and not from the change only in the waveform.

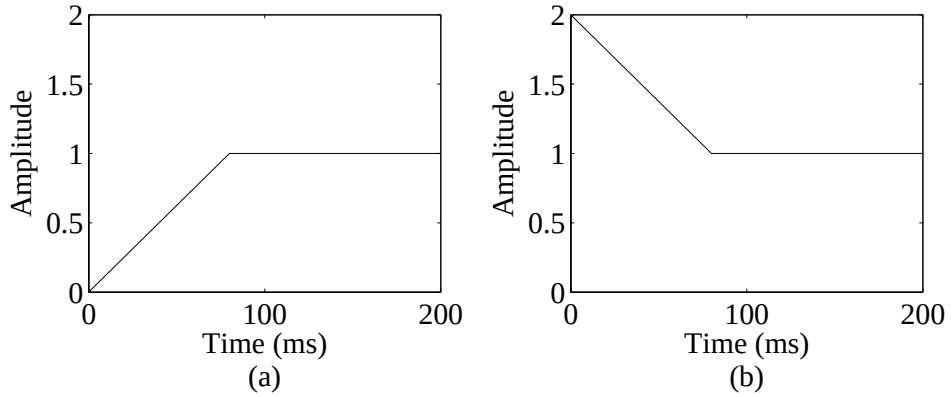

Figure 3.13: The envelope curves for (a) lengthening ($x_{\mathrm{a}} = -0.08$) and (b) shortening ($x_{\mathrm{a}} = 0.08$) the attack time.

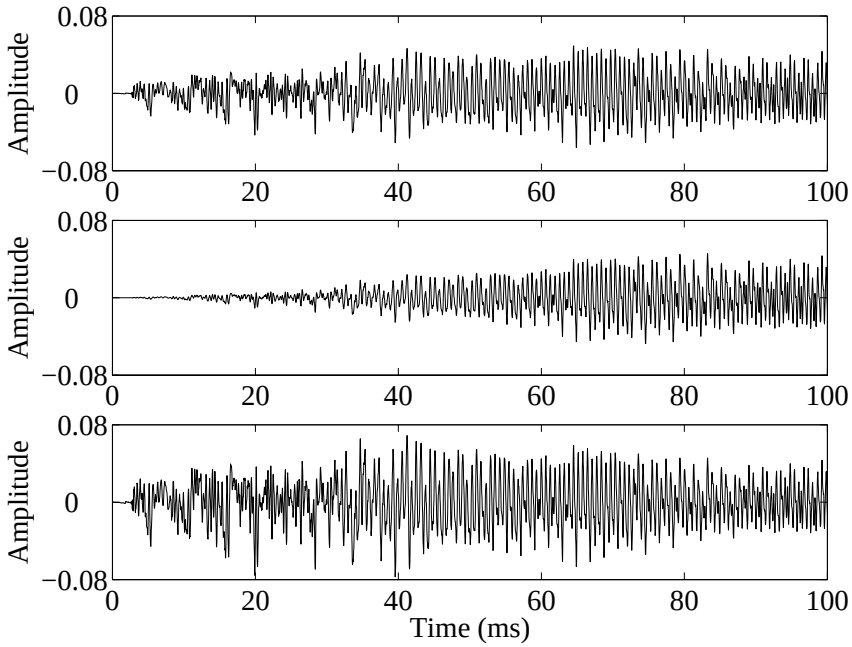

Figure 3.14: Top to bottom: the waveform of the reference sound, of the sound with longer attack time ($x_{\mathrm{a}} = -0.08$), and of the sound with shorter attack time ($x_{\mathrm{a}} = 0.08$).

3.6 Modification of Noise

The fourth dimension of the sound matrix is the amount of noise. This is the only dimension where the separated harmonics of the original sound are not processed. The background noise is extracted from the original sound by removing all harmonics so that only noise is remaining. The background noise is extracted by using the same

inverse comb filter as in the separation of harmonics (see Section 3.2, Equation (3.3)). The only difference is that now there is no resonator picking up the harmonics because all harmonics are required to be removed. The magnitude response of the extracted noise is shown in Figure 3.15. Comparison against the spectrum of the original tone in Figure 3.1 reveals that the harmonic components are quite much attenuated and only the background noise is left in the signal.

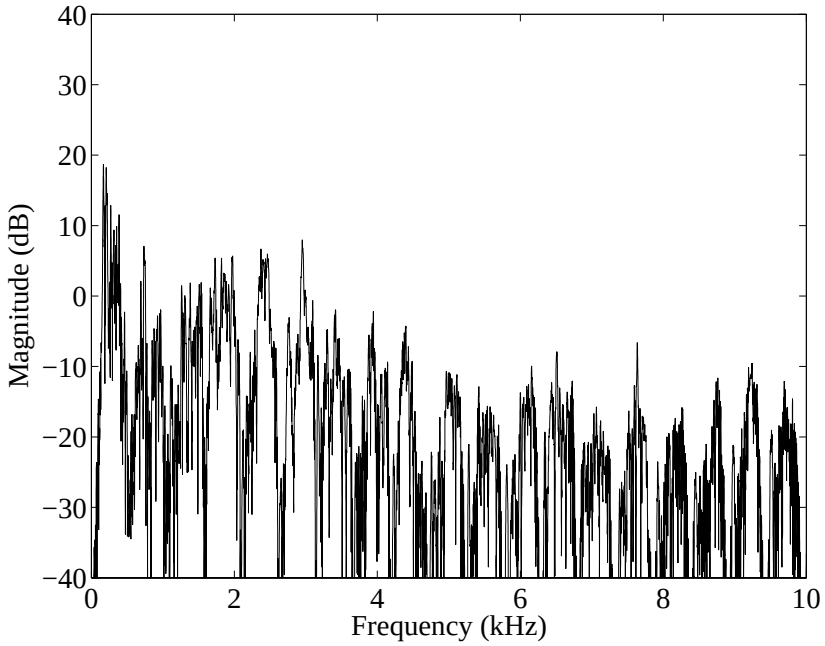

Figure 3.15: The magnitude response of the background noise.

The background noise is either reduced from the sound or added to the sound, according to the timbre parameter for noise dimension x_{noise}. Noise parameter describes the required amount of noise in the signal. It is required that $x_{\mathrm{noise}} \geq 0$.

The noise coefficient c_{noise} tells which amount of noise has to be reduced from or added to the sound signal. The noise coefficient is defined as

$$c_{\mathrm{noise}} = x_{\mathrm{noise}} - 1 \ . \tag{3.14}$$

The noise coefficient is negative if the noise parameter is below 1. Then noise is reduced from the sound signal. If the noise parameter is above 1, the noise coefficient is positive and noise is added to the sound. If the parameter is exactly 1 the sound has its original background noise. The value of the noise parameter in the reference sound is 1, which means that the amount of noise is not modified.

When the noise parameter and the noise coefficient are specified, the extracted noise signal is multiplied by the noise coefficient. This processed noise is summed to the reference sound, which results to reducing or adding of noise depending on whether the noise coefficient is negative or positive, respectively. Since separated harmonics are not processed at all no summing of harmonics needs to be done, unlike in the other dimensions. The implementation of noise modifications is expressed as

$$s_{\mathrm{noisy}}(n) = s_{\mathrm{ref}}(n) + c_{\mathrm{noise}} s_{\mathrm{noise}}(n) \ , \tag{3.15}$$

where $s_{\mathrm{noisy}}(n)$ is the sound processed in noise dimension and $s_{\mathrm{noise}}(n)$ is the extracted background noise.

3.7 Combining Dimensions

The sound matrix consists of the reference sound and the sounds that differ from the reference in one or more timbre dimensions. Thus, the dimensions have to be combined.

Timbre parameters for harmonic, brightness, attack time, and noise dimensions in the reference sound are x_{harm} = 0, x_{b} = 1, x_{a} = 0, and x_{noise} = 1, respectively. The sounds are reconstructed from the extracted harmonics and noise according to the timbre parameters and coefficients. When generating sounds and combining dimensions, each harmonic is multiplied by harmonic coefficient (c_{even} or c_{odd}), brightness coefficient (c_i), and curve for modifying the attack time ($c_{\mathrm{a}}(n)$). The processed harmonics are summed up in order to reconstruct the sound. The noise is reduced from or added to the combination sound according to the noise coefficient (c_{noise}). If the sound is processed both in attack time and noise dimensions so that the attack time is required to be longer ($x_{\mathrm{a}} < 0$) the noise is operated with the attack time curve, too. The implementation of the reference sound and the sounds that are modified in one to four dimensions is expressed as

$$s(n) = \begin{cases} c_{\mathrm{noise}} s_{\mathrm{noise}}(n) c_{\mathrm{a}}(n) + \sum_{i=1}^{M} c_{\mathrm{harm}} c_m c_{\mathrm{a}}(n) h_i(n) & \text{if } x_{\mathrm{a}} < 0 \text{ and} \\ & \text{if } x_{\mathrm{noise}} \neq 1 \ , \\ c_{\mathrm{noise}} s_{\mathrm{noise}}(n) + \sum_{i=1}^{M} c_{\mathrm{harm}} c_m c_{\mathrm{a}}(n) h_i(n) & \text{else} \ , \end{cases} \tag{3.16}$$

where $s(n)$ is the processed sound, c_{harm} is either c_{even} (for even values of i) or c_{odd} (for odd values of i), and c_m is either c_i (for $x_{\mathrm{b}} \leq 1$) or c_{15} (for $x_{\mathrm{b}} > 1$).

Since four dimensions of timbre are considered within the sound matrix, the sounds can be described by four timbre parameters. If a sound is modified, e.g., in harmonic

dimension, the harmonic parameter of this sound is not equal to that of the reference sound unlike all other timbre parameters.

