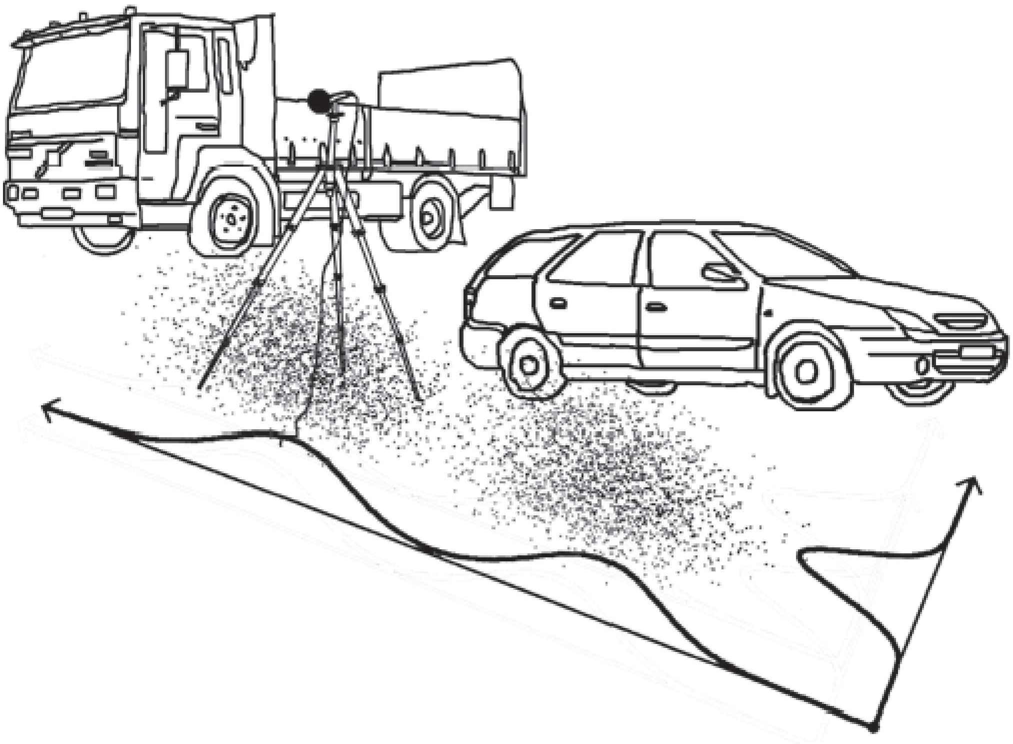
CHALMERS



Acoustic Traffic Classification using an Artificial Neural Network

Thesis for the degree of Master of Science in Engineering Physics

RASMUS ELOFSSON BERNSTEDT

Department of Civil and Environmental Engineering
Division of Applied Acoustics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2005

Master´s Thesis 2005:112

MASTER'S THESIS 2005:112

Acoustic Traffic Classification using an Artificial Neural Network

RASMUS ELOFSSON BERNSTEDT

Department of Civil and Environmental Engineering
Division of Applied Acoustics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2005

Göteborg, Sweden 2005

Acoustic Traffic Classification using an Artificial Neural Network

Master's Thesis 2005:112

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-41296 Göteborg
Sweden

Tel. +46 (0)31 772 1000

Cover: The outlines of a car and a truck along with an illustration of the distributions of their respective class in feature space (concept: Rebeka Hansson).

Reproservice / Department of Civil and Environmental Engineering
Göteborg, Sweden 2005

Acoustic Traffic Classification using an Artificial Neural Network
Master's Thesis in the Master's programme in Sound and Vibration
RASMUS ELOFSSON BERNSTEDT
Department of Civil and Environmental Engineering
Division of Applied Acoustics
Chalmers University of Technology

Abstract

Traffic noise and/or community noise can be measured with an unmanned measurement station which continuously records the sound pressure level (e.g. Symphonie measurement system). If wanted or needed, the measurement equipment can be configured to record all sounds exceeding a previously defined trigger level. For labeling or classification of the source type, from which the recorded sound originates, the recording must be listened to and manually classified. The desire to render this classification less time consuming suggests the development of an automatic method for sound source classification. In this thesis, the development of such a method is aimed at. The choice of an Artificial Neural Network as a classifier is motivated by its design model; the human brain and nervous system, and furthermore; the human ability to accurately distinguish between different sounds.

Sounds from heavy and light traffic (e.g. trucks and cars respectively) have been recorded, preprocessed and successfully classified. The preprocessing techniques used are filtering, resampling, signal modeling (ARMA-model) and Principal Components Analysis. The Neural Network employed for source type selection is a Multi Layer Perceptron with one hidden layer. One key issue is the extraction of features which defines and separates the different source types.

Method performance is validated by simulation of new measurements and classification thereof. The results show that the classification is 94 % accurate for the specific measurement situation. For assessment purposes, the performances of two reference methods are compared with the artificial classification. Manual classification of the recorded sounds was 96 % accurate and a method utilising the euclidean distance from new, unkown vehicles to the class average in feature space was 83 % accurate.

Keywords: traffic classification, artificial neural network, ARMA signal model, principal component analysis

Akustisk klassificering av trafik med ett artificiellt neuralt nätverk
Examensarbete inom Ljud och vibrationer
RASMUS ELOFSSON BERNSTEDT
Institutionen för bygg- och miljöteknik
Avdelningen för Teknisk Akustik
Chalmers tekniska högskola

Sammanfattning

Trafikbuller och/eller samhällsbuller kan idag mätas med en obemannad mätstation som registrerar ljudtrycksnivå kontinuerligt (t.ex. mätsystemet Symphonie). Om så behövs kan systemet konfigureras till att spela in ljud som överstiger en viss, i förväg inställd nivå. För att bestämma källan till de inspelade ljuden krävs att de avlyssnas och manuellt klassificeras. För att minska tidsåtgången vid denna klassificering skulle en metod för automatisk klassificering vara önskvärd. Målet med detta examensarbete är utvecklingen av en sådan metod. Valet av ett artificiellt neuralt nätverk som klassificeringsmetod motiveras av dess förlaga; den mänskliga hjärnan och dess förmåga att särskilja olika ljud.

Ljud från tung och lätt trafik (lastbilar respektive personbilar) har spelats in, dataförbehandlats och framgångsrikt klassificerats. Dataförbehandlingen inbegriper tekniker såsom filtrering, omsampling, signalmodellering (ARMA-modell) och principalkomponentanalys. Det neurala nät som använts för särskiljning av källtyp är en perceptron med ett dolt lager. En viktig aspekt är framtagningen av karaktäristiska egenskaper vilka kan definiera och åtskilja de olika källtyperna.

Metodens förmåga till korrekt klassificering är validerad med hjälp av simulerat nya mätningar som klassificeras. Resultaten visar att metoden är till 94 % korrekt för den situation i vilken mätningarna utförts. För att bedöma träffsäkerheten jämförs den artificiella klassificeringen med två referensmetoder. Manuell klassificering av de inspelade ljuden visar sig vara 96 % korrekt, och en metod som utnyttjar det euklidiska avståndet från nya, okända fordon till de olika klassernas egenskapsmedelvärde var till 83 % korrekt i klassificeringen.

Nyckelord: klassificering, trafik, artifiellt neuralt nätverk, ARMA signal modell, principal komponent analys

Preface

This Masters Thesis touches upon the acoustical areas of community noise, signal processing, vehicle noise and psycho acoustics. It is written at Ingemansson Technology AB, Gothenburg and the department of Applied Acoustics at Chalmers, Gothenburg. Swedish Road Administration *(Vägverket)* has decided to reward the author upon completion of the thesis.

Acknowledgments

I would hereby like to acknowledge the help and support I have received from the following people:

Georgios Natsiopoulos, my supervisor at Chalmers

Martin Almgren, my supervisor at Ingemansson Technology AB

Urban Hjorth, at the Dep. of Mathematical Statistics, Chalmers, for help with the statistical analysis

Josef Barsi, at Ingemansson Technology AB, for help with the measurement equipment

Göteborg, December 15, 2005

RASMUS ELOFSSON BERNSTEDT

CONTENTS

1 INTRODUCTION

Background

Road noise and/or community noise can currently be measured with an unmanned measurement station which continuously records the sound pressure level[1]. If wanted or needed, the measurement equipment can be configured to record all sounds exceeding a previously defined trigger level. These recorded sounds are subsequently listened to to determine the sound source type. This work is however rather time consuming and since the desire to reveal the source type exists, a quicker and cheaper method is strived for. Different types of source discrimination are requested such as recognition of different vehicles or recognition of non-vehicle sources.

The present method of recognising sources by listening involves human perception, memory and understanding the source behaviour. The fact that the mathematical models of neural networks are inspired by human perception and nervous system leads to the belief that acoustical classification can be achieved by such models. It is however not a aim of this thesis to create an identification method which resembles the human process in any qualitative way, meaning that no conclusions is to be made on the human process of classification based on the function of the neural network.

Aim of the Thesis

A method for sound source classification is to be developed. Preferably, many different approaches to this task are to be tested and evaluated. Due to the idea of using neural networks for classification, the choices of possible approaches is limited. The chosen framework for further work is illustrated in Figure 1.1. A recorded signal from some event is assumed to be available for further evaluation. The process of classification is then subdivided into i) pre-processing, ii) characteristics/feature extraction and iii) neural network classification. Feasibly, this entire process is ultimately fully automatic and implemented in different applications such as community noise measurement. Apart from previously mentioned aims, a quite simple method is desired.

Previous Work

Searching the scientific databases INSPEC and COMPENDEX for articles on sound classification and artificial neural networks one finds fairly many articles concerning

[1]Symphonie measurement system

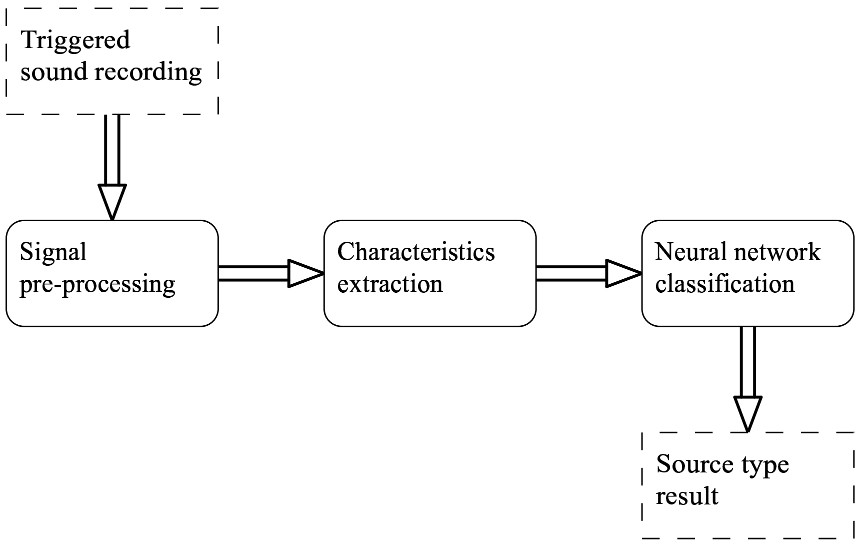


Figure 1.1. Flow diagram illustrating the method framework.

classification of, for instance, insect sounds, heart sounds, respiratory sounds, underwater sounds, earthquakes and nuclear bombs. One article on traffic classification written by A. Y. Nooralahiyan and H. R. Kirby at the University of Leeds is to be found.

Only a selection of articles within the realm of interest are studied with respect to their chosen methods and specifically the three steps illustrated in Figure 1.1. Regarding the neural network classification, most authors have chosen to use a Multi Layer Perceptron (MLP) as a neural network classifier (Nooralahiyan and Kirby 1998, Coggins and Principe 1998, Ham and Park 2002). The Multi Layer Perceptron is trained according to the back-propagation algorithm or some modified alternative thereof. Only one exception; a Gaussian classifier (Tzanetakis *et al.* 2001) and one addition; a Self-Organising Map (SOM) used for pre-processing (Coggins and Principe 1998) are found . SOM are also referred to as an unsupervised neural networks or Kohonen network.

Characteristics extraction techniques encountered in the selected articles include:

I) Fast Fourier Transform (FFT) magnitudes (Greene and Field 1991),

II) Cepstral coefficients (Ham and Park 2002),

III) Linear Predictive Coding (LPC) coefficients (Nooralahiyan and Kirby 1998),

IV) Discrete Wavelet Transform (DWT) coefficients (Tzanetakis *et al.* 2001).

LPC is also referred to as Auto Recursive (AR) model or all pole model.

Novelties

For feature extraction, AR or LPC coefficients are commonly used, not only by Reference (Nooralahiyan and Kirby 1998), but for various tasks of classification. In this thesis, the AR signal model is expanded to a ARMA (Auto Recursive Moving Average) signal model.

Some have used Principal Component Analysis (PCA) as a tool for condensing the amount of data, but the combined approach of Principal Component Analysis and a perceptron neural network has not been tested in the context of traffic classification. Moreover, the result of PCA is more thoroughly analysed in thesis and a method for sorting data by a quality measure is introduced. Any equivalent data sorting method has not been encountered in the mentioned articles.

Thesis Overview

Chapter 1 **Introduction**: introduces the reader to the research area by giving a background and some examples of work of others. The used methods are touched upon and how this work is different from the works of others.

Chapter 2 **Notations**: accounts for the most commonly used variables and notations used in the thesis.

Chapter 3 **Measurements**: describes the conducted measurements aiming to create a training set for the neural network.

Chapter 4 **Preprocessing**: deals with the task of rectifying and condensing the recorded data to a set of quantities with characteristic features which describes the source of the recorded signal sufficiently detailed to enable the subsequent classification.

Chapter 5 **Principal Component Analysis**: aims to describing and explaining PCA as a tool for data reorganisation. A method for information quality assessment is also described.

Chapter 6 **Artificial Neural Networks**: The algorithm for decision-making in the classification process is described along with training methods and chosen network design. Method performance is presented.

Chapter 7 **Listening Test**: describes a test aiming to provide a reference to the developed classification method.

Chapter 8 **Discussion**: The results are discussed and analysed relative to other peoples results and the performance of the reference methods.

Chapter 9 **Conclusion**: The conclusion briefly describes the problem, the methods used and the results, followed by the conclusions made on basis of the results.

2 NOTATIONS

Abbreviations

ANN	Artificial Neural Network
AR	Auto Recursive
ARMA	Auto Recursive Moving Average
HGV	Heavy Goods Vehicle
MA	Moving Average
MC	Motor Cycle
MLP	Multi Layer Perceptron
PCA	Principal Component Analysis
PV	Personal/Private Vehicle
SOM	Self Organising Map
UV	Utility Vehicle

Latin Abbreviations

cf.	confer
e.g.	exempli gratia
et al.	et alii
etc.	et cetera
i.e.	id est
N.B.	nota bene
No.	numero
q.v.	quod vide
vs.	versus

Capital Letters

B	Binomial distribution
F_s	Sampling frequency [Hz]
I	Identity matrix
L	Level [dB]
M	Number of individual observations
M	Mach number
N	Normal distribution
N	Number of nodes
O	Neural network output
T	Period time, signal length [s]
V	Projection / transformation matrix

Small Letters

f	Frequency [Hz]
i	The imaginary unit $i = \sqrt{-1}$
p	Number of feature vector elements
$w_{i,j}$	Synaptic weight

Greek Letters

α	Momentum term coefficient
β	Neural network noise (temperature) factor
Φ	Distribution function
ϕ	Neural network activation function
φ	Frequency function of a distribution
γ	Weight decay factor
η	Steepest descent step size
μ	Distribution expectancy value
θ	Neuron threshold level
ξ	Pattern; vector of characteristics
ζ	Neural network class key ("correct" answer)

Subscripts

$eq.$	equivalent
i, j, k	Neuron index

Superscripts

μ	Pattern number
n	Neural network layer index
T	Transpose

Diacritical marks

$\tilde{}$	transformed
$\vec{}$	vector
$\{\ \}$	vector
$\overline{}$	estimation
$\langle\ \rangle$	average
$\sim$	distributed
$\times$	matrix multiplication
$\cdot$	scalar multiplication

3 MEASUREMENTS

To provide useful data for further analysis, measurements have been conducted in three sessions. The purpose of collecting data is to setup a set of vehicles for training of the neural network. Circumstances and setup vary between the three sessions, and this is accounted for along with other conditions in the sections named by which month the measurements were done.

A measurement is a procedure of sampling in a set of sample points or a population. The set, in this case, consists of all possible vehicles that travel the roads. By choice of measurement location and date (point of time) the set of all vehicles can no longer be fully represented. The reader must bear in mind that all further analysis is constrained by this statistical discrepancy.

Moreover, the sampling is carried out such that it complies with the adopted framework as described in Chapter 1. The practical consequences of this framework, e.g. that a recorded sound from *one* event is available for analysis, further limits the possibilities of generalised sampling. Put differently, sound recordings of *two* vehicles passing the microphone simultaneously are *not* included in the sampled subset and the further analysis.

The following description shows the subsets chosen to define classes for the population of vehicles: During the measurement, vehicle class is visually identified and

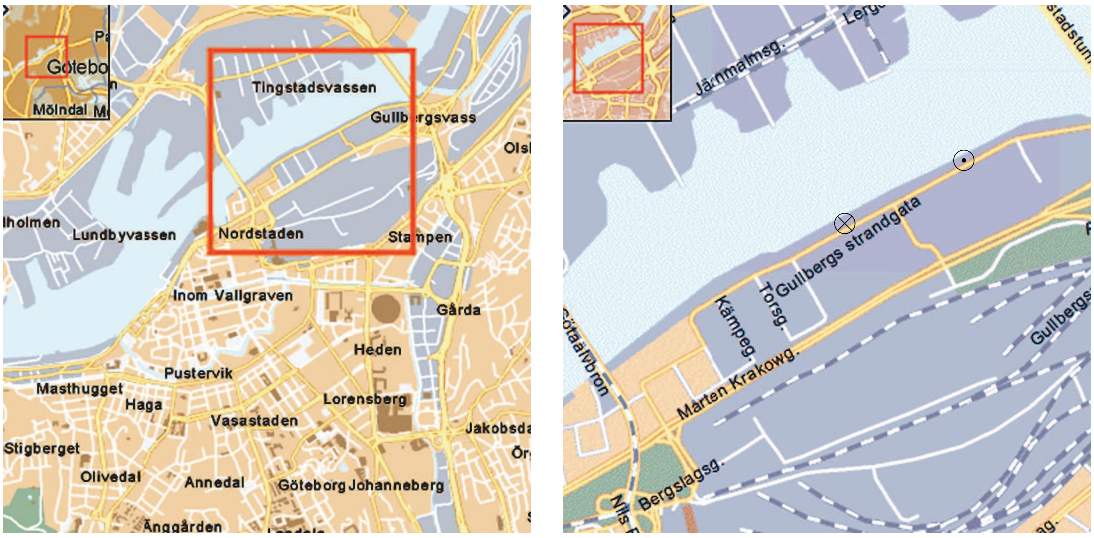
Class	*Description*
PV	Personal/Private Vehicle: small and medium size cars *(Swe: personbil)*.
UV	Utility Vehicle: vans, light trucks *(Swe: lätt lastbil)*.
HGV	Heavy Goods Vehicle: heavy trucks, tractors, buses *(Swe: tung lastbil)*.
MC	Motor Cycle: motor cycles, mopeds *(Swe: motorcykel)*.

noted with reference to the recording of the vehicle sound signal.

