Advanced Neural Network Head Movement Classification for Hands-Free Control of Powered Wheelchairs

Philip Taylor

This thesis is presented as part of the requirements for the award of the degree of Doctor of Philosophy from the University of Technology Sydney

2007

Certificate of Authorship/Originality

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student

Production Note:
Signature removed prior to publication.

Acknowledgements

First, I acknowledge the contribution of my supervisor, Professor Hung T. Nguyen. Undoubtedly, this research would not have reached its current point without his support, direction and knowledge. I also acknowledge the appreciable contribution of my co-supervisor, Professor Ashley Craig.

I acknowledge my colleagues from the UTS Faculty of Engineering and the Centre for Health Technologies, in particular Rob Mitchell, Sean Williams, Van Minh Tri Nguyen, Son Nguyen, Leslie King, Bradley Skinner and Daniel Craig for their camaraderie and assistance.

I acknowledge the input, assistance and cooperation received from Dr James Middleton and the staff and patients of the Moorong Spinal Injuries Unit (Royal Rehabilitation Centre, Sydney).

Finally, yet perhaps most importantly, I acknowledge my parents, Alan and Caralyn, my sister, Kathryn, my extended family and my friends for the assistance, support and encouragement they have provided.

Tables of Contents

Table of Figures

Table of Tables

Abstract

Assistive technology is increasingly being used to allow people with disabilities to be more engaged in personal, social and vocational activities. However, people with high-level disabilities are still affected by barriers to independence and full inclusion into broader society. The focus of this thesis is to advance the development of a wheelchair control system for highly disabled people, providing a means for satisfying some of the mobility needs of people who are have difficulty or are unable to achieve mobility through existing assistive technologies.

A control system allowing hands-free wheelchair control is proposed, advancing from systems previously described in literature. The proposed control system allows the operation of a powered wheelchair by using an artificial neural network (ANN) classifier to recognize head gesture commands. The feasibility of this control system is tested on its ability to correctly recognise command gestures of both able-bodied and disabled people in real-time.

Techniques for improving the ability of the head gesture classifier to recognise gestures performed by people with disabilities are investigated. The effectiveness of these techniques is evaluated for highly disabled people. The effect of empirically selecting an optimal ANN architecture and training algorithm using training data from a general population is considered, as is the marginal benefit of additional training data from such a population. The benefit of adapting the classifier using data from the specific end user is investigated as a means of further improving performance.

While demonstrating the feasibility of the proposed control system, the results presented showed that the performance of the system was lower for people with disabilities than for able-bodied users. It was found that selection of the ANN architecture, training algorithm and training set size all had significant effects of some degree on the ability of the classifier component to recognise command gestures by people with disabilities in real time. It was also found that data from a specific end user to train the ANN can significantly improve classifier performance.

It was found that adapting the classifier ANN using a combination of user specific and generic data could improve the performance of the classifier for that end user while minimising or avoiding any reduction in classifier performance for other people. It was also found that retraining such an ANN with user specific data alone improves the performance of the classifier for that end user but is detrimental to the classification performance for other people.

Notation and terminology

Table 1 Terminology used throughout this thesis

Term	Definition
SCI	Spinal Cord Injury
ANN	Artificial Neural Network
Delta rule	A training algorithm for artificial neural networks, also known as the gradient descent (Hagan 1995) or backpropagation (Haykin 1999) algorithm.
Tetraplegia	Paralysis of all four limbs, also known as quadriplegia
Generic dataset	A set containing recorded gestures or input-output pairs observed from the data provided by a group of people.
Specific dataset	A set containing recorded gestures or input-output pairs observed from the data provided by a single, known person.
PWM	Pulse Width Modulation, a modulation scheme where duty cycle of a square-wave carrier encodes the signal.

Chapter 1. Introduction

Head movement has been utilised by people with motor impairments in many applications, including communication and mobility. Assistive technology is increasingly being used to enhance the ability of people with motor impairments in these applications and to allow such people to perform tasks that would otherwise be impossible. Mobility is an important prerequisite for many key activities that contribute to a person's quality of life (Routhier et al. 2003). Mobility has been considered one of the major factors that determines a person's level of independence (CG Warren 1990), and in many circumstances, impairment in mobility can restrict participation in domestic, vocational and recreational activities (Brandt, Iwarsson & Ståhle 2004).

Wheelchairs are one of the most commonly used mobility aids. A manually propelled or powered wheelchair can be used to provide mobility to people with permanent conditions, such as spinal cord injury or brain injury, and progressive conditions, such as multiple sclerosis or muscular dystrophy (Routhier et al. 2003). Manually propelled wheelchairs are the most prevalent form. By contrast, powered wheelchairs are generally more expensive to procure and maintain, harder to transport and have limited range. However, for highly disabled people, the nature of their condition may preclude the use of a manually propelled wheelchair.

Powered wheelchairs have been used to meet the mobility needs of people with high-level disabilities for many years. Early appearances of powered wheelchairs in the literature can be traced back at least as far as the 1930's, employing electrical ('Gangway!: A New Style Vehicle Comes to Warm Springs.' 1932) and oil ('Gasoline Custer Chair' 1933) power. Polio was one of the most common conditions leading to the use of these early powered wheelchairs. Due to the high mortality rate for those people who would otherwise have developed the most severe forms of motor impairment and the cumbersome medical equipment used to treat those who survived, few of the population served by these early powered wheelchairs were affected by complete tetraplegia. These early powered wheelchairs were controlled using hand-operated mechanisms and more closely resembled modern motorised scooters than modern powered wheelchairs.

The proportion of the population affected by high-level motor impairments has steadily increased since the 1940s, as advances in medical technology and practices have reduced mortality rates due to injury and illness (Zola 1988). This has been followed by advances in assistive technology for this group of people, many using head movement employed mechanical or electrical methods to support or extend a user's ability to perform particular tasks. These advances in assistive technology have been incremental, reflecting both the gradual increase in the size and composition of the highly disabled population, changes in the target applications and the development of enabling technologies. For example, head controlled computer interfaces, such as those proposed by Evans, Drew and Blenkhorn (2000) and LoPresti and Brienza (2004) have evolved from head controlled typewriter adaptations proposed in the 1950s (Barg 1959; Ziskind 1959) to provide vocational tools to people with tetraplegia.

The use of assistive technology has allowed people with high-level disabilities to be more engaged in personal, social and vocational activities. However, people with high-level disabilities are still affected by barriers to independence and full inclusion into broader society (Wattenberg 2004). Apart from the need for assistance in many domestic activities, people with disabilities are less likely to be in fulltime employment and, if employed, are likely to earn a lower than average income (Kruse & Schur 2003; Wattenberg 2004). This can be attributed to several potential reasons other than employer discrimination (Schur 2003). One of these is that employment options for people with disability are restricted by the degree of mobility that is available to the individual, both geographically and vocationally.

The inability of existing powered wheelchair control systems to provide effective mobility for some disabled people can be attributed to the interaction of several factors. Some of these factors are related to the capabilities of the technology to allow a potential user to interact with the control system. Some other reasons arise from the socio-technical environment in which the individual lives. This thesis focuses on the interaction between the user and the control system.

Modern powered wheelchairs primarily use a joystick interface to control speed, direction and auxiliary features of the chair. Existing, widespread alternatives to a joystick interface include sip and puff controls, modified joysticks and switch arrays. However, some people find it difficult or impossible to use these existing interfaces to achieve sufficient control of a powered wheelchair in a domestic or office environment (Cooper 1995; Fehr, Langbein & Skaar 2000; Tzafestas 2001).

1.1. Aims and contribution of this thesis

The central aim of the research in this thesis is to develop a power wheelchair control system for highly disabled people. This wheelchair control system is to provide a method for the needs of people who are unable to easily operate a power wheelchair using existing techniques.

The main requirement of interfacing techniques is that they must use the abilities of the user to allow them to communicate their intentions. Existing alternatives to the joystick interface utilise the user's ability to move their head or control their breathing, such as the chin stick, mouth stick, switch arrays and sip-and-puff systems. Although existing interfaces are useful for many people, there remain significant numbers of people who are unable to use them. For such people, the nature of their impairment renders existing control systems unsuitable for a particular person due to the physical or cognitive loads necessary to achieve the individual's mobility needs, or due to lack of reliability, rate of communication, interference from the user's environment, or the inability of the control system to satisfy practical or cost-benefit requirements.

The control system that is the focus of this thesis is designed to allow people with high-level spinal cord injuries or motor impairments presenting similar symptoms to operate a powered wheelchair. People with such conditions represent a portion of those who are unable to use existing wheelchair control systems to meet their mobility requirements. By designing for the needs of this group, the findings of this research may also provide enhanced mobility options for people with less severe forms of disability.

The first major objective towards achieving this central aim is to demonstrate the feasibility of a control system for this target population. The design of a control system is proposed, advancing from systems previously described in literature. The proposed control system makes use of an artificial neural network (ANN) to recognize commands given by head gestures. Assuming that the user is able to adequately perceive their operating environment and is able to determine commands necessary to reach their destination, the suitability of the proposed control system is determined by the ability of the user to perform command gestures and the ability of the control system to recognise such gestures. The ability of the control system to recognise command gestures performed by disabled people is used to experimentally examine the feasibility of the control system.

The second major objective towards achieving this central aim is to show that the performance of the control system can be improved using data from a group of people that does not include the end user. The degree of control with which a person is able to control the course of a power wheelchair using the proposed control system depends on the accuracy with which command gestures are recognised. This consequently affects the range of navigation tasks that a user can perform in a safe and practical manner. To maximise the utility of the control system, it is therefore important to optimise the components of the system that perform the recognition of commands. The effect of such optimisation will naturally differ between individual users due to physiological and behavioural differences, but there is a range of techniques that can be applied to optimise the performance of the control system in spite of these differences.

The third major objective towards achieving the central aim of this thesis is to show that the performance of the control system can be improved by adapting the control system for the end user. As noted previously, physiological and behavioural differences between users cause the variations in the performance of the control system between users. When optimising the performance of the control system without using user specific feedback or data, these physiological and behavioural variations tend to lead to degradation of system performance, relative to the performance that could be obtained by fully customising the system to the end user. Instead of pursing the full customisation of the control system to the condition of each potential user, this thesis investigates the benefit that can be obtained by a partial customisation.

1.2. Structure of this thesis

Chapter 2 reviews literature relevant to the formation of the research questions investigated in this thesis. From an examination of the features necessary for a practical wheelchair control system, this chapter goes on to consider the properties of existing and proposed wheelchair control methods.

Chapter 2 also states the research questions considered in this thesis in more detail and summarises the background and methodology of the experiments presented later chapters. It aims to establish the specific research questions posed in this thesis and theoretical framework by which these questions can be addressed.

Chapter 3 presents a prototype wheelchair control system designed for users with high-level motor impairments. The control system presented advances from systems previously at UTS, such as that described by Joseph and Nguyen (1998). The prototype control system allows a user to issue commands to the control system by gestures of the head. The current literature indicates a lack of experimental results for disabled users for this type of control system, leaving an open question as to whether head gestures are an appropriate technique for the target population of disabled people. To resolve this question, experimental results are presented examining the ability of the prototype to correctly identify the head gestures of able-bodied and disabled people in real-time. These results seek to demonstrate that the prototype control system is able to provide a practical interface for some members of the target population. It also examines the degree to which the performance of the prototype control system differs between able-bodied and disabled people.

There are many factors affecting the ability of the prototype to accurately detect and classify head gestures. Chapter 4 investigates the degree to which the real-time classification performance of the prototype can be improved by making changes to several key factors. This chapter focuses on the classifier component of the prototype control system, seeking to optimise the performance of the classifier for disabled users. In particular, Chapter 4 examines the degree to which classifier performance can be improved using methods that do not require user-specific data or feedback prior to the implementation of the controller. Experiments to determine the effect of several factors for disabled users are presented.

Results in Chapters 3 and 4 demonstrate the performance of classifiers implemented without the use of user specific data or feedback. The effect of utilising user specific data in the implementation of the classifier is not considered in these chapters, nor is there sufficient evidence in the literature to support or discourage the use of user specific data in this application. It remains an open question whether the use of data from a specific user in the implementation of the classifier can be associated with a change in the classifier performance for that user. Further, if such a difference can be found to exist, it is also an open question as to whether it is significantly different from the change can be found to be associated with the other factors. These questions are addressed in Chapter 5. Experiments are presented showing the effect of the use of user specific data during the classifier implementation on classifier performance for two highly disabled users.

Chapter 6 discusses the experimental results presented in the preceding chapters and summarises the conclusions drawn from these results. It considers the extent to which the questions addressed in the preceding chapters can be considered to be answered, summarises the contribution of this thesis and proposes several avenues of investigation which extend from the results presented in this thesis.

Several appendices are included at the end of this thesis. These appendices deal with matters that are not central to the research questions considered in this thesis, but are nonetheless pertinent to the methodology or results contained in the main text. The first four appendices provide background on specific issues in the classifier optimisation and experiment design in Chapters 5 and 6. The fifth appendix relates to the human research ethics approval under which the experiments in this thesis were conducted and the following appendices provide specifics of the statistical results obtained in Chapters 5 and 6.

1.3. Publications

At the time of writing, ten papers have been published on the prototype wheelchair controller described in Chapter 4 and subsequent advancements.

1.3.1. Fully refereed papers

Taylor, P.B. & Nguyen, H.T. 2003, 'Performance of a head-movement interface for wheelchair control', *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, Cancun, Mexico, vol. 2, pp. 1590-1593 Vol.1592.

Nguyen, S., Nguyen, H. & Taylor, P. 2004, 'Hands-free control of power wheelchairs using Bayesian neural network classification', *Cybernetics and Intelligent Systems, 2004 IEEE Conference on*, Singapore, vol. 2, pp. 746-750.

King, L.M., Nguyen, H.T. & Taylor, P.B. 2005, 'Hands-free Head-movement Gesture Recognition using Artificial Neural Networks and the Magnified Gradient Function', *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, Shanghai, China, pp. 2063-2066.

Nguyen, S.T., Nguyen, H.T. & Taylor, P.B. 2006, 'Bayesian Neural Network Classification of Head Movement Direction using Various Advanced Optimisation Training Algorithms', *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*, Pisa, Italy, pp. 1014-1019.

Nguyen, S.T., Nguyen, H.T. & Taylor, P. 2006, 'Improved Head Direction Command Classification using an Optimised Bayesian Neural Network', *IEEE*

International Conference of the Engineering in Medicine and Biology Society, New York, USA.

Nguyen, H.T., Nguyen, S.T., Taylor, P.B. & Middleton, J. 2007, 'Head Direction Command Classification using an Adaptive Optimal Bayesian Neural Network', *International Journal of Factory Automation, Robotics and Soft Computing*, In Press, Paper ID 365.

1.3.2. Other conference papers

Taylor, P.B., Nguyen, H. & Craig, A. 2002, 'Head Movement Recognition for Power Wheelchair Control', *Engineering and Physical Sciences in Medicine*, Rotorua, New Zealand, p. 135.

Nguyen, H.T., Legaspi, S., Knight, G., Ekanayke, R., Taylor, P.B. & Martinez-Coll, A. 2002, 'A head movement system for environmental control units', *Engineering and Physical Sciences in Medicine*, Rotorua, New Zealand, p. 157.