Chapter 4

Subjective Evaluation of Timbre Modifications

4.1 Introduction

For a brain research task the sounds of the sound matrix are required to have an equal perceived distance from the reference sound. The sounds are modified in several timbre dimensions according to timbre parameter values and so these values have to be chosen carefully. A listening test was arranged in order to determine timbre parameter values for each dimension in two directions: one that is smaller than in the reference sound and another that is greater than in the reference. The method for evaluating the psychoacoustic distances and obtaining equal perceived differences within modified sounds through a listening test has been accepted for publication in an international conference [10].

Before the actual listening test the test values for timbre parameters were determined. In each dimension four parameter values that were smaller than in the reference sound and four values that were greater than in the reference were used. It turned out in informal listening that in attack time dimension it was not possible to attain as large psychoacoustic distances in sounds as in other dimensions and so in attack time parameter the whole value range was used ($x_{\mathrm{a}} \in [-0.08, 0.08]$). In harmonic, brightness, and noise dimensions the parameter values were determined so that there would be as large perceived differences as in attack time dimension. The timbre parameter values used in the listening test are given in Table 4.1. The values of the reference sound are bolded in the table.

The listening test was conducted so that the sounds differed from each other only

Table 4.1: Timbre parameter values used in the listening test.

Index	Harmonics	Brightness	Attack Time	Noise
1	−0.4	0.75	−0.08	0
2	−0.3	0.91	−0.06	0.3
3	−0.2	0.95	−0.04	0.5
4	−0.1	0.97	−0.02	0.75
5	**0**	**1**	**0**	**1**
6	0.1	1.04	0.02	1.5
7	0.2	1.06	0.04	1.75
8	0.3	1.09	0.06	2.1
9	0.5	1.15	0.08	3

in timbre. The definition of the timbre is that timbre allows a listener to discriminate sounds that have the same pitch, duration, and loudness. Now each sound has the frequency of 524 Hz and the duration of 200 ms but also the loudness of the sounds had to be equalized before the listening test. At first two automatic equalization methods which are based on Moore's loudness model [18] were tried out. It turned out that these methods were not suitable for loudness equalization of these sounds. Therefore the sounds were finally equalized for perceived loudness so that two listeners adjusted the loudness of the sounds to be equal with interactive software. The whole equalization process is described in the reference [9].

4.2 Listening Test Method

A listening test with six subjects (two women and four men) was arranged in order to determine the perceived difference of the sounds compared to the reference sound and so to determine convenient timbre parameter values for each dimension. The subjects were 23–30 years old with an average age of 26.3 years. Four of the six subjects were the personnel of the Laboratory of Acoustics and Audio Signal Processing. Two other subjects were the students of Helsinki University of Technology, both having experience either in acoustics or string instruments. All subjects reported normal hearing. The test was implemented using the experimental listening test software GuineaPig [8] and it was arranged in the listening room of the Acoustics Laboratory. The sound samples were played through Sennheiser HD 580 headphones.

In the test altogether 33 different sounds including the reference sound, i.e., eight

sounds in each dimension of timbre, were used. Each of the sounds differed from the reference with respect to one dimension only. Any combination sounds were not included in order to restrict the number of test samples. The listening test was an A/B scale (hidden reference) test. In the test each sound and the reference were played in pairs but the subjects did not know which one the reference was. Also a pair with two reference sounds was included. The subjects were allowed to listen to the sounds as many times as they wanted and they could also have breaks whenever they wanted. The subjects were told that the sounds differ from each other only in timbre. They were asked to evaluate how different sample B is compared to sample A. The subjects evaluated the psychoacoustic distance of each pair on a continuous scale from 0 (sounds are the same) to 10 (sounds are totally different). The evaluation was done by moving a slider on a computer screen. The user interface of the listening test is shown in Figure 4.1.

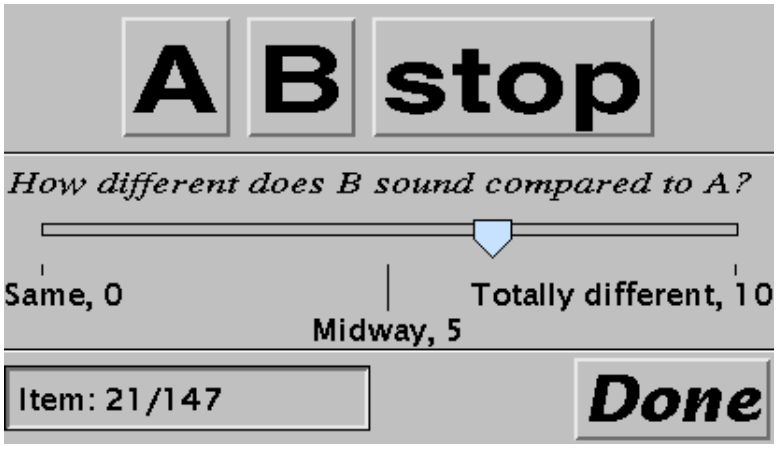

Figure 4.1: A graphical user interface of the listening test.

In the listening test 33 different sound pairs were included. Each pair was presented four times so that sample A was the reference twice and sample B was the reference also twice. Thus, altogether 132 sound pairs were presented in the test. Two different lists of sound pairs were generated from which three subjects had one and three subjects the other. The sound pairs were presented in random order.

In the beginning of the test the subjects had the possibility to hear some of the sounds and note differences between them. A sound pair with two reference sounds and eight pairs including the reference and the most different sounds in each dimension considering the parameter values (see Table 4.1, indices 1 and 9) were presented to the subjects. It was told if the sounds were the same or totally different. After this instruc-

tion there was a training session of six sound pairs where the subjects could practice the evaluation before the actual test. The subjects were asked if the volume level was optimal for them or if they wanted to adjust it. The listening test with instruction and training sessions took about 40 minutes per subject. After the test the subjects were asked for comments on the test and sounds.

4.3 Results

In this section the results of the listening test are presented. The results tell how the timbre modifications were perceived and how the psychoacoustic distances between the sounds were evaluated.

For each sound the average of the perceptual distances from the reference sound was calculated. The upper L_{u} and lower L_{l} bounds for a 95% confidence interval were also calculated according to [16]

$$L_{\mathrm{u}} = \mu + 1.96\left(\frac{\sigma}{\sqrt{N_{\mathrm{s}}}}\right) \quad , \tag{4.1}$$

$$L_{\mathrm{l}} = \mu - 1.96\left(\frac{\sigma}{\sqrt{N_{\mathrm{s}}}}\right) \quad , \tag{4.2}$$

where μ is the mean value of psychoacoustic distances, σ is the standard deviation and N_{s} is the sample size.

The averages of the psychoacoustic distances evaluated by comparing the sounds to the reference sound are given in Table 4.2. The results are presented separately for each dimension so that the timbre parameter value increases with increasing index number. The index numbers correspond to the ones in Table 4.1 and so the index number 5 corresponds to the reference sound paired with itself. The reference sound compared to the reference sound itself gave the psychoacoustic distance of 0.7. This reveals the noise level of the test and the accuracy of the listeners. In the scale of 0–10 the psychoacoustic distance of 0.7 is very small, which means that the subjects have evaluated the perceptual differences quite accurately.

The psychoacoustic distances with a confidence interval of 95% are shown in Figure 4.2. The results are presented separately for each dimension so that the parameter value controlling the timbre is increasing for each point, being smaller than in the reference sound for the first four points (see Tables 4.1 and 4.2, indices 1–4) and greater than in the reference for the last four points (see Tables 4.1 and 4.2, indices 6–9). The middle point of each dimension is the reference sound paired with itself (see Tables 4.1 and

Table 4.2: Psychoacoustic distances of the sounds compared to the reference sound.

Index	Harmonics	Brightness	Attack Time	Noise
1	8.7	9.8	5.2	7.1
2	7.2	8.5	4.8	7.2
3	4.9	6.9	3.8	5.6
4	2.1	4.3	2.6	1.5
5	0.7	0.7	0.7	0.7
6	1.3	5.0	2.0	4.6
7	4.0	6.8	2.5	6.1
8	5.7	8.3	4.0	6.8
9	8.6	9.3	2.6	8.4

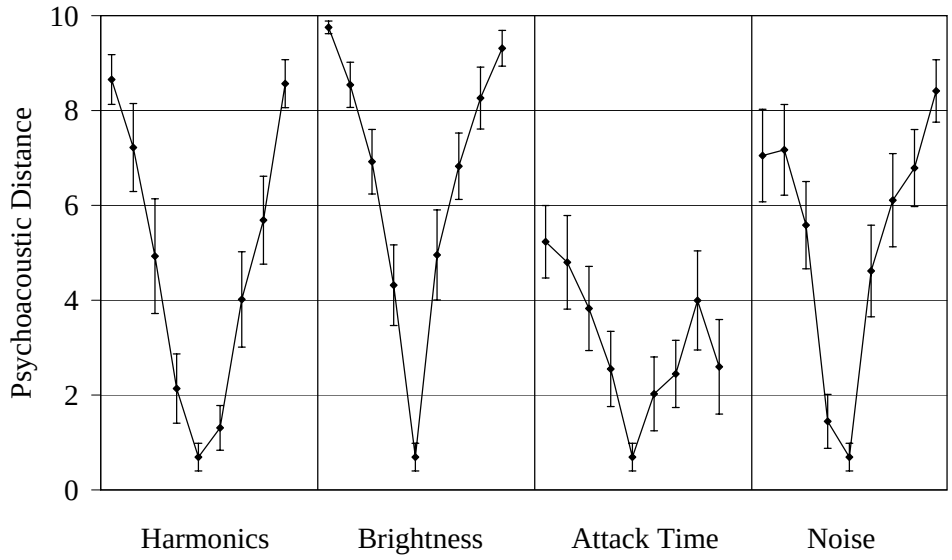

Figure 4.2: Psychoacoustic distances of the sounds compared to the reference sound with a confidence interval of 95%.