The difficulty of obtaining sufficiently many measurements of motor cycles and mopeds has led to the exclusion of the MC- class from classification attempt. Furthermore, UV- class is excluded due to the difficulty of manually classifying such vehicles during the actual measurements. The remaining two classes, PV and HGV, cover light and heavy vehicles respectively, with PV including small and medium size cars such as sedans and station wagons and HGV including all vehicles weighing more than 3.5 metric tons.

A map pointing out the location for the measurements is provided for in Figure 3.1 (Västtrafik 2005).

Table 3.1 shows the total result of the sampling. Data collection has yielded a useful set of 141 heavy vehicles and 141 cars.



(a) Large scale (b) Small scale

Figure 3.1. Measurement location, indicated by ⊗ (June) and ⊙ (August, September).

Class	*Samples*		
	June	August	September
PV	125	0	16
UV	63	0	0
HGV	31	47	63
MC	7	5	6

Table 3.1. Sampled population

There are many variables in a measurement situation that can be considered, for instance weather, distances and traffic situation. The measurements conducted for this thesis however, does not aim to control all possible variables, but rather to assimilate differing variables into the classification method with the intention of achieving a more robust method than what might be the case, where all variables fully controlled. In other words, the classification method is more likely applicable to a “real life” measurement situation if all recorded signals are not from vehicles traveling by the same speed.

A Note on the Equipment

Instruments are calibrated according to the Ingemansson quality standards which comply with the demands stated in SS-EN ISO/IEC 17025. Dates for the latest calibrations are listed in Ingemansson’s calibration log.

3.1 Measurement Details

Measurements June 2005

Dates:	2005-06-07	2005-06-20	2005-06-22
Weather:	14° C, NW 5 m/s	20° C, S 6 m/s	19° C, W 6 m/s
Trigger level:	70-75 dB(A)		
Recorded sample length:	4-6 s		
Sampling frequency:	51.2 kHz		
Location:	Gullbergs strandgata, q.v. Figure 3.1		

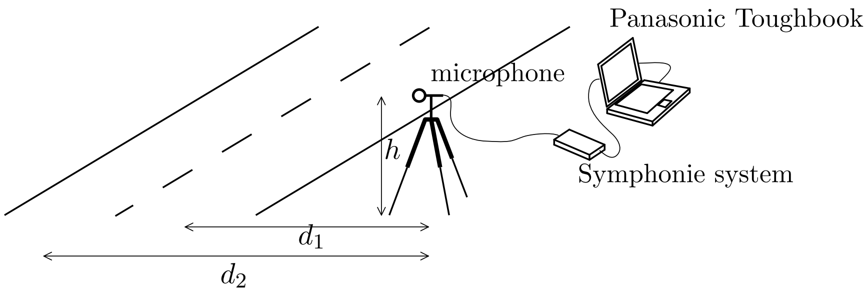
Equipment and Setup

The measurement system (Symphonie) used for data collection is the very same as the one intended for implementation of the classification method. The symphonie measurement system registers sound pressure level over a long strech of time. It is configured to record sound when the level exceeds a predefined trigger level. The system buffers the sound signal continuously, which allows for collecting data before and after an event is established by exceeding the trigger level.

Item description	*Manufacturer*	*Type*	*Internal notation*
Symphonie measurement system	Spektrum GmbH, 01dB		AL134
Microphone	G.R.A.S.	26AF	MK070
Rugged notebook	Panasonic Toughbook		D016

Table 3.2. Equipment, measurements June 2005

(a) Photograph of actual setup



(b) Schematic picture of the setup

Figure 3.2. Measurement Setup, June 2005

microphone height, h:	1.1 m ± 10 %
distance to closest lane, d_1:	2.6 m ± 10 %
distance to farthest lane, d_2:	7.4 m ± 10 %

Measurements August 2005

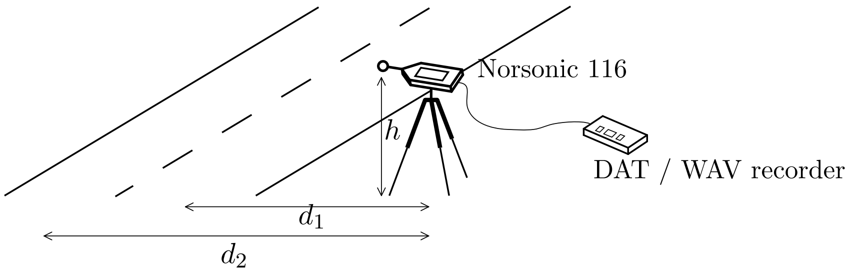
Dates:	2005-08-23	2005-08-24
Weather:	19–22° C, SE 2 m/s	17–20° C, S–SW 1-3 m/s
Trigger:	manual trigger	
Recorded sample length:	2-6 s	
Sampling frequency:	44.1 kHz	
Location:	Gullbergs strandgata, q.v. Figure 3.1	

Equipment and Setup

Item description	*Manufacturer*	*Type*	*Internal notation*
DAT recorder	Sony		B066
Microphone calibrator	Brüel & Kjær	4231	KU051
Sound Level Meter	Norsonic	116	LM058

Table 3.3. Equipment, measurements August 2005

(a) Photograph of actual setup



(b) Schematic picture of the setup

Figure 3.3. Measurement Setup, August 2005

microphone height, h:	1 m ± 10 %
distance to closest lane, d_1:	2.6 m ± 10 %
distance to farthest lane, d_2:	7.4 m ± 10 %

Measurements September 2005

Dates:	2005-09-05	2005-09-06
Weather:	17–20° C, SW 1–3 m/s	20-22° C, S 2–4 m/s
Trigger:	manual trigger	
Recorded sample length:	2-6 s	
Sampling frequency:	44.1 kHz	
Location:	Gullbergs strandgata, q.v. Figure 3.1	

Equipment and Setup

Item description	*Manufacturer*	*Type*	*Internal notation*
WAV/MP3 recorder	EDIROL	R-1	B072
Microphone calibrator	Brüel & Kjær	4231	KU047
Sound Level Meter	Norsonic	116	LM058

Table 3.4. Equipment, measurements September 2005

Setup for measurements in September is equivalent to the setup in August as seen in Figure 3.3, only the DAT recorder is replaced by the WAV/MP3 recorder.

3.2 The Acoustics of a Moving Source

The behaviour of a moving source is different from a stationary one. Firstly, since sound pressure is inversely proportional to distance, a moving source, or vehicle, will be perceived as sounding more strongly at close range than at short range. Secondly, the speed of the source relative to the receiver affects the sound signal envelope and amplitude as well as the sound frequency. The latter is called Doppler effect [1] and will not be considered in the following analysis.

Given the conditions illustrated in Figure 3.4: a source at position $\vec{x}_s$ moving along the x_1 axis at speed $\vec{v}_s$ and a receiver at position $\vec{x}_r$, the perceived sound signal at the receiver position depends on the angle θ, the distance d according to:

$$p\left(\vec{x}_r, t\right) = \frac{A(\tau)}{4\pi \, r(\tau) \left|1 - M_s \cos(\theta(\tau))\right|} \tag{3.1}$$

$$M_s = \frac{\vec{v}_s}{c} \tag{3.2}$$

$$r(\tau) = \left|\vec{x}_r - \vec{x}_s(\tau)\right| = c \cdot (t - \tau) \tag{3.3}$$

$$\sin(\theta(\tau)) = \frac{d}{r(\tau)} \tag{3.4}$$

where c is the speed of sound, $A(\tau)$ is the signal amplitude at source time τ, M_s is the source Mach number as defined in Equation (3.2) and $r(\tau)$ fulfills the requirements of Equation (3.3) (Ehrenfried 2004).

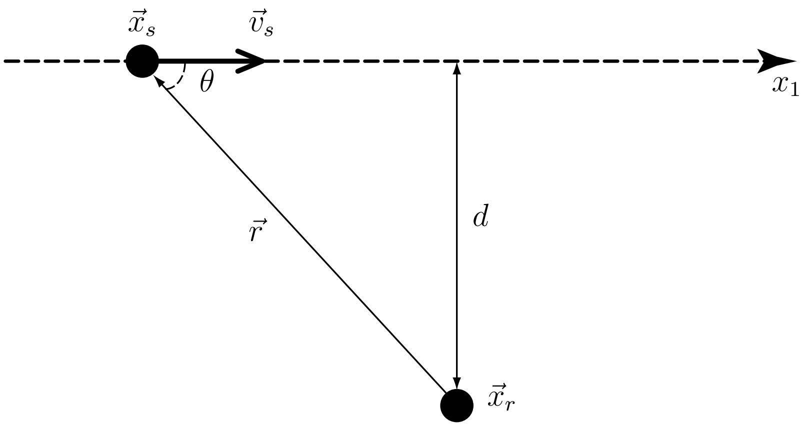
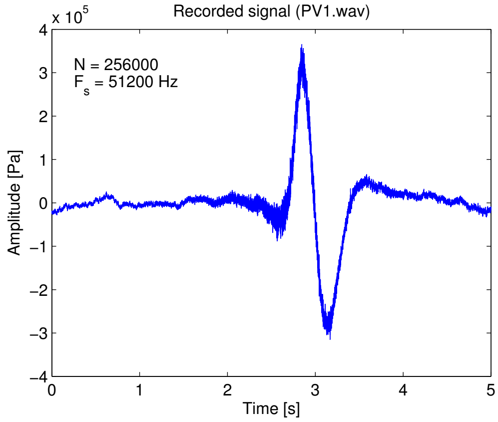


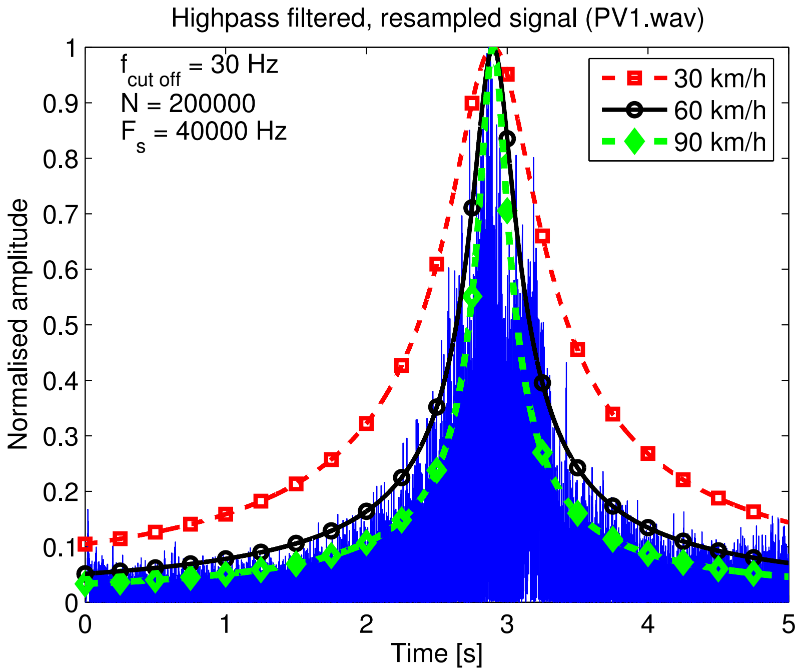
Figure 3.4. Illustration for the acoustics of a moving source.

Figure 3.5(a) shows a recorded signal (*PV1.wav*) from a light vehicle of class *PV* and 3.5(b) shows the theoretical signal amplitude according to Equation (3.1) for three different speeds; 30, 60 and 90 km/h.

[1] Named after its discoverer, the Austrian mathematician Christian Doppler.



(a) Signal



(b) Signal envelope

Figure 3.5. A recorded signal and signal envelope for three different vehicle velocities according to Equation (3.1)

Plotted in Figure 3.5(b) is the absolute value of the highpass filtered recorded recorded signal. The reason for filtering is to eliminate the effects on signal appearance of the overpressure in front of the vehicle and the underpressure after it on the appearance of the signal amplitude.

The 60 km/h envelope displays a good resemblance to the recorded signal and it is likely that the vehicle traveled at approximately 60 km/h.

The characteristics of a traveling sound source in terms of source speed influences the human perception of the sound signal and therefore it is also likely to influence the possibility or performance of an artificial classification method. The choice to disregard variables such as source speed renders further analysis of sound properties in terms of source speed impossible. For this reason, the slightly varying speeds of the vehicles at the measurement location is looked upon as a inherent variation in the subset of recorded vehicle sounds.

4 PREPROCESSING

A major area of concern is the preprocessing of data and especially the means of producing a very limited amount of data for each individual vehicle and still preserve its characteristics. This chapter is focused on the basic signal processing tools used in the classification method. Apart from reducing the amount of data for each individual, the aim is to attain a general framework by which signals of new and unknown vehicles can represented. In other words, trifles such as a sampling frequency chosen differently than those of the method should not affect the performance of the classification method.

4.1 Filtering and Resampling

Firstly, it is assumed that a sound event has occurred (a vehicle passing by) and that it has been recorded. Such a signal is illustrated in Figure 4.1(a), and it is measured according to circumstances described in the Measurement chapter (q.v. page 7).

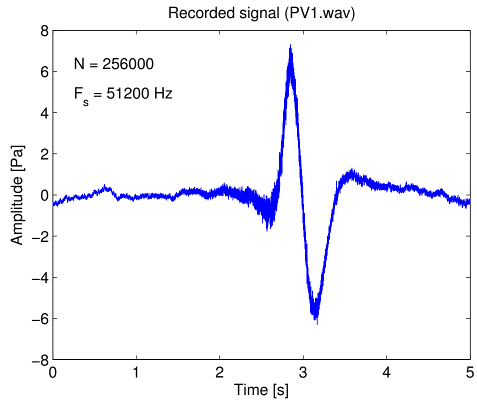
The signal y is a function of time $y = y(t)$ and the sampled signal can be represented as a discrete vector, such that y_n is the function value at time t_n for $n = 1, 2, ..., N$. Sampling frequency and recording length, T determines the vector length $N = F_s \times T$.

The acquired signal is then bandpass filtered with cut-off frequencies $f_{low} = 100$ Hz and $f_{high} = 8000$ Hz and resampled with sampling frequency $F_s = 8$ kHz; Figure 4.1(b). For this purpose a Finite Impulse Response, FIR filter is used. Matlab function *fir1.m*, which implements a classical method of windowed linear-phase FIR digital filter design (Mat 2004), is used to produce filter coefficients. The response length of the filter is set to $2^{11} = 2048$ samples. The number of data points in this case is reduced from $N = 256000$ to $N = 40000$.

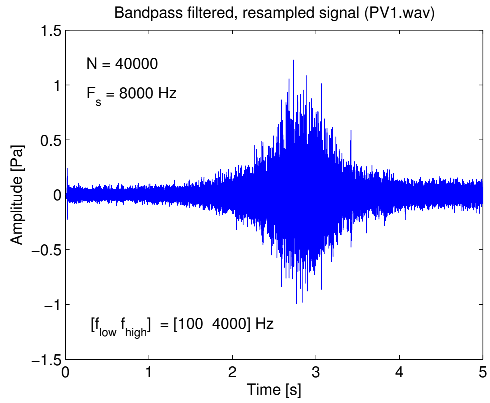
Signal Spectrum and Information Content

To confirm the choice of filter cut-off frequencies average spectra of PV- and HGV vehicles are calculated and plotted in Figure 4.2. Signal spectra in Figure 4.2(a) are presented in standardised 1/3 octave bands[1] and plotted versus a logarithmic scale. In Figure 4.2(b), the signal spectrum is A-weighted[2] and the frequency axis logarithmic.

[1] Upper and lower frequency limits from Reference (Bodén 2001)
[2] Corrections for A-weighting from Reference (Fahy 2001)



(a) Recorded signal

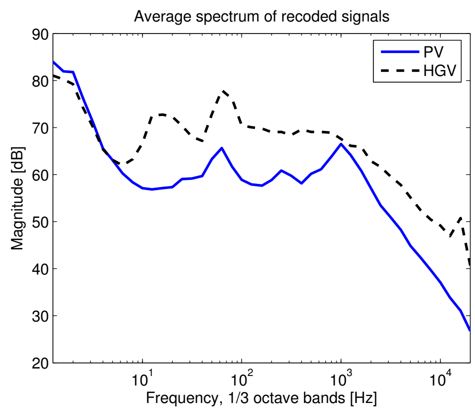


(b) Bandpass filtered, resampled signal

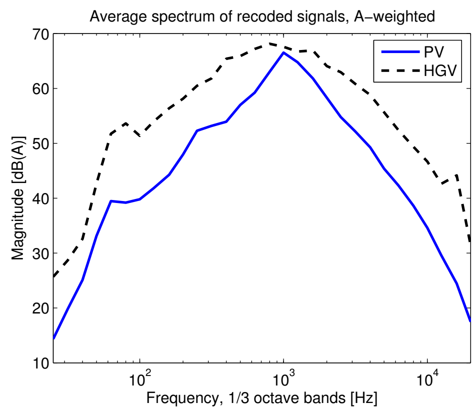
Figure 4.1. Recorded signal and preprocessing; bandpass filtering and resampling.

Previous work including measurements on "coast by"[3] vehicles show that the major part of the sound energy is located between frequencies 100 Hz and 3 kHz in the frequency domain (Andersson 2005).

It is believed that very little useful information can be extracted from the signal for frequencies larger than 4 kHz. The properties of the human ear, with its decreased perceptibility for lower frequencies, and the ability of humans to distinguish between cars and trucks implies that vehicle classification is possible without signal information below a certain frequency. Hence, the lower bound for information extraction is chosen to 100 Hz.



(a) 1/3 octave bands



(b) Linear bands

Figure 4.2. Average spectrum of recorded signals for PV- and HGV classes.

[3]Cars traveling with the engine turned off; all sounds generated by the tyres.

4.2 Characteristics extraction

Although the signal has been resampled and now contains a lesser number of data points than the original signal, data for a single vehicle is still too large to be effectively used in further computations, i.e. the neural network classification. The aim now is to find means of representing the signal in an adequate form. Most importantly, it is imperative that this form of signal representation preserves any characteristics which defines and separates the chosen vehicle classes (definitions of vehicle classes, q.v. Chapter 3).

Defining an Event

As described in the Introduction, Chapter 1, a prerequisite for further evaluation is a recorded signal of some event.