Taylor, P.B. & Nguyen, H. 2003, 'Neural network classification of head-movement for wheelchair control', *World Congress on Medical Physics and Biomedical Engineering*, Sydney, Australia, pp. 3953 (CD-ROM).

Taylor, P.B. & Nguyen, H. 2004, 'Adaptive training of neural network classifiers for power wheelchair control', *Engineering and Physical Sciences in Medicine*, Geelong, Australia, p. 012OSF.

Chapter 2. Literature review

This chapter reviews literature on the inadequacies of existing powered wheelchair control strategies. Alternatives to existing control strategies are also reviewed, identifying the factors preventing the widespread use of these techniques in the present environment. The use of head movement as a control strategy is given particular examination, with the objective of identifying literature related to the research questions that will be addressed in the chapters that follow.

This chapter also states the research questions considered in this thesis in more detail and summarises the background and methodology of the experiments presented later chapters. It aims to establish the specific research questions posed in this thesis and theoretical framework by which these questions can be addressed.

2.1. Wheelchair control strategies for people with high level disabilities

The hand-operated joystick interface is widely held to be the standard interface for a powered wheelchair. The unsuitability of joystick interfaces for a significant number of potential wheelchair users has been commented on widely in the literature. Lack of mobility, lack of muscular force and excessive spasticity were listed by Pruski and Bourhis (1992) as factors that prevent the use of joysticks or other classical proportional sensors. In addition to these factors, LaCourse and Hludik (1990) noted that people with severe handicaps are unable to effectively use switches and scanning devices due to the motor coordination required and slow response times. Miller and Slack (1994) state that for users who have difficulty using a conventional joystick to operate a wheelchair, the limited rate of communication between the user and the wheelchair control system limits the speed at which they can safely travel.

More recently, Cooper et al (2000) state that current powered wheelchair interfaces do not provide adequate mobility for some people with physical disabilities. Fehr, Langbein & Skaar (2000) state the conclusion that "no independent mobility options exist at this time" for a significant number of persons with disabilities, including high level spinal cord injury and nervous system disease. In addition to this, Fehr et al conclude that there is evidence suggesting that people who are able to operate a wheelchair using existing techniques have difficulty with steering tasks, including some for whom such tasks are impossible to perform without assistance.

Jonkers et al (2004) state that there are few functional scales available for the use of powered wheelchairs, and further state

> "There appears to be no literature reporting motor requirements to operate a joystick interface for steering a powered wheelchair." (Jonkers et al. 2004, p. 930)

The results reported were that fine motor control tasks took significantly longer to perform and were described by the users as the most difficult to perform. Jonkers et al used surface electromyography to measure the muscle effort and fatigue occurring across a set of reference and functional tasks performed by 10 subjects diagnosed with multiple sclerosis. These measurements identified profiles for the activity of several muscle groups during the tasks. Although the results presented indicate muscle groups employed by wheelchair users with multiple sclerosis to operate a joystick, the results do not directly identify factors that reduce a user's ability to use such an interface. Jonkers et al do not set out a scale or criteria regarding the motor requirements to use a powered wheelchair.

Although the use of joysticks, switch arrays and alternate sites provide viable options for some people, there remains a significant number who are unable to effectively use a wheelchair using these interfaces, as noted by Fehr, Langbein & Skaar (2000). In order to advance from these existing interfaces, a range of interfaces has been proposed to allow the user to issue commands by gestures. The type of movement required varies among the proposed interfaces, as do the techniques used to measure the location of the relevant part of the user's body and the identification of command gestures.

In the most general terms, a wheelchair control system must harness some form of voluntary action of the user to determine the user's intention. This action or set of actions can be used to generate analogue control signals, such as those produced by a joystick, or discrete control signals, such as those produced by a switch (CG Warren 1990). In a review considering the set of characteristics which have been identified as important determinants of the quality of an assistive technology, Thorkildsen (1994) listed 17 features from 29 papers. Eight of these characteristics were flagged as being most important, based upon the frequency with which they were proposed in the review, largely coinciding with a list of characteristics identified by Batavia and Hammer, which was also included in Thorkildsen's review.

Dependability and durability are characteristics defined by Thorkildsen (1994) as being due to the technology itself, rather than the user's interaction with it. According to Thorkildsen, dependability is used to refer to the extent to which a device operates with repeatable or predictable accuracy, across the range of conditions in which it can be expected to operate, and durability is used to refer to the extent to which a device can be expected to operate for extended periods of time. In defining dependability, Thorkildsen refers only to the properties of the device and variations in its operating environment. In the case of wheelchair control, the concept of dependability can be extended to include short-term variations in the abilities or behaviour of the user, as is advocated by Dewsbury et al (Dewsbury et al. 2003; Dewsbury, Taylor & Edge 2001, 2002). Similarly, the concept of durability can be extended to include long-term variations in the abilities or behaviour of the user.

The extent to which the user can easily operate a system was the most frequently cited desirable characteristic identified by Thorkildsen (1994). Thorkildsen termed this characteristic operability, although it is synonymous with several terms, such as "convenience", "ergonomics" and "simplicity of use". Operability is highly dependent on the particular needs of the set of users for whom a system is designed. The physical and mental requirements on the user are key in determining the population who will be capable of using a particular control system to operate a wheelchair. The physical requirements of many forms of wheelchair control system make it impossible for some people to produce the necessary control signals, due to the level of motor skills required. Depending upon the condition of the user, particular types of action may be difficult, uncomfortable or impossible to perform due to lack of strength, coordination or reproducibility, or pain or fatigue.

The physical requirements of a control system set a limit on the rate at which a user can issue commands or adjust the value of a continuous control signal, thereby affecting the degree of control that the user is able to achieve. The mental requirements placed on the user arise from the need for a user to communicate their intention to the control system. Many types of control system use a communication protocol involving a series of control signals in order to allow a greater number of selectable commands or to allow differentiation between actions by the user that are intended to be control signals and those that are not. The degree to which the actions performed by the user to communicate with the control system are a natural extension of communication methods, the provision of proprioceptive or external feedback, the precision with which control signals must be timed and the rate at which signals must be produced in order to maintain control of the wheelchair are also factors that determine the mental load on the user.

Affordability, being the extent to which the user is able to incur the initial or going costs of a device without financial difficulty or hardship, was also a frequently cited characteristic identified by Thorkildsen (1994). In the context of wheelchair control systems, the cost of the wheelchair itself is generally quite significant. Warren (1990) noted the sensitivity of third parties to the price of wheelchair systems, relating this to the tendency for systems to be less effective for some users than was expected and the resulting reduction in the benefits of the expenditure on the system.

Ease of maintenance, relating to the actions necessary by a user of attendant to keep a device safe and operable, is included by Thorkildsen (1994). Although linked with durability, the ease of maintenance differs in that it is related to the actions required by or on behalf of the user, rather than the variation in performance over an extended period.

The compatibility of a wheelchair control system with devices that a disabled user may use and the degree of flexibility available for the selection of optional features for a device are also identified by Thorkildsen (1994) as important features. A wheelchair control system using discrete commands or continuous control signals could conceivably be generalised to a wide range of applications. Computer interfacing and environmental control are two commonly cited related applications. Consequently, in the context of this review, these characteristics are outside the scope of the issues to be considered as many aspects of compatibility and flexibility are specific to particular implementations of a technology.

2.2. Hands-free powered wheelchair control strategies

In order to review the wheelchair control interfaces proposed in the literature, the systems proposed will be categorised by the method employed to allow the user to issue signals to the control system.

Some authors have proposed the combination of several methods in order to allow a greater rate of information transfer between the user and the control system, such as Coyle (1995), or to create redundancy, such as Simpson and Levine (1997). For the purposes of this review of interface methods, each method will be considered separately.

2.2.1. Modified joystick interfaces

Modifications to the construction or processing of signals have been proposed in order to improve the operability of the conventional joystick interface. Common to each of these interfaces is a requirement that the user has some degree of voluntary movement, although the techniques vary in the reproducibility, range and type of movement required.

Alternative sites for joystick and switch array input devices have been identified by several authors and are common in conventional existing interfaces. In these alternative sites, the input device is modified to allow the user to operate the wheelchair by moving a different part of the body, such as the chin, mouth, shoulder or foot. Several authors have identified disadvantages relating to the use of a joystick or switch array in an alternative site. Coyle (1995) stated that physically activated devices positioned around the users face or head intrude upon the user, creating problems with the operability of the interface. Further, Coyle (1995) and Chen (2001) noted that orally activated devices have sanitation risks, requiring that the parts in contact with the user must be regularly cleaned or replaced. Despite the problems related to the use of these sites, these sites are employed in existing control systems for many users, as is indicated by Warren (1996), due largely to the need for some degree of mobility and the lack of existing alternatives available in the market.

Isometric joysticks have been proposed and tested by several authors as an alternative type of joystick that may benefit some users. This type of joystick produces an output related to the force applied to the joystick rather than the conventional output related to the position of the joystick, hence requiring a smaller range of movement to operate. Rao, Seliktar and Rahman (1999) found that positional joysticks provide significantly superior control relative to isometric joysticks when tested on cursor pointing tasks. In discussing their findings, it was suggested that the isometric joystick increased tremor related artefacts in the signal, while the friction and momentum of the users arm and hand reduced these artefacts for the positional joystick.

Cooper et al (2000) reported that an isometric joystick was used to achieve superior performance for some driving tasks, both for able-bodied and disabled users. The task for which the superior performance was reported for isometric joysticks were those involving driving in a straight line and turning in a full circle. Despite this, superior performances were reported for positional joysticks during other tasks. In a separate paper Cooper et al (2000) used a different methodology and reported results that a significant difference existed only for movements involving driving forwards along a curving path. Since the subjects in these papers were able to use either positional or isometric joysticks in a handheld configuration, it is difficult to meaningfully extrapolate to potential subjects for whom handheld joysticks are not a viable control interface. Guo et al (2002) reported that isometric joysticks can be used in a chin-operated configuration, but did not present results measuring the performance of such interfaces, nor compare this interface with others.

In order to allow people with tremor to operate a wheelchair via a joystick, a number of signal processing techniques have been proposed. Corbett and Martinez (1998) state that a common technique to deal with mild tremor is to apply a time-delay mechanism or low-pass filter to the input from the user. This method was noted to slow the wheelchair down and adversely affect the ability of the user to perform some navigation tasks. Corbett and Martinez also state that, for joystick interfaces, more severe tremors are primarily dealt with by mechanically restricting the range of directions in which the user can move the joystick, which has the disadvantage of making it difficult to navigate the wheelchair smoothly or consistently.

Fuzzy logic based filtering techniques have been reported to improve performance for users with light or moderate tremors, but are not useful for severe tremors (Corbett & Martinez 1998; van der Zwaag, Corbett & Jain 1999). Corbett and Martinez suggest an extension of the technique they propose would be to combine an artificial neural network with fuzzy logic to create a neuro-fuzzy controller, with the purpose of allowing the fuzzy membership set to adapt to a particular user. Ding, Cooper and Spaeth (2004) also presented results that fuzzy logic can be used to filter out tremor signals, and also indicated that there is a need to adjust the parameters of such filters.

Joystick interfaces have many benefits, including a low cost, durable construction and a low requirement for maintenance in most sites. Provided that the user is capable of meeting the physical requirements of operating a joystick, the control action can be intuitive and accurate. The primary factor preventing or limiting the use of joysticks for potential users is that the physical requirements of such an interface are too high, as the condition of many highly disabled individuals prevents the range of movement necessary to effectively operate the wheelchair, as was noted earlier with regard to existing wheelchair control systems.

2.2.2. Bioelectric input signals

Several types of bioelectric signals have been considered as a means of issuing command signals to a wheelchair control system. Electroencephalography (EEG) has been explored as a means to achieve a brain-computer interface (BCI) in many applications, including wheelchair control. Electromyography (EMG) and surface electromyography (sEMG) have been also been investigated for the purpose of wheelchair control.

Electroencephalography (EEG) has been speculatively considered as a potential mechanism for communication and control for many years. Research on EEG-based BCI has grown rapidly since 1995, although the topic has been actively investigated for communication and control since at least the 1970's (Wolpaw et al. 2000). Wolpaw et al cite the work of Vidal (1977) as an example of this early research, which highlighted the importance of being able to distinguish between EEG signals, which are typically measured on the scalp in the range of 0 to 200 μV, from EMG signals arising from scalp or facial muscles.

Most literature on BCI focuses either on the development of a generic communication tool, equivalent to one or more binary switches, or on applications such as text entry and environmental control. The same techniques can be generalized to use in a wheelchair application in a manner similar to the use of a switch array, such as the system proposed by Tanaka, Matsunaga and Wang (2005). In comparing BCI communication tools, Wolpaw suggests the use of bit rate as a measure of performance, as this reflects both the speed and the accuracy of the interface, as well as the number of control actions available to the user. Several functional BCI have been proposed in the literature, although Wolpaw notes these as having a relatively low bandwidth, in the order of 5 to 25 bits per minute. One method for the implementation of a BCI is to train users to control their brain waves, for example, by performing particular mental tasks or by using biofeedback. An alternative to this is to use the response of the brain to external stimuli, such as the use of evoked potentials or event related potentials.

Anderson, Devulapalli and Stolz (1995) propose a BCI allowing the user to control a binary selection by performing two mental tasks: a relaxed baseline and performing mental multiplication. Anderson, Devulapalli & Stolz justified the use of this method by noting that contemporary systems using event related potentials were unsuitable due to the device creating the external stimuli being excessively slow. The interface proposed by Anderson, Devulapalli and Stolz was not customized for particular users, using an ANN trained on all subjects to classify EEG data. The results reported by Anderson, Devulapalli and Stolz showed that the performance of the ANN depended heavily on the pre-processing of the EEG data, with some representations of the input patterns resulting in error rates of up to 49%, close to the 50% error rate attributable to chance alone. The best results obtained were from a frequency-based representation, which was reported to have achieved an error rate of 26%. This BCI was later refined (Anderson, Stolz & Shamsunder 1998), where it was noted that the optimal classification method was not the same for all users. The BCI tested by Anderson, Stolz and Shamsunder used the same mental tasks as the Anderson, Devulapalli and Stolz, and was reported as capable of an error rate of less than 10%, which was described as being marginally acceptable for use in a real-time system.

McIsaac et al (2002) proposed a BCI using a single control action, the closure of the eyes, in an environmental control system (ECS). The system described by McIsaac et al allowed the user to select and control devices using a scanning protocol, having evolved from a system designed for activating or deactivating a single device. This system was tested with both able-bodied and highly disabled subjects. It was reported that both able-bodied and disabled subjects were able to operate the ECS in performing a set of tasks. The error-rate of the interface was reported to be 1.8 per 5 trials for disabled subjects, although it was noted that this was reduced by almost half as subjects become more experienced with the system. Time taken to perform an environmental control task was reported to be approximately 30s, much of which was due to the time required for the scanning protocol to display the necessary option. McIsaac et al indicated that more efficient noise suppression and signal processing would improve the BCI. It was also indicated that the effect of drugs, fatigue and variation in environmental conditions warranted further investigation, a comment that could be applied generally to BCI systems.