4.2, index 5). More detailed results are presented in Appendix A. The general impression and comment on the test was that it was hard to compare the sounds because they were different in terms of so different attributes. The oral comments on the listening test given by the subjects are presented in Appendix B.

4.4 Discussion

The listening test was arranged in order to be able to generate the sound matrix where the sounds are perceived as equally different compared to the reference. From Figure 4.2 it is evident that the required sound matrix can be generated by choosing sounds and timbre parameter values from the same level of psychoacoustic distance. It seems that in harmonic, brightness, and noise dimensions the sounds were successfully modified and the timbre differences in the sounds were observable. It seemed already earlier in informal listening that in attack time dimension it is not possible to attain as large differences as in other dimensions. From the listening test results it is evident that the psychoacoustic distances in attack time dimension were not large enough. Obviously the method of modifying the attack time by processing the signal waveform as described in Section 3.5 was not efficient enough. The processing of the sound waveform modified also the spectrum of the sound in these cello tones and the perceived differences may have resulted from this change both in the waveform and the spectrum and not from the change only in the waveform. It might be that in the sounds, in which the spectrum would not have been modified too, the timbre modifications would not have been perceived at all. However, since three dimensional sound matrix with 3^3 = 27 sounds is sufficient to the brain research task, the three ”best” dimensions, i.e., harmonic, brightness, and noise dimensions, were chosen to be used in the brain recordings.

We aimed at choosing two sounds, or two timbre parameter values, for each dimension: one parameter value that is smaller than in the reference sound and the other that is greater than in the reference. Since the aim was to generate a set of sounds that have an equal psychoacoustic distance from the reference, each dimension provides only two suitable sounds. In all these three dimensions (harmonics, brightness, and attack time) there is a sound whose timbre parameter value is smaller than in the reference sound and which has been evaluated to have psychoacoustic distance of about 7. Also in two dimensions (brightness and noise) there is a sound that has a greater parameter value than the reference sound and psychoacoustic distance of about 7. So the sounds and parameter values were chosen from the level of 7 for harmonic, brightness, and noise dimensions. For a sound of harmonic dimension that is required to have a greater timbre parameter value than the reference sound the parameter value is interpolated because none of the sounds has psychoacoustic distance of about 7. Linear interpolation is done according to parameter values and psychoacoustic distances of two sounds that are above and below the distance level of 7 in harmonic dimension

(see Tables 4.1 and 4.2, indices 8–9)

$$x_{\mathrm{harm}} = 0.3 + (0.5 - 0.3)\left(\frac{7.0 - 5.7}{8.6 - 5.7}\right) = 0.39 \ . \tag{4.3}$$

The sounds having the psychoacoustic distance of about 7 have the following index numbers (see Table 4.2). For harmonic dimension the index is 2, the other parameter is interpolated between indices 8 and 9. For brightness dimension the indices are 3 and 7, and for noise dimension they are 1 and 8. The final parameter values of the sounds to be used in the brain recordings are shown in Table 4.3. The parameter values of the reference sound are bolded in the table.

Table 4.3: Final timbre parameters.

Index	**Harmonics**	**Brightness**	**Noise**
1	-0.3	0.95	0
2	**0**	**1**	**1**
3	0.39	1.06	2.1

The sound matrix was generated as described in Chapter 3 using the timbre parameters given in Table 4.3. Since only three dimensions instead of four were chosen to be used in the brain recordings, the size of the matrix is $3 \times 3 \times 3$. The sounds of the sound matrix differ from each other in one to three timbre dimensions. The matrix consists of 27 sounds, i.e., the reference and all possible combinations. Each sound has an equal fundamental frequency and duration. The sounds were also equalized for perceived loudness. The loudness equalization was carried out like within the test sounds of the listening test. The sound matrix as a function of the timbre parameters is illustrated in Figure 4.3. The dots in the figure present the sounds and their locations in timbre space. The reference sound is marked in the center of the figure.

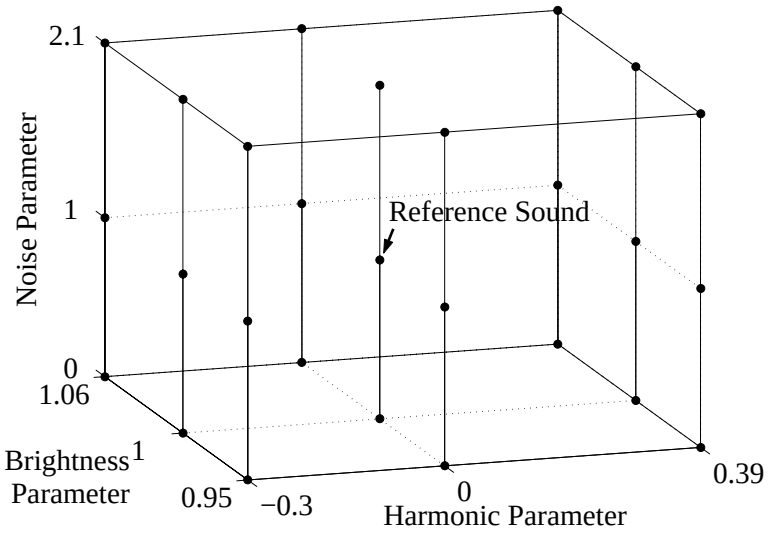

Figure 4.3: The sound matrix as a function of the timbre parameters.

Chapter 5

Brain Recordings

5.1 Introduction

The sounds of the sound matrix were used as test stimuli in an ERP study. The brain recordings were conducted by presenting the test stimuli to the subjects, recording EEG from the scalp, and extracting ERPs from the ongoing EEG. In the recordings the main purpose was to study the effect of the standard probability on the MMN response when the stimuli used differed from each other only in timbre and each deviant differed from the standard with an equal step. The amplitude and latency of the MMN responses were studied.

In earlier studies the effect of standard and deviant probabilities on the MMN response have been examined [29]. It has been found that the MMN amplitude is dependent on the probabilities of the standard and deviant. Frequently presented standard tones generate a memory trace with which the deviant tones are compared. The memory trace represents the features of the standard and becomes stronger the more often the standard occurs. Increasing the strength of the memory trace increases the MMN amplitude elicited by the deviant, too. Thus, the higher the standard probability the higher the MMN amplitude. However, it is also found that whether a sound functions as standard or deviant is not based on its relative probability but it depends on its relationship to all the other sounds in the stimulus sequence and the longer-term context of the sequence. So the absolute value of the standard or deviant probability does not necessarily tell whether the MMN response is evoked or not by the deviant but the relationship between the standard and deviant probabilities could be a better measure.

5.2 Method

The EEG recordings were performed at the Cognitive Brain Research Unit (CBRU) at the University of Helsinki. Six subjects (two women and four men) with reportedly normal hearing participated in the recordings. The subjects were 23–30 years old with an average age of 28.0 years. One of the subjects participated also in the listening test where the timbre modifications were evaluated. The test stimuli were played through the same headphones (Sennheiser HD 580) as in the listening test. The sound pressure level of the stimuli was 60 dB above the subject's absolute hearing threshold. The threshold for each subject was determined before the EEG recordings. In the ERP study the stimuli were played with BrainStim v0.43 software and EEG was recorded with the NeuroScan system.

The test stimuli used were the sounds of the sound matrix. The reference sound was chosen to be the standard stimulus (frequent sound) because in timbre space it is in the middle of the other sounds (infrequent deviants). The psychoacoustic distance of each deviant compared to the standard was approximately the same as what was found in the listening test to correspond to the level of 7 (see Chapter 4). Five different stimulus sequences were presented to the subjects. The probability of the standard was varied in different sequences. Since there were 26 different deviants in each sequence and they all were presented almost equally many times, the probability of each individual deviant was relatively small in each sequence. However, it should be noted that because the number of the different deviants was equal in all sequences the probability of the individual deviant in the sequence increased with decreasing standard probability. A relatively large number of different deviants was required in order to make sure that the individual deviant probability would be smaller than the standard probability also in the sequences with low standard probability, i.e., the standard would occur clearly more often than each individual deviant. Equal perceived difference between each deviant and standard ensures that the MMN response elicited by the deviants is not based on how large the psychoacoustic distance is between an individual deviant and the standard but it depends on how often the standard and deviants occur in the stimulus sequence. The sequence information is given in Table 5.1. In the table the sequence number, probabilities of the standard and deviants, the number of stimuli, and the probability of each individual deviant are presented.

The standard and deviants were presented in pseudorandom order with the following requirements. It was required that in sequences 1 and 2 there would be at least two deviants between two consecutive standards and in sequence 4 at least two stan-

Table 5.1: Information about stimulus sequences used in the EEG recordings.

Sequence Number	**Standard Probability (%)**	**Deviant Probability (%)**	**Number of Stimuli**	**Individual Deviant Probability (%)**
1	10	90	1000	3.4–3.5
2	30	70	1000	2.6–2.7
3a	50	50	1000	1.9–2.0
3b	50	50	1500	1.9
4	80	20	1000	0.7–0.8

dards between two consecutive deviants. In sequences 3a and 3b nothing like this was required. Also in sequence 1 it was required that at least 18 stimuli would be presented before the same deviant could occur again. In other sequences this number of stimuli between the same deviants was required to be 20. With these requirements we tried to avoid developing a memory trace also for deviants by obtaining the stimulus sequences where the standard probability would be the same over the whole sequence and the same individual deviant would not be repeated too often.

The EEG was recorded from the subject's scalp using an electrode cap with 32 channels. The cap was placed to the head according to an individually measured location of the electrode at Cz. Only 30 channels of the cap were used, which includes both the left (LM) and right mastoids (RM). In addition to these channels the electrooculogram (EOG) measuring eye movements was recorded. Both horizontal (HEOG) and vertical (VEOG) EOG were recorded with electrodes placed above and next to the right eye. The EEG was recorded towards a reference electrode attached to the tip of the nose.