Even though an event can be established through triggering at predefined sound pressure level[4], the more well defined event of the precise passage of the vehicle is wanted. This event is taken to be some time before and after the sound pressure level is at its maximum value. To obtain the time of maximum SPL, a sliding average is applied to the signal according to

$$y_n = \frac{1}{2k+1} \sum_{i=n-k}^{n+k} y_i \ , \ n = k+1, ..., N-k \qquad (4.1)$$

in which the sliding average length is set to $k = \frac{N}{100}$. Signal data are collected both before and after the signal envelope maximum and treated separately: signal parts Δy_1 and Δy_2 in Figure 4.3. In this case Δy_1 and Δy_2 consists of $N = 2^{13} = 8192$ data points each. The dashed, vertical lines in Figure 4.3 defines the selected data. With a sampling frequency of 8 kHz, each of the two signal portions has a duration of $\Delta t = 1.0240$ seconds.

Signal Model

Linear Predictive Coding (LPC) coefficients, also referred to as Auto Recursive or AR coefficients have been successfully used by Reference (Nooralahiyan and Kirby 1998) for the purpose of acoustical classification. In the context of signal modeling, an Auto Recursive process uses N signal samples to predict the subsequent sample number $N+1$. In terms of frequency analysis and filter design, the AR coefficients are referred to as a spectrum estimator or an all-pole model.

A more general approach to signal modeling than an AR- process is the Auto Recursive Moving Average or ARMA- process. Apart from modeling all poles of a

[4]Symphonie measurement system (described more thoroughly in the Measurements chapter on page 9)

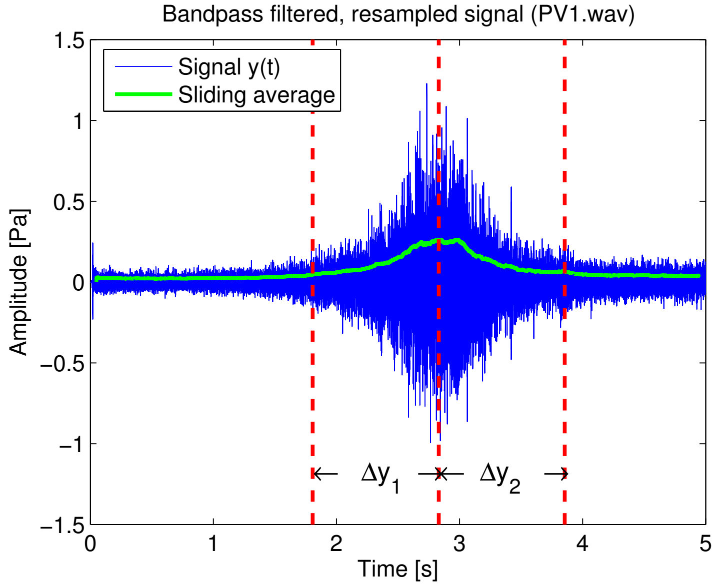


Figure 4.3. Sliding average, as defined in Equation 4.1, applied to signal to find signal envelope maximum.

system or signal, the ARMA process includes a model of the zeros. The ARMA approach yields a better model approximation to the signal than its components AR and MA (Hayes 1996).

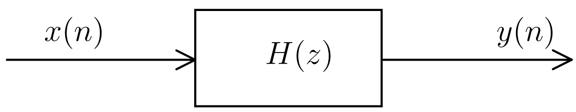


Figure 4.4. System representation of an ARMA model.

The relation between the input, x, and output, y, variables for an ARMA process is defined as:

$$y(n) = -\sum_{k=1}^{p} a_p(k)\, y(n-k) \; + \sum_{k=0}^{q} b_q(k)\, x(n-k) \tag{4.2}$$

for discrete time steps t_n, $\{n : t_{n+1} > t_n\}$ [5]

Applying the z transform to Equation (4.2), the system transfer function in Figure 4.4 can be rewritten as:

$$H(z) = \sum_{n=0}^{\infty} h(n) z^{-n} = \frac{\sum_{n=0}^{q} b_q(n) z^{-n}}{\sum_{n=0}^{p} a_p(n) z^{-n}} = \frac{B(z)}{A(z)} \tag{4.3}$$

The spectrum of the signal model is obtained by setting $z = e^{i\omega}$ in Equation (4.3),

[5]Implies causality.

which yields:

$$H(\omega) = \sum_{n=0}^{\infty} h(n) e^{-in\omega} \tag{4.4}$$

Prony's Method

To calculate the signal coefficients $a_p(k)$ and $b_q(k)$ in Equation (4.2) numerous methods have been developed, e.g. Least Mean Square, LMS- method and the Padé Approximation. Prony's method[6] to calculate the poles of the system, $a_p(k)$, involves solving a set of linear equations which satisfy the Least Mean Square error criterion. Once the poles have been calculated, the zeros, $b_q(k)$, are found by setting the residue error to zero for q samples. Further details on Prony's methods and more is found in References (Parks and Burrus 1987) and (Hayes 1996). The algorithm for calculating ARMA coefficients is taken from Reference (Hayes 1996).

ARMA- model of signal

After defining the two signal parts Δy_1 and Δy_2 one ARMA model for each part is calculated using Prony's method. For this application, it is found practical to model the AR process with 32 poles, i.e. 32 coefficients and the MA process with 16 zeros, and a amplification factor, i.e. 49 feature vector elements for each of the signal parts. In total, each of the two signal parts is modeled with a ARMA(48) model. Using the notation in Section 4.2: $p = 32$ and $q = 16$.

To validate the signal model, the frequency response is superimposed onto the spectrum of each signal part, see Figure 4.5. The models show a generally good agreement to the signal spectrum. Whether the details in the model frequency response are related to actual physical quantities or not will not be evaluated.

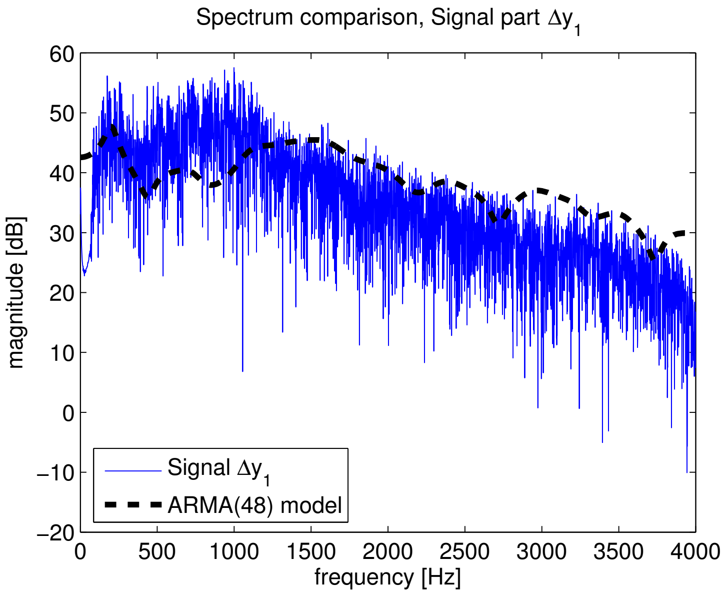
The two ARMA(48) models are now combined to a single vector along with the ratio $L_{max}/L_{eq.}$ for each recording to create a feature vector, ξ, for every vehicle:

$$\xi = \left[\ \{a_1\}_k \quad \{b_1\}_l \quad \{a_2\}_k \quad \{b_2\}_l \quad L_{max}/L_{eq.} \ \right]^T \tag{4.5}$$

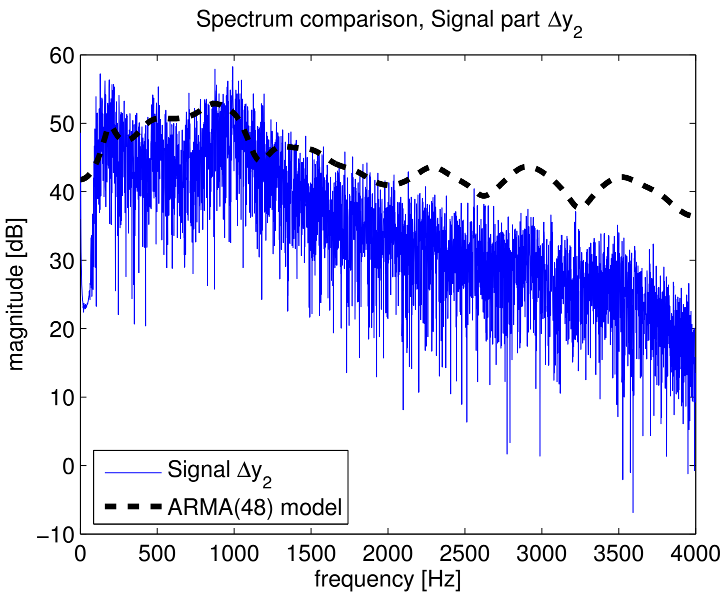
in which subscript 1 denotes signal part 1, subscript 2 denotes signal part 2, $k = 1, 2, \ldots, p$ and $l = 1, 2, \ldots, q$.

The reason for including the ratio $L_{max}/L_{eq.}$ and not the actual factors L_{max} and $L_{eq.}$ is to acquire a method fairly independent of the distance between the source (vehicle) and the receiver (microphone or recording system). Based on Equations (3.1) to (3.4), one can say the ratio $L_{max}/L_{eq.}$ is a measure of source speed.

[6]In 1790, Prony derived formulations for the analysis of elastic properties in gases which produced linear equations (Parks and Burrus 1987)



(a) Part 1



(b) Part 2

Figure 4.5. Comparison between the spectrum of the recorded signal *PV1.wav* and the ARMA model spectrum for the two parts Δy_1 and Δy_2.

5 PRINCIPAL COMPONENT ANALYSIS

The purpose of Principal Component Analysis, PCA, is to identify the most significant information within a set and extract a lesser amount of data, which still contains a greater part of the information. Any set of statistically sampled data is distributed around some ideal or theoretical value. Apart from the underlying physical process this distribution depends on, for instance, sampling noise, natural occurring variance within the set and most importantly, the variance due to true variations between subsets or groups. The prerequisite for PCA is a matrix of individually sampled vectors or signals (also referred to as feature vectors or patterns), which defines the sampled set, or population. In this case, the matrix consists of vectors of ARMA coefficients calculated using Prony's method and the ratio $L_{max}/L_{eq.}$, one vector for every recorded vehicle.

Figure 5.1. PCA illustration.

Figure 5.1 shows an illustration of principal component analysis. OA is the first principal component direction of the distribution that generated the cloud of points. The projection onto OA shows up more structure than the projection onto the other direction OB (Hertz 1991)

5.1 PCA Calculation

Principal Component Analysis relies on eigenvalue and eigenvector decomposition of a matrix and it is used to project a signal or vector onto a new basis. PCA is computed through a Singular Value Decomposition, SVD. A matrix, X, size $m \times n$, $m > n$ with m rows corresponding to observed variables and n columns corresponding to individual observations. The first step of PCA is to subtract variable mean from all individual observations in the following manner:

$$\vec{x}_i \rightarrow \vec{x}_i - \frac{1}{n}\sum_{j=1}^{n} x_{i,j} \tag{5.1}$$

where $x_{i,j}$ are the elements of X and $\vec{x}_i$ is vector of equally computed or selected variables (ARMA coefficients) in X for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Put differently, the feature vectors are transformed from

$$\vec{x}_i \sim N(\vec{\mu}, \vec{\sigma}^2) \quad \rightarrow \quad \vec{x}_i \sim N(0, \vec{\sigma}^2) \tag{5.2}$$

In terms of signal analysis, the bias or offset is removed.

Now, the transformed matrix X is decomposed using a Singular Value Decomposition, such that

$$X = U \times S \times V^T \tag{5.3}$$

Here, U is an $m \times n$ matrix and V is an $n \times n$ square matrix, both of which have orthogonal columns so that

$$U^T \times U = V^T \times V = I \tag{5.4}$$

in which I is the identity matrix and S is an $n \times n$ diagonal matrix consisting of the eigenvalues of X in the order of descending magnitude. The PCA transformation of matrix X is obtained by either multiplying $U \times S$ or $X \times V$.

The projection $\tilde{X} = X \times V$ can subsequently be truncated by removing the rows of $\tilde{X}$ starting with the last one; $\tilde{X}^{m \times n} \rightarrow \tilde{X}^{m_t \times n}$, $m_t < m$. Through this, information is conserved to a large extent. This truncation is equivalent to removing the smallest eigenvalues in S firstly and then multiplying $U \times S$. The matrix V is referred to as the transformation matrix.

The projected and truncated matrix $\tilde{X}$ provides the best linear projection onto a subspace in terms of preserving the signal energy. Preserving signal energy is equivalent to preserving the largest fluctuations of the signal, or in other words: the variance.

5.2 Performing PCA

For each recorded vehicle, one feature vector or pattern, ξ^μ, is compiled as described in the Preprocessing chapter. Every pattern has $p = 99$ elements and it consists of two ARMA(48) models and the ratio $L_{max}/L_{eq.}$ according to

$$\xi = \left[\ \{a_1\}_k \quad \{b_1\}_l \quad \{a_2\}_k \quad \{b_2\}_l \quad L_{max}/L_{eq.}\ \right]^T \tag{5.5}$$

in which subscript 1 denotes signal part 1, subscript 2 denotes signal part 2, $k = 1, 2, \ldots, 32$ and $l = 1, 2, \ldots, 17$. All M patterns are put into the matrix A according to:

$$A = \begin{bmatrix} | & | & & | & & | \\ \xi^1 & \xi^2 & \cdots & \xi^\mu & \cdots & \xi^M \\ | & | & & | & & | \end{bmatrix}^{p \times M} \tag{5.6}$$

Performing PCA on A, the projection $\tilde{A}$ is obtained. For visualisation purposes, two components or dimensions of the projected matrix $\tilde{A}$ are plotted in a two dimensional scatter plot in Figure 5.2. Each point (x, y) is determined by projected feature vector elements $(x, y) = (\xi_\alpha, \xi_\beta)$, where α and β are two chosen dimensions 1 and 2.

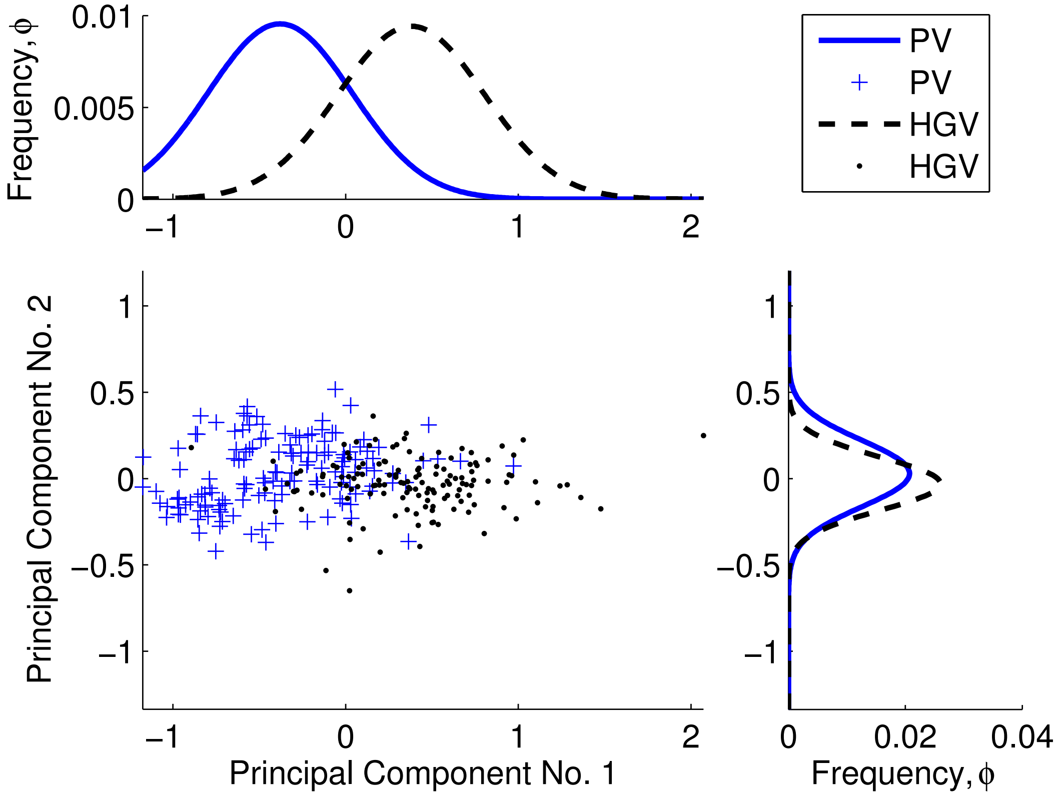


Figure 5.2. Scatter plot of the two first principal dimensions.

Figure 5.2 clearly shows that individual patterns of the two different classes group and form clusters. The plots above and to the left of the scatter plot show the distribution of the two vehicle classes for each of the PCA dimensions. Comparing the two principal dimensions, the two distributions are more separated for principal component number 1 than for number 2. The overlap of one distribution onto the other will make the classification process more difficult later in the process.

Were the two distributions in the first PCA dimension completely separated, i.e. no overlap, perfect classification would be possible just by separating the classes with a simple lie or plane. The intention with PCA is to obtain as much separation as possible, but to achieve this the global variation is maximised. This means that even though the data points of the two classes in one PCA dimension contain much information (widely spread), their distributions could overlap totally rendering this particular PCA dimension useless for identification purposes. A large information content, large global variation or energy, does not automatically provide information which defines and separates the two vehicle classes.

To assess whether a principal component might be useful in the classification process, the distribution overlap is integrated over and compared to the overlaps of the other components. Figure 5.3 shows the overlap for one of the PCA dimensions and the filled regions in the plot defines the overlap thereof.

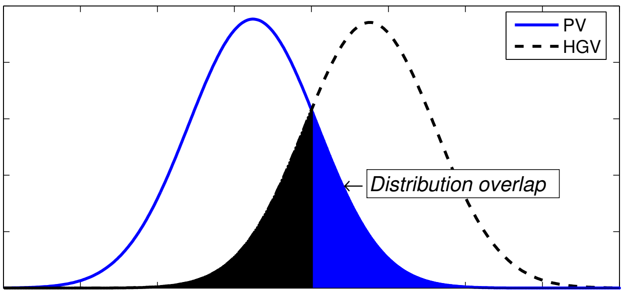


Figure 5.3. Illustration of the distribution overlap for one of the PCA dimensions.

Figure 5.4 shows the distribution overlap for all PCA components, and it shows that for many of the PCA dimensions, data from the two vehicle classes overlap substantionally. Also plotted is the overlap for the PCA dimensions, sorted by increasing overlap (dashed line). Henceforth, the pattern components are ordered by increasing overlap.

5.3 PCA Convergence

A reqirement for the successful application of PCA is the ability of generalisation, which means that the projection obtained by the Singular Value Decomposition must be applicable for patterns of ARMA coefficients not included in the original decomposition. Also the mean value as given in Equation (5.1) must converge to some quantity valid for the entire subset of vehicles.

To ensure PCA is a valid method for generalisation, the set of recorded vehicles is divided into one training set and one validation set. The training set is now utilised for the PCA decomposition and the validation set is used to ensure the validity of the projection.