Customisation of a BCI to a particular user has been proposed in order to deal with the variations in physiology and psychology between individuals (del R. Milan et al. 1998). Unlike the systems proposed by McIsaac et al and Anderson, Devulapalli and Stolz , which used signal processing and classification techniques designed to cater for the general population of users, the system proposed by del R Milan et al used a hierarchical structure of ANN, trained entirely on data specific to the user. The function of this hierarchical structure was to improve the specificity of the BCI. Del R Milan et al commented that an overall error rate of 20% to 30% would be acceptable if the false positive rate were negligibly small, thus resulting in a specificity close to 1. The overall results of the BCI proposed by del R Milan et al were not reported, although it was shown that the method described were able to achieve an error rate of 25% with a false positive rate of 1.6% for at least one of the three mental tasks. Del R Milan et al remarked on the potential to apply these methods to the control of a wheelchair, using the three mental tasks to generate discrete commands for forward, left and right, although no details nor results were presented.

In presenting a prototype wheelchair controller, Tanaka, Matsunaga and Wang (2005) commented that there were no reports of wheelchair control solely using EEG in the literature. The system proposed by Tanaka, Matsunaga and Wang used two mental actions, corresponding to left and right, to control the wheelchair. EEG signals were classified using a recursive training algorithm with error rates ranging from 9% to 53% on individual patterns, with an average error rate of 21%. An experiment on the ability to control a wheelchair using this BCI was also reported. This experiment was conducted in highly structured conditions, being to test the ability of the subject to navigate a wheelchair to one of two nearby goals in an otherwise empty room. The results reported for this experiment were the subjects were able to successfully perform this task in approximately 80% of trials, although there was significant variation between subjects.

Serby, Yom-Tov and Inbar (2005) propose a BCI using event related potentials in a text entry application. A computer screen was used as the source of external stimulus, comprising a grid of 36 symbols. The rows and columns of this matrix were intensified in a random sequence. To select a symbol, the user was required to mentally count the number of times that the symbol was intensified, with the BCI classifying EEG signals to recognise the resulting event-related potentials and correlate these with the intensification of the symbols in the matrix. In an experiment involving 6 able-bodied subjects, it was reported that the BCI was able to recognise the correct symbol with an error rate of 9% in offline experiments and 20% in online experiments, with approximately 4.5 symbols per minute in both cases. The bit rate reported for the online experiment was 15 bits per minute. Serby, Yom-Tov and Inbar used event related potentials as a means to reduce the degree of training required to allow users to operate the BCI while still allowing sufficient performance for the application. Although aiming to remove the need for the user to be trained to use the system, the results presented indicated that the error rate and bit rate of the BCI varied with the number of repetitions performed by the user.

The physical load placed on the user by an EEG based wheelchair control system is the lowest of all approaches to wheelchair control reviewed in this chapter. For this reason, EEG has been proposed as a control mechanism for severely disabled individuals, such as those with high-level spinal cord injuries or amyotrophic lateral sclerosis. However, this low physical load does not make EEG interfaces suitable for all potential users. The difficulty of compensation for EMG and EOG artefacts in analysing EEG signals prevents the use of such interfaces by individuals whose condition results in an excessive degree of involuntary action of facial muscles. Further, the condition of some individuals, such as stroke victims, results in changes to the physiology of the brain, thus affecting their EEG patterns.

The operability of EEG based control systems depends heavily on the bit rate that the interface is able to achieve and the mental load placed on the user. The response time of EEG systems reported in the literature is slower than other interface techniques, and is generally too slow to be used in a real-time wheelchair control system without some form of navigational assistance in order to operate in all but the most tightly structured operating environments. The mental load placed on the user is most significant in systems requiring users to exercise control of their brain waves, as doing so requires several stages of action. Since it is difficult to classify the EEG signals resulting from mental actions that intuitively correspond to particular commands, the user is required to memorise the mental actions used to encode these commands. This adds a layer of complexity to the operation of a wheelchair using this form of interface.

The dependability of EEG based wheelchair control systems is affected by changes in the physiological and psychological state of the user. As was observed by several of the above authors, the performance of the interface can change as the user adapts to the system, as well as changes in the user's condition or operating environment. Wolpaw et al (2000) and Tanaka, Matsunaga and Wang (2005) state that adaptation of the interface during its operation can allow the performance of the system to be maintained, although long term adaptive protocols were not proposed. Despite being recognised as a problem as far back as 1977 (Vidal), interference from EMG, EOG, EEG signals other than command actions and external sources is still a significant issue. Although sophisticated filtering and signal processing techniques have been proposed to eliminate or mitigate the effects of these sources of interference, these techniques are not completely effective. Anderson, Stolz and Shamsunder (1998) noted that these artefacts do not necessarily degrade the performance of the interface, and in fact may be helpful if they are correlated with the signals being recognised, although no indication was given as to how often this was the case. Tanaka, Matsunaga and Wang (2005), taking a pragmatic but limited approach to this problem, requested that users minimise all forms of voluntary movement while operating the wheelchair interface.

The durability, cost and ease of maintenance of EEG based wheelchair control systems are not discussed in the literature, due to the lack of systems that are sufficiently developed to be considered close to being a practical implementation.

Electromyography (EMG) has also been proposed as a bioelectric input signal for wheelchair control systems, although it has received considerably less attention than EEG. EMG measures the electric potential generated by the contraction of muscle cells, and can be measured by implanted, transcutaneous or surface electrodes. The magnitude of surface EMG (sEMG) signals from muscle in the forearm was used as a continuous control signal in a computer pointer application by Rosenberg (1998), using an ANN to translate the signals to motion of the pointer. Rosenberg noted that the performance of this interface was markedly lower than other pointing devices, but did not discuss the appropriateness of the device for people with disabilities. Since the response time and bit rate of EMG-based systems is generally superior to those based on EEG, the need for assisted navigation would be reduced.

Muscles in the face and neck have been more popular in research on aids for people with disables, as voluntary control over these muscle groups is more likely to be available to highly disabled individuals. SEMG from the sternocleidomastoid muscles have been used in research on wheelchair control by Martinez-Coll, Papacosta & Nguyen (2003a; 2003b) and Han et al (2003), while Williams and Kirsch (2004) used facial muscles in a generic neural prosthesis interface.

The use of EMG as a control input requires relatively low levels of physical ability on the part of the user. While it is necessary that the user is capable of voluntary, controlled contraction of the relevant muscle groups, this does not require the user to be capable of controlled movement using that muscle group. Han et al (2003) noted that fatigue can affect the level of control an individual is able to achieve using EMG, particularly when using control protocols requiring that muscles be contracted for extended periods.

Selection of muscle groups that are used in conventional gestures or pointing movements allows EMG to be used with control actions that are more intuitive than other bioelectric interfaces (Han et al. 2003). Muscle groups can also be selected to avoid or minimize interference between the use of the control interface and other actions by the user, although the scope for this is diminished for highly disabled users (Williams & Kirsch 2004). A limiting factor in the use of EMG as a control interface is in the translation of EMG signals into controller outputs. Although the relationship between the magnitude of EMG signals and the force exerted or level of contraction of the muscle has been noted to be static (Han et al. 2003) or even "obvious" (Rosenberg 1998), the processing of EMG signals to accurately determine the appropriate control output remains a problematic issue, affecting the dependability of such interfaces.

Several sources are noted to interfere with the measurement and classification of EMG signals, including the use of nearby muscle groups (Williams & Kirsch 2004), fatigue (Han et al. 2003) and electrocardiograph (ECG) potentials (Ragupathy et al. 2004). Some authors have used ANN classifiers or filters to remove these artifacts, as with interference on EEG signals, but the effects of interference are still noted to affect the performance of EMG classifiers in control applications. Rosenberg (1998) and Han et al (2003) both state that it is practical to develop low cost control systems using EMG. The instrumentation required to measure and process EMG signals is cheaper and less complicated when compared with EEG, but more costly than systems measuring physical movements.

2.2.3. Eye movement

Eye movement has been used as a source of both discrete commands and continuously varying control signals. Several different methods to determine the position of the eye have been reported in the literature. Electro-oculography (EOG) is the most common method, estimating eye position by measuring the electric potential created on the skin by the orientation of the eye, typically ranging from 0.05 to 3.5 mV. Other methods of measuring eye position are infrared oculography (IROG) and video-oculography (VOG). Although no complete control systems are presented in the literature using IROG and VOG, these methods are proposed in principle as alternative sources of control input.

LaCourse and Hludik (1990) proposed a control system using EOG to monitor the position of the user's eyes. This system allowed the user to select discrete commands by moving their eyes towards nine specific target locations on a template. The results presented by LaCourse and Hludik were obtained from a single subject (that subject being Hludik), and showed that for the target locations used, error rates varied between 0 and 60%. LaCourse and Hludik suggested that the high error rate reported for some targets was due to movement of the subject's head, which changed the angle to which the subject's eye would move in order to look at the target and therefore changed the EOG potential. Bahi, LaCourse and Hludik (1991) explored this issue, proposing a modified control system and an experimental method to test the effect of a tremor on the ability of the system to identify the target that the subject was looking at. The results presented by Bahi, LaCourse and Hludik showed, under a particular selection of parameters, the error rate was reduced to below 30%.

EOG was later investigated by Barea et al (2002; 2000) to develop a system designed to allow highly disabled users to directly select navigational commands. Two models for the generation of discrete commands were discussed in detail, being the direct selection of commands from a fixed display, and the selection of commands from a scanning display. Recognising that the rate at which a user would be capable of issuing commands would be too low to safely allow direct control over the movement of the wheelchair, the proposed system included navigational aids, such as infrared and ultrasonic sensors for collision avoidance. Mazo et al (Mazo 2001; Mazo et al. 2002) presented an extension of this system, incorporating more advanced navigational assistance, providing higher level command signals, such as the selection of a destination.

The control protocols presented using oculography as a control interface require low levels of physical ability to allow a user to achieve control of a wheelchair. However, as reported by Barea et al (2002), fatigue depended on the rate at which eye movements must be made to achieve control of the wheelchair, thus affecting the operability of the control system. Further to this physical load, they noted that the concentration required also influenced the development of fatigue, depending heavily on the ability of the users to control their ocular actions and the control protocol used.

Operability of controls systems based on oculography is also affected by what is termed by Barea et al (2002) as the Midas Touch problem: that the user's eye is always active and so long as it is open, it is always looking somewhere. In the case of EOG based systems, the user's eye has a position even when it is closed. However, not all ocular actions are intended by the user to control the wheelchair, and it is necessary to have some degree of free eye movement to allow the user to navigate or perform other activities. Neither Barea et al (2002) nor the articles by Mazo et al (Mazo 2001; Mazo et al. 2002) on the same project propose a solution to this problem other than to implement emergency stop and collision avoidance features, and by only conducting experiments on oculographic control under supervision in structured environments.

Mazo (2001, p. 49) noted that the EOG signal is "seldom deterministic, even for the same person in different experiments", therefore complicating the recognition of command actions. The high error rates reported by La Course and Hludik (1990) and Bahi, LaCourse and Hludik (1991) were obtained from classifying oculographic signals using a threshold crossing method. The systems presented by Barea et al (2002) and Mazo et al (Mazo 2001; Mazo et al. 2002) used more sophisticated signals processing, including an ANN to estimate gaze angle, and reported that moderately disabled subjects were able to achieve sufficient control of the wheelchair to perform a set of tasks. However, no details were given on the accuracy or reliability of these control systems.

It was noted in these articles that there are several signals that interfere with the processing of electro-oculographic potentials. The direction of a user's gaze is affected by both the position of the eyes and the position of the head. Since EOG signals only reflect the position of the eyes, it is necessary to compensate for movement of the user's head to use fixed targets for the selection of commands or path tracking relative to overall gaze direction. Electromyographic signals from the movement of the head, face and eyelids and the effect of electrode placement are also sources of interference. VOG is mentioned by Barea et al (2002) potentially as being able to resolve these issues. However, as with other camera-based techniques, interference from lighting conditions, camera occlusion and the processing load are introduced.

The authors above proposing the use of EOG do not discuss the long-term effects of the placement of electrodes on the skin. This issue is significant due to the extended periods during which a user could be expected to use such a control system and the absence of alternative sites for electrodes for EOG. The typical sites used to measure EOG potentials, shown in Figure 1, create problems for the preparation of the skin using either electrolyte gel or abrasion of the outer layer of skin (Griss et al. 2001). The authors proposing the use of EOG did not propose solutions for these issues.

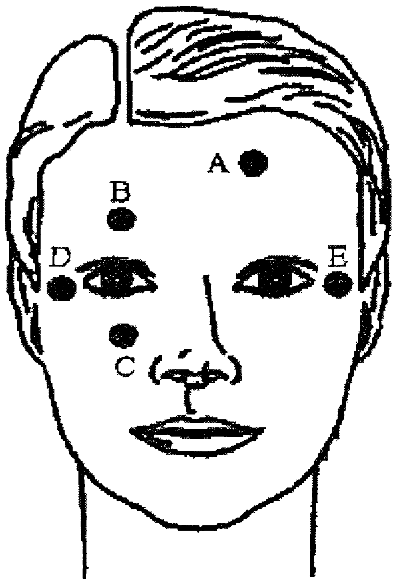

Figure 1 Electrode placement for electro-oculography (Barea et al. 2000)

EOG is described as a low-cost control interface (Barea et al. 2000), although this does not include the cost of the assistive navigation technology several authors (Barea et al. 2002; Mazo 2001; Mazo et al. 2002) associate with the use of EOG for wheelchair control. Barea et al (2002) suggests that VOG using relatively low-cost cameras may further reduce the cost, although this was not investigated through to a comparable system.

2.2.4. Speech recognition

Voice control and speech recognition have been investigated as a source of discrete commands requiring a low physical capability. The difficulty of processing voice signals in a robust manner and problems regarding operability and dependability have been noted by many authors, such as Newell and Barr (1971), Chauhan et al (2000). and Komiya et al. (2000).

Although speech recognition has been used successfully in systems that are not safety critical, such as computer interfacing, its successful use in wheelchair control has been largely restricted to laboratory conditions. Noyes and Starr (1996) presented a case study of the use of speech recognition in an environmental control system. The results reported came from a six-week study involving subjects with a range of high-level disabilities, and showed that the reliability of the system was too low, limiting its operability and dependability for many users. The recognition rate was reported to be as low as 50%, a problem which led to feedback processes whereby a user's voice patterns would become altered as they became irritated or fatigued, further degrading the performance of the classification.

Shortly after this paper, Rockland and Reisman (1998) presented a design for a wheelchair control system using a small vocabulary speech recognition kit. Although it was stated by Rockland and Reisman that an individual for whom the speech classifier was customised could theoretically achieve accuracies in excess of 95%, there was a lack of detailed experimental data supporting this claim.

2.2.5. Head position and head movement

The position of the head has been used to generate both discrete commands and continuously varying control signals. In addition to the use of head position, gestures formed by changes in head position have been used as discrete commands. Some degree of voluntary control of head position is available to many disabled individuals who are unable to control the movement of other body parts. In addition to the wheelchair application, head position and head movement interfaces have also been proposed for other applications, including environmental control and computer interfaces.