The stimulus sequences were presented to the subjects in the following order. To the subjects S1–S3 the sequences presented were 2 (twice), 3a (twice), and 4 (twice) and to the subjects S4–S6 they were 1 (twice), 2 (twice), 3b (once), and 4 (twice). The stimulus onset asynchrony (SOA), i.e., the time between the onsets of two stimuli in the sequences was 505 ms. The duration of the EEG recordings including short breaks between the sequences was approximately 55 minutes for the subjects S1–S3 and 65 minutes for the subjects S4–S6. During the recordings the subjects watched a movie that was shown with subtitles and without sounds. The subjects chose the movie themselves. The subjects were asked to concentrate on the movie and try not to listen to the stimuli in order to study MMN in the absence of attention. They were also asked to avoid muscle movements and blinks if possible but still feel themselves comfortable and relaxed without any muscle tensions.

5.3 Data Analysis and Results

The recorded EEG was amplified and digitized. Then the EEG was bandpass filtered between 1 and 20 Hz with the slope of 12 dB/octave. Epochs containing artifacts created by muscle movements or blinks and resulting a change of over 100 μV in the EEG were rejected. Epochs of 500 ms with a 100 ms prestimulus interval were used. The epochs that were not rejected were averaged separately for standard and deviants in the cases of each standard probability. With standard probability of 80% only those standards that were preceded by at least two standards were accepted to the averaging procedure in order to obtain as good and stabile responses as possible. With other standard probabilities all standards were accepted. All different deviants were averaged together in each case of standard probability. After correcting the baseline in ERPs to both the standard and the deviants, the ERP difference curves were obtained by subtracting the standard average from the deviant average. All sequences except 3b were presented twice to the subjects and also the resulting two difference curves were averaged for each subject and for each standard probability.

In Figure 5.1 the electrode locations and the difference curves of one subject (S4) in 30 channels with standard probability of 80% are shown. The EOG channels are not included in the figure. The time interval of 100–300 ms is marked to show the most significant time period when considering the MMN response. Also the time instant of the largest MMN amplitude (i.e., the most negative peak) in channel 6 (midline frontal location) is shown in the figure. It seemed that the largest MMN was elicited in channel 6 and that is why the data analysis was further on conducted by the responses recorded in this channel. In Figure 5.2 the MMN amplitudes averaged from the difference curves of all subjects in channel 6 are shown. The MMNs with all standard probabilities are shown in the figure.

The linear derivation (LDR) distribution was defined for each subject according to the latency of the largest MMN amplitude in channel 6. The standard probability of 80% was used because the largest MMN was elicited with that standard probability. The time interval considered was 100–300 ms. The amplitudes and latencies of the largest MMN in channel 6 for each subject are given in Table 5.2. Also the averages of the amplitudes and latencies are given. The LDR is a spatial template and it includes the amplitudes of each channel at the latency of the largest MMN and so describes the best possible or the perfect MMN response distribution across the electrodes for each subject. The difference curves of each subject and each standard probability were filtered with the individual spatial filter. The signals of interest were retained

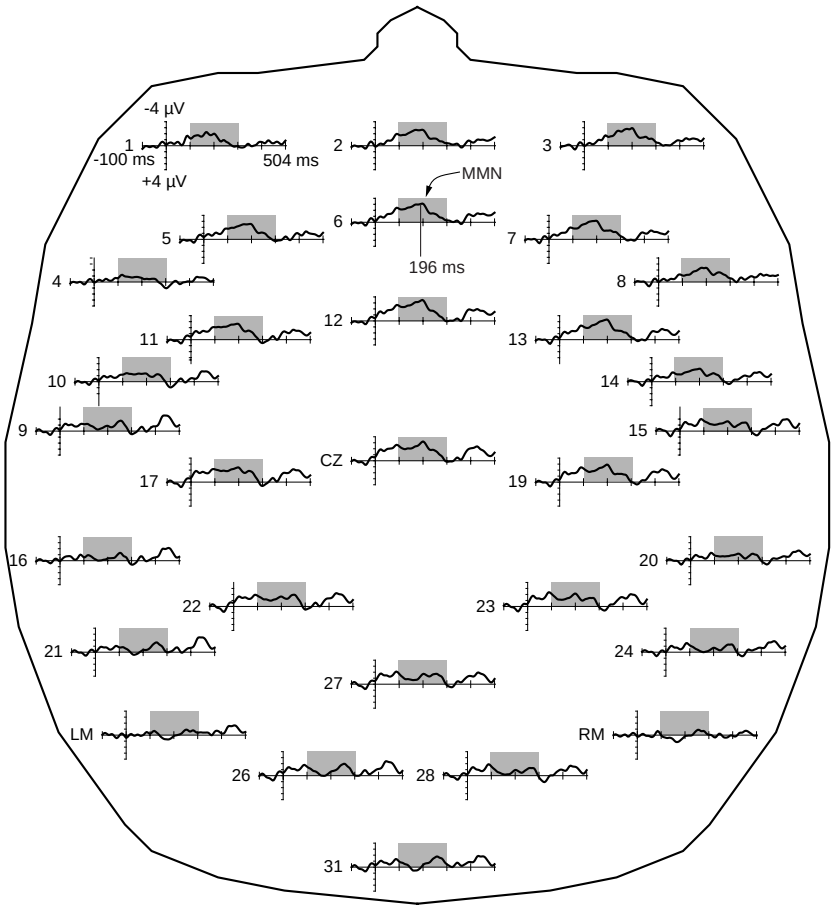

Figure 5.1: The electrode locations and the ERP difference curves of one subject in 30 channels with standard probability of 80%.

according to the LDR distribution so that for each subject their own LDR was used for each standard probability. This means that the perfect MMN was searched from each ERP difference curve resulting in a correlation measure between these signals.

The results of spatial filtering are shown in Figures 5.3–5.7. In Figures 5.3–5.6 the spatially filtered responses (difference curves filtered according to the LDR distribution) are presented separately for each standard probability. In each figure the response for each subject (S1–S6) and also the average are shown. In Figure 5.7 the averages of the spatially filtered responses for each standard probability are presented.

In Table 5.3 the peak amplitudes and latencies of the averages of the spatially filtered responses for each standard probability are given. The spatially filtered response's relative amplitude of 1 corresponds to the MMN amplitude of -1.46 μV (i.e., the mean MMN amplitude in channel 6). For standard probability of 10% the actual peak

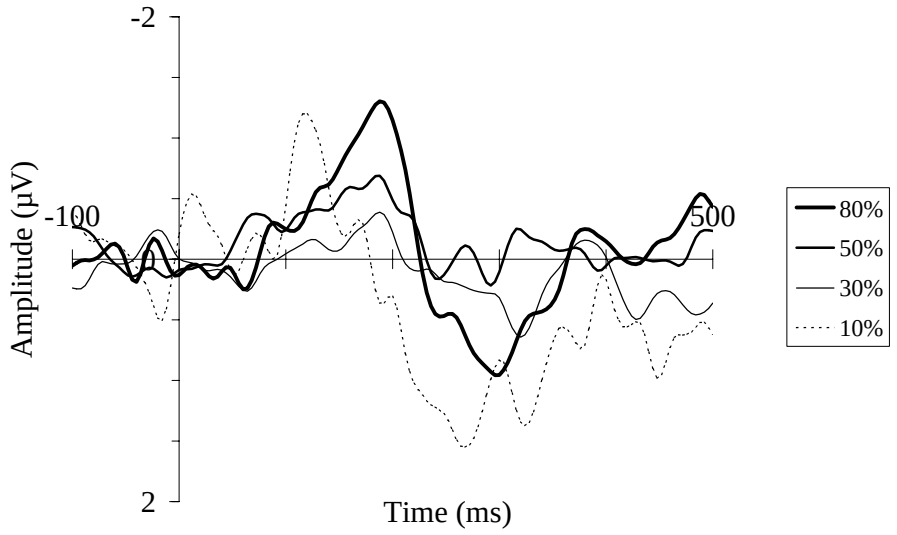

Figure 5.2: The ERP difference curves in channel 6 with different standard probabilities. Averages of all subjects.

Table 5.2: The MMN peak amplitudes and latencies elicited by the deviants in channel 6 with standard probability of 80%.

Subject	**Amplitude (μV)**	**Latency (ms)**
S1	-0.38	144
S2	-1.73	196
S3	-1.29	192
S4	-3.12	196
S5	-0.54	172
S6	-1.68	184
Mean	-1.46	181

latencies were 108 ms and 364 ms (see Figure 5.7) which are too early and too late, respectively, for the MMN response. Since there was no appropriate latency for the MMN, the amplitude given in the table was taken at the same latency as with the standard probability of 30% (188 ms).

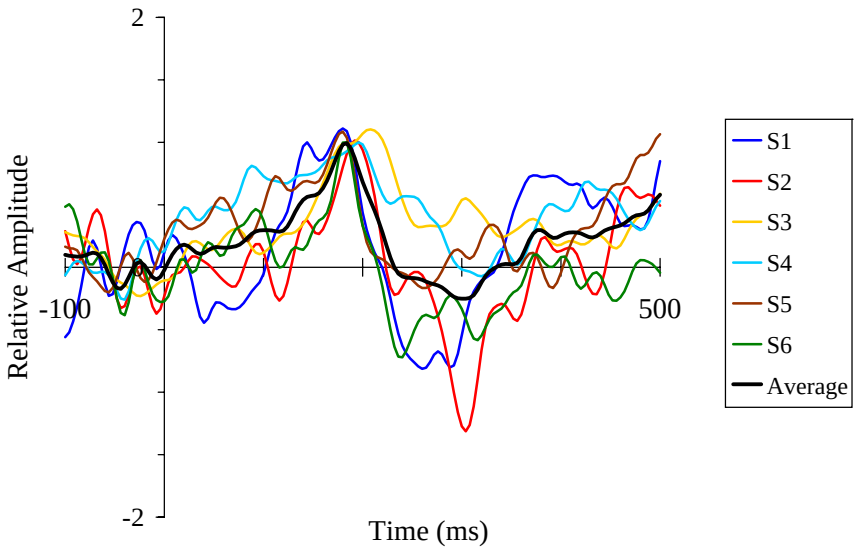

Figure 5.3: Spatially filtered responses for each subject and the average, standard probability of 80%.