Figure 5.5 displays a scatter plot for which 75 % of the patterns are used for the decomposition, and the remaining 25 % are projected onto the principal components

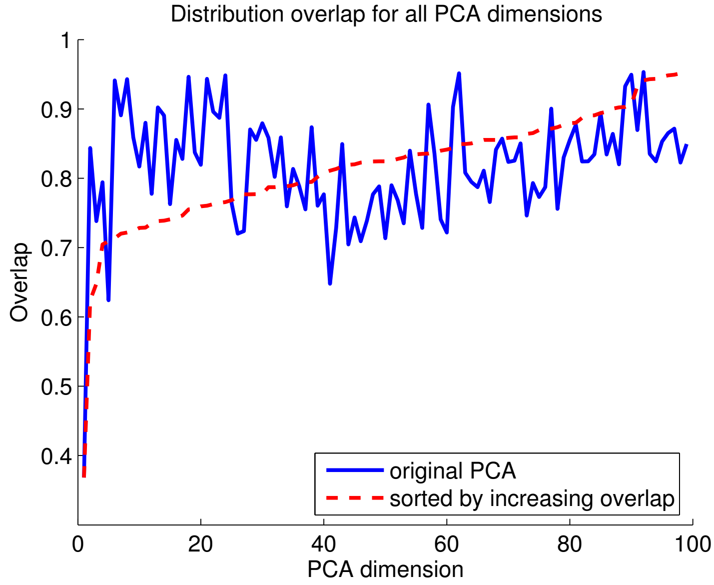


Figure 5.4. Distribution overlap for all PCA components.

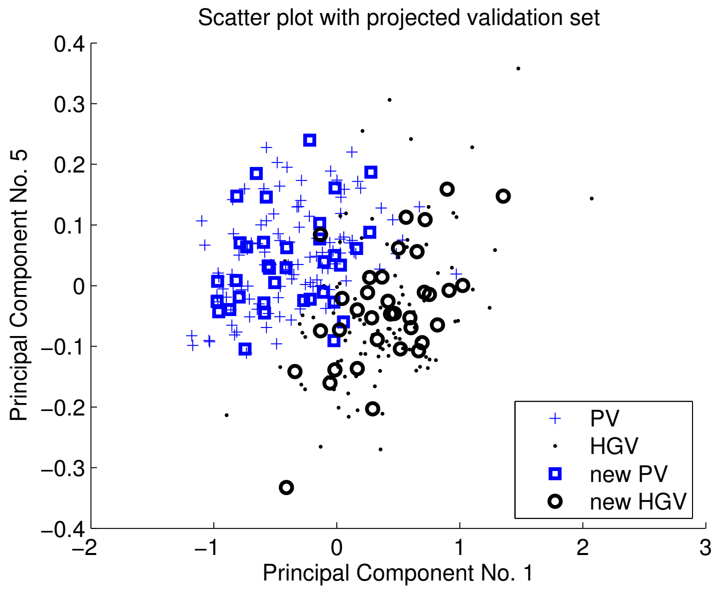


Figure 5.5. Scatter plot with 25 % of the patterns used as validation set.

axis. *new PV* and *new HGV* denotes light vehicles and heavy vehicles respectively, projected onto the PCA axis compiled with the remaining 75 % of the patterns, but not used in the compilation. It shows that patterns not used in the decomposition are projected to similar "positions" as the patterns used for obtaining the PCA transformation. This means that a new pattern of ARMA coefficents from a heavy good vehicle, HGV, pre-processed accordingly, will project close to the cluster of the other HGVs.

N.B. Points $(x, y) = (\xi_\alpha, \xi_\beta)$ in Figure 5.5 are projected feature vector elements

$(\alpha, \beta) = (1, 5)$, and *not* $(\alpha, \beta) = (1, 2)$ as in Figure 5.2.

An additional means of assessing or analysing the convergence and generalisation of PCA is to gradually increase the amount of patterns used for the PCA decomposition and evaluate the difference between two transformations obtained with differently many patterns. By this procedure it would be possible to see whether some additional or new patterns affects the transformation or not. Apart from the projection matrix, the variable mean subtracted from the feature vectors in Equations (5.1) and (5.2) is required to converge. This evaluation is made in the same manner as for the projection matrix, i.e. by increasing the amount of patterns utilised for PCA.

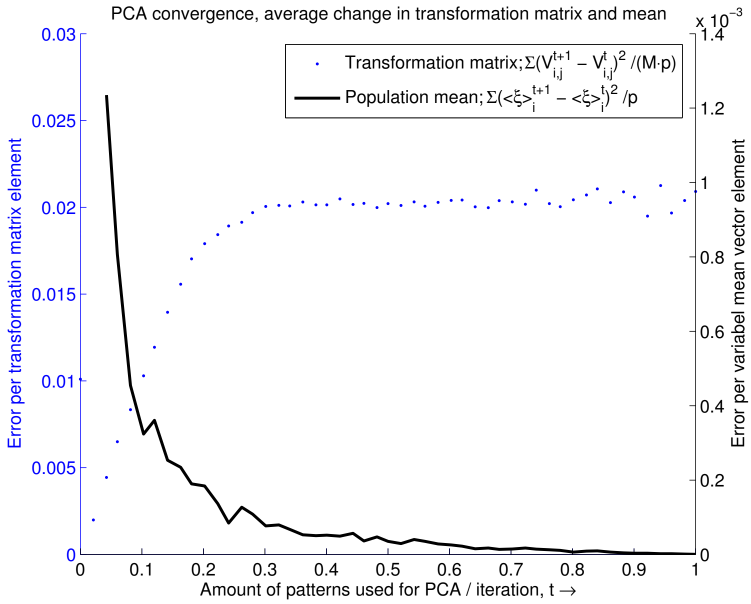


Figure 5.6. Evaluation of PCA convergence.

For this purpose, two errors are evaluated. Firstly, the error between the elements of two projection matrices, obtained using PCA with differently many patterns: Equation (5.7), and secondly the variable mean of differently many patterns: Equation (5.8).

$$e_V^{t+1} = \frac{1}{p \cdot M} \sum_{i,j}^{p,M} \left(V_{i,j}^{t+1} - V_{i,j}^{t}\right)^2 \tag{5.7}$$

$$e_{\langle\xi\rangle_i}^{t+1} = \frac{1}{M} \sum_{j=1}^{M} \left(\langle\xi_j\rangle_i^{t+1} - \langle\xi_j\rangle_i^{t}\right)^2 \tag{5.8}$$

The superscript t denotes the iteration for which the amount of patterns used for PCA is kept constant. For increasing t, the number of patterns used for obtaining the transformation matrix is increased. Subscript i denotes the pattern variables (the ARMA coefficients), j denotes an individual pattern or vehicle. Finally, $\langle\xi\rangle_i$ is the variable mean for all patterns as defined in Equation (5.1).

In Figure 5.6, the two errors are plotted versus two different y-axis. The dots represent the transformation matrix error as defined in Equation (5.7) and the solid line shows the variable mean error as defined in Equation (5.8). The figure shows the average error of 50 individual PCA decompositions, all with randomly ordered and selected patterns.

It shows that as the amount of utilised patterns increase to more than 0.35 the transformation matrix error converges to some error around 0.02 per matrix element. The reason for this is that, when less than $0.35 \cdot M$ of the patterns are employed for PCA, the matrix rank of V is less than the vector space dimensionality. The number of patterns needed to completely define the vector space is equal to the number of variables in each feature vector. Put differently, it seems that as soon as $M > p$ the transformation matrix converges.

The variable mean error is a steadily decaying function, and when 75 % of the patterns are used for PCA, the error is less than $1.5 \cdot 10^{-5}$ per vector element.

5.4 Euclidean Distance Classification

With pictures as in Figures 5.2 and 5.5 showing the two classes of vehicles forming nearly separable clusters, the idea of using a geometrical rule for classification arises. In the p- dimensional space defined by the elements of the vehicle patterns, it is assumed that each of the two classes has a centre of gravity. Now, if the euclidean distance from a pattern of an unknown vehicle to the centre of gravity of the light vehicle class is smaller than the distance to the centre of gravity for the heavy vehicle class, the vehicle is assumed to belong to the set of light vehicles and vice versa. For statistical data the centre of gravity equals the expectancy value which is estimated by the mean value.

For a pattern of an unknown vehicle, ξ^{new}, the selection rule is based of which of the euclidean distances in Equations (5.9) and (5.10) is the smallest.

$$\|\langle\xi_i\rangle_{PV} - \xi_i^{new}\| \tag{5.9}$$

$$\|\langle\xi_i\rangle_{HGV} - \xi_i^{new}\| \tag{5.10}$$

where $i = 1, 2, \ldots, p$.

To test this approach for classification, 75 % of the patterns, randomly selected after PCA decomposition, are used to calculate the centre of gravity, and the remaining 25 % are used as a validation set. The procedure of choosing patterns randomly for validation is repeated for 50 times to produce results that keeps well from a general point of view. This method of classification is on average 83 % accurate, and the probability of erroneous classification is then 17 %. The distribution of the probability of erroneous classification is providied in Figure 5.7. Figure 5.8 shows

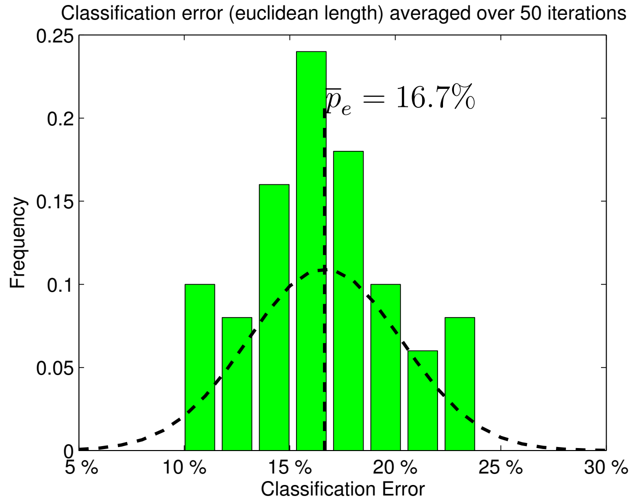


Figure 5.7. Distribution of the probability of erroneous classification using all PCA components (dimensions) and 25 % of the patterns as validation set.

how the classification performance is related to the number of PCA components used in the selection rule defined in Equations (5.10) and (5.9). Apparently, the number of components used in this classification approach does not affect the performance significantly.

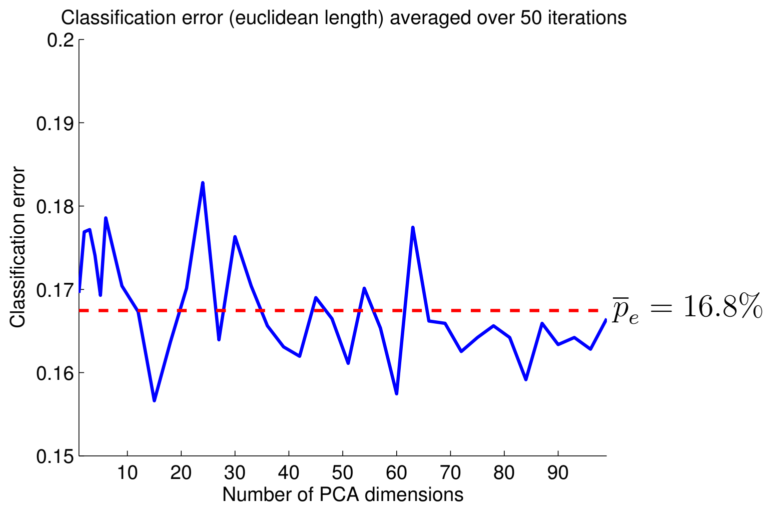


Figure 5.8. Probability of erroneous classification using differently many PCA components (dimensions) and 25 % of the patterns as validation set.

6 ARTIFICIAL NEURAL NETWORKS

In the field of neural computations the perceptron neural network is widely used as a pattern classifier. Apart from the perceptron, the realm of neural network computing include features as Self Organising Maps, or Kohonen networks, the Hopfield model, Boltzmann machines and recurrent neural networks.

The perceptron, or Multi Layer Perceptron, MLP, is a feed-forward neural network with simple processing elements or neurons whose connectivity resembles that of the brain. Each neuron compiles a weighted sum of all its inputs and passes on a signal through a non-linear activation function. In a layer, every neurons processes all inputs from the previous layer, and by this the signals propagate from input layer to hidden layers and finally to the output layer in a forward pass.

A Multi Layer Perceptron has the ability to learn arbitrarily complex non-linear regressions by adjusting the synaptic weights using a training algorithm. The resulting output of the perceptron is compared to a desired target output and errors are propagated backwards. In this back-propagation, the synaptic weights are adjusted according to their contribution to the overall error. The algorithm employed in the backward pass to minimise the output error involves gradient descent with a momentum term.

The structure and complexity of the network is determined beforehand by choosing number of input nodes, number of nodes in the hidden layer(s) and network noise etc. The Multi Layer Perceptron tries to determine the best hyperplane to separate partitions in the input feature space.

6.1 Neural Computation

The input to a perceptron is referred to as a pattern, one for each vehicle, which consists of the PCA transformed ARMA coefficients. The different patterns are denoted by the superscript $^{\mu}$. To every pattern a desired output is assigned, a "correct answer", referred to as ζ^{μ}.

Given pattern ξ^{μ}, a neuron in the first hidden layer receives an input

$$v_j^{n=1} = -\theta_j^{n=1} + \sum_{k=1}^{N_{n=1}} w_{j,k}^{n=1} \xi_k^{\mu} \tag{6.1}$$

where $w_{j,k}^{n=1}$ are the synaptic weights between input layer and the first hidden layer connecting input unit k to neuron j, $\theta_j^{n=1}$ is the threshold level of neuron j and $N_{n=1}$ is the number of neurons in the first hidden layer $n = 1$. Neuron j then produces an output of

$$V_j^{n=2} = \varphi(v_j^{n=1}) \tag{6.2}$$

in which $\varphi = \varphi(v)$ is the non-linear activation function

$$\varphi(v) = \tanh(\beta v) \tag{6.3}$$

The parameter β is a measure of how much noise is present during the neural network computations. It is closely related to the Boltzmann factor for which β equals the inverse temperature of a system of units (or neurons in this case).

For a neuron in any succeeding layer, n, the synaptic input is defined as

$$v_j^n = -\theta_j^n + \sum_{k=1}^{N_n} w_{j,k}^n V_k^n \tag{6.4}$$

hence, the output of this neuron will be

$$V_j^n = \varphi(v_j^{n-1}) = \varphi\left(-\theta_j^{n-1} + \sum_{k=1}^{N_{n-1}} w_{j,k}^{n-1} V_k^{n-1}\right) \tag{6.5}$$

Finally, the output of the neural network, the perceptron, is denoted O_j^μ and defined as

$$O_j^\mu = \varphi(v_j^{n=N_l}) \tag{6.6}$$

where N_l is the total number of hidden layers.

At the output of the neuron, the activation function, φ, is applied to produce the output signal. Such activation functions include the Heaviside step function and the Signum function, but applications such as this require the activation function to be continuous and differentiable. Commonly used activation functions are:

I) the Logistic function

$$\varphi(v) = \frac{1}{1 + \exp(-\beta v)}\ ,\ \beta > 0 \tag{6.7}$$

II) the Hyperbolic tangent function

$$\varphi(v) = \tanh(\beta v)\ ,\ \beta > 0 \tag{6.8}$$

The amount of noise introduced by the factor β will influence the perceptron's ability to generalise and additionally ensure quick learning. Choosing a large beta means less temperature noise which results in a more decisive perceptron, but also less able to generalise. For this application, the hyperbolic tangent function is chosen as activation function.

6.2 Neural Network Training

The energy of the network is formed by the deviation of the output from the desired or correct output squared. By training the network, a steepest descent minimisation of this energy is performed with respect to network weights and thresholds. The classification error propagates backwards, updating all the synaptic weights.

Given the perceptron output, O_j^μ, the energy cost function $E = E(W)$ is introduced as a measure of error accordning to

$$E(W) \quad = \quad \frac{1}{2}\sum_{\mu,j}\left(\zeta_j^\mu - O_j^\mu\right)^2 \tag{6.9}$$

utilising the desired or correct output, in this case class, ζ_j^μ. The dependent variable W includes all synaptic weights and thresholds in the network. For this application, the training is batch type, meaning that weight update is made for all patterns simultaneously, hence the summation over μ. The alternative is sequential updating, which may have advantages for some applications involving large sets of redundant training data.

To minimise the classification energy (∼error) in the multi dimensional space defined by W the gradient descent rule is applied according to (Råde and Westergren 2001):

$$\vec{x}_{t+1} = \vec{x}_t - \lambda \nabla f\left(\vec{x}_t\right) \tag{6.10}$$

A correction of the weights in the preceding layer is formed by

$$\Delta w_{j,k} = -\eta\frac{\partial E}{\partial w_{j,k}} = \eta\sum_\mu\left(\zeta_j^\mu - O_j^\mu\right)\varphi'(v_j^\mu)V_k^\mu = \eta\sum_\mu \delta_j^\mu V_k^\mu \tag{6.11}$$

where the following definition of the local gradient of an output node is made:

$$\delta_j^\mu \quad = \quad \left(\zeta_j^\mu - O_j^\mu\right)\varphi'(v_j^\mu). \tag{6.12}$$

For the hidden layers in the perceptron, the chain rule for derivatives/differentiation is used, forming a general expression for the weight corrections of the synaptic weights:

$$\delta_j^{n-1} \quad = \quad \varphi'(v_j^{n-1})\sum_{k=1}^{N_n}\delta_k^n w_{j,k}^n \tag{6.13}$$

$$\Delta w_{j,k}^n \quad = \quad -\eta\frac{\partial E}{\partial w_{j,k}^n} = \eta\delta_j^n V_k^n \tag{6.14}$$

Subsequent to the calculations of weight corrections,$\Delta w_{j,k}$, the synaptic connections are updated according Equation (6.10):

$$w_{j,k}^{new} \quad = \quad w_{j,k}^{old} + \Delta w_{j,k} \tag{6.15}$$

Neuron thresholds are updated analogously:

$$\Delta\theta_j^n = \eta \sum_{k=1}^{N_n} \delta_j^n V_k^n \tag{6.16}$$

$$\theta_j^{new} = \theta_j^{old} + \Delta\theta_j \tag{6.17}$$

To achieve a faster convergence rate and avoiding local minima during training a *momentum term*[1] is added to the correction term. This addition also counteracts any ocillating behaviour of the descent algorithm. The added momentum or inertia is simply an addition of the correction term from the previous training iteration according to

$$\Delta w_{j,k}^t = \eta \left(\Delta w_{j,k}^t + \alpha \Delta w_{j,k}^{t-1} \right). \tag{6.18}$$

where α is the momentum coefficient and t is the training iteration (epoch).