Each of the three degrees of freedom of the head has been used, either individually or in combination with the others. Although the terminology used varies, the control actions performed by the user in all systems based on head position are comprised of the same basic movements. For consistency, the anatomical terms used by Hamill and Knutzen (2003) will be used in this section, as illustrated in Figure 2 and Figure

3. Flexion of the head refers to an anterior rotation about the mediolateral axis, that is, bringing the face forwards, towards the chest. Extension refers to the opposite movement, returning to the anatomical starting position. Hyperextension is used to describe the posterior rotation about the mediolateral axis from the anatomical starting position. Lateral flexion describes rotation about the anteroposterior axis, either to the left or to the right. This type of movement is sometimes referred to in the engineering literature as a head tilt, although this usage is not uniform. Rotation describes a rotation about the longitudinal axis.

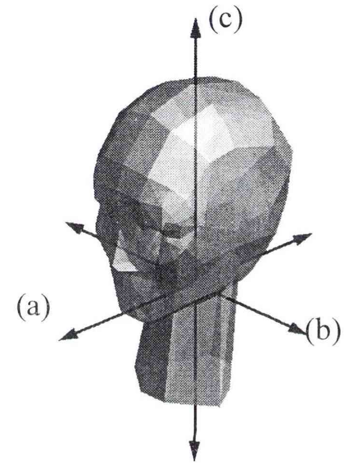

Figure 2 Anatomical reference axes – (a) anteroposterior axis; (b) mediolateral axis; (c) longitudinal axis. (Hamill & Knutzen 2003; Tordoff & Mayol 2001)

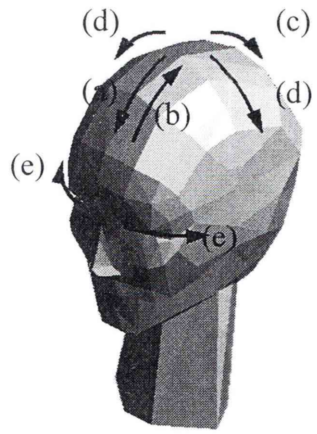

Figure 3 Anatomical descriptors of head movements – (a) flexion; (b) extension; (c) hyperextension; (d) lateral flexion; (e) rotation (Hamill & Knutzen 2003; Tordoff & Mayol 2001)

The position of the head without examination of any temporal features in the signal is the simplest method of using the head as a control input. Coyle (1995) used head position to provide a variable input signal to a wheelchair. In this system, head positions within a dead band region were ignored. Outside the thresholds creating this region, the speed and direction of wheelchair movement was determined as a function of the position of the user's head, although no details were given as to what form this function took nor which movements were used. Salagnicoff (1995) used head position, in conjunction with the position of the torso, to control a robot arm with 6 degrees of freedom. The torso is capable of the same three types of movement as described for the head, although it was not stated whether all types were necessary, nor how much movement was necessary to achieve control using this method. The position and attitude of the end effector of the robot followed the position of the user's head, amplifying the user's movements. Force-feedback was provided to the user via a robotic arm connected to the user's head, reflecting the force sensed at the robot's end effector.

The results presented by Salagnicoff et al (1995) showed that several able-bodied users were able to perform a set of precision control tasks using this control system. It was noted that one of the subjects was only able to perform one of the tasks if force feedback was supplied, which was concluded to suggest that the proprioceptive and tactile feedback allowed the users to achieve greater precision in operating the arm than using visual feedback alone. No results for disabled users were presented.

Head position has also been used to implement a computer interface. Takami et al (1996) proposed the use of flexion, extension and rotation to control the position of the cursor on a computer monitor. The method used to translate head position to cursor position was that the velocity of the cursor was proportional to the degree of flexion or rotation of the user's head. The system was tested with three highly disabled subjects, although clinical information regarding the subjects' disabilities was not presented. Although results regarding the performance of interface were not reported in detail, it was stated that all three subjects were able to satisfactorily operate the system.

Min et al (2002) proposed the use of head position for wheelchair control using multiple thresholds to reduce the precision with which the user was required to have over voluntary head movements. Flexion, hyperextension and lateral flexion were used as control actions, with the command being recognised after the user had moved past a preset threshold. A second, lower threshold was to recognise the conclusion of the command. The aim of this hysteresis was to prevent jitter in the case that the user moved to the threshold but was unable to move sufficiently past that point that interference did not cause the head position signal to repeatedly cross the threshold. Results reported were that of six disabled subjects, five were able to use the wheelchair with a median recognition accuracy of 75%. Min et al indicated that this accuracy is lower than would be necessary to provide sufficient control without assisted navigation, and attributed at least part of the inaccuracy to the difficulty experienced by users when making movements close to the limits of their range of motion, particularly for lateral flexion.

The use of current head position as a control input in the systems described above provides an intuitive interface, minimising the mental load placed on the user by utilising familiar movements as control actions. Both Salagnicoff et al (1995) and Takami et al (1996) note that very little or no training was required for subjects in the tests performed. A drawback to the use of head position in this manner is that the condition of many disabled individuals prevents them from exercising precise voluntary control over the position of their head. As a result, the number of control actions is limited, therefore limiting the transfer of information between the user and the system. Most systems using head position as the control input use thresholds to determine the user's intention. Where a user lacks sufficient control of their head movement, this leads to a reduction in the specificity of the control system, thus degrading the operability and dependability of the system. The durability, cost and maintenance requirements of these systems depend largely on the technique used to measure the position of the head.

Temporal information has been used widely to assist in the interpretation of head position data. Morimoto, Yacoob and Davis (1996) proposed the use of Hidden

Markov Models (HMMs) for the classification of head movement gestures for use in a human-computer interface. The classifier proposed identified four movements from a stream of data measuring the position of the user's head in each of the three axes of movement. Morimoto, Yacoob and Davis note that since the gestures used were relatively simple and there were only four gestures, it would have been feasible to develop a classifier using less sophisticated methods but dismissed this approach on the grounds that the vocabulary of the interface could be expanded more easily with the use of a HMM classifier.

Fuzzy logic and heuristic methods were employed by Adachi et al (1998) in a wheelchair control application. The system proposed by Adachi et al used head position as the input source, using thresholds to create a dead band region similar to that proposed by Coyle (1995). In a variation from the system proposed by Coyle, a fuzzy rule set was then used to filter out fast movements. This approach was based on several assumptions about the user's intention and communication method and the nature of signals and user actions that would be considered interference. Anecdotal results were presented, showing that the proposed system could be used to guide a wheelchair around a large room. No details were given on the nature of the subject or subjects, nor were quantitative results presented.

Joseph and Nguyen (1998) proposed a wheelchair control system using head gestures as directional commands. An ANN classifier was used to recognise eight commands, being nodding movements involving flexion, hyperextension and lateral flexion, or a combination of these. A ninth gesture, head shaking, was used as an interrupt or emergency stop signal, but was recognised using a heuristic method. Joseph and Nguyen noted that the performance of the control system is highly dependant on the quality of the classifier used to recognise commands. The results reported were that the ANN was able to correctly classify almost 92% of gestures performed by an able bodied subject.

Bergassa et al (2000; 1999) also classified gestures as commands, with gestures comprising flexion, hyperextension and rotation. Instead of using an ANN, Bergassa et al used a fuzzy logic classifier. Inputs to the classifier were the position and velocity of the user's head in the directions parallel to the mediolateral and longitudinal axes. Bergasa et al (2000) mention an experiment involving 5 disabled subjects in a laboratory and corridor, but present no quantitative results and give little detail on the experimental procedure. Bergassa et al also used facial gestures as command actions, one of which was noted to adversely affect the performance of the control system, as the specificity of the classifier for this component was inadequate.

Y-L Chen et al (2003; 2002) proposed the use of thresholds as a means of recognising gestures in head position data. Although the classification of gestures was based solely on head position, the system proposed by Chen et al differed in that the movement of the wheelchair was governed by a finite state machine such that the wheelchairs speed could be changed by subsequent commands, thus making use of addition temporal information. As with Coyle (1995), Adachi et al (1998) and Min et al (2002), a dead band region was created using thresholds. Discrete commands were generated by flexion, hyperextension and lateral flexion. In a clinical evaluation, seven disabled subjects performed a series of driving tasks, comparing the difference in time taken to complete the tasks between a joystick interface and the head movement controller. The results presented showed no significant difference between the two controllers, which was interpreted as showing the head movement controller to be a suitable alternative interface.

By including temporal information in addition to head position, head gesture based control systems provide a greater number of possible control actions for a given physical ability. In addition to allowing command actions to be given using a larger number of combinations of the basic types of movement, processing head position with temporal information allows more sophisticated classification techniques, such as the ANN classifier used by Joseph and Nguyen (1998) and the HMM proposed by Morimoto et al (1996). As with head position, several authors noted that head gestures are a natural, intuitive methods of communication and thus create a low mental load for many potential users. The physical requirements are greater than bioelectric control systems, as it is necessary for the user to be able to exercise a reliably reproducible gesture. However, the increase in flexibility of the form of the control actions facilitated by the use of signal processing techniques can provide a mechanism for the removal of interfering signals, such as those due to tremor or other problems with control over voluntary head movement, therefore making the system operable by more highly disabled individuals.

The ability to identify command actions in the presence of interference and noise determines the dependability of head gesture control systems. The use of thresholds and similar heuristic based approaches is appropriate only when the user is capable of maintaining precise control of the spatial and temporal aspects of their head movement. From the results presented in the literature discussed above, techniques such as ANN, fuzzy logic and HMM classifiers provide reliable performance to a greater number of potential users.

As with head position control systems, the method used to measure head position is a key factor in the durability, cost and maintenance requirements of head gesture control systems. Three general methods appear in the literature for measuring head movement: image processing, ultrasonic proximity sensors and sensors placed in connection to the user's head. Capacitive and Hall effect methods have also been used in commercially available control systems, such as the discontinued Peachtree proportional head control system (ABLEDATA 2003) and ASL Proximity Head Array Package (Adaptive Switch Laboratories Inc 2004). The ASL system is claimed have been used successfully with developmental disabilities as it is able to measure head position without the application of mechanical pressure, a feature shared with all of the above methods. Capacitive and Hall effect based methods are not common in recent engineering literature, and are thus not discussed in further detail in this review.

The system developed by Bergasa et al (2000; 1999) is one of the more notable image processing approaches. Bergassa et al propose a "vision-based command generation system", which allows a user to issue commands to a wheelchair by a set of facial and head movements. Commands are identified by the interface by processing images of the users head. Bergasa et al (1999) describe the global system architecture, methodology and some experimental results for the proposed system. A colour CCD camera is placed about 80 cm in front of the user, acquiring images of their face that are digitised by a frame grabber. The head is located in the image using a skin colour segmentation algorithm and, using this location, face tracking is used to govern a state machine whose outputs set the linear and angular velocities of the wheelchair. Facial features, such as the eyes and mouth, are distinguished as hollows in the face by being a different colour, allowing the system to recognise movements of the user's head, eyewinks and hiding of the user's lips as command actions. It was reported to work well indoors where suitable illumination is available. However, the performance was reported to worsen in poorer light conditions, particularly in outdoor settings where light conditions are not uniform.

Morimoto, Yacoob and Davis (1996), Adachi et al (1998) and Matsumoto, Ino and Ogsawara (2001) used similar video based techniques. By using processing of colour images, the number of command actions was increased by being able to detect facial gestures using the same processing instrumentation. Matsumoto, Ino and Ogsawara also noted difficulty in outdoor settings. Although the system proposed by Matsumoto, Ino and Ogsawara was successfully used by many subjects, in particular conditions, the saturation of the camera image by sunlight prevented the use of any signal processing, rendering the wheelchair inoperable.

The demands of computation and instrumentation have been identified as significant drawbacks to the use of image processing. An alternative method to using algorithms such as that described by Bergassa et al (1999) is proposed by Takami et al (1996) for a computer based environmental control unit, which could conceivably be adapted for wheelchair control. This system simplifies the task of locating the position and orientation of the user's head by mounting 3 infrared LEDs on a glasses frame worn by the user in a triangular configuration. Two LEDs were mounted close to the user's face at the edge of the glasses frame, while the third was mounted several centimetres in front of the frame, between the user's eyes. The image-processing algorithm proposed by Takami et al was reduced to finding the relative locations of the 3 LEDs. A heuristic estimate of the orientation of the user's head was determined from the deviation of the centre LED from the midpoint of the two distal LEDs. Tests of the precision of the head position measurement reported by Takami et al showed that the absolute error was less than 4 degrees across a 60 degree range, reproducible within 0.5 degrees.

Coyle (1995) proposed a system using several interface techniques, including head movement measured by ultrasonic transducers mounted in a headrest the behind the user. Head position was measured using the time taken for sound to be reflected back to each transducer. Lo Presti and Brienza (2004) also used ultrasonic instrumentation in a computer interface. This system involved placing a transducer on the user's head, with the receiver mounted in a fixed position on the monitor. Ultrasonic measurement tends to be cheaper than video processing, but has several drawbacks. It cannot be used in noisy environments due to interference with the signals. Similarly, performance of ultrasonic instruments is degraded in outdoor environments.

Electromechanical sensors are the most common instruments used to directly measure head movement, although several other types of angle sensor have also been used.

Joseph and Nguyen (1998), Knight (1999), Y-L Chen et al (2003) and S-H Chen et al (2003) used microelectronic accelerometers to measure the angle of the user's head. The method involves fixing a multi-axial accelerometer to the user's head such that one axis was parallel with the mediolateral axis and the other with the anteroposterior axis. In a position where the user's head is level, the force of gravity is orthogonal with these axes, and no acceleration is measured. When the user performs flexion, hyperextension or lateral movements, the force of gravity is projected onto one or both of these axes, which is detected as an intransient acceleration in that direction. This enables the angle of the user's head to be determined, although rotation could not be detected apart from small transient signals related to the movement of the head rather than its position.

Although Joseph and Nguyen (1998) used a tri-axial accelerometer, Knight (1999) presented similar results using a dual-axis accelerometer. This measurement technique provides acceptable accuracy across a wider range than ultrasonic sensors and is largely free from external sources of interference that affect image processing and ultrasonic techniques (1999a). The cost of instrumentation and processing is considerably lower than image processing and ultrasonic techniques. The main disadvantage of using microelectronic accelerometers is the requirement that a sensor be mounted on the user's head, which can cause problems with fitting, comfort, obstruction and aesthetics. These problems can be mitigated by using telemetry and concealing the sensor, thus improving the aesthetic qualities.

2.3. Discussion

From the literature reviewed above, no particular method presented can be considered superior in all characteristics for development of a wheelchair control system. The overall suitability of each method is determined by a trade off between the desirable characteristics identified earlier, and it is clear that this depends heavily on assumptions made about the population of potential users. Although each of the characteristics identified by Thorkildsen (1994) can be a limiting factor on the suitability, particular attention is paid to operability, dependability and cost in the literature proposing wheelchair control systems. Durability and maintenance are rarely discussed, possibly as these are properties that are considered more relevant to commercial systems rather than laboratory based research.