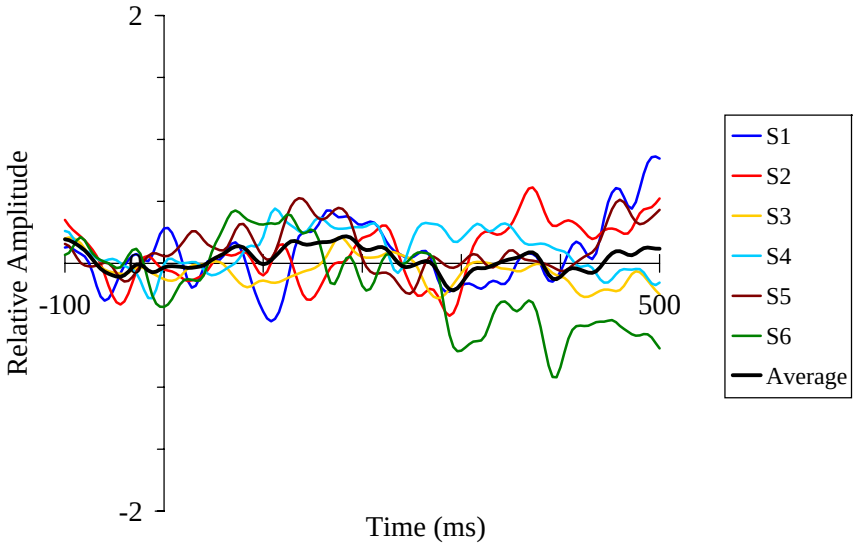

Figure 5.4: Spatially filtered responses for each subject and the average, standard probability of 50%.

5.4 Discussion

The main goal of the brain recordings was to study the MMN response and the effect of the standard probability on it. The MMN response was studied by using spatially filtered difference (deviant−standard) curves.

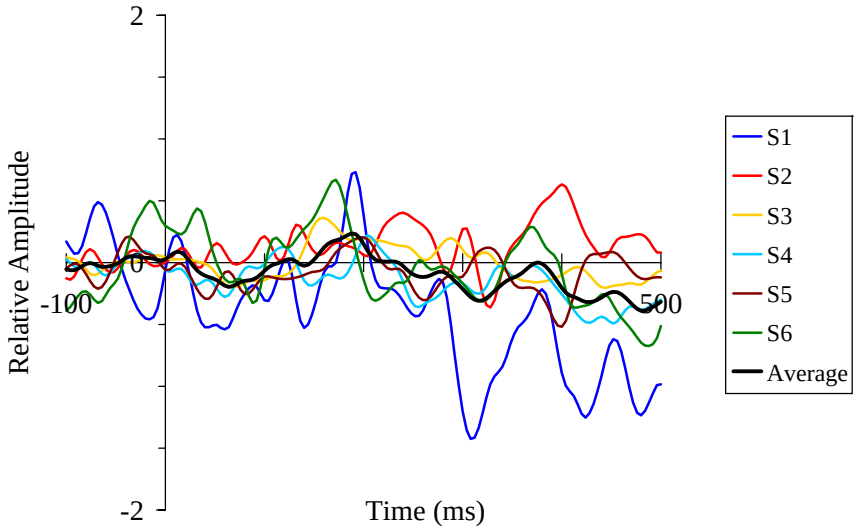

Figure 5.5: Spatially filtered responses for each subject and the average, standard probability of 30%.

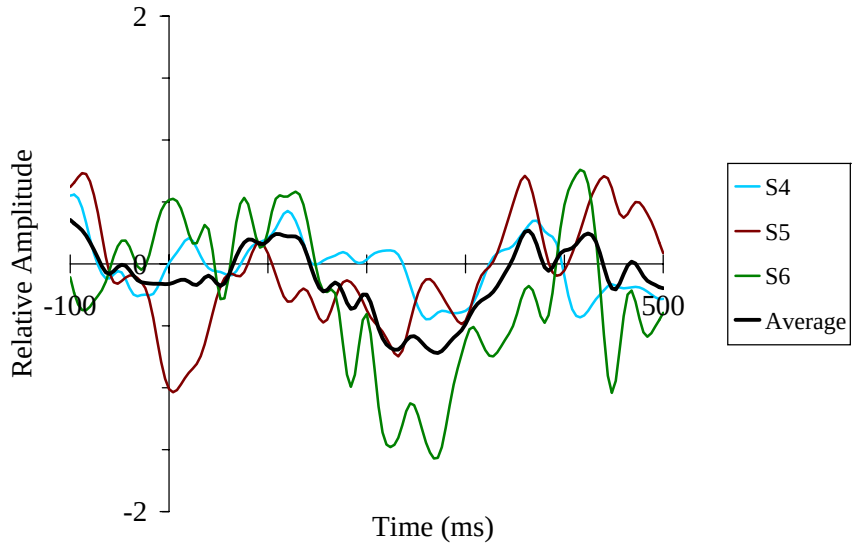

Figure 5.6: Spatially filtered responses for subjects S4, S5, and S6 and the average, standard probability of 10%.

From the difference curves in Figure 5.2 (difference curves of all subjects averaged, channel 6) it is evident that the memory trace of the standard was strong enough and the MMN response was elicited by the deviants for standard probabilities of 80%, 50%, and 30%. For the standard probability of 10% the MMN response was not elicited. With spatial filtering the shape of the perfect MMN response for each subject, or the

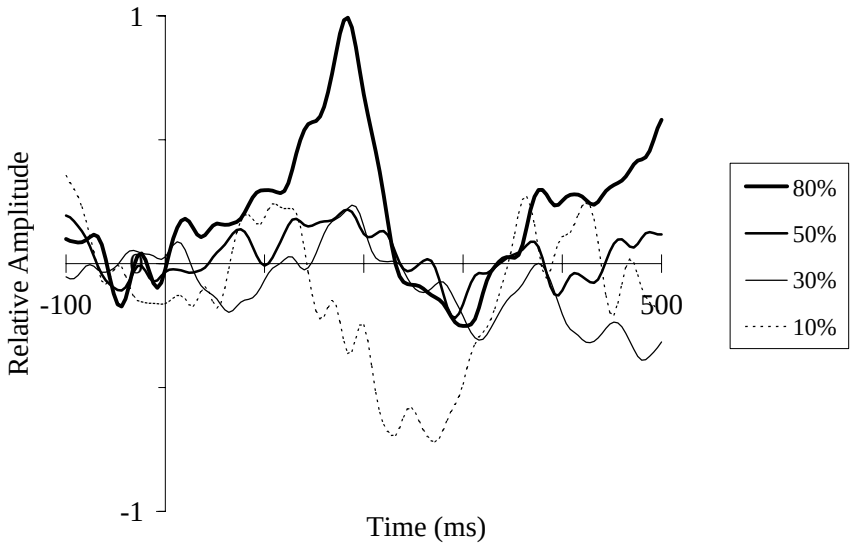

Figure 5.7: Spatially filtered responses, the average for each standard probability.

Table 5.3: The peak amplitudes and latencies of the spatially filtered MMN response averages for each standard probability.

Standard Probability (%)	**Relative Amplitude**	**Amplitude (μV)**	**Latency (ms)**
80	0.99	-1.45	184
50	0.22	-0.32	184
30	0.24	-0.35	188
10	-0.35	0.51	188

typical shape of each subject's MMN response, was taken into account. The perfect MMN response was searched from the difference curves resulting the spatially filtered MMN response. The elicited MMN response and the spatially filtered MMN are connected to each other so that the more positive the relative amplitude of the spatially filtered response the larger the elicited MMN. From Figure 5.7 and Table 5.3 (averages of the spatially filtered responses for each standard probability) it is observed that in the case of standard probability being 80% the MMN response was elicited with the spatially filtered response's relative amplitude of 0.99 and the latency of 184 ms. With the standard probability of 50% the spatially filtered MMN relative amplitude is clearly smaller (0.22) but it exists at the same latency as with 80% (184 ms). With the standard probability of 30% the spatially filtered MMN relative amplitude is approximately the same (0.24) as in previous case. Also the latency is only a bit greater (188 ms). For

standard probability of 10% the relative amplitude is negative (-0.35), which confirms the observation that the MMN response was not elicited.

From these results it can be concluded that the task of developing a memory trace for the standard and discriminating deviants from it becomes harder when the standard probability is decreased. The results with the standard probability of 80% are quite obvious considering the definition of the MMN response. However, it should also be noted that even in the case where the majority of the stimuli in the sequence were deviants the MMN was elicited if the standard probability was large enough (i.e., 30%). Obviously the memory trace for the standard tone cannot be developed if the standard occurs with probability of only 10%.

Since the MMN response elicitation is not based only on the absolute value of the standard or deviant probability, these results should also be analyzed by considering the relationships between the standard and individual deviant probabilities. From Table 5.1 we get that the individual deviant probabilities corresponding the standard probabilities of 80%, 50%, 30%, and 10% are 0.7–0.8%, 1.9–2.0%, 2.6–2.7%, and 3.4–3.5%, respectively. Combining the relationships between these probabilities with the MMN response information given in Table 5.3 it is observed that the ratios of the probability of the individual deviant and the probability of the standard (i.e., $0.8/80 = 0.01$, $2.0/50 = 0.04$, $2.7/30 = 0.09$, and $3.5/10 = 0.35$) were small enough to elicit the MMN response by the deviant with the standard probabilities of 80%, 50%, and 30% but not with the standard probability of 10%. It is possible that the differences in the MMN elicitation between the test cases where the majority of the stimuli were deviants were based not only on the standard probabilities but also on the relationships between the standard and individual deviant probabilities.