To further enhance the learning algorithm, weight elimination is introduced. The elimination is referred to as pruning in literature (Haykin 1999, Hertz 1991) and works as a complexity penalty which removes small, insignificant connections.

$$\epsilon_{j,k} = 1 - \gamma\eta \frac{1}{1 + w_{j,k}^2} \tag{6.19}$$

$$w_{j,k}^{new} = \epsilon_{j,k} w_{j,k}^{old} + \Delta w_{j,k} \tag{6.20}$$

When training is completed, i.e. desired or best performance is achieved, the updating of the synaptic weights and thresholds is terminated.

6.3 Network Design

The design of the neural network highly influences the classification performance, and there are many parameters that can be altered and tweaked to achieve the best performance. The parameters that can be chosen and/or optimised are:

I) η – update rule step size, $\eta \in [0, 1]$,

II) α – momentum term coefficient, $\alpha \in [0, 1]$,

III) γ – weight decay coefficient, $\gamma \in [0, 1]$,

IV) β – network noise factor, $\beta \in [0, \infty]$

V) the number of nodes in the input layer, hidden layer(s) and output layer and

VI) the number of hidden layers in the perceptron.

[1]Gradient descent algorithms with a momentum term are referred to as *conjugate* gradient descent methods.

Theoretically, any function or rule, and thereby classifier, can be realised using two hidden layers, provided the underlying physical process is governed by this rule. However, it shows that the best perceptron performance is obtained using only one hidden layer.

Regarding the number of output nodes, two possibilities arise: i) training one perceptron per class, or ii) training one perceptron to classify both classes. In this report only the latter is considered.

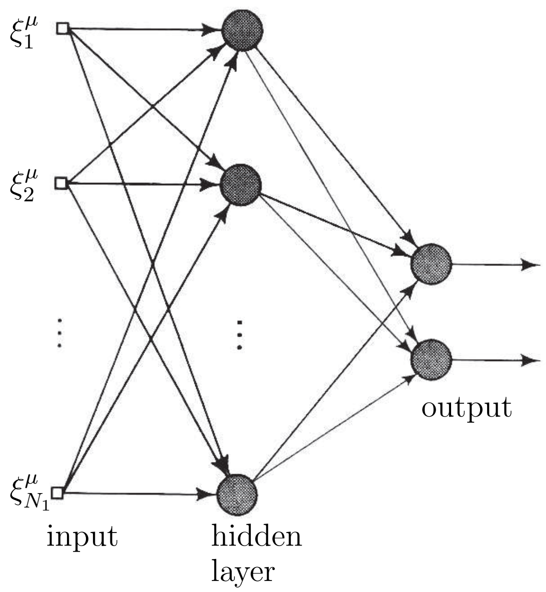


Figure 6.1. Neural network design with one hidden layer.

Figure 6.1 shows the chosen layout of the perceptron classifier. Input data is μ-dimensional and the output is two-dimensional.

Perceptron Input and Output

The choice of activation function determines how output data are to be represented. The nonsymmetric Logistic function, Equation (6.7), requires output data in the range $[0, 1]$, whilst the antisymmetric Hyperbolic tangent function, Equation (6.8), requires output data to be $[-1, 1]$. The desired outputs for the two classes are:

$$\begin{array}{lll} [1 \quad -1]^T & , & \text{for class PV} \\ [-1 \quad 1]^T & , & \text{for class HGV} \end{array} \tag{6.21}$$

and thus, the output O_j^μ is two dimensional.

Perceptron input is scaled to have zero mean and unit variance

$$\vec{\xi} \sim N(\mu, \sigma^2) \quad \rightarrow \quad \vec{\xi} \sim N(0, 1) \tag{6.22}$$

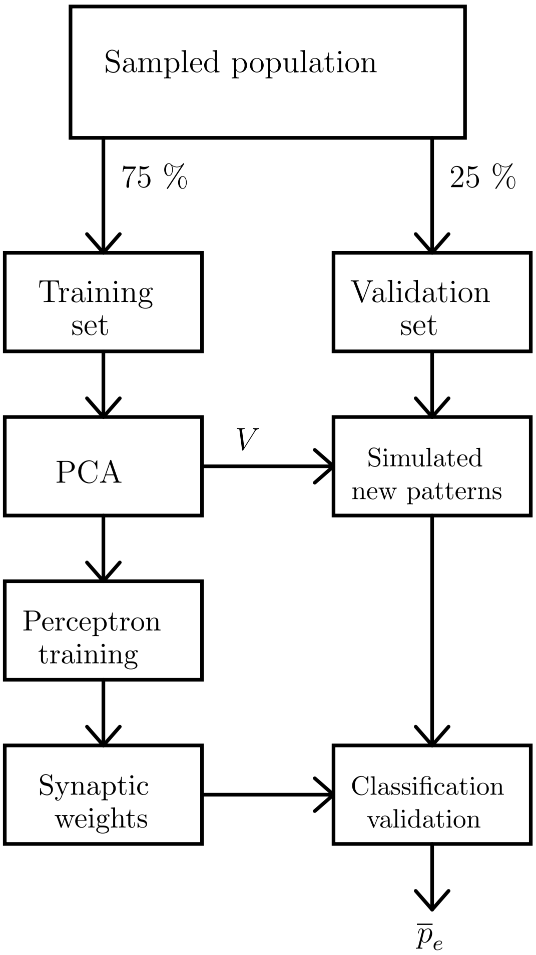


Figure 6.2. Flow diagram for ANN calculations.

From the collected data set, 75 % randomly selected patterns are used for training of the perceptron, and the remaining 25 % are used to simulate new patterns. These patterns are not contributing to the training of the perceptron and therefore, they can be used to validate the classification performance in a general sense.

This manner of classifying a validation set is basicly equivalent to recording new vehicles, preprocessing the recordings according to Chapter 4 and finally presenting them to a readily trained perceptron with fixed synaptic connections.

A flow diagram of this procedure is presented in Figure 6.2, in which V is the transformation matrix obtained from the PCA used to transform the simulated new vehicles. $\overline{p}_e$ denotes the probability of erroneous classification, i.e. perceptron performance.

Parameter Optimisation

To choose the parameters listed above wisely is imperative for successful training and classification. The number of input nodes is limited by the number of available components in the pattern and the number of output nodes is set to two. Apart from the number of outputs, the other parameters could be optimised with respect to classification performance.

For the purpose of optimising the parameters, the classification error, i.e. the probability of misclassified vehicles, is analysed during training for a number of different values for each parameter. Other parameters are kept constant during the evaluation. A small classification error for a certain choice of parameter value suggests that the performance is optimal. However, for a correct optimisation, this implies that the other parameters are strictly independent of the evaluated one, which is not the case. Hence, the result of the optimisation is decisive, but only taken into consideration when choosing parameter values.

Implementing a certain parameter value, the perceptron is trained for a period of 500 training epochs. The procedure is then repeated 30 times and averaged over.

Table 6.1 shows the perceptron parameter values chosen by aid of the optimisation procedure. Enclosed in Appendix C are graphs of the evaluation results.

Parameter	*Value*
α	0.95
β	0.38
γ	0.05
η	0.075
N_{hidden}	100
N_{input}	90

Table 6.1. Parameter optimisation: chosen parameter values.

The parameter N_{input} refers to the number of input elements presented to the perceptron and N_{hidden} denotes the number of nodes in the hidden layer.

6.4 Classification Results

For deciding which class a pattern (vehicle) is adherent to, a decision rule is formed. Given the binary nature of the classification answers, classification performance is binomially distributed. For n independent trials of x with probability p, the binomial distribution is defined as:

$$B(n,p) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{for } x = 0, 1, \ldots, n \tag{6.23}$$

The probability of incorrect classification is estimated by

$$np_e = N_e \Leftrightarrow p_e = \frac{N_e}{n} \tag{6.24}$$

where p_e denotes the probability of erroneous classification, N_e are the number of incorrectly classified vehicles and n are the number of trials.

Figure 6.3 shows the estimated probability of incorrect classification according to (6.24) during training. Training is executed for 400 epochs and during each training epoch all of the 75 % patterns in the training set are used. The remaining 25 % of the patterns, the validation set, are tested as training progresses, q.v. the dashed line in Figure 6.3.

For proof of the performance of the perceptron, training on the randomly selected is done 250 times, each time with a differently chosen training set. The results of these 250 independent runs are shown as a an average misclassification probability

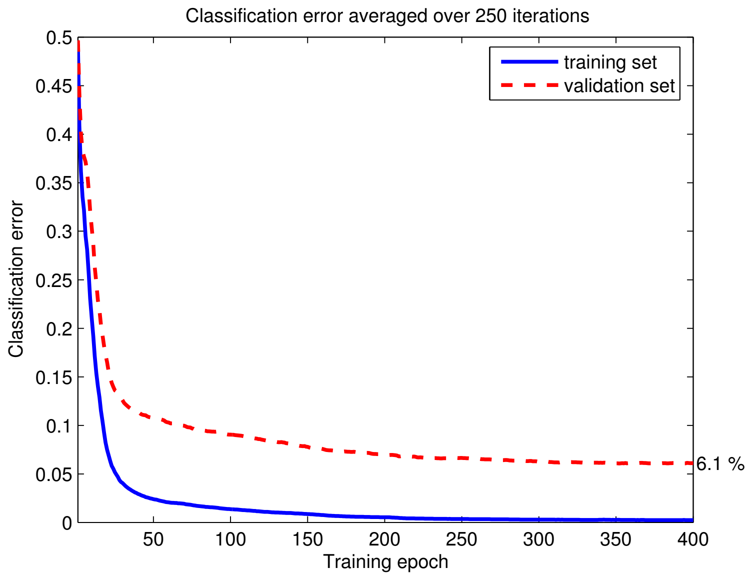


Figure 6.3. Classification error during training averaged over 250 independent trainings.

in 6.3 and as a histogram in 6.4.
N.B. For each and every independent training, a new validation and training set is randomly selected from the sample set and patterns in the validation set are not included in the training.

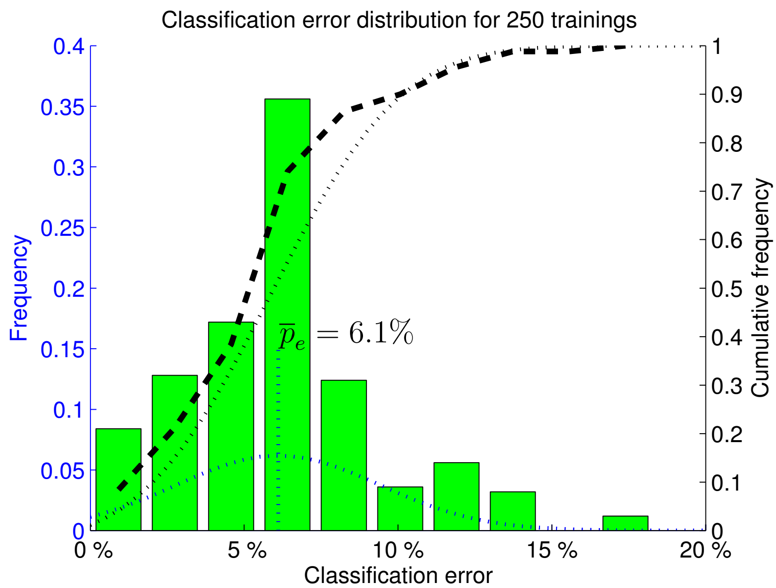


Figure 6.4. Distribution for the probability of erroneous classification.

Also seen in Figure 6.4, as the dotted line, is a normal distribution. This distribution is only plotted as a visual aid. Distribution and histogram values are read on the left y-axis and cumulative frequency functions are read on the right y-axis. It shows that if an error of 10 % is accepted, 90 % of all perceptron trainings will comply with this demand.

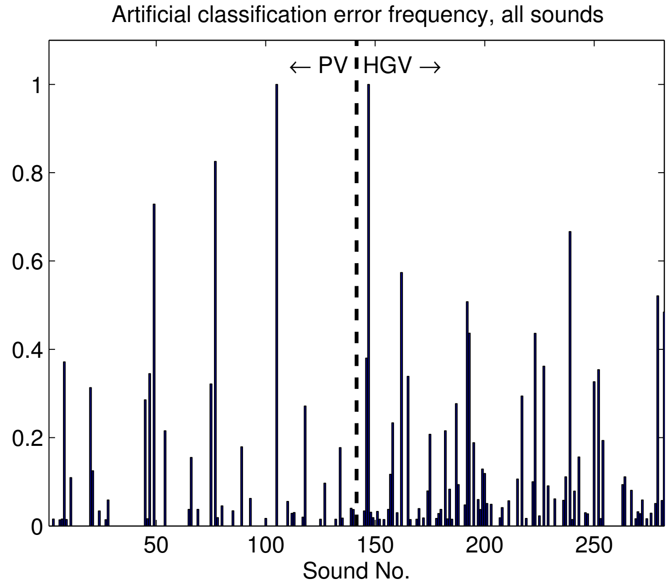


Figure 6.5. Classification error frequency for all recorded sounds.

For evaluation of which class is the more difficult to classify, the error frequency for vehicle sounds in the validation set is displayed as a histogram in Figure 6.5. Apparently, attempting to classify heavy vehicles provides for 64 % of the errors and light vehicles for 36 %. The probability of erroneously classifying a heavy vehicle as a light is $\overline{p}_{e,HGV} = 7.7\%$ and vice versa $\overline{p}_{e,PV} = 4.4\%$. The combined probability of misclassification is $\overline{p}_e = 6.06\%$. On average, all vehicles occure 63 times in the validation set for the 250 independent trainings computed to produce Figures 6.3, 6.4 and 6.5.

7 LISTENING TEST

To provide a reference to the developed classification method, a listening test is designed. In the test, participants are asked to identify heavy and light vehicles when presented recorded sounds thereof. The test sounds are the very same as the ones used in the training of the artificial neural network. Apart from the main goal to acquire a reference, the purpose of the test is to investigate the accuracy of manual sound classification. It is also desired to obtain a measure of the time needed for classification.

7.1 Test Design

To fulfill all the stated requirements, a test is to be designed for the purpose of testing manual classification accuracy and measuring the necessary time consumption thereof. To attain these goals, an individual test is needed, in which the participants perform the test autonomously[1]. For this type of test, the participants are tested separately. Such a test has the advantages of: i) a good measure of accuracy, ii) a good measure of the time needed to perform classification, iii) minimal influence of sound sequence ordering on test outcome, and the disadvantage of iv) extensive time usage for test execution.

However, to lessen the time needed for execution of the test, a test design which tests all participants simultaneously is chosen. In this type, all participants execute the test simultaneously. They will listen to the same sounds, in the same order and they are asked to perform the vehicle classification simultaneously in a predefined amount of time. The anticipated properties of such a test include: i) a good measure of accuracy, ii) very little time usage for test execution, iii) no control over sound sequence effects and iv) a restricted or limited measure of the individual time needed to perform classification.

Preceding the test execution, participants are presented a response sheet and given appropriate instructions. The response sheet is enclosed in the thesis as Appendix A. To limit the effects of a steep learning curve, and accommodate the participants to the test soundscape[2], 4 test sounds are played, and the adherent class is given.

Apart from the 4 test sounds, the test consists of 160 sound recordings, randomly selected from the subset of recorded vehicles. All sounds are preprocessed by bandpass i) filtering, ii) re-sampling and iii) cropping to produce a set of uniform samples. The preprocessing is performed by the same manner as in the Preprocessing section 4.1 on page 17. The parameters used are:

[1]By performing the test autonomously it is implied that participants are tested isolated, for example at a PC workstation.

[2]The sonic environment; background noises at the recording site. (Schafer 1994)

I) filter cut-off frequencies $f_{low} = 50$ Hz and $f_{high} = 22$ kHz,

II) sampling frequency $F_s = 44$ kHz,

III) sound clip length $T = 3$ seconds.

Succeeding every sound sample is a quiet period of 5 seconds in which the participants are required to decide and note their decision. Each sound sample is also preceded by a short tone to portend the next sound. The test procedure is described by Table 7.1.

	3 s	5 s	3 s	5 s	
...	sound i	quiet + tone	sound $i+1$	quiet + tone	...

Table 7.1. Test procedure and time distribution during the test, $i = 1, 2, 3, ..., 159$

7.2 Sound Properties

As indicated in the previous sections, the test purpose is not to evaluate how manual classification is achieved, but simply to attain a measure of accuracy. In this context, a thorough analysis of the test sounds is not self-evident, but a general overview in terms of psycho-acoustic measures can be clarifying. Physically and psycho-acoustically, the test sounds can be described by a number of different quantities, for instance:

I) maximum sound pressure level, L_{max} [dB],

II) equivalent sound pressure level, $L_{eq.}$ [dB],

The maximum sound pressure level, L_{max} is defined as the decibel value of the largest pressure amplitude during a certain period of time, $T = 3$ seconds in this case. The equivalent sound pressure level, $L_{eq.}$ is defined as the mean pressure squared during a time interval, also 3 seconds. Mathematical definitions provided in Equation (7.1) with $p_{ref.} = 20\mu Pa$.

$$\begin{aligned} L_{max} &= 20\ \log_{10}\left(\max_p(|p|)\ /p_{ref.}\right) \\ L_{eq.} &= 10\ \log_{10}\left(\frac{1}{T}\int_0^T p^2(t)dt\ /p_{ref.}^2\right) \end{aligned} \tag{7.1}$$

For the statistical description of the sounds in the test, some parameters which characterises the sampled population are used. Which parameters that are useful can vary between different distributions, but the most commonly used include:

I) expectation or mean

II) variance

III) median

IV) range

All but items III and IV, of the statistical analysis is considerably simplified if one assumes normally distributed quantities. A variable with a general normal distribution, $\phi(x) \sim N(\mu, \sigma^2)$ is described by the expectation, μ and the variance, σ^2 according to:

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{7.2}$$

For statistical data with n observations $x_1, x_2, \ldots, x_n$, the mean (expectation) is defined:

$$\overline{x} = \frac{x_1, x_2, \ldots, x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \tag{7.3}$$

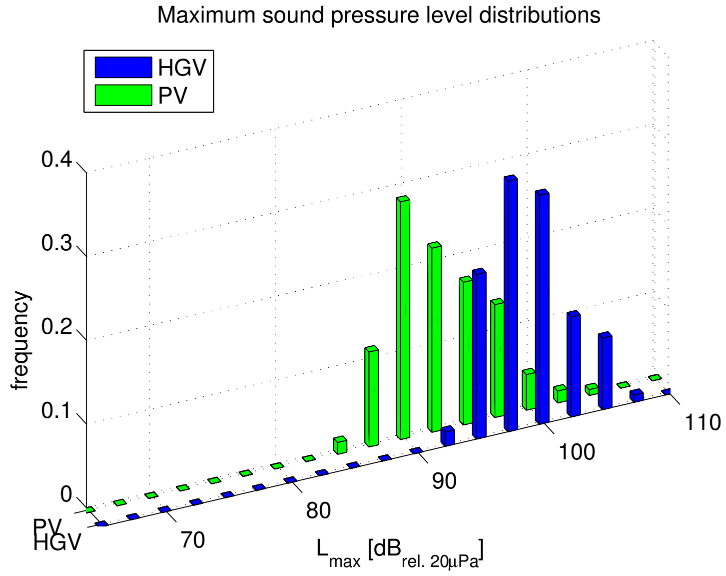
and the variance:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \quad . \tag{7.4}$$

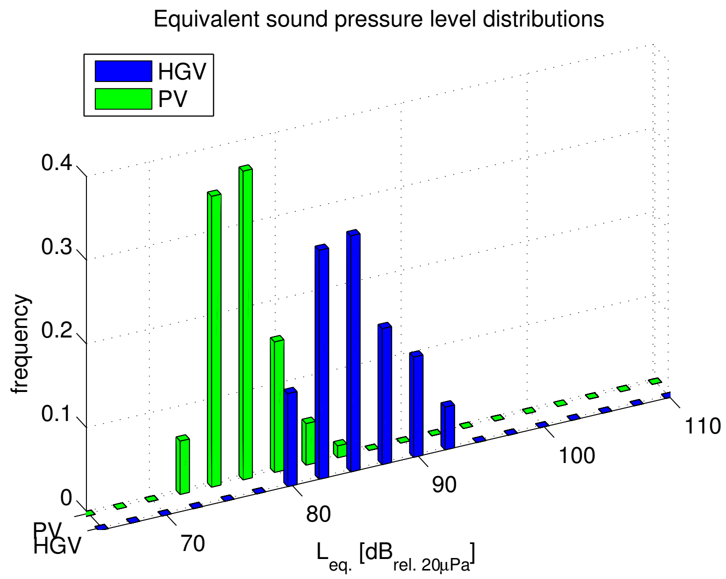
The estimated mean, $\overline{x}$, is $t(n-1)$ distributed, with $n-1$ degrees of freedom, and the estimated variance $\sigma^2 \sim \chi^2(n-1)$, also with $n-1$ degrees of freedom.