In terms of operability, bioelectric and eye movement based systems have the lowest physical requirements, making them accessible to that largest number of potential users. However, the high mental load, low bit rate and interference between control actions and a user's other activities make bioelectric control systems less operable than alternative systems for all but the most highly disabled potential users. Further, assisted navigation is required to make a bioelectric control system sufficiently dependable, the cost of such systems becomes much higher than the alternatives.

Despite being useful in applications that are not safety critical, the low dependability of speech recognition control systems is a significant barrier to their use in wheelchair applications. Chauhan et al. (2000) noted that although speech recognition systems have been developed with very high accuracy in laboratory conditions or in other quiet settings, these systems have been found to lack sufficient robustness to use in practical environments for wheelchair control. As with bioelectric control systems, using assistive navigation to improve the overall system dependability increases the cost of the system.

Head position and head movement based control systems have the highest physical requirements of the systems reviewed above. The requirement for some degree of reproducible voluntary head movement excludes individuals with the highest levels of spinal cord injury from the target population. For many disabled people who use a powered wheelchair and have sufficient control of their head movement, the joystick is an acceptable control interface. These set the limits for the population for who head movement control systems will be appropriate. Despite these limits on the population of potential users, the size of this population is still sufficiently large to justify the development of control systems, as noted by O'Conner (2001; 2003a; 2003b; 2004) and Cripps (2003; 2004). The dependability and operability of the systems based on head position and head movement have been shown to be sufficient to achieve wheelchair control in laboratory and production settings in a number of systems. Although systems using head movement rather than head position are offset by a marginal increase in cost, the improvement in operability and dependability of systems incorporating both spatial and temporal data justifies this expense.

Based on the greater operability and dependability of head movement systems when compared to bioelectric, eye movement and speech recognition based system, head movement was selected as the most appropriate interface method for further investigation in this thesis. The measurement of head movement also involves a trade off, between image processing, ultrasonic and contact based sensors. For systems of similar price, electromechanical sensors of the type used by Joseph and Nguyen (1998) and Knight (1999) have a greater accuracy and require less power when compared with video and ultrasonic measurement.

The system architecture proposed by Joseph and Nguyen (1998) has been selected for the control system developed in this thesis.

2.4. Research objectives

This section outlines the specific objectives of the research presented in Chapters Chapter 3, Chapter 4 and Chapter 5. It aims to establish the key research questions posed in this thesis and theoretical framework by which these questions can be addressed.

The control system proposed in Section 3.1 advances from head movement control systems previously proposed in the literature, particularly the architecture proposed by Joseph and Nguyen (1998). The ability to reliably classify data from a head position sensor to identify commands is the most significant problem outstanding in the use of head movement measured by electromechanical sensors for wheelchair control. Although Joseph and Nguyen (1998) reported a classification accuracy of almost 92%, this performance was observed for a single able-bodied subject using a relatively small set of test movements and did not measure the performance on the real-time gesture classification task.

Several different pattern recognition techniques are used in the literature directly relating to wheelchair control, including ANN, HMM, fuzzy logic and heuristic techniques, and many techniques exist in other areas of literature. There are several advantages to the use of ANN classifiers for head gesture classification for wheelchair control. Artificial neural networks with at least one hidden layer are noted to be capable of representing input-output mappings with arbitrary accuracy, provided that certain constraints are met (Irie & Miyake 1988). This property of ANN allows complicated decision boundaries to be implemented, providing the potential for several forms of optimisation of the classifier for highly disabled users.

Results from the optimisation of simple classifiers, such as the threshold based EMG gesture classifier described by Moon et al. (2003), are not directly comparable to ANN classifiers because in these simple classifiers, the parameter being optimised has a direct, linear relationship with the decision boundary of the classifier. In optimising the ANN classifier, the parameters being optimised for the ANN head gesture classifier are not directly related to the decision boundary of the classifier, but rather influence the way in which the decision boundary is derived from exemplar data.

More complex gesture classification systems have been proposed in other applications, such as ANN classification of EMG (Han et al. 2003; Hiraiwa, Shimohara & Tokunaga 1989; Matsumura et al. 2002; Reischl, Groll & Mikut 2004; Thompson, Picton & Jones 1996) and EEG (D Coyle, Prasad & McGinnity 2005; Garrett et al. 2003; Gope, Kehtarnavaz & Nair 2005; Maiorescu, Serban & Lazar 2003). Although such applications have some similarities with head movement classification as considered in this thesis, there are substantial differences that exist between the applications in the nature of the source signals and the objectives of classification. Consequentially, the results presented on such applications cannot be appropriately extended to head gesture classification.

The experiment presented in Chapter 3 determines the accuracy and classification delay that can be expected from an ANN head gesture classifier for disabled users. The primary objective of this investigation is to establish whether it is feasible to train an ANN classifier for a head movement wheelchair system that is able to provide sufficient generalisation for dependable performance of the system. The ability of ANN training to generalise classification rules directly from data is a particularly useful feature of ANN classifiers for this application. However, there is insufficient existing evidence to expect *a priori* that a classifier trained on data from one individual or group will generalise appropriately to any particular user, given the difference in the levels of physical and mental ability between users. This is particularly the case when considering the variation between able-bodied and disabled subjects. Many studies reviewed in the literature used able-bodied users to both train and test the performance of wheelchair control systems. In order to test the validity of this method, a classifier trained to classify gestures performed by able bodied subjects is tested on both able bodied and disabled subjects.

By relying on the level of similarity between the gestures performed by different individual members of the target user population, it is possible to create classifiers without reference to data from the specific end user. The term ‘generic classifier’ is used to describe such a classifier in this thesis, to differentiate from classifiers created using user specific data. The results presented in Chapter 3 relate to the performance of one generic classifier.

A range of techniques can be used for the optimisation of generic classifiers. Several of these techniques are investigated in Chapter 4. The aim of this investigation is to improve the ability of the control system to recognise head gestures performed by members of the target user population, thereby improving the utility of the control system. The techniques investigated focus on the performance of the ANN used in the classifier. The effect of the classifier structure and training algorithm on classifier performance and training time is examined, as is the marginal effect of additional training data. These effects are investigated using recorded gestures from several able bodied and disabled people to train ANN classifiers, which are then tested on separate, more highly disabled users to determine whether the classifier generalises appropriately.

The optimisation carried out in Chapter 4 is important for several reasons. Its primary contribution is to further advance the head gesture classifier described in Chapter 3. In addition to developing a classifier with superior accuracy, the results in this chapter provide insight into the distribution of results that can be attributed to several factors in the ANN training process, thus indicating ways in which the classifier may be further optimised.

The level of generalisation, that is, the ability of a classifier to correctly classify input patterns which were not contained in that classifier's training set, can be expected to vary between users and over time for a particular user. This variation is a consequence of the characteristics of command gestures performed by an individual being affected by short-term factors, such as fatigue and emotional state, and long-term factors, such as physiological condition and behaviour of the user. As noted by Lin and Lee (1996), there is no practical, universal procedure or rule that can be used to guarantee that a particular set of training data will result in arbitrary levels of generalised performance.

Variations between the characteristics of gestures performed by an individual in particular conditions to the general characteristics of gestures performed by members of the target user population can be expected to tend to have a negative effect on the performance of the classifier component of the control system. The performance of ANN classifiers can be unreliable when the inputs being classified are substantially separated from those used to train the classifier (D Chakraborty & Pal 2003). Although generic classifiers can be optimised by methods including, but not limited to, those investigated in Chapter 4, the differences between the characteristics of gestures by an individual in particular conditions to the general characteristics of gestures performed by members of the target user population lead to a theoretical, if unknown, bound on the performance of the classifier.

A procedure is proposed in Chapter 5 to improve the performance of a classifier trained using generic data by adapting it for a specific user. By using user specific data to adapt a classifier, the characteristics of that user's gestures are represented in the ANN training set. Therefore, it is less likely that inputs to the classifier while operating will be substantially separated from those in the training set. The methodology and results of experiments testing the effectiveness of the adaptive training for the wheelchair application are presented in Chapter 5. Classifiers trained solely on generic data are used for comparison, to allow the effect of the use of user specific data to be considered.

In each of the experiments detailed in this thesis, classifier performance is measured on the real-time classification of gestures performed by disabled users. The measurement of classifier performance for disabled people in this thesis differ from the results from those presented by Joseph and Nguyen (1998), Nguyen, Knight and Ekanayke (2003) and Nguyen et al (2002), which present results obtained from a group of able bodied subjects. The experiment in Chapter 3 expands on these results to demonstrate the effect of using an ANN trained on data from able-bodied people to classify gestures performed by disabled users. This difference is significant to the conclusions that can be drawn from the experiments, as the use of results from able-bodied subjects in these papers relies on the assumption that results from disabled subjects will be similar, without having empirical evidence to support this assumption.

The examination of the ability of the ANN to classify head gestures in real-time, using a sliding window input, is a difference between the results in this thesis and the results presented by Nguyen, King and Knight (2004) and Nguyen, Knight and Ekanayke (2003). In these papers, classifier performance was measured solely on the ability to classify a single input pattern without considering real-time effects.

King, Nguyen and Taylor (2005), Nguyen et al (2007) and Nguyen, Nguyen and Taylor (2004; 2006; 2006) present results on the effect of training of ANN head gesture classifiers using different algorithms. These results are comparable in some aspects to the results in this thesis, but differ in several regards. The focus of Nguyen et al and King, Nguyen and Taylor was on the effect of training classifiers using sophisticated ANN training algorithms that are relatively novel in this type of application, and are not strictly based on gradient descent. Methodology described by King, Nguyen and Taylor used a magnified gradient training procedure, while Nguyen et al and Nguyen, Nguyen and Taylor employed Bayesian techniques. Although the results presented in Nguyen et al, Nguyen, Nguyen and Taylor and King et al include performance for disabled users, these papers differ significantly from those in this thesis in that the real-time aspect of the classification task is not considered.

Results presented in Taylor, Nguyen and Craig (2002) compare two gradient descent algorithms, delta rule and recursive least squares, but are separate from the results in this thesis. The key differences between these results and the results presented in this thesis are in the use of able-bodied subjects in assessing the test set performance and in the absence of results on real-time classification performance. The results presented in 2002 are consistent with those presented in Chapter 4.

Several papers examine particular issues relating to the implementation of assistive technology devices using head gesture interfaces (King, Nguyen & Taylor 2005; H.T. Nguyen, King & Knight 2004; H. T. Nguyen, Knight & Ekanayke 2003; H. T.

Nguyen et al. 2002). These papers, however, do not focus on the optimisation of the ANN classifier, as is the case in this thesis.

Chapter 3. Performance of a head-movement interface for wheelchair control

This chapter presents the design and testing of a wheelchair control system to demonstrate the feasibility of a control system for the target user population. The control system proposed advances from systems previously described in the literature. It further develops the control system described by Nguyen, Knight and Ekanayke (2003). The suitability of the control system depends directly on the ability of the system being able to accurately recognise command gestures performed by the user. The feasibility of the control system is tested by experimentally measuring the real-time classification accuracy and delay of the control system. The experiment described in this chapter assesses the performance of the classifier component of the prototype and examines the differences observed between able bodied and disabled users. The chapter concludes with a discussion of the significance of these results and inferences that can be made from them.

Head movement and head position were used to allow the user to communicate with the prototype control system. Head gestures were used as commands, governing the transitions of a finite state machine. The outputs of the state machine set the speed and rate of turn of the wheelchair. The state machine gave the control system five speed settings: backwards at low speed, stationary, forwards at low speed, forwards at medium speed and forwards at full speed. Forward and backward nodding gestures were used to select between these states. A forward nod was defined as comprising flexion from a neutral starting position followed by extension back to the original position, while a backward nod was defined as being hyperextension from a neutral starting position and a return to the original position.

Lateral nodding or tilting gestures were used to turn the wheelchair. These gestures were defined as a lateral flexion from a neutral starting position. Lateral gestures could be held in a laterally flexed position for an extended period before returning to the neutral position to allow the user to control the duration of turns. The degree of lateral flexion sustained during a lateral gesture was used to control the rate of turn.

Head shaking was used as a command gesture to immediately stop the wheelchair. This command reset the finite state machine governing the speed and direction of the wheelchair to its initial state. Although rotational movements can be detected by the linear portion of such movements parallel with the axes of the sensor, the sensor is more sensitive to the projection of gravity onto these axes during flexion, hyperextension and lateral flexion. To make use of this sensitivity, head shaking gestures were defined as rapid, repeated laterally flexing movements, alternating from side to side. Details of the state machine and control logic are described in more detail in Section 3.1.7.

The prototype control system is an extension of the system described by Nguyen, Knight and Ekanayke (2003). The system of Nguyen, Knight and Ekanayke was based on an architecture proposed by Joseph and Nguyen (1998), and several features of this earlier control system are retained. A block diagram of this architecture is shown in

Figure 4 below. Central to the prototype is a notebook computer, which implements the signal processing, control calculation, software calibration and graphical feedback to the user.

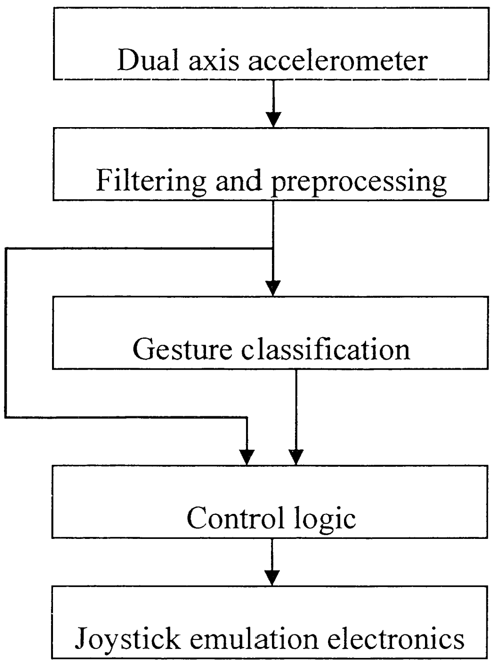

Figure 4 System architecture of prototype control system

The head position sensor used for the prototype control system was implemented using the Analog Devices ADXL202 dual axis accelerometer. The ADXL202 is chosen due to its having sufficient range and precision on two orthogonal axes, low cost and relatively simple interface electronics. The same instrument was used by Nguyen et al (2003) and later by Nguyen, Knight and Ekanayke (2003) and Nguyen, King and Knight (2004). The acquisition of data from the sensor was performed using the ADXL202EB-232A interface board and method proposed in documentation provided by the manufacturer (Analog Devices 1999b) and further detailed in Section 3.1.1.

Filtering and pre-processing, gesture classification, control logic and graphical display were implemented on a notebook computer using LabView software. The filtering and pre-processing processes applied to the acquired sensor data are described in Section 3.1.2.

An ANN classifier was used to detect command gestures in the data collected from the head position sensor and to determine which of the command gestures had been observed. The architecture and weights of the ANN used in this classifier applied those described by Nguyen, Knight and Ekanayke (2003). The classifier component is discussed further in Section 3.1.3.