From Figure 5.2 it is also obvious that in addition to the MMN response the P3 response was evoked by the deviants. This positive component of the ERP is observed approximately 300 ms after the stimulus onset. The P3 response is related to the spontaneous attention switching from a task to a deviant tone. This means that in this experiment the subjects have unconsciously detected the difference in deviant compared to the standard and their attention has spontaneously switched from the movie to the deviant tone for a short period of time. The P3 response was most clearly evoked with the standard probability of 80% but also with the standard probabilities of 50% and 30%.

To sum up this discussion it is observed that in this experiment the MMN response was elicited with standard probabilities of 80%, 50%, and 30% but not with standard probability of 10%. The percentage differences between the standard and individual

deviant probabilities were large enough to elicit the MMN in the three previous situations but not in the latter situation. Decreasing the standard probability decreased clearly the amplitude of the MMN response but did not have a great effect on the latency. In addition to the MMN response, also the P3 response was evoked. These results with 6 subjects were not statistically analyzed because the number of subjects should be at least 12 in order to verify the results statistically on the group level.

Chapter 6

Conclusions and Future Work

Within this thesis a set of sounds, or a sound matrix, was generated by modifying the timbre of a recorded cello tone in different timbre dimensions. The sounds differ from the reference sound, or the middle sound of the matrix, with an equal step in perceptual sense. The equal psychoacoustic distances compared to the reference sound were evaluated through a subjective listening test. The sounds of the matrix were used as test stimuli in an MMN study.

The 30 lowest harmonic components were extracted from the original cello tone and the sounds of the sound matrix were modified by processing these harmonics. This was a suitable method for timbre modifications and resulted in natural-sounding tones. Although the sounds were reconstructed from the modified harmonics and their spectra were processed, they did not sound synthetic but they sounded quite natural. The method worked well in harmonic, brightness, and noise dimensions where timbre modifications were observable and the perceived differences between the sounds were large enough.

Also in the attack time dimension observable differences were obtained but they were not large enough. This reveals that the method of modifying the harmonic waveforms was not efficient enough. The linear envelope curve for modifying the attack time could be better optimised. Also the short duration of the attack may have affected so that the perceptual differences were not large and observable enough. However, in three dimensions the sounds with an equal psychoacoustic distance from the reference sound were successfully defined. By combining the different dimensions a three dimensional sound matrix with 27 sounds differing with equal steps in timbre space was obtained. Each sound has an equal fundamental frequency and duration. The sounds were also equalized for perceived loudness.

In the brain recordings with a standard tone (i.e., the reference sound) and 26 different deviants (i.e., the other sounds of the matrix) the effect of the standard probability on the MMN response was studied. It was observed that the MMN was elicited by the standard probabilities of 80%, 50%, and 30% but not with the standard probability of 10%. It can be concluded that decreasing the standard probability results to the decrease of the MMN amplitude but does not affect much the MMN latency. It can also be thought that in addition to the standard probability the elicitation of the MMN depends on the relationship between the probabilities of the standard and the individual deviants.

This thesis provides information about audio signal processing, timbre, evaluation of the psychoacoustic distances, and the MMN response. According to the results obtained within this work it is possible to construct natural-sounding tones with timbre modifications by processing the harmonic components. It is also possible to generate a set of sounds with an equal psychoacoustic distance from the reference. The results of the brain recordings provide information about the short-term auditory memory and automatic sound discrimination. The task of discriminating sounds automatically becomes more difficult when the probability of the repeated sound is decreased. This is because the memory trace of the standard gets stronger when the standard occurs more often. When the memory trace is stronger, it is easier to observe that the deviant differs from the standard. However, even if the majority of the incoming sounds differ from the repeated sound, a strong enough memory trace for this repeated sound can also be developed and different sounds can be discriminated from it. The repeated sound can be recognized and different sounds can be discriminated from it even with a quite low standard probability.

Typically the MMN has been studied by using pure sine waves as test stimuli. However, those stimuli do not sound very natural. Since the test stimuli used in these brain recordings sounded quite natural, these results are thought to provide more accurate information about the human cognition in everyday life than the results obtained by using pure sine waves.

The sounds and results obtained within this thesis provide a good basis for further studies. Both the method for extracting and modifying the harmonics and the method for evaluating psychoacoustic distances within the sounds will be published in international conferences [10, 33]. If the modification of the attack time had produced large enough differences within the sounds, a four dimensional sound matrix would have been obtained. This means that the number of different deviants would have been larger and the probabilities of the individual deviants would have been smaller. With a

larger sound set more accurate results about the effect of the standard and deviant probabilities could have been obtained. In the future, one challenge could be the generation of the fourth dimension to the sound matrix.

Although these results are quite accurate and reliable, a larger number of subjects both in the listening test and in the brain recordings would have provided a better reliability. Also the results obtained within the brain recordings could have been statistically verified if more subjects had participated in the recordings. In the future, this study will be continued. The number of subjects will be increased in order to obtain more reliable results and also to be able to analyze the results statistically.

In the future, the sounds of the sound matrix can be used in order to study also other features of the MMN, e.g., how the temporal probability (i.e., the probability that a deviant will occur within a given time period) affects the MMN response.

Bibliography

[1] G. Agostini, M. Longari, and E. Pollastri. Musical instrument timbres classification with spectral features. *EURASIP Journal on Applied Signal Processing*, 1:5–14, 2003.

[2] American Standards Association. *USA Standard Acoustical Terminology*. S1.1-1960, American Standards Association, New York, 1960.

[3] A. T. Cacace and D. J. McFarland. Quantifying signal-to-noise ratio of mismatch negativity in humans. *Neuroscience Letters*, 341:251–255, 2003.

[4] M. W. Eysenck and M. T. Keane. *Cognitive Psychology*. Psychology Press, 4th edition, 2000.

[5] E. B. Goldstein. *Sensation and Perception*. Wadsworth, 6th edition, 2002.

[6] J. M. Grey. Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5):1270–1277, May 1977.

[7] C. Hourdin, G. Charbonneau, and T. Moussa. A multidimensional scaling analysis of musical instruments' time-varying spectra. *Computer Music Journal*, 21(2):40–55, 1997.

[8] J. Hynninen and N. Zacharov. GuineaPig — A generic subjective test system for multichannel audio. *In Proceedings of the 106th Audio Engineering Society Convention*, Munich, Germany, May 1999.

[9] M. Ilmoniemi. *Calculation and Equalization of Loudness*. Special Assignment, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 2004.

[10] M. Ilmoniemi, V. Välimäki, and M. Huotilainen. Subjective evaluation of musical instrument timbre modifications. *Accepted for Publication in the Baltic-Nordic Acoustics Meeting*, Mariehamn, Åland, June 8–10, 2004.

[11] R. Ilmoniemi. *Aivojen rakenne ja toiminta*. Tfy–99.247 Course material, Helsinki University of Technology, Fall 2003.

[12] P. Iverson and C. L. Krumhansl. Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, 94(5):2595–2603, November 1993.

[13] M. Karjalainen. *Kommunikaatioakustiikka*. Teknillinen korkeakoulu, Otamedia Oy, 2000.

[14] J. B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29:115–129, 1964.

[15] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine. Splitting the unit delay — Tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 13(1):30–60, January 1996.

[16] J. S. Milton and J. C. Arnold. *Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences*. McGraw-Hill, 2nd edition, 1990.

[17] S. K. Mitra. *Digital Signal Processing: A Computer-Based Approach*. McGraw-Hill, 2001.

[18] B. C. J. Moore and B. R. Glasberg. A revision of Zwicker's loudness model. *Acustica — Acta Acustica*, 82:335–345, 1996.

[19] R. Näätänen. Mismatch negativity (MMN): perspectives for application. *International Journal of Psychophysiology*, 37:3–10, 1999.

[20] R. Näätänen, A. W. K. Gaillard, and S. Mäntysalo. Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42:313–329, 1978.

[21] R. Näätänen, A. Lehtokoski, M. Lennes, M. Cheour, M. Huotilainen, A. Iivonen, M. Vainio, P. Alku, R. J. Ilmoniemi, A. Luuk, J. Allik, J. Sinkkonen, and K. Alho. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(30):432–434, January 1997.

[22] R. Näätänen, M. Tervaniemi, E. Sussman, P. Paavilainen, and I. Winkler. 'Primitive intelligence' in the auditory cortex. *Trends in Neuroscience*, 24(5):283–288, May 2001.

[23] F. Opolko and J. Wapnick. *McGill University Master Samples*. McGill University, Montreal, Quebec, Canada, 1987.
http://www.music.mcgill.ca/resources/mums/html/ (March 30, 2004).

[24] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, R. Johnson, Jr. and G. A. Miller, W. Ritter, D. S. Ruchkin, M. D. Rugg, and M. J. Taylor. Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, 37(2):127–152, 2000.

[25] R. L. Pratt and P. E. Doak. A subjective rating scale for timbre. *Journal of Sound and Vibration*, 45:317–328, 1976.

[26] T. D. Rossing. *The Science of Sound*. Addison-Wesley Publishing Company, 2nd edition, 1990.

[27] J. Sinkkonen and M. Tervaniemi. Towards optimal recording and analysis of the mismatch negativity. *Audiology & Neuro-Otology*, 5:235–246, 2000.

[28] K. Steiglitz. *A Digital Signal Processing Primer with Applications to Digital Audio and Computer Music*. Addison-Wesley Publishing Company, 1996.