The median is defined as the middle value of the observations and the range simply consists of the difference between the largest, $\max(x_i)$, and the smallest, $\min(x_i)$, values of the observations. Provided in Figure 7.1 are the distributions of L_{max} and $L_{eq.}$ of the sound recordings in the test subset. Not surprisingly, heavy vehicles (HGV) generally have larger values than light vehicles (PV) for both properties, L_{max} and $L_{eq.}$. The distribution overlap seems to be larger for L_{max} (Figure 7.1(a)) than for $L_{eq.}$ (Figure 7.1(b)), which could indicate that light vehicles generally travel faster than heavy ones at the measurement location. A high maximum sound pressure level and a low equivalent level implies a narrower, steeper signal envelope.

For the purpose of describing the population, the statistical quantities described earlier are calculated for the properties L_{max}, $L_{eq.}$, and accounted for in Table 7.2 for the subsets PV and HGV.



(a) L_{max} distribution



(b) $L_{eq.}$ distribution

Figure 7.1. Distributions of L_{max} and $L_{eq.}$ properties describing listening test sounds.

	L_{max} [dB]		$L_{eq.}$ [dB]	
	PV	HGV	PV	HGV
mean	92.9	99.1	77.0	85.1
median	82.4	98.6	76.7	84.4
range	18.1	14.9	14.0	13.6
variance	14.2	9.9	6.4	11.0

Table 7.2. Statistical quantities describing the properties of the test sound recordings.

7.3 Test Result

With 22 test participants (students of ages 21 to 30) and 160 sounds, the test comprises a total of 3520 answers. Included in the test are 80 randomly selected sounds from each class. Due to the binary nature of the response alternatives, i.e. a choice of either PV or HGV, the classification responses can be assumed to be binomially distributed. For n independent trials of x with probability p, the binomial distribution is defined as:

$$B(n,p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x} \qquad \text{for } x = 0, 1, \ldots, n \tag{7.5}$$

In Figure 7.2, the classification error for all participants is plotted in a histogram. The skewed appearance of the distribution adheres from the properties of the binomial distribution. It is also possible that the distribution in Figure 7.2 actually is a skew gaussian or even an exponential distribution.

However, with 139 errors out of 3520 classifications, one could assume the errors to be normally distributed by support of the Grand Mean Theorem, although the superimposed normal distribution in Figure 7.2 should not be seen as estimate of the true distribution, but as a visual aid only. It is of course impossible to achieve less probability of error than 0, as the left "tail" of the normal distribution indicates. No further analysis of the test result in terms of distributions will be done.

The probability of incorrect manual classification for one participant is estimated by

$$np_e = N_e \Leftrightarrow p_e = \frac{N_e}{n} \tag{7.6}$$

where n is the number of classification trials and N_e is the number of erroneous classifications.

In Figure 7.3 the performance of all participants is displayed. The performance ranges from 2 to 12 erroneous classifications with an average of $\overline{N}_e = 6.32$. Judging by the differently coloured patches in the histogram, classification of cars is more difficult than of heavy vehicles. The total number of incorrectly classified cars is 94, $\overline{p}_{e,PV} = 0.053$, whilst heavy vehicles are erroneously classified 45 times, $\overline{p}_{e,HGV} = 0.026$. Light vehicles seem twice as difficult to classify as heavy.

Figure 7.4 shows the total number of classification errors for each sound during the listening test sequence. In this graph, learning- or fatigue effects on test performance would be visible, where there any. No effects of learning during the test is prominently visible, nor any effects of weariness or fatigue. It is however clear that some sounds are much more difficult to classify correctly than others, sound No. 67 for instance, was incorrectly classified as a heavy vehicle by 13 participants ($13/22 = 59\%$). The occurance of sound No. 67 in the listening test is justified by the fact that the method, for which this test is a reference, might suffer from the same difficulties.

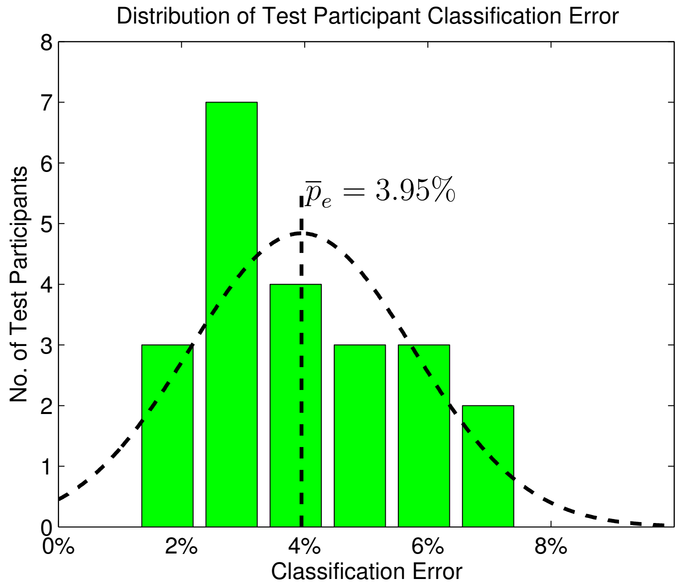


Figure 7.2. Distribution of the classification error for the test participants.

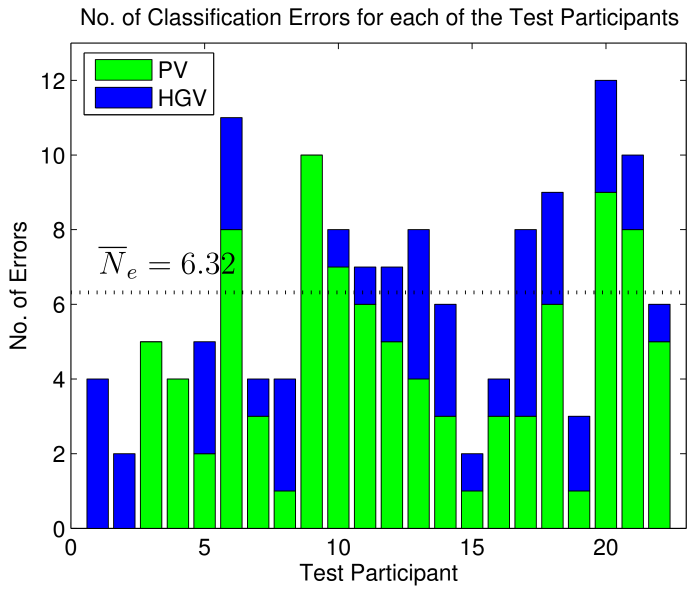


Figure 7.3. Number of classification errors for the test participants.

Manual Classification Conclusions

The performed listening test had the purpose of providing a reference to the neural network classification method. It shows that manual classification is 96 % accureate for the present circumstances.

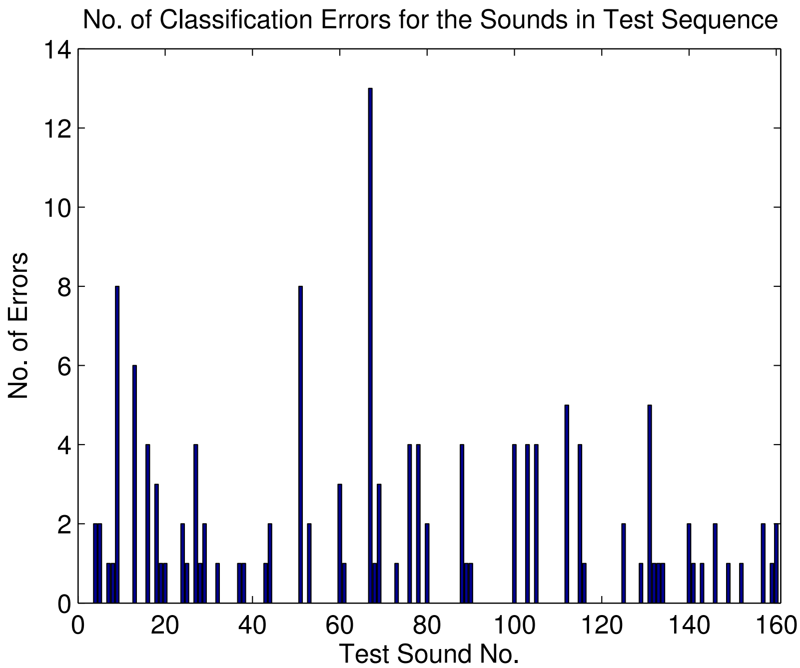


Figure 7.4. Number of classification errors for each sound in the test sequence.

8 DISCUSSION

The developed method for acoustical classification of traffic, heavy and light vehicles, is proved viable for the adopted framework with a performance of ca 94 % correct classifications. Method performance is validated such that it applies from a general point of view for the described conditions and circumstances. The statistical sampling described in Chapter 3 and the method framework described in Chapter 1 limits the applicability of the results on an arbitrary traffic situation or location.

The conducted measurements have proven adequate in terms of measurement techniques for the utilisation of data in the classification process. The influence of varying vehicle velocities on classification performance is not evaluated but believed to be negligible. Neglecting to control circumstances and variables such as wind and weather does not influence the performance of the developed method.

Data reduction as described in Chapter 4 significantly reduced the number of data points for each individual observation by a factor $> 10^3$. The successful application of the neural network for the classification task proves that the data reduction preserves the characteristics which separates the two classes.

Principal Component Analysis seems to be a viable tool for reorganising data and bring out sets of data which are intrinsically separated. The resulting principal components are however not evidently ordered by some measure of separation. It seems that even though all distributions of pattern elements overlap, the perceptron can find a decision rule utilising many components to attain the goal.

The employment of an artificial neural network, a perceptron, for the task of finding a decision rule for the patterns of principal components proved quite successful.

Classification Performance

Regarding the classification performance, some comparable results can be found in the articles mentioned in the introduction chapter, page 1. The article most closely related to the work in this thesis, i.e. traffic classification, by Reference (Nooralahiyan and Kirby 1998), presents classification performance of 84 %.

Classification of underwater sounds was roughly 70 % accurate (Greene and Field 1991), whilst infrasound classification was 100 % accurate (Ham and Park 2002).

Comparison to Reference Methods

The reference methods considered in the thesis are i) manual classification (listening test), q.v. Chapter 7 and ii) classification by distance to centre of gravity, q.v. Section 5.4.

Not surprisingly, manual classification is more successful than its artificial adversary. The difference in performance is however remarkably small; the probability of erroneous classification is $\overline{p}_e = 3.95\%$ for manual classification and $\overline{p}_e = 5.55\%$ for the artificial. The more linear approach to classification, utilising the euclidean distance, rendered the much poorer result of $\overline{p}_e = 16\%$.

Artificial classification performance is compared to manual for those sounds included in the listening test, q.v. Figure 8.1(a) and Figure 8.1(b). The correlation histogram in Figure 8.1(c) clearly shows that the artificial classification method is not qualitatively comparable to manual classification

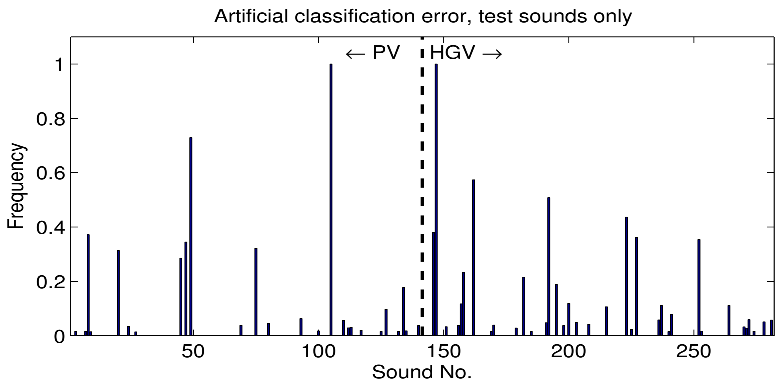
Method	$\overline{p}_e$
Manual	4 %
Euclidean length	17 %
Neural Network	6 %

Table 8.1. Probability of erroneous classification for the different methods.

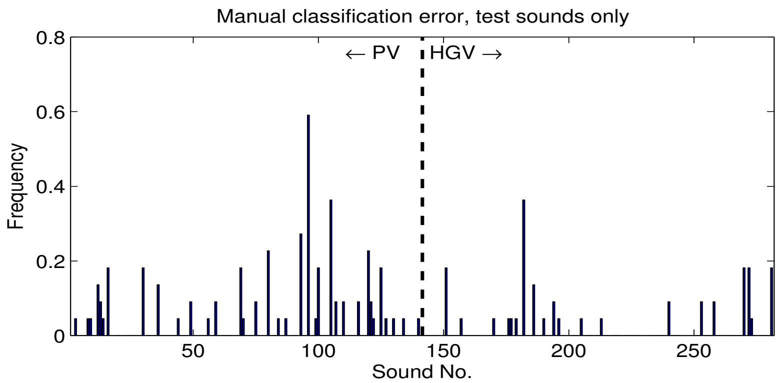
Convergence of training

Figure 8.2 shows for which training epoch index the minimum classification error, i.e. best performance, is obtained as a function of the number of input elements and number of neurons in the hidden layer. To produce the plot, the perceptron is trained 25 times and averaged over. Training terminates after 200 epochs as decided a priori. The training epoch index, for which the best performance is obtained is saved and assigned to a colour shown in colourbar left of the square plot. The figure shows under which conditions training converges. If minimum error is obtained close to training termination, the training can be assumed to have converged. The similarly coloured region on the left hand side of the plot marks for which combination of inputs and hidden neuron convergence is achieved. The light coloured curve in the figure indicates the convergence region limits and it is given by the condition in Equation (8.1).

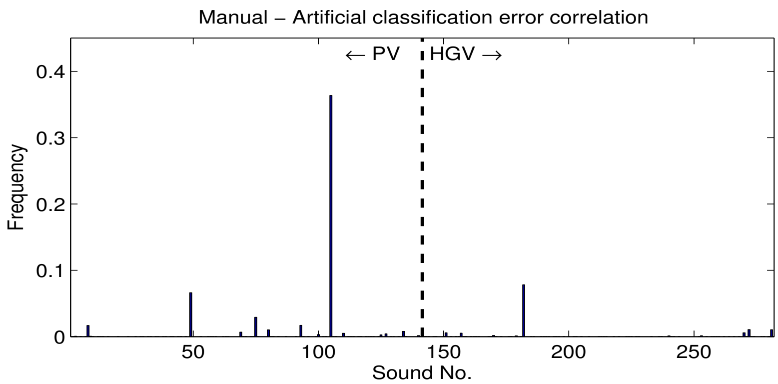
$$
\begin{aligned}
N_{input} &\geq 50 \cdot \left(1 + \frac{10}{N_{hidden}}\right) \quad \Leftrightarrow \qquad (8.1) \\
N_{hidden} &\geq \frac{500}{N_{input} - 50}
\end{aligned}
$$



(a) Artificial classification



(b) Manual classification



(c) Classification error correlation

Figure 8.1. Classification error for test sounds only

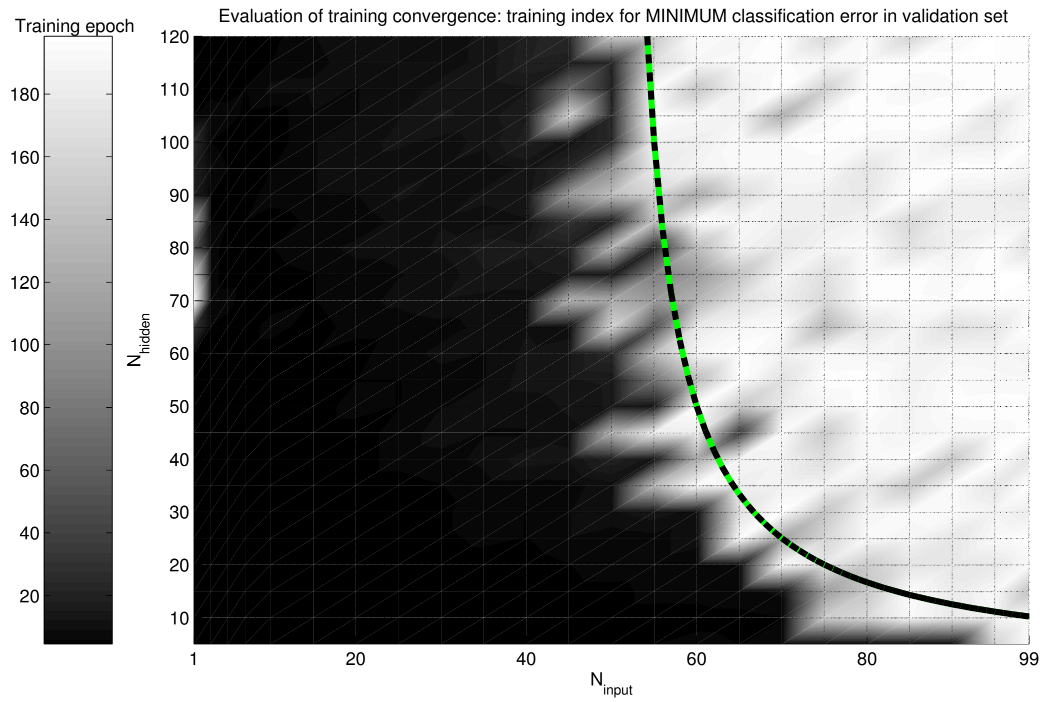


Figure 8.2. Evaluation of perceptron training convergence: training epoch index for minimum classification error in the validation set.