The control logic of the prototype control system is an extension to that proposed by Nguyen, Knight & Ekanayke (2003). Nguyen, Knight & Ekanayke used a finite state machine to control the speed and direction of the wheelchair, where state transitions are governed by the outputs of the ANN classifier. The prototype presented in this chapter combines the use of instantaneous head position with a finite state machine to provide greater flexibility in the control of the wheelchair. Details of the control logic are provided in Section 3.1.7.

Many aspects of the prototype control system can be modified at the implementation stage to suit the needs of a particular user, such as altering the location of the head position sensor or the maximum speeds associated with each state of the control logic. The classifier component of the control system as proposed by Nguyen, Knight and Ekanayke (2003) cannot be readily modified in this way and therefore becomes a critical component in order for the system to provide a feasible wheelchair control system for people with high level disabilities. An experimental investigation of the performance of the classifier component is presented in Section 3.2.

3.1. Prototype wheelchair control system

3.1.1. Sensor

The prototype control system uses a dual-axis accelerometer to measure the position of the user's head. The device chosen for this task is the ADXL202, produced by Analog Devices. The ADXL202 is a low-cost, low-power, accelerometer, measuring linear acceleration on two orthogonal axes, both of which are parallel to the plane of the integrated circuit package. The ADXL202 has a full-scale range of $\pm 2\ g$, and can measure static acceleration forces, such as the force of gravity (Analog Devices 1999b, 2000).

The angle between an axis of the accelerometer and the vertical is determined by measuring the projection of gravitational force onto that axis. The output for each axis of the accelerometer is a duty cycle modulated square-wave signal. The ratio T_1 / T_2 is proportional to the measured acceleration, where T_1 is the length of time that the signal is high during a period and T_2 is the length of the total cycle. At an acceleration of $0g$, the duty cycle is nominally 50%, and changes by 12.5% per g. If the axis were parallel to the vertical, gravitational force would therefore cause a 12.5% change in duty cycle. The angle can be found from the durations T_1 and T_2 using Equation 1.

$$\theta = \arcsin\left(\frac{T_1/T_2 - 0.5}{0.125}\right)$$

Equation 1 Calculation of the angle between an accelerometer axis and the vertical

The accelerometer was connected to the computer implementing the processing, classification and control logic components using an ADXL202EB-232A and communication between the devices conducted using an RS-232 protocol provided in (Analog Devices 1999b). The protocol for communication between the computer and the ADXL202EB-232A is for the computer to send a single byte, an ASCII 'G' (0x47), to request that the data is transmitted, following which the ADXL202EB-232A returns a 4-byte packet. This packet contains 2 unsigned numbers, each comprised of a most-significant-byte (MSB) and a least-significant-byte (LSB), for the X and Y channel data.

Although the data sheet indicates that each byte can carry a value from 0 to 255, since the signal is a duty cycle, it is impossible for the value to be greater than 100%. Further, since the range of the accelerometer is approximately ±2*g* and the change in duty cycle per *g* is 12.5%, the duty cycle on each channel has a range of 25% to 75%. The resolution of the sensor is therefore approximately 12.3 bits, rather than the 16 bits of data transmitted. The resulting quantization noise is 74.0 dB smaller than the signal from the sensor.

The ADXL202EB-232A derives its power supply from the RS232 serial port to which the device is connected. Many common RS232 serial ports are able to supply the 6-12V required by the device through the RTS signal line. However, as noted in the device data sheet (Analog Devices 1999b), some RS232 ports are unable to source sufficient current. As a consequence, this can cause unreliability with some serial port designs as the current drawn by the sensor may place an excessive load on the port, causing the voltage supplied to drop below the minimum supply voltage of the sensor. This unreliability was resolved in this control system by the use of a

USB-serial adaptor, making use to the ability of a USB host controller to supply 100mA (Compaq et al. 1998).

The sensor hardware is mounted in the crown of a hat worn by the user. This places the sensor close to the longitudinal axis of the user's head, minimising interference resulting from the user performing rotational movements. To achieve maximum sensitivity to flexion, hyperextension and lateral flexion, the accelerometer is mounted so that the axes of the sensor are parallel with the mediolateral and anteroposterior axes of the user when in a neutral position.

Some noise is observed between individual packets of sensor data and some variations are also observed in the sample rate, due to the operating system under which the control software runs. To mitigate these effects, the sensor is oversampled, a moving average filter is applied and downsampled. The length of the moving average filter is adjustable at run time. It has been found that in practice a filter length of 5 samples provides sufficient noise suppression.

3.1.2. Pre-processing

To reduce the complexity of the classification task, several operations are performed to pre-process the data. The major pre-processing operations are to adjust the offset and scale the magnitude of the acquired data, compensate for the misalignment of the sensor axes with the user's mediolateral and anteroposterior axes and to apply a dead-band to the data. Parameters for the offsetting and scaling operations are determined during the calibration process by having the user maintain a neutral head position, and by having the user move their head through their full range of flexion, lateral flexion and hyperextension.

The offset adjustment removes the steady state signal present due to the tilt of the sensor when the user's head is in a level position. This offset arises from the mounting of the sensor within the crown of a hat, due to difference between the angle of the crown of the hat and the user's transverse plane and the need for the sensor to be mounted such that it causes minimal interference with the comfort of the user. Since the user is able to choose the angle at which the hat sits on the head, there can be a significant offset on each axis between the angle of the user's head and the axes of the accelerometer. A baseline value is obtained by averaging data over 10s while the user maintains a neutral position and the offset adjustment is performed by subtracting this from all subsequent samples.

Scaling is performed to adjust the data to allow for the range of movement available to the user. The maximum sensor readings in each axis recorded while the user performs full range flexion, lateral flexion and hyperextension movements are used to convert the offset adjusted pulse-width modulated (PWM) values to a normalised value corresponding to the proportion of maximum head tilt, in the range of ± 1.

Dead-banding effectively reduces the resolution of the sensor when the user's head is close to the neutral position, so that small, spurious signals are removed. Using the method proposed by Knight (1999), all values of normalised head tilt below a threshold value of 0.03 are reduced to 0.

3.1.3. Classifier

The classifier component of the prototype control system uses an ANN to detect and determine the type of command gestures. As each new sample of head position is obtained from the sensor, the data from each channel enters a tapped delay line. The contents of the delay line are used as the inputs to a multilayer feed-forward ANN, the output of which indicates the recognition or otherwise of each type command gesture. The classifier operates in the manner of a nonlinear finite impulse response digital filter, in the sense that it performs a mathematical function on a sliding window of sensor data.

In the head gesture classifier application, the ANN can be considered to be performing an input-output mapping (Hornik, Stinchcombe & White 1989). That is, that each point in the input space of the ANN is deterministically associated with a single value in the output space of the ANN. The input of the ANN is made up of the contents of the delay line and output of the ANN is a numerical representation of the class of gesture. The input space and output space correspond to the set of possible values taken by the input and output respectively. The input-output mapping performed by the ANN is determined by the network architecture and the weights of the network. The architecture is fixed during training, so the output of the ANN is a function of the ANN weights and inputs during this process. In operation, the weights of the ANN are also fixed and the output of the ANN becomes a function of the ANN inputs.

This section describes the implementation of the ANN classifier and the process by which the weights were determined. The first subsection provides background material on ANN classifiers and may be superfluous for readers already familiar with ANN theory. The second subsection details the use of these methods for the head gesture classifier.

The input output mapping performed by an ANN is the result of the combination of relatively simple elements. These elements are often called neurons, in reference to the biological cells that inspired their design. The most common form of neuron in used in ANN literature is the non-linear adaptive filter. This comprises a linear combiner and an activation function, as depicted in Figure 5. The linear combiner produces a weighted sum of the bias and input values to the neuron, while the activation function limits the amplitude of the output of the neuron (Haykin 1999). Definitions of the activation functions mentioned in this section are provided in Appendix B.

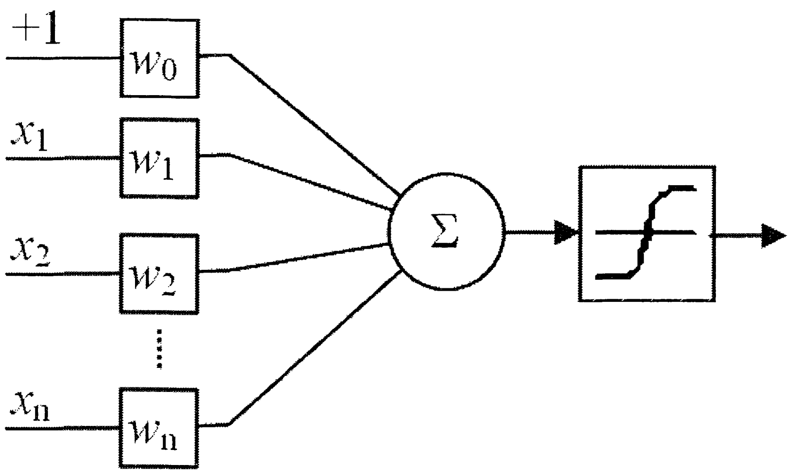

Figure 5 Non-linear adaptive filter model of an artificial neuron (Widrow & Lamego 2002)

The output of such a neuron can be expressed as $y(\mathbf{x}) = f(\sum_{i=0}^{n} w_i x_i)$, where f(•) is a the activation function of the neuron, x_k is the k-th input to the neuron ($k \in [1,2,\ldots,n]$), x_0 is a constant, arbitrary bias value, n is the number of inputs to the neuron and w_k is the k-th weight ($k \in [0,1,\ldots,n]$) (Haykin 1999).

Using a threshold function as the activation function of a neuron, such as the signum or Heaviside step functions, results in a dichotomous decision boundary on a hyperplane through the input space of the neuron. This type of neuron is sometimes referred to as the McCulloch-Pitts model, in reference to the seminal paper by McCulloch and Pitts (1943). Such a neuron is sufficient to classify between two linearly separable classes by associating each class with a region of input space mapping to one side of the decision boundary.

Multiple neurons can be used to create classifiers for more than two classes. In such classifiers, the neuron outputs represent a point in the output space of the network. In the case of each neuron using a threshold function, each neuron creating a distinct decision boundary will exponentially increase the number of points that can be represented in the output space of the ANN (Looney 1997). Each class can be associated with a set of points in the output space of the ANN and, therefore, with a region of the input space.

In the above cases, the input to each neuron comprises the input to the ANN and the output of the ANN comprises the output of each neuron. It is also possible to construct an ANN in such a way that some neurons receive inputs from the outputs of other neurons, or that the output of some neurons is not directly represented in the output of the ANN. Neurons whose output is not directly represented in the output of the ANN are known as hidden neurons, and a collection of hidden neurons with topologically identical connections to other neurons in the network is known as a hidden layer. An example of such a network is depicted in Figure 6.

A McCulloch-Pitts model neuron has the limitation that it is unable to create a decision boundary between two classes unless they are linearly separable, as was demonstrated by Minsky and Papert (1969). The further conjecture of Minsky and Papert that this result would extend to multilayer ANN was later contradicted, and it has been shown to be theoretically possible for an ANN containing a single layer of hidden neurons to approximate any input-output mapping with arbitrary accuracy, provided that certain constraints are met (Hornik, Stinchcombe & White 1989; Irie & Miyake 1988).

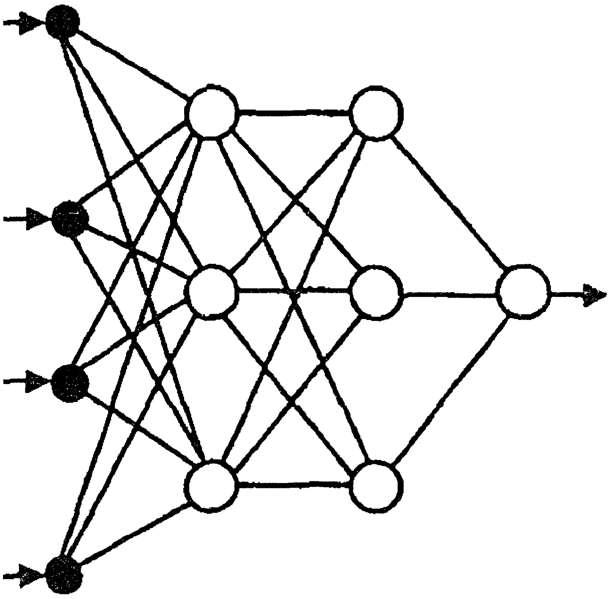

Figure 6 An example of a feedforward ANN architecture with two hidden layers (Widrow & Lamego 2002)

By applying a threshold to output neurons, an ANN using sigmoidal activations can be used to implement a classifier in the same way as for the single perceptron. In this case, the decision surface will not be comprised of hyperplane segments but will instead be a continuous, differentiable function through the input space of the ANN. Multilayer ANN are often implemented using sigmoidal activation functions, such as the sigmoid or bipolar sigmoid functions. This class of activation functions does not create dichotomous boundaries in the same manner as the McCulloch Pitts model. One of the key reasons for the use of these functions is that, unlike the threshold functions, sigmoid functions are differentiable for all real valued inputs (Haykin 1999), a property which becomes significant in the training of the ANN.

Since it has been shown that an ANN with a single hidden layer is capable of approximating any input-output mapping with arbitrary accuracy, the selection of an adequate architecture for an ANN to separate the examples in the training set relies upon the presence of a sufficiently large number of neurons in the hidden layer. The ANN architecture determines the composition of the function implemented by the ANN, and therefore determines the possible input-output mappings (Haykin 1999). Use of an inappropriate ANN architecture can lead to poor generalisation, as the function implemented by the ANN may be incapable of implementing the mapping if there are too few hidden nodes, or that weights providing a good estimate of the underlying ideal input-output mapping cannot be found using the chosen training data and training algorithm.

Throughout this thesis, ANN architecture is restricted to that of the single hidden layer feedforward network with bipolar sigmoidal activation functions. This restriction is made on the grounds that such an ANN is able to approximate any input-output mapping with arbitrary accuracy (Hornik, Stinchcombe & White 1989; Irie & Miyake 1988), and the extensive literature available on the training and implementation of such ANN. There is no *a priori* evidence suggesting that the accuracy or efficiency of head movement classification is enhanced by using more than one hidden layer. The architecture used in this chapter is a single hidden layer of three neurons, as proposed by Nguyen, Knight and Ekanayke (2003).

3.1.4. ANN training

Supervised learning is used to determine the function implemented by the ANN. Supervised learning produces an estimate of the weights necessary to approximate the input-output mapping by using input points for which the target ANN output is known. By approximating the function in this manner, the ANN is able to implement a non-linear mapping based directly on recorded examples of gestures, without using an explicit model defining rules or characteristics of each class (Haykin 1999).

A wide range of supervised ANN training algorithms exist in the literature, based on various numerical optimisation techniques. In general, supervised ANN training algorithms use a set of input samples paired with known, correct ANN outputs. Such a set is known as a training set. The algorithms seek to minimise the error between the actual output of the ANN and the expected output for each of the input-output pairs.

One or more separate sets of input-output pairs may be used to test the performance of the ANN during or after the training process to measure the extent to which the optimisation of ANN weights for the training set error improves the performance of the ANN on the task for which it is being trained. When such a set is used during the training of the ANN, it is known as a validation set.