[29] E. Sussman, K. Sheridan, J. Kreuzer, and I. Winkler. Representation of the standard: Stimulus context effects on the process generating the mismatch negativity component of event-related brain potentials. *Psychophysiology*, 40:465–471, 2003.

[30] R. Takegata, P. Paavilainen, R. Näätänen, and I. Winkler. Preattentive processing of spectral, temporal, and structural characteristics of acoustic regularities: A mismatch negativity study. *Psychophysiology*, 38:92–98, 2001.

[31] M. Tervaniemi and M. Huotilainen. Neuroscience of music — methods and discoveries. In: M. Leman (Ed.) *Tendences, Perspectives, and Opportunities for Systematic (Cognitive) Musicology*. Submitted.

[32] P. Toiviainen, M. Tervaniemi, J. Louhivuori, M. Saher, M. Huotilainen, and R. Näätänen. Timbre similarity: Convergence of neural, behavioral, and computational approaches. *Music Perception*, 16(2):223–241, 1998.

[33] V. Välimäki, M. Ilmoniemi, and M. Huotilainen. Decomposition and modification of musical instrument sounds using a fractional delay allpass filter. *Accepted for Publication in the 6th Nordic Signal Processing Symposium*, Espoo, Finland, June 9–11, 2004.

[34] D. L. Wessel. Timbre space as a musical control structure. *Computer Music Journal*, 3(2):45–52, June 1979.

[35] I. Winkler, E. Schröger, and N. Cowan. The role of large-scale memory organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience*, 13(1):1–13, 2001.

Appendix A

Listening Test Results

In this appendix more detailed results of the listening test than in Section 4.3 are presented. In Figures A.1–A.6 each subject's averages of psychoacoustic distances with a confidence interval of 95% are presented separately for each dimension. In Figures A.7–A.10 the results of all subjects are presented separately for each dimension. In these figures S1–S6 denote the six subjects and indices correspond to those in Tables 4.1 and 4.2.

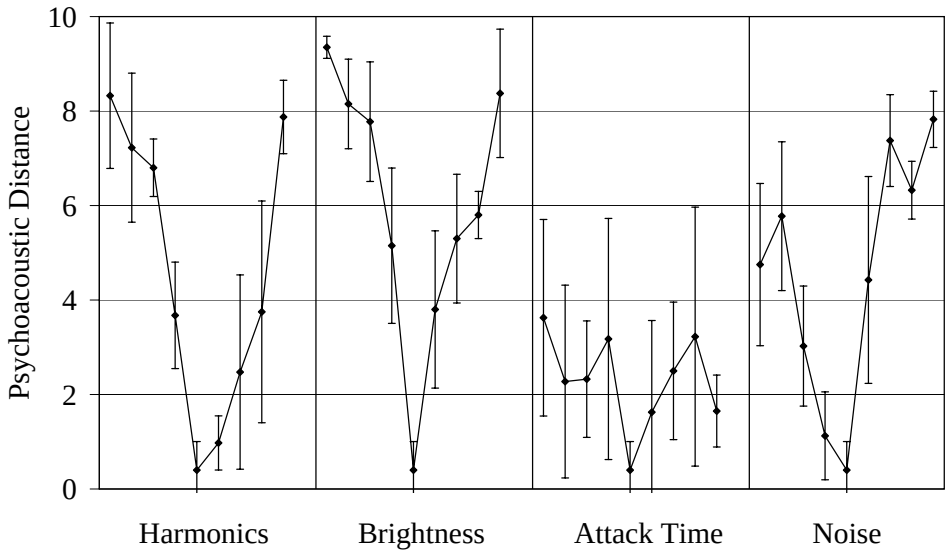

Figure A.1: Psychoacoustic distances of the sounds compared to the reference sound. Results of subject 1.

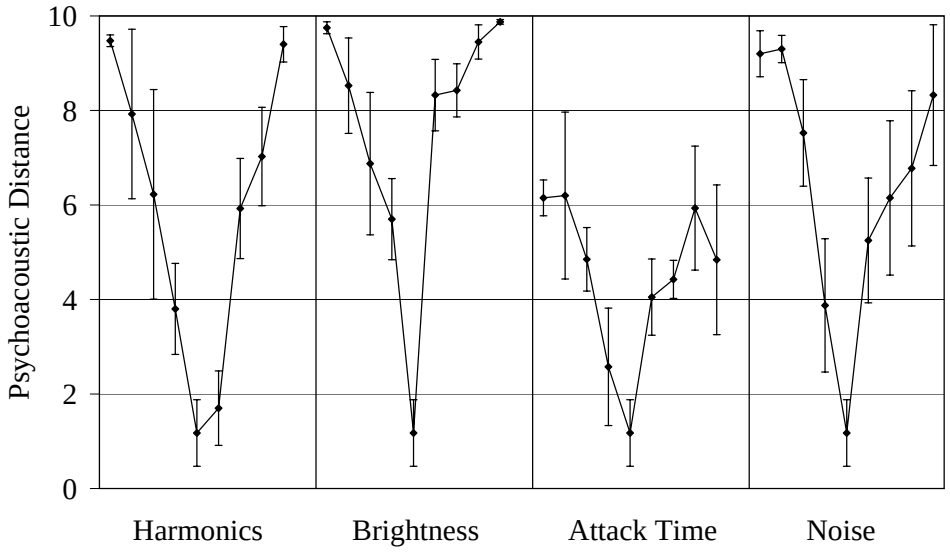

Figure A.2: Psychoacoustic distances of the sounds compared to the reference sound. Results of subject 2.

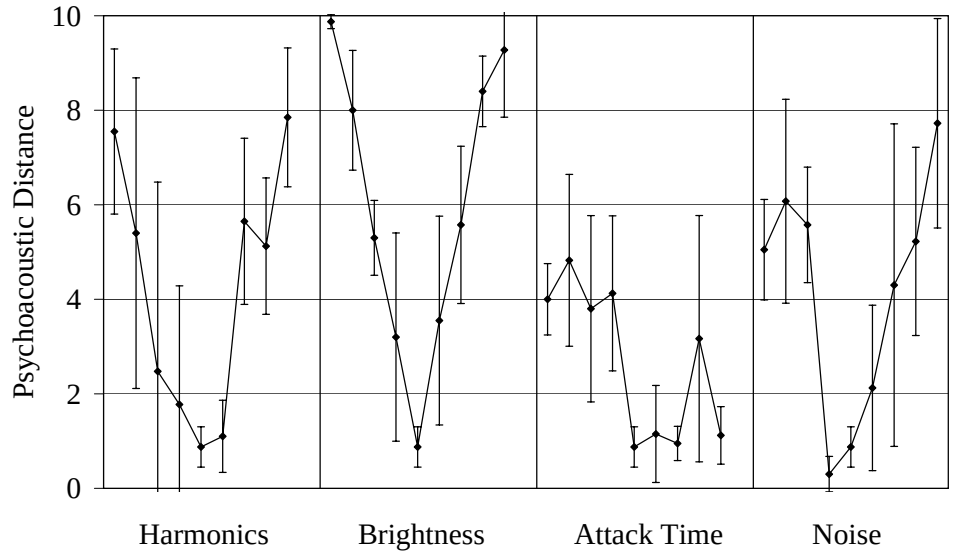

Figure A.3: Psychoacoustic distances of the sounds compared to the reference sound. Results of subject 3.

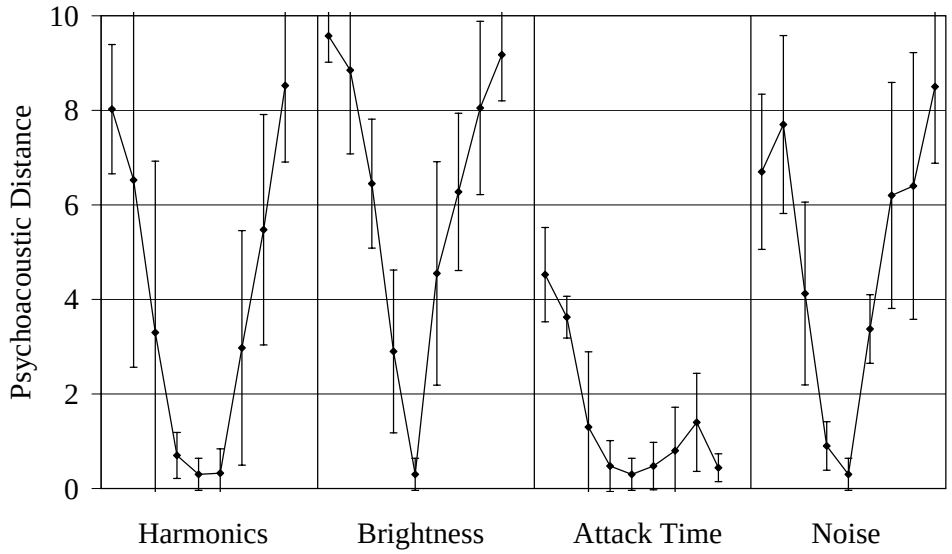

Figure A.4: Psychoacoustic distances of the sounds compared to the reference sound. Results of subject 4.