The vertical asymptote in Figure 8.2 shows that if less than roughly 50 input elements (PCA components) are provided for the classification task, training cannot attain convergence unless the neural network design is altered.

Classification of Non Traffic Sounds

To test the redundancy of the classification method, non traffic sounds are recorded and employed. These extra, non traffic sounds include sounds of laughter, shouting, music from a radio and sound effects such as sirens and dial tones. Measurement details are included in Appendix D.

For this test, the recorded, extra sounds are preprocessed in the same manner as the traffic sounds and presented to a readily trained perceptron with fixed synaptic weights. There is no change in the network design, which means that if a non traffic sound is to be correctly classified, the perceptron response must be negative for both classes.

The dash-dotted line in Figure 8.3 shows the classification error for sounds from sources other than traffic. Apparently, classification of those sounds fails disastrously when the perceptron chooses to label them as either heavy or light traffic. The

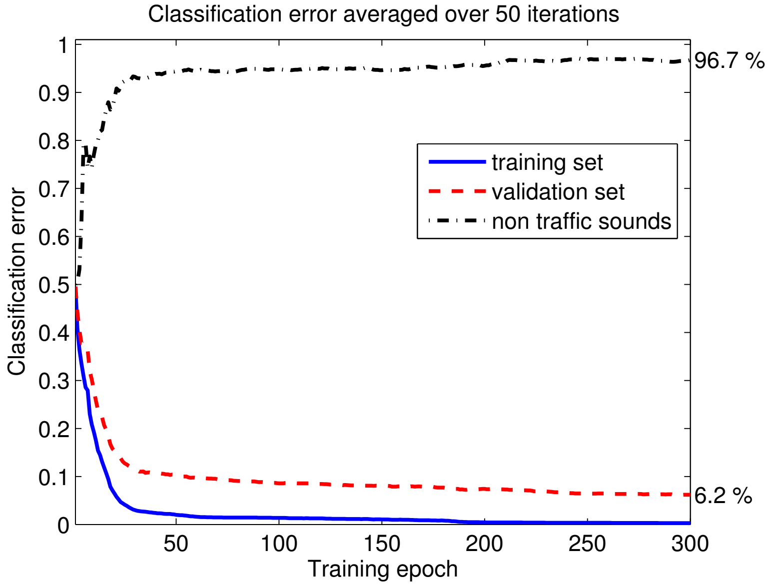


Figure 8.3. Classification error including non traffic sounds.

probability of erroneous classification is approximately [1] 97 % as training terminates after 300 epochs. This result is believed to be inherent from the fact that such non-traffic sounds are not included in the training and not assigned to any class of their own. Introducing a class for secondary, non traffic sounds and including them in the training does not seem to result in significantly better performance at this stage however. It seems that the introduction of a secondary class must be accompanied by modifications in the classification method in whole. Such modifications is believed to include a more accurate sampling of secondary sounds and modified characteristics extraction.

When a secondary class of non traffic sounds is incorporated, classification accuracy of those sounds improve at the expense of classification accuracy of vehicle sounds. This indicates that the feature extraction is insufficient to separate the classes in the feature space. A slight improvement is can be expected if better and more extensive sampling is undertaken.

At this point it seems as the developed classification method is most accurately applied when all possible sound events can be sampled and included in the training.

It seems that the major disadvantage with this method is the deficiencies in the statistical sampling and its inherent uncertainties.

[1]Not entirely reliable with an average of 50 independent trainings.

9 CONCLUSION

Striving for a low cost method of obtaining source information from unmanned noise measurements, an attempt to classify different source types using an artificial neural network is made. The work conducted includes the stages of traffic noise measurements, preprocessing, characteristics extraction and finally neural network classification. Techniques considered in these stages are filtering and resampling, signal modeling using an ARMA model, Principal Component Analysis, statistical analysis, neural computation using a perceptron and supervised learning thereof. Sounds from vehicles passing by a measurement station are recorded and subsequently classified as heavy or light traffic. A listening test is carried out to provide a reference to the developed method.

Method performance is validated using simulated new measurements, which means that a portion of the measured sounds are not used in the training of the perceptron. The developed stages in the classification method and the results are however restricted to one certain, or an equivalent location, where the measured sounds were recorded. Results show that the perceptron classification performs better than the linear attempt for which the simulated new measurements are compared to the average features of each class (Euclidean length), and slightly worse than manual classification, see Table 9.1.

Method	Accuracy
Manual	96 %
Euclidean length	83 %
Neural Network	94 %

Table 9.1. Accuracy of classification for the different methods.

On basis of these results, it is concluded that the preprocessing preserves vehicle sound characteristics and the signal ARMA model contains sufficient information for separating the two considered vehicle classes. Using an artificial neural network decreases the number of erroneous classifications by 65 % compared to the Euclidean length method. It seems that artificial and manual classification are of similar merit, since classification of the same sounds is attempted by both.

Bibliography

Andersson, Patrik (2005). Modelling Interfacial Details in Tyre/Road Contact - Adhesion Forces and Non-Linear Contact Stiffness. PhD thesis. Chalmers University of Technology, Division of Applied Acoustics.

Bodén, H. et al. (2001). *Ljud och Vibrationer.* Institutionen för Farkostteknik, Kungliga Tekniska Högskolan.

Coggins, Kevin M. and Principe, Jose (1998). Detection and classification of insect sounds in a grain silo using a neural network. In: *1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227).* pp 1760–1765.

Ehrenfried, Klaus (2004). *Strömungsakustik.* Mensch & Buch Verlag.

Fahy, Frank (2001). *Foundations of Engineering Acoustics.* Academic Press.

Greene, Ronald L. and Field, Robert L. (1991). Classification of underwater acoustic transients by artificial neural networks. In: *IEEE Conference on Neural Networks for Ocean Engineering (Cat. No.91CH3064-3).* pp 275–281.

Ham, Fredric M. and Park, Sungjin (2002). A robust neural network classifier for infrasound events using multiple array data. In: *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290).* pp 2615–2619.

Hayes, Monson H. (1996). *Statistical Digital Signal Processing and Modeling.* John Wiley & Sons, Inc.

Haykin, Simon (1999). *Neural Networks - A Comprehensive Foundation.* Prentice Hall International, Inc.

Hertz, John A. et al. (1991). *Introduction to the Theory of Neural Computation.* Addison-Wesley Publishing Company.

Mat (2004). *MATLAB Helpdesk.*

Nooralahiyan, A. Y. and Kirby, H. R. (1998). Vehicle classification by acoustic signature. *Mathl. Comput. Modelling*, **27**, 205–214.

Parks, T. W. and Burrus, C. S. (1987). *Digital Filter Design.* John Wiley & Sons, Inc.

Råde, Lennart and Westergren, Bertil (2001). *Mathematics Handbook.* Studentlitteratur.

Schafer, R. Murray (1994). *The Soundscape: our sonic environment and the tuning of the world.* Destiny Books.

Tzanetakis, George, Essl, Georg and Cook, Perry (2001). Audio analysis using the discrete wavelet transform. In: *In. Proc. WSES Int. Conf. Acoustics and Music: Theory and Applications (AMTA 2001)*.

Västtrafik (2005). Reseplaneraren. `http://www.vasttrafik.se`. Accessed on November 14 2005.

A LISTENING TEST RESPONSE SHEET

Instructions

Welcome to this vehicle classification test. You are kindly asked to turn off your cellphone before the test starts.

You will hear recorded sounds of vehicles; cars, trucks and buses et cetera, followed by a 5 second period of silence, during which you must select vehicle class. The next sound will follow automatically. The two classes you must choose between are PV and HGV:

Class	*Description*
PV	Light Vehicles; Personal/Private Vehicle: small and medium size cars.
HGV	Heavy Vehicles; Heavy Goods Vehicle: heavy trucks, tractors, buses.

Preceding each sound is a brief tone to call upon attention. Before every 10$^{\text{th}}$ sound the number of the sound to be played next will be announced. The test consists of 160 sounds, and the estimated time needed for the test is 22 minutes.

	3 s	5 s	3 s	5 s	
...	sound i	quiet + tone	sound $i+1$	quiet + tone	...

Test procedure and time distribution during the test, $i = 1, 2, 3, ..., 159$

Before the actual test, a short example sequence with two light and two heavy vehicles is played. The purpose of this test sequence is to let you get acquainted with how typical sounds of each class will sound during the test. No answering is needed during the test sequence, the correct answers are provided for in the table below.

	PV	**HGV**
1	x	
2		x
3	x	
4		x

Thank you for your participation!

Classification Table

	PV	HGV		PV	HGV		PV	HGV
1								
2								
3								
4								
5								
10			40			70		
20			50			80		
30			60			90		

Table continues on one additional page.
Next page *not* included in Appendix.

B LISTENING TEST SOUNDS

The following table shows which sound were played during the listening test and in which order.

Seq. No.	Class	Sound No.	Seq. No.	Class	Sound No.
1	PV	39	41	HGV	83
2	HGV	116	42	HGV	22
3	PV	7	43	HGV	49
4	HGV	112	44	PV	107
5	PV	49	45	PV	124
6	HGV	54	46	PV	41
7	PV	8	47	PV	112
8	PV	87	48	HGV	133
9	HGV	41	49	HGV	27
10	HGV	30	50	HGV	111
11	HGV	87	51	PV	105
12	PV	76	52	PV	63
13	PV	93	53	HGV	53
14	PV	57	54	PV	73
15	HGV	136	55	PV	45
16	HGV	131	56	HGV	100
17	HGV	12	57	PV	27
18	HGV	45	58	HGV	101
19	HGV	29	59	HGV	67
20	HGV	64	60	PV	12
21	HGV	23	61	HGV	132
22	PV	103	62	PV	58
23	PV	35	63	PV	10
24	HGV	117	64	HGV	118
25	HGV	35	65	PV	29
26	HGV	96	66	PV	135
27	HGV	10	67	PV	96
28	PV	122	68	PV	140
29	HGV	99	69	PV	36
30	PV	47	70	HGV	103
31	HGV	28	71	HGV	17
32	HGV	16	72	PV	82
33	PV	64	73	PV	44
34	HGV	21	74	PV	42
35	PV	71	75	HGV	14
36	PV	20	76	PV	100
37	PV	127	77	PV	2
38	PV	134	78	PV	69
39	HGV	119	79	HGV	71
40	PV	91	80	PV	110

Continued on the next page.

Seq. No.	Class	Sound No.
81	PV	141
82	PV	129
83	HGV	121
84	HGV	73
85	PV	119
86	HGV	48
87	PV	24
88	PV	16
89	HGV	36
90	HGV	55
91	HGV	82
92	PV	33
93	PV	81
94	HGV	114
95	HGV	5
96	PV	53
97	HGV	74
98	HGV	51
99	HGV	20
100	HGV	129
101	PV	32
102	PV	114
103	HGV	140
104	HGV	50
105	PV	125
106	HGV	137
107	HGV	2
108	PV	62
109	PV	94
110	PV	1
111	HGV	59
112	PV	80
113	PV	102
114	HGV	79
115	PV	30
116	HGV	72
117	HGV	139
118	HGV	57
119	HGV	77
120	HGV	92

Seq. No.	Class	Sound No.
121	PV	117
122	PV	113
123	HGV	44
124	HGV	9
125	PV	116
126	PV	74
127	PV	106
128	HGV	32
129	PV	14
130	PV	132
131	PV	120
132	PV	84
133	PV	70
134	PV	99
135	HGV	84
136	HGV	6
137	HGV	125
138	HGV	86
139	PV	38
140	PV	75
141	HGV	38
142	HGV	15
143	PV	130
144	HGV	40
145	HGV	127
146	PV	121
147	HGV	61
148	HGV	123
149	PV	3
150	PV	5
151	HGV	62
152	PV	56
153	HGV	130
154	HGV	18
155	HGV	95
156	HGV	3
157	PV	59
158	HGV	1
159	PV	9
160	PV	13

C PARAMETER OPTIMISATION

For the purpose of optimising the parameters, the classification error, i.e. the probability of misclassified vehicles, is analysed during training for a number of different values for each parameter. Other parameters are kept constant during the evaluation. A small classification error for a certain choice of parameter value suggests that the performance is optimal. However, for a correct optimisation, this implies that the other parameters are strictly independent of the evaluated one, which is not the case. Hence, the result of the optimisation is not decisive, but only taken into consideration when choosing parameter values.

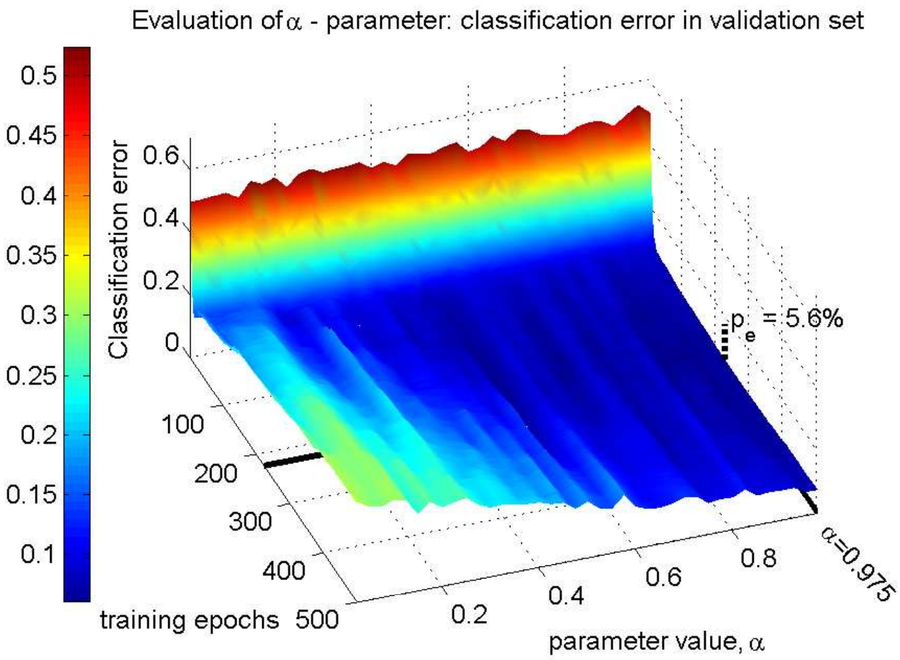
Implementing a certain parameter value, the perceptron is trained for a period of 500 training epochs. The procedure is then repeated 30 times and averaged over. The parameter N_{input} refers to the number of input elements presented to the perceptron and N_{hidden} denotes the number of nodes in the hidden layer.

Table C.1 shows at which value remaining parameters are kept constant, unless stated differently in the graph and figure legend.

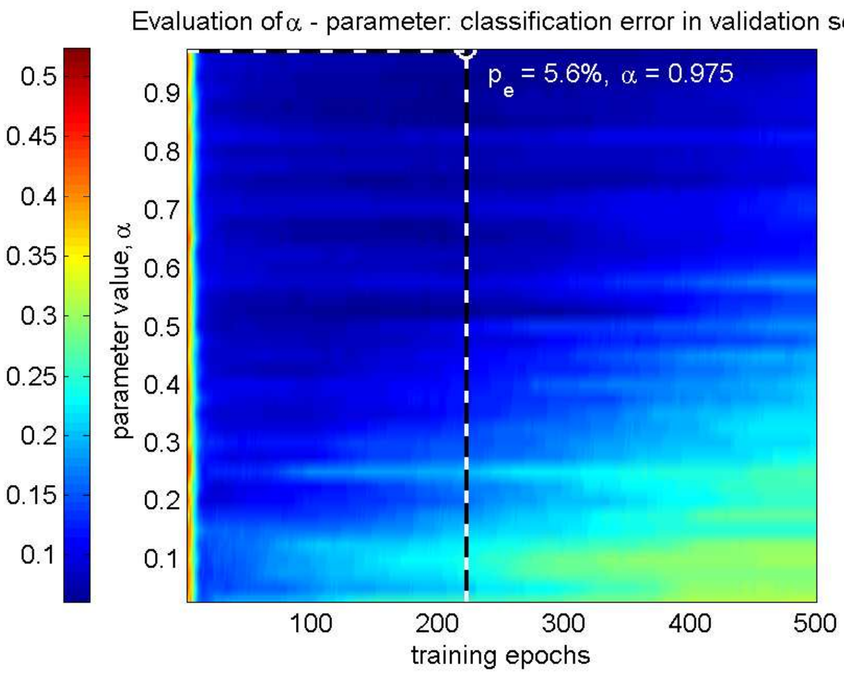
Parameter	*Value*
α	0.9
β	0.36
γ	0.05
η	0.08
N_{hidden}	60
N_{input}	70

Table C.1. Parameter optimisation: constant parameter values.

Note that figures denoted (a) contain the same information as figures denotd (b), only representation is changed (3-dimensional and 2-dimensional).