Three common supervised training algorithms are summarised below. The delta-rule gradient descent algorithm was chosen for the purposes of training the classifier used in this chapter, based on its widespread use and the absence of evidence showing that other algorithms provide superior results for the optimisation of classifier performance. The Levenberg-Marquardt and scaled conjugate gradient algorithms are listed here for comparison and reference, as they are significant to the material appearing in subsequent chapters. The algorithms below are described widely in the literature, although there is some variation in minor points of the algorithms between different sources and a wide variation in the notation used to describe the algorithms. The algorithms, as presented below, are primarily referenced from Hagan (1995), Hagan and Menhaj (1994) and Møller (1993; 1997), with some modifications to reflect the Matlab Neural Network Toolbox (Demuth, Beale & Hagan 2005) implementations.

3.1.4.1. Delta-rule gradient descent

The delta-rule gradient descent algorithm is arguably the most common and widely used procedure for the training of multilayer neural networks. Haykin (1999) attributes the popularity of the algorithm to Rumelhart, Hinton and Williams (1986) for proposing the use of the algorithm for machine learning and demonstrating its use, although it is noted that the algorithm was also independently discovered several times, including Werbos (1974).

The change in weights at each iteration is proportional to and in the direction of the gradient of the performance metric with respect to each weight. The algorithm is also known by several other names, including generalised delta rule (Chan 1990), Widrow-Hoff learning (Hagan 1995), steepest descent (Barnard 1992), and backpropagation learning (Haykin 1999). As Haykin notes, backpropagation refers to part of the method by algorithm by which error signals for output nodes of the network are used to determine error signals for neurons in preceding layers. In this thesis, the term 'backpropagation' will be restricted to this latter meaning, as a similar processing of the error signal is performed in the second order algorithms presented in the sections that follow. The backpropagation of error requires that the function of the network be differentiable for all inputs. One of the key advantages of the delta-rule gradient descent algorithm is that each iteration can be performed in a computationally efficient manner.

Procedure

Given:

- a training set T, containing Q input-output pairs $[\mathbf{x}_q, \mathbf{t}_q]$;
- an ANN architecture of L layers, with each layer containing N^l neurons and using activation functions $f^{(l)}(x)$;
- a small positive real-valued scalar learning rate, α;
- termination parameters C_v, E_{max}, and k_{max}.

1. The algorithm is initialised by setting the all of the weights to small random values and iteration counter, k, to 1.
2. The first step of the algorithm is to propagate the input for a pattern, q, forward through the network to determine the resulting output with the existing weights.

$$\mathbf{y}^{(0)} = \mathbf{x}_q \qquad \text{Equation 2}$$

$$\mathbf{n}_q^l = \mathbf{W}^l \mathbf{y}_q^{(l-1)} \qquad \text{Equation 3}$$

$$\mathbf{y}^l = f^{(l)}(\mathbf{n}^{(l)}) \qquad \text{Equation 4}$$

3. The error of the ANN is measured by the difference between the ANN output, $y_q^{(L)}$, and the expected output for $\mathbf{x}_q$, $\mathbf{t}_q$,

$$\mathbf{e}_q = \mathbf{t}_q - \mathbf{y}_q^L \qquad \text{Equation 5}$$

4. The sensitivity of the error to the weighted sum of the input to each node in the output layer is determined by

$$\begin{aligned}\boldsymbol{\delta}_q^L &= -f^{(L)\prime}(\mathbf{n}_q^L) \bullet \mathbf{e}_q \\ &= -\dot{\mathbf{F}}^{(L)}(\mathbf{n}_q^L) \times \mathbf{e}_q\end{aligned} \qquad \text{Equation 6}$$

5. The sensitivity, by definition, is the first derivative of the error with respect to the weighted sum of the inputs to that node for the q-th input, $\boldsymbol{\delta}_q^l = \dfrac{\partial E_q}{\partial \mathbf{n}_q^l}$. The gradient of the error with respect to each weight can be found using the chain rule as

$$\frac{\partial E_q}{\partial w_{j,i}^l} = \frac{\partial E_q}{\partial n_{q,j}^l} \times \frac{\partial n_{q,j}^l}{\partial w_{j,i}^l} = \delta_{q,j}^l \times y_{q,i}^{(l-1)} \qquad \text{Equation 7}$$

6. The sensitivity for preceding layers is found by propagating the output layer sensitivities back through the network by applying the recurrence relation

$$\begin{aligned}\boldsymbol{\delta}_q^l &= f^{(l)\prime}(\mathbf{n}_q^l) \bullet (\mathbf{W}^{l+1})^T \times \boldsymbol{\delta}_q^{l+1} \\ &= -\dot{\mathbf{F}}^{(L)}(\mathbf{n}_q^L) \times (\mathbf{W}^{l+1})^T \times \boldsymbol{\delta}_q^{l+1}\end{aligned} \quad \text{for } 1 \le l \le L\text{-}1. \qquad \text{Equation 8}$$

7. The change to the weights and biases due to the q-th pattern is proportional to the magnitude of the gradient of the error and in the direction of steepest descent. In matrix form, this is

$$\Delta \mathbf{W}_q^{(l)} = -\alpha \boldsymbol{\delta}_q^l \times \mathbf{y}_q^{(l-1)T}$$ Equation 9

$$\Delta \mathbf{b}_q^{(l)} = -\alpha \boldsymbol{\delta}_q^l$$ Equation 10

8. Once the changes to weights and biases have been calculated for all input-output pairs in T, the weights are updated by

$$\mathbf{W}^{(l)} = \mathbf{W}^{(l)} + \sum_{q=1}^{Q} \Delta \mathbf{W}_q^{(l)}$$ Equation 11

9. The sum of the squared error across all patterns in the training set is used as a performance metric,

$$\begin{aligned} E &= \frac{1}{2}\sum_{q=1}^{Q}\sum_{i=1}^{N^L}(t_{q,i} - y_{q,i}^L)^2 \\ &= \frac{1}{2}\sum_{p=1}^{Q}\mathbf{e}_q{}^T\mathbf{e}_q \end{aligned}$$ Equation 12

10. Performance of the validation set, E_v, is calculated using the updated weights and biases determined in step 7, using Equation 2 to Equation 5, and Equation 12.

11. The algorithm terminates when any of the following criteria are met:

- The training set error drops below the given threshold, $E < E_{max}$;
- The maximum number of iterations has been reached, $k \geq k_{max}$; or
- Performance measured on the validation set, E_v , has increased more than C_v times since the last decrease.

Otherwise, the iteration counter is incremented, $k = k + 1$, and the algorithm returns to step 2.

The weights returned are those obtained in the iteration for which the optimal observed validation set performance occurred.

3.1.4.2. Levenberg Marquardt

The Levenberg Marquardt algorithm is a variation on the Gauss-Newton method, interpolating between the Gauss-Newton and delta-rule gradient descent algorithms (Hagan 1995). Newton's method minimises the error of the network by approximating the error as being a quadratic function of the weights. It uses the inverse of the second derivative of the error, or Hessian matrix, to determine the change to weights in each iteration. However, the computational and storage requirements for calculating the Hessian matrix for non-trivial networks prevent its widespread use. Further, Newton's method can provide poor results when the error function is not convex, as it may converge to a local minimum or saddle point (Lin & Lee 1996).

The Gauss-Newton method uses an approximation of the Hessian matrix to reduce the computational complexity. The Gauss-Newton method requires a modification to ensure that the approximated Hessian is invertible, and shares the property of poor convergence in problems where the quadratic approximation is poor (Hagan 1995).

The Levenberg Marquardt algorithm introduces a damping factor to the Gauss-Newton algorithm that makes the convergence properties more robust. This damping factor regulates the extent to which the change to weights follows the path of steepest descent or Gauss-Newton. The damping factor is adjusted at each iteration, allowing the algorithm to obtain the benefits of fast convergence from Gauss-Newton when the quadratic approximation holds and the stability of delta-rule gradient descent when it does not hold (Hagan 1995).

The Levenberg-Marquardt algorithm requires significantly more storage and computation when compared to delta-rule gradient descent, but has been shown to converge in fewer iterations for some tasks (Hagan & Menhaj 1994).

Procedure

Given:

- a training set T, containing Q input-output pairs $[\mathbf{x}_q, \mathbf{t}_q]$;
- an ANN architecture of L layers, with each layer containing N^l neurons and using activation functions $f^{(l)}(x)$;
- a small positive real-valued scalar initial damping factor, μ_1;
- a positive real-valued scalar adjustment factor, σ;
- termination parameters C_v, E_{max}, E_v, k_{max} and μ_{max}.

1. The algorithm is initialised by setting the all of the weights to small random values and iteration counter, k, to 1.
2. The first steps of the algorithm are to propagate an input pattern forward through the network using Equation 2, Equation 3, Equation 4 and Equation 5.

   All patterns are presented to the network and the error calculated to produce the overall error vector, $\mathbf{e}^*$, and the sum of square errors over all inputs, E.

   $$\mathbf{e}^* = \left[e_{1,1}, \cdots, e_{1,N^L}, e_{2,1}, \cdots, e_{Q,N^L},\right]^T \qquad \text{Equation 13}$$

3. The sensitivity of error to the weighted sum at each node used in the steepest descent algorithm is extended in the Levenberg Marquardt algorithm so that rather than being a vector, a matrix is obtained. The Marquardt sensitivity is propagated back through the layers of the network using a recurrence relation, with the output layer initialised as:

$$\widetilde{\mathbf{S}}_q^L = \begin{bmatrix} s_{1,1}^L & \cdots & s_{1,N^L}^L \\ \vdots & \ddots & \vdots \\ s_{N^L,1}^L & \cdots & s_{N^L,N^L}^L \end{bmatrix} = \begin{bmatrix} -f^{L\prime}(n_{q,1}^L) & 0 & \cdots & 0 \\ 0 & -f^{L\prime}(n_{q,2}^L) & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -f^{L\prime}(n_{q,N^L}^L) \end{bmatrix} = -\dot{\mathbf{F}}^{(L)}(\mathbf{n}_q^M)$$ Equation 14

$$\widetilde{\mathbf{S}}_q^l = \dot{\mathbf{F}}^{(l)}(\mathbf{n}_q^l) \times (\mathbf{W}^{l+1})^T \times \widetilde{\mathbf{S}}_q^{l+1} \text{ for } 1 \le l \le \mathrm{L-1}$$ Equation 15

$$\widetilde{\mathbf{S}}^l = \begin{bmatrix} \widetilde{s}_{1,1}^l & \cdots & \widetilde{s}_{1,Q\times N^l}^l \\ \vdots & \ddots & \vdots \\ \widetilde{s}_{N^l,1}^l & \cdots & \widetilde{s}_{N^l,Q\times N^l}^l \end{bmatrix} = \left[\ \widetilde{\mathbf{S}}_1^L \,\middle|\, \widetilde{\mathbf{S}}_2^L \,\middle|\, \cdots \,\middle|\, \widetilde{\mathbf{S}}_Q^L\ \right]$$ Equation 16

4. The Marquardt sensitivities are used to compute the elements of the Jacobian matrix

$$[\mathbf{J}]_{r,c} = \begin{cases} \widetilde{s}_{i,h}^l \times y_{q,j}^{l-1} & \text{where } w_c^* \text{ is a weight} \\ \widetilde{s}_{i,h}^l & \text{where } w_c^* \text{ is a bias} \end{cases}$$ Equation 17

5. The change to weights is computed as

$$\Delta\mathbf{w}_k^* = -\left[\mathbf{J}^T\mathbf{J} + \mu_k\mathbf{I}\right]^{-1}\mathbf{J}^T\mathbf{e}_k^*$$ Equation 18

6. The performance metric is recalculated using the new weights and compared to the value obtained in step 1. If the performance is improved, that is, $E(\mathbf{w}_k^* + \Delta\mathbf{w}_k^*) < E(\mathbf{w}_k^*)$, then the new weights are adopted, the damping factor is decreased.

$$\mathbf{w}^{*}_{k+1} = \mathbf{w}^{*}_{k} + \Delta\mathbf{w}^{*}_{k}$$ Equation 19

$$\mu_{k+1} = \frac{\mu_k}{\sigma}$$ Equation 20

Performance of the validation set, E_v, is calculated using the updated weights, using Equation 2 to Equation 5, and Equation 12.

Otherwise, the damping factor is increased. If the limit on the damping factor has not been reached, that is $\mu_k < \mu_{max}$, algorithm returns to step 5.

$$\mu_k = \mu_k \sigma$$ Equation 21

7. The algorithm terminates when any of the following criteria are met:

- The training set error drops below the given threshold, $E < E_{max}$;
- The maximum number of iterations has been reached, $k \geq k_{max}$;
- Performance measured on the validation set, E_v , has increased more than C_v times since the last decrease; or
- μ_k exceeds a limit, μ_{max}, which implies a succession of failures to find weights that reduce the error from the weights vector.

Otherwise, the iteration counter is incremented, $k = k + 1$, and the algorithm returns to step 2.

The weights returned are those obtained in the iteration for which the optimal observed validation set performance occurred.

3.1.4.3. Scaled Conjugate Gradient

The Scaled Conjugate gradient algorithm is a member of the broader class of conjugate gradient algorithms. Møller (1997) regards this class of algorithm as being "somewhat intermediate between the method of gradient descent and Newton's method." The conjugate gradient algorithms also approximate the error as a quadratic function of the weights, but do not require the storage and computational overheads of evaluating or inverting the Hessian matrix (Hagan 1995).

Most conjugate gradient algorithms involve a repeated line search. The algorithm determines the optimal step size along a particular direction in weight space, and the repeats this optimisation for other directions. If each step is conjugate to those preceding it, then the algorithm is guaranteed to find the minima of a quadratic function in a finite number of steps.

The scaled conjugate gradient algorithm avoids the need for a line search by a scaling of the step size depending on the success of a step in reducing the error function and the quality of the quadratic approximation of the error function.

Procedure

Given:

- a training set T, containing Q input-output pairs $[\mathbf{x}_q, \mathbf{t}_q]$;
- an ANN architecture of L layers, with each layer containing N^l neurons and using activation functions $f^{(l)}(x)$;
- a small positive real-valued scalar initial regulating factor and initial scaling factor, μ_1, σ_1, in the ranges $\begin{array}{l} 0 \le \sigma \le 10^{-4} \\ 0 \le \lambda_1 \le 10^{-6} \end{array}$;
- termination parameters C_v, E_{max}, E_v, k_{max} and μ_{max}.

1. The algorithm is initialised by setting the all of the weights to small random values and variables initialised as $k = 1$, $success\_flag = true$ and $\overline{\lambda}_k = 0$.

2. The first steps of the algorithm are to propagate an input pattern forward through the network using Equation 2, Equation 3, Equation 4 and Equation 5

to propagate the input for a pattern, q, forwards through the network to determine the resulting output with the existing weights..

All patterns are presented to the network and the error calculated to produce the overall error vector, $\mathbf{e}^*$, and the sum of square errors over all inputs, E, as per Equation 13

The gradient of the error with respect to the weights is determined using the same procedure as that described for the delta-rule gradient descent algorithm, with the gradient of error with respect to each weight over all input output pairs in the training set being the sum of the gradients for each pair using Equation 6, Equation 7, Equation 8, Equation 9 and Equation 10.

For convenience of notation in describing the SCG algorithm, the gradient for all weights and biases in the network will be manipulated as a single vector, $\nabla E(\mathbf{w}_k^*)$.

3. The search direction vector variables, **r** and **p**, are set to the direction of steepest descent.

$$\mathbf{p}_k = -\nabla E(\mathbf{w}_k^*) \qquad \text{Equation 22}$$

$$\mathbf{r}_k = -\nabla E(\mathbf{w}_k^*) \qquad \text{Equation 23}$$

4. If *success_flag* is set, the second order information is calculated

$$\sigma_k = \frac{\sigma}{|\mathbf{p}_k|} \qquad \text{Equation 24}$$

$$\mathbf{s}_k = \frac{\nabla E(w_k^* + \sigma_k) - \nabla E(w_k^*)}{\sigma_k} \qquad \text{Equation 25}$$

$$\delta_k = \mathbf{p}_k^T \times \mathbf{s}_k \qquad \text{Equation 26}$$

5. δ_k is scaled,

$$\delta_k = \delta_k + (\lambda_k - \overline{\lambda}_k)|\mathbf{p}_k|^2 \qquad \text{Equation 27}$$

6. If δ_k is less than or equal to 0, then λ_k must be increased to make the Hessian positive definite by

$$\overline{\lambda}_k = 2(\lambda_k - \frac{\delta_k}{|\mathbf{p}_k|^2}) \qquad \text{Equation 28}$$

$$\delta_k = -\delta_k + \lambda_k|\mathbf{p}_k|^2 \qquad \text{Equation 29}$$

$$\lambda_k = \overline{\lambda}_k \qquad \text{Equation 30}$$

7. The step size is calculated as

$$\mu_k = \mathbf{p}_k^T\mathbf{r}_k \qquad \text{Equation 31}$$

$$\alpha_k = \frac{\mu_k}{\delta_k} \qquad \text{Equation 32}$$

8. The comparison parameter, $\Delta_{k,}$, is calculated as

$$\Delta_k = \frac{2\delta_k(E(\mathbf{w}_k^*) - E(\mathbf{w}_k^* + \alpha_k\mathbf{p}_k))}{\mu_k} \qquad \text{Equation 33}$$

9. If $\Delta_k \geq 0$, then the step calculated in step 7 reduces the error, so the weight change is adopted for a further iteration:

$$\mathbf{w}_{k+1}^* = \mathbf{w}_k^* + \alpha_k\mathbf{p}_k \qquad \text{Equation 34}$$

$$\mathbf{r}_{k+1} = -E'(\mathbf{w}_{k+1}^*) \qquad \text{Equation 35}$$

$$\overline{\lambda} = 0, \; \textit{success\_flag} = \text{true} \qquad \text{Equation 36}$$

a. If the number of iterations performed, k, is divisible by the number of weights, Ω, then the search for conjugate directions is restarted:

$$\mathbf{p}_{k+1} = \mathbf{r}_{k+1} \qquad \text{Equation 37}$$

b. Otherwise, a new conjugate direction is found

$$\beta_k = \frac{|\mathbf{r}_{k+1}|^2 - \mathbf{r}_{k+1}^T \mathbf{r}_k}{|\mathbf{r}_k|^2} \qquad \text{Equation 38}$$

$$\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_k \mathbf{p}_k$$

10. If $\Delta_k \leq 0$ then no reduction of error is possible,

$$\bar{\lambda}_{k+1} = \lambda_k\,,\ \ success\_flag = \text{false} \qquad \text{Equation 39}$$

11. If $\Delta_k \geq 0.75$, the scale parameter is reduced,

$$\lambda_k = \frac{1}{4}\lambda_k \qquad \text{Equation 40}$$

Otherwise, if $\Delta_k \leq 0.25$, the scale parameter is increased

$$\lambda_{k+1} = \frac{\lambda_k + \delta_k (1 - \Delta_k)}{|\mathbf{p}_k|^2} \qquad \text{Equation 41}$$

Otherwise, for $0.25 \leq \Delta_k \leq 0.75$, the scale parameter is unchanged.

$$\lambda_{k+1} = \lambda_k \qquad \text{Equation 42}$$

8. The algorithm terminates when any of the following criteria are met:

- The training set error drops below the given threshold, $E < E_{max}$;
- The maximum number of iterations has been reached, $k \geq k_{max}$;

- Performance measured on the validation set, E_v , has increased more than C_v times since the last decrease; or

- The steepest descent direction is flat, (i.e. $\mathbf{r}_k = 0$);

Otherwise, the iteration counter is incremented, $k = k + 1$, and the algorithm returns to step 2.

The weights returned are those obtained in the iteration for which the optimal observed validation set performance occurred.

3.1.5. ANN Implementation

The inputs to the ANN are taken from a sliding window of the pre-processed sensor data. This sliding window is comprised of the most recent 20 samples of the anteroposterior and mediolateral channels. As each new sample is obtained, the new pair of values is added to the start of the window, the oldest pair of values is shifted out of the window, and the intermediate pairs of values are shifted along the window. Sensor data is sampled at 10Hz, so the sliding window contains the most recent sample and those samples obtained in the preceding 2s. The sliding window is implemented using a Serial-In, Parallel-Out, First-In-First-Out data structure, as depicted in Figure 7. The windowed data was converted to vectors of real-valued numbers able to be input to the ANN by appending data from the mediolateral axis to the data from the anteroposterior axis.

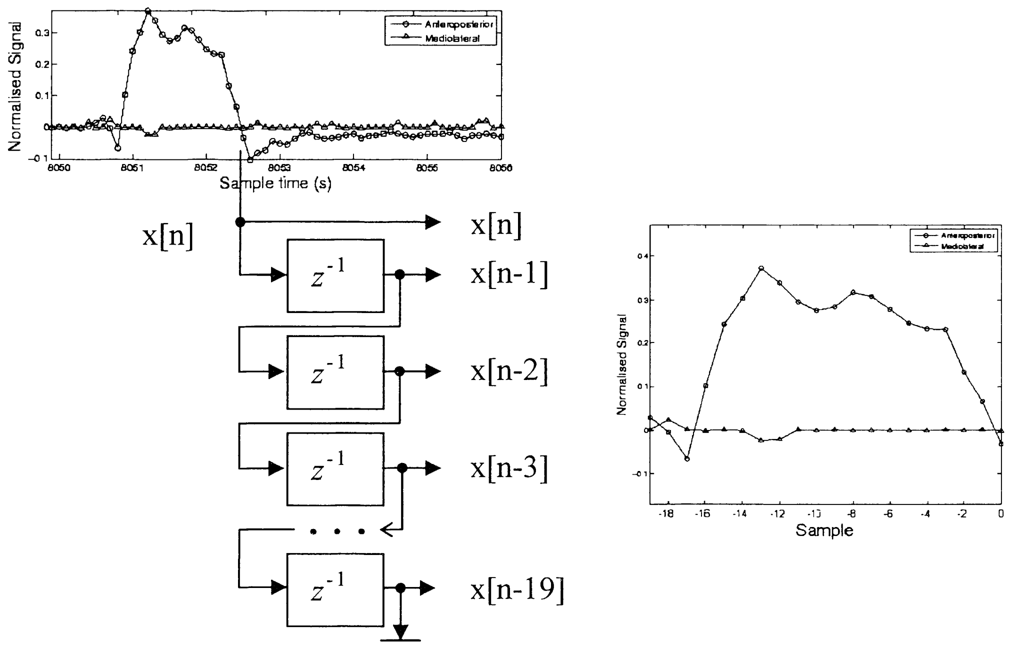

Figure 7 Implementation of sliding window data structure

The outputs of the ANN are selected to indicate the presence of command movements using a one-hot-one encoding. That is, each output node indicates the presence or absence of one of the command gestures in the input data. An output value close to 1 indicates that the gesture associated with that node has been recognised, while a value close to –1 indicates that the gesture is not recognised. If all output nodes are close to –1, then it is interpreted as meaning that no recognisable gesture is present. A threshold of 0.75 was used to determine whether an output node's value is sufficiently close to 1 to indicate the presence of a gesture.

The output space of the ANN is dimensionally smaller than the input space. Each point in the output space of the ANN therefore maps to a set of points in the ANN input space. In order to implement a classifier, the ideal mapping is for the set of input points corresponding to sensor data in which a particular type of gesture is present to map to a particular output point, and for only that set of input points to map to that output point. In practice, however, the set of input points that correspond to sensor data containing a particular type of gesture is unknown. In order to train the ANN, a sample of points from the input space is assigned output values. This set of sampled input-output pairs is known as the training set. The input-output mapping for these sampled points approximates the ideal mapping. The weights of the ANN are then adjusted during the training procedure so that the function implemented by the ANN approximates the sampled mapping (G Chakraborty, Shiratori & Noguchi 1993).

It is commonly expected in training ANN classifiers that by approximating the input-output mapping represented by the training data, the ANN approximates the ideal mapping. When this occurs, the ANN will accurately classify inputs that were not included in the training set and is said to generalise (Haykin 1999). An ANN must generalise to some acceptable extent to be useful as a head gesture classifier, as the set of input points included in the training set is much smaller than the set of inputs not included in the training set.

When a network uses features that are present in the training data, but are not true of the underlying classes from which the examples are drawn, the generalisation of the network is reduced. The existence of such features may arise due to systematic bias in the sampling method used to create the training set, or be due to natural variation that occurs in any random sampling process. Among other terms, this is phenomenon is known as "overfitting", "overtraining", "specialising" or "memorising" (Haykin 1999). Haykin argues that this happens when the network learns too many input-output examples, while Looney (Looney 1997) held that overfitting of an ANN occurs when the network is "underdetermined", meaning that number of weights in the network is greater than the number of sample patterns used to train it. In this case, there are multiple functions that can be implemented by the ANN approximating the mapping in the training set with the required accuracy, but the training algorithm does not have sufficient information to eliminate functions that do not approximate the ideal mapping.

Although the arguments of Haykin (1999) and Looney (1997) seem to imply a paradoxical result whereby poor generalisation is a consequence of both too many and too few training samples, these apparently contradictory statements are simply a reflection of the case that in some applications, no function can approximate the ideal input-output mapping with complete accuracy. The arguments of Looney and Haykin can be considered congruent in that if an ANN with an underdetermined architecture approximates an the mapping in the training set to a high accuracy, but the mapping in the training set does not adequately reflect the underlying ideal mapping.

Generalisation error can be considered to be comprised of two components: error due to the difference between the sampled training set and the underlying ideal input-output mapping, and error due to the difference between the ideal input-output mapping and the function implemented by the ANN. Although expressed in different terms, this decomposition of generalisation error can be shown to be consistent with the bias and variance discussed by Haykin (1999).

Selection of an appropriate training set is important for achieving good generalisation in an ANN classifier. Exemplar points are generally not drawn uniformly across the input space of the ANN. They are instead taken from inputs observed from the operating environment of the classifier, or an approximation of it. Consequently, the exemplar points tend to be clustered and information contained in some exemplars can become redundant. The input-output mapping of the ANN in regions of the input space where observed data is sparse is difficult to control using the most common implementations of ANN classifiers (D Chakraborty & Pal 2003).

Haykin (1999) asserts that the complexity of the classification task, meaning the complexity of the ideal input-output mapping, cannot be controlled in the design of the classifier. Although it can be shown through counter-example that this is not entirely the case, such as the use of pre-processing techniques to transform an input space into a more easily mapped space, it remains that some form of complexity is unavoidable in order to create an appropriate input-output mapping in non-trivial classification tasks once the input and output spaces are fixed. It is beyond the scope of this investigation to consider alternative formulations of the classification task.

The selection of training samples for the real-time head gesture classification task has several features that must be considered when compiling the training set. In a stream of real-time data from the operating environment of the classifier, windows of data that contain a particular type of gesture are much less prevalent compared to windows that do not contain that type of gesture. Since all gestures are defined as beginning in a neutral position, the points in the ANN input space in which no command gesture is present and in which the user's head is in a close to a neutral position are observed most often.

Although it would be possible to create a training set by assigning a target output mapping for every observed window of data in a recording of head gestures, there are significant reasons not to do this. The possibility that the set of ANN input points corresponding to windows of data containing a particular type of gesture and the sets of ANN input points corresponding to windows of data containing other types of gesture, or no gesture at all, are not mutually disjoint implies that more than one label could validly be applied to some points in the ANN input space. Since the point at which the user intentionally acted is unknown, it would be necessary to determine a method for selecting the transition point. Any such method would necessarily be arbitrary, thus introducing a source of bias. This transition bias was avoided by Joseph and Nguyen (1998) by selecting a single window per gesture.

The difference in the characteristics of gestures performed by different people is a potential cause of poor generalisation in the head gesture classification task. The distribution of ANN inputs observed from the gestures of a particular user may correspond to a region of the input space that is sparsely covered by the data in the training set. Several authors, such as Chakraborty and Pal (2003) and Looney (1997), argue that accurate classification of a point not included in the training set depends on the proximity of that point to clusters of points within the training set. These authors contend that input points in the training set can be grouped into clusters according to their proximity to other points of mapping the same output. The surface defined by the function implemented by the ANN crossing the threshold set for the output approximately passes between the edges of different clusters. Consequently, generalisation can be expected to be good for an input point that is not a member of the training set if it lies inside the edges of a cluster of points of the same class in the training set.

The ANN used in the classifier examined further in this chapter was trained on 20 input-output pairs exemplar of each class of gesture, with a validation set of 10 input-output pairs. These patterns were manually labelled by examining recordings of gestures performed by an able-bodied subject. The delta-rule gradient descent algorithm described in Section 3.1.4.1 was used to train the ANN, with a learning constant, α, of 0.7, weights initialised randomly in the range ±1 according to a uniform distribution and termination criteria $C_v = 5$ iterations, $E_{max} = 0.0$ and $k_{max} = 20{,}000$ iterations. This ANN is derived from Nguyen, Knight and Ekanayke (2003), and employs the same weights as were developed for that paper.

Three hidden nodes were used and bias nodes were included in both the input and hidden layers, with each of the nodes using bipolar sigmoidal activation functions.

3.1.6. ANN output decoding

To interpret the outputs of the ANN as classifications, it is necessary to deal with each of the possible output states of the ANN, and with the circumstances arising because data from each gesture will be processed several times by the ANN as the data passes through the sliding window. The ANN does not produce the same output values for each input pattern containing some data from a gesture. Some input patterns may contain too small a portion of the gesture to be recognised or may contain data from multiple gestures, so that the ANN output may change several times during the transition of data from a single gesture through the classifier's sliding window.

To ensure that the classifier only recognises one command from each gesture, the outputs of the ANN are interpreted such that upon recognising a gesture, all subsequent ANN outputs are ignored for a period of 2.5s. This period is set as the minimum separation between commands, during which it is assumed that the ANN output is due to the original gesture and so any change in the ANN output value is spurious. In the original design of this interpretation (Knight 1999), Knight made the assumption the first positive classification from the ANN is more likely to be correct than those that follow it.

The exception to this interpretation is the recognition of the headshake gesture. Since this gesture is used as an emergency stop command, it is given priority over the other gestures and