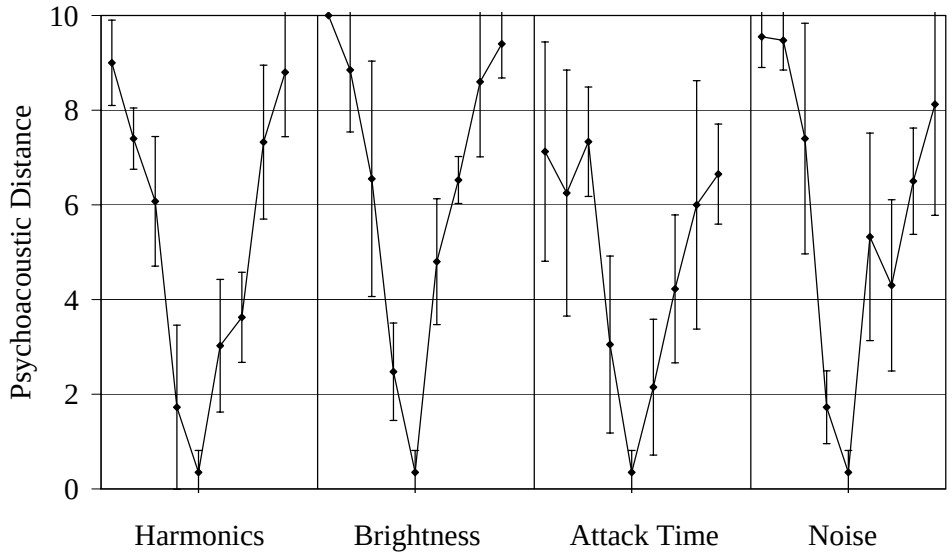

Figure A.5: Psychoacoustic distances of the sounds compared to the reference sound. Results of subject 5.

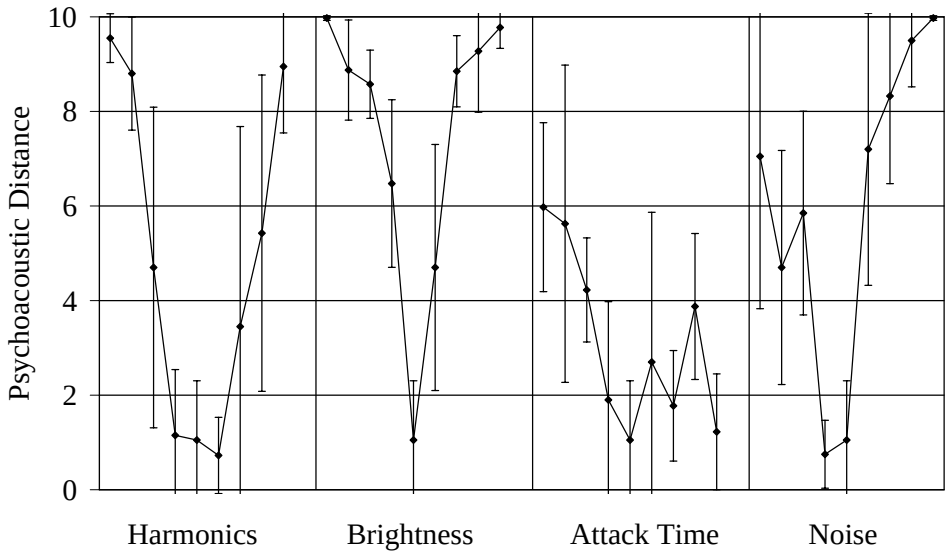

Figure A.6: Psychoacoustic distances of the sounds compared to the reference sound. Results of subject 6.

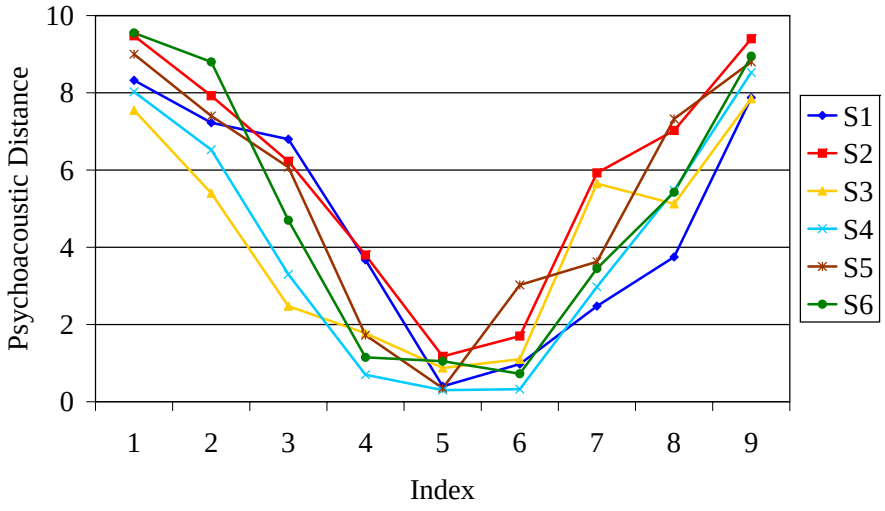

Figure A.7: Psychoacoustic distances of the sounds modified in harmonic dimension compared to the reference sound.

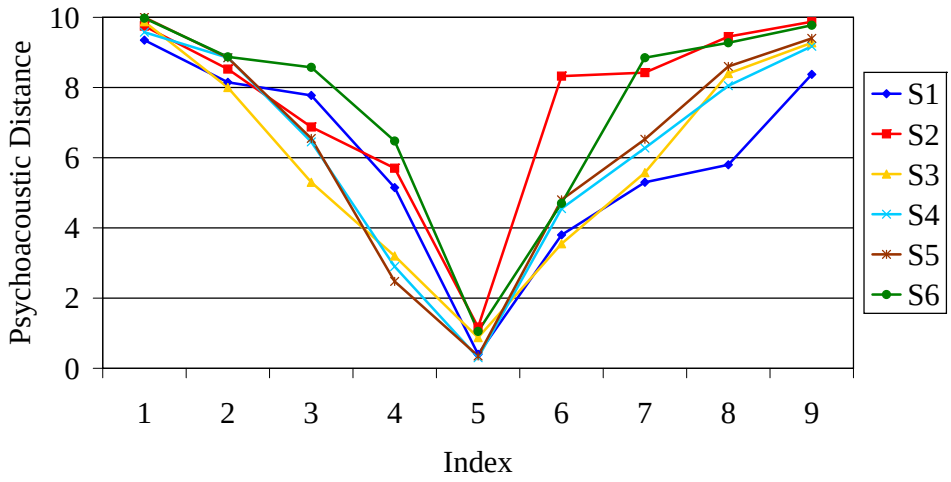

Figure A.8: Psychoacoustic distances of the sounds modified in brightness dimension compared to the reference sound.

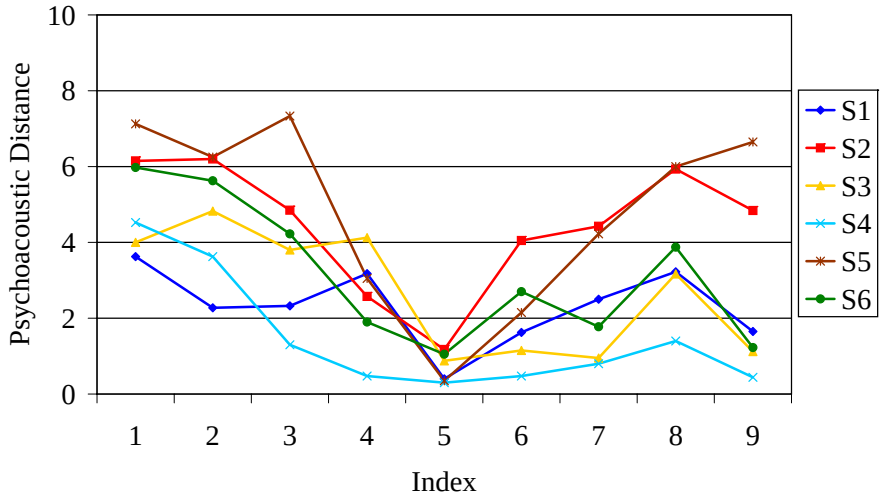

Figure A.9: Psychoacoustic distances of the sounds modified in attack time dimension compared to the reference sound.

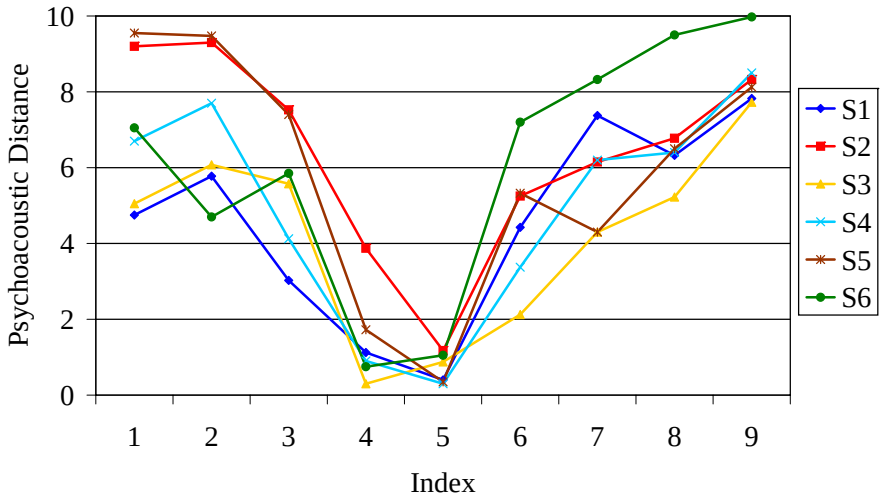

Figure A.10: Psychoacoustic distances of the sounds modified in noise dimension compared to the reference sound.

Appendix B

Comments on the Listening Test

In this appendix the oral comments on the listening test given by the subjects are presented. The subjects were interviewed after the test and their comments were written down.

Oral comments

- It was hard to create the scale for oneself.
- It was easy to evaluate differences in those sounds that had differences.
- Evaluation was hard because the sounds were so different.
- The sounds with small differences were easy to categorize but the sounds with larger differences were not.
- It was hard to compare the sounds because they were different in different ways.
- There were a lot of differences in sounds.
- Those sounds that had only small differences had to be listened to many times.
- Although there were instruction and training sessions in the beginning of the test it was hard to remember what ”totally different” means.
- Criteria for evaluation changed during the test because of learning and getting accustomed to the sounds.