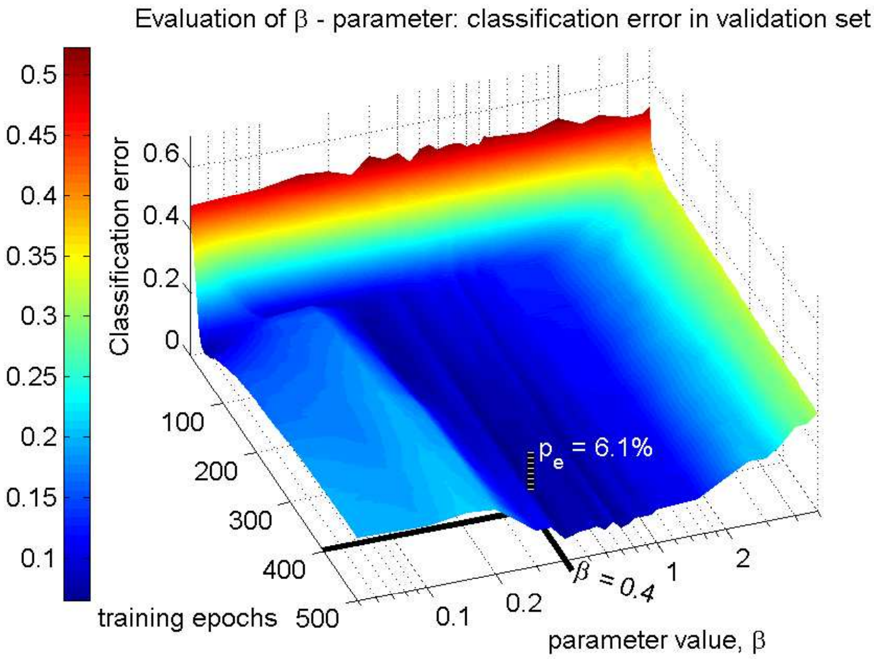


(a) 3-D

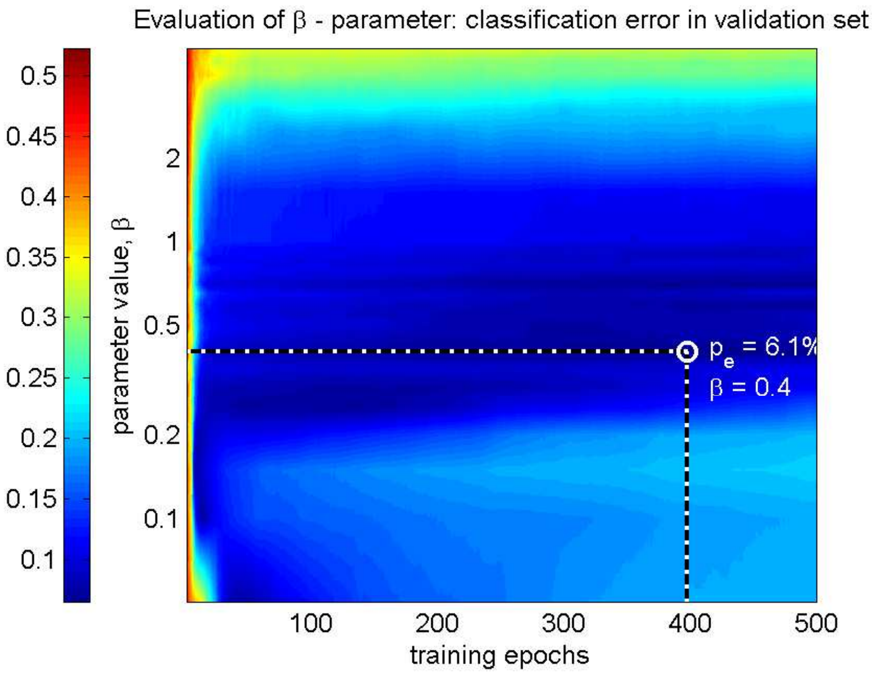


(b) 2-D

Figure C.1. Evaluation of parameter α; classification error in validation set.

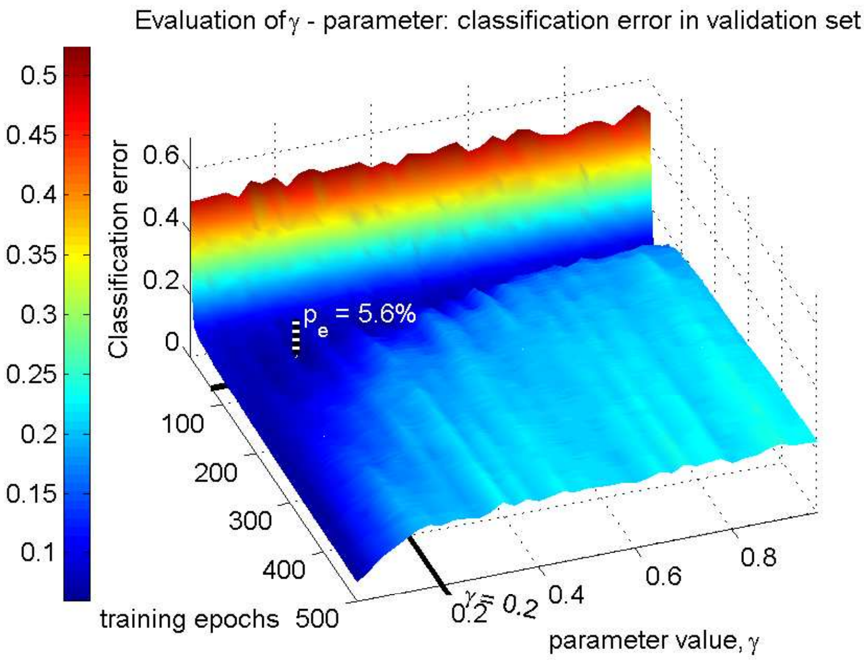


(a) 3-D

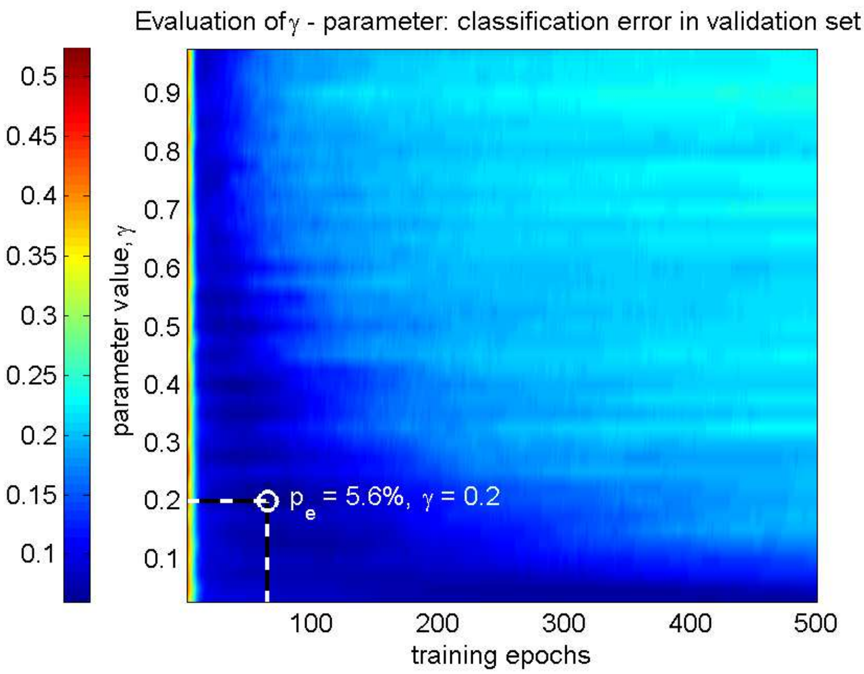


(b) 2-D

Figure C.2. Evaluation of parameter β; classification error in validation set. (N.B. the logarithmic scale on the β-axis.)

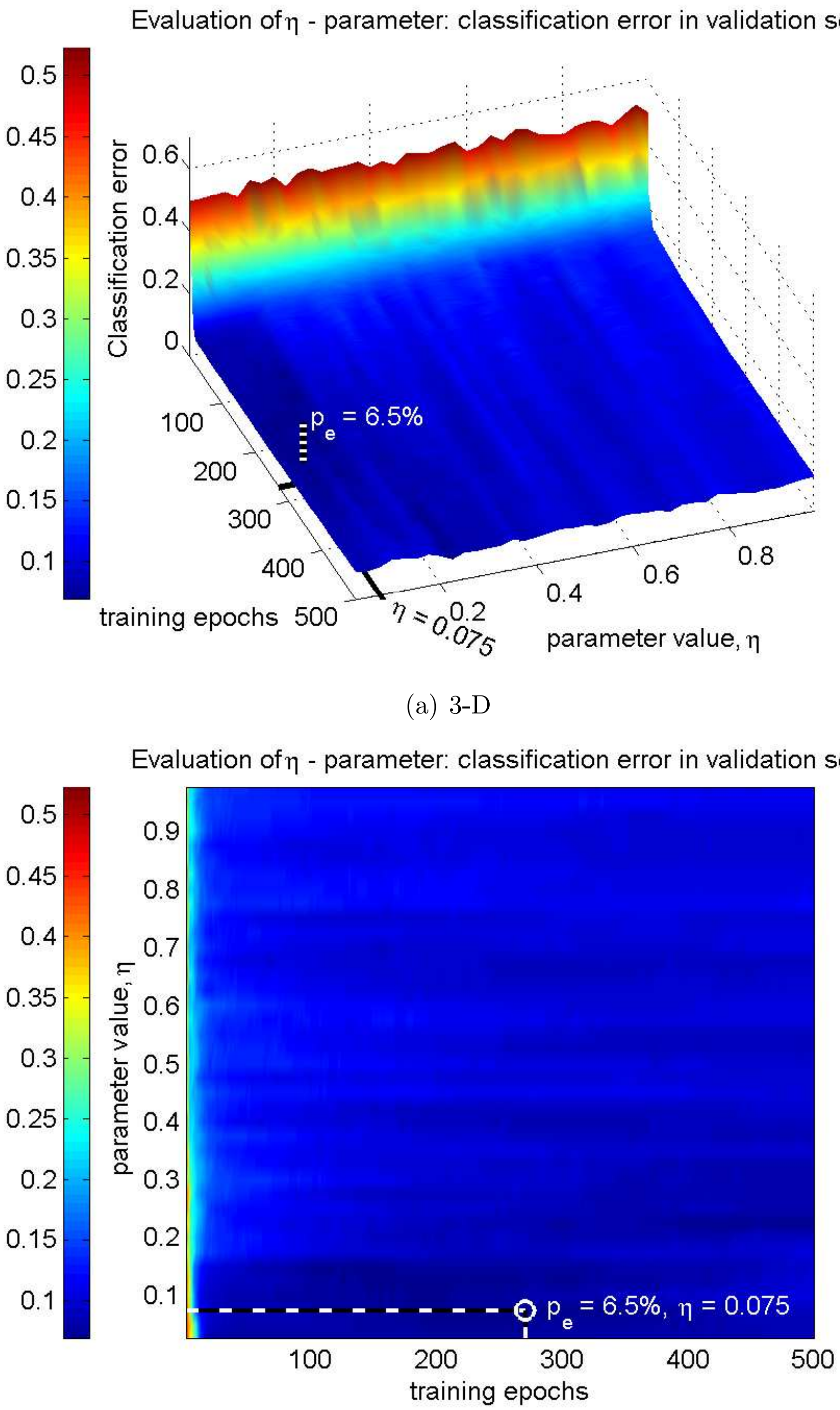


(a) 3-D



(b) 2-D

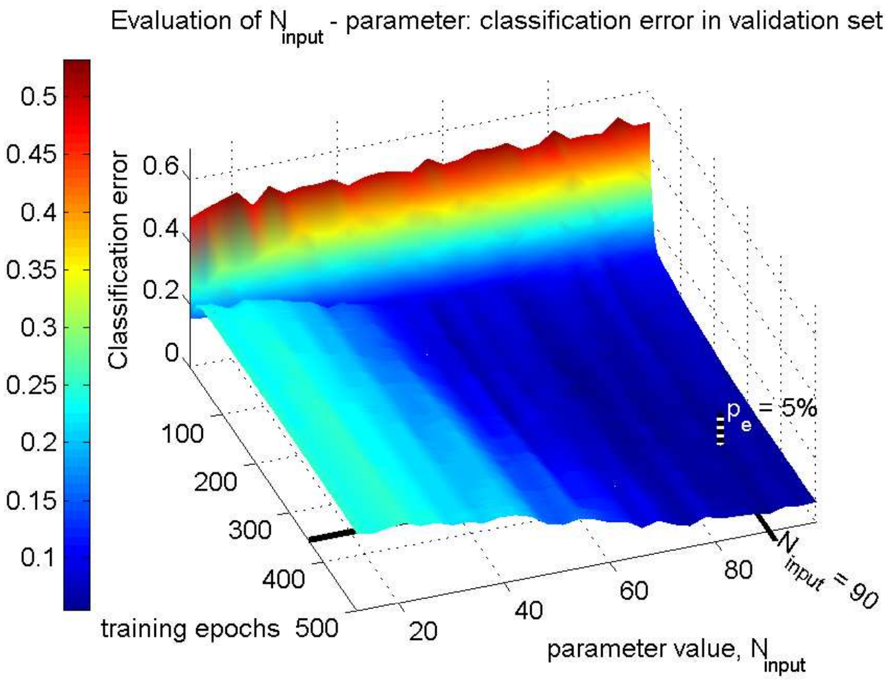
Figure C.3. Evaluation of parameter γ; classification error in validation set.



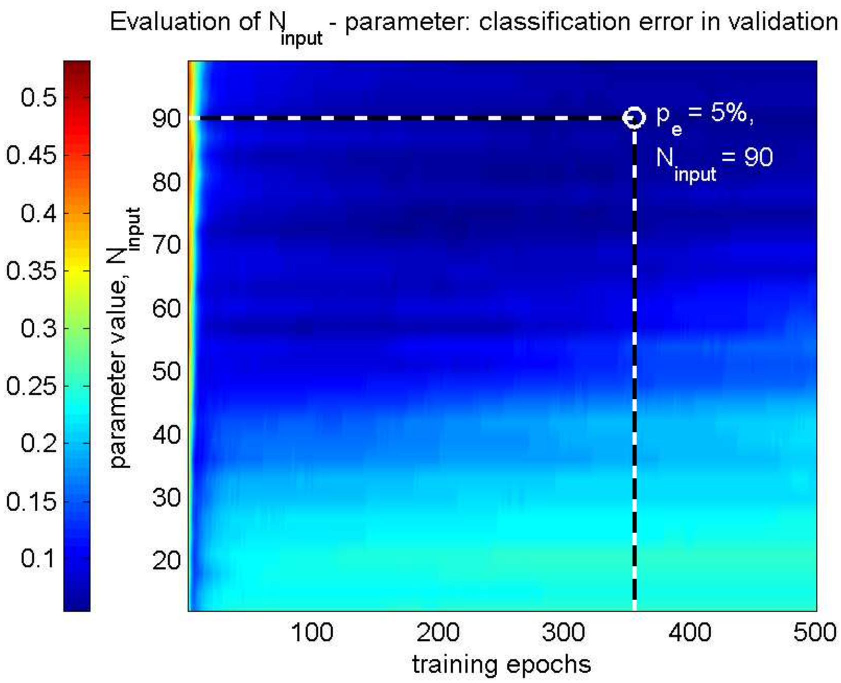
(a) 3-D

(b) 2-D

Figure C.4. Evaluation of parameter η; classification error in validation set.

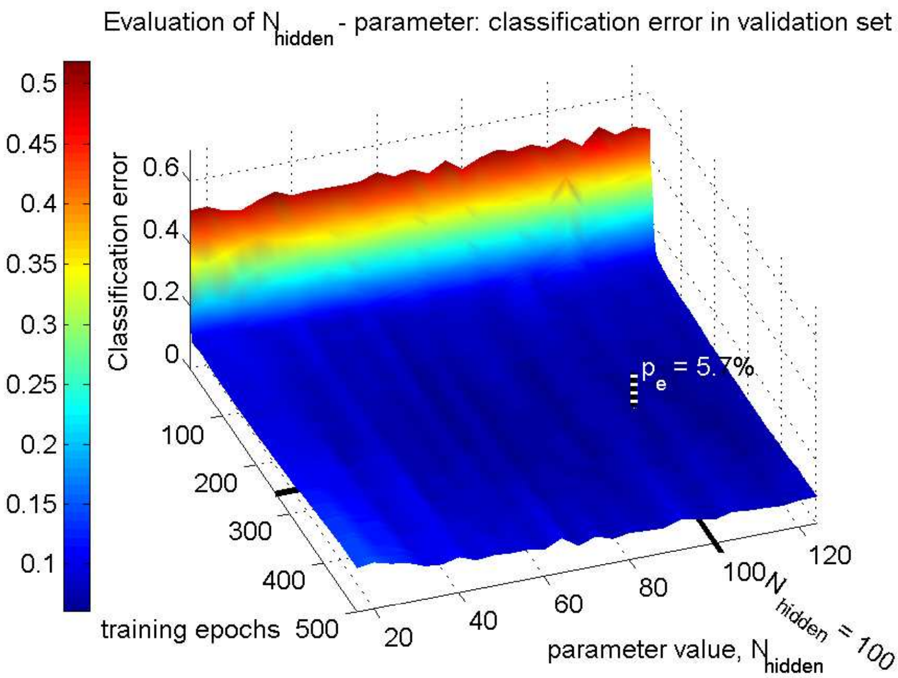


(a) 3-D

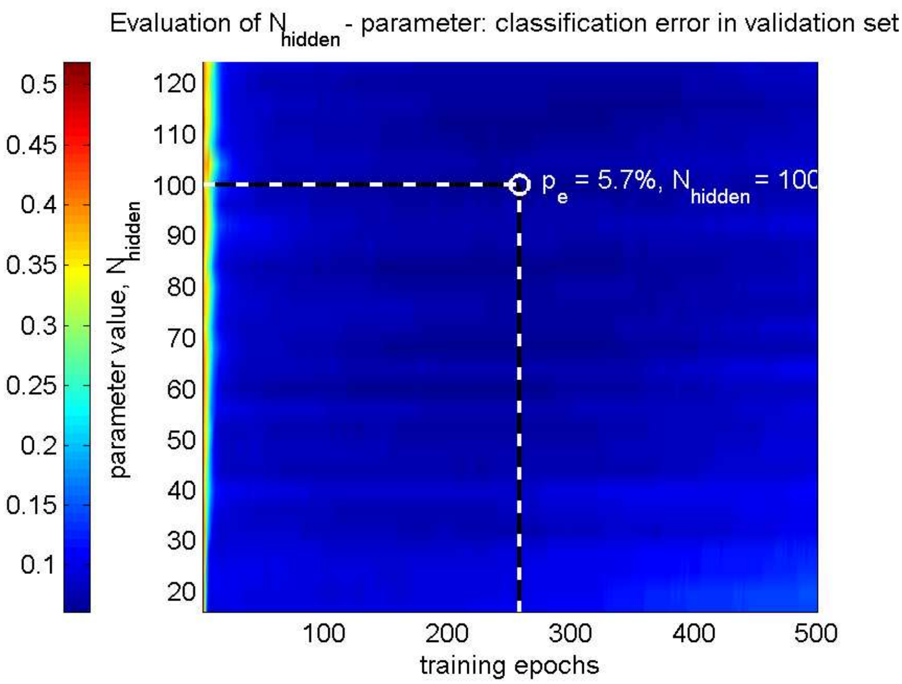


(b) 2-D

Figure C.5. Evaluation of parameter N_{input}; classification error in validation set.



(a) 3-D



(b) 2-D

Figure C.6. Evaluation of parameter N_{hidden}; classification error in validation set.

D NON TRAFFIC SOUNDS

To test the redundancy of the classification method, non traffic sounds are recorded and employed. These extra, non traffic sounds include sounds of laughter, shouting, music from a radio and sound effects such as sirens and dial tones. Recordings are made indoors with the Symphonie measurement system. A set of 39 extra sounds is sampled.

Date:	2005-11-29
Weather:	Indoors
Trigger level:	70 dB(A)
Recorded sample length:	5 s
Sampling frequency:	51.2 kHz
Location:	Ingemansson Technology AB

Equipment

Item description	*Manufacturer*	*Type*	*Internal notation*
Symphonie measurement system	Spektrum GmbH, 01dB		AL134
Microphone	G.R.A.S.	26AF	MK070
Rugged notebook	Panasonic Toughbook		D016
Calibrator	Brüel & Kjær		KU47

Table D.1. Equipment

Instruments are calibrated according to the Ingemansson quality standards which comply with the demands stated in SS-EN ISO/IEC 17025. Dates for the latest calibrations are listed in Ingemansson's calibration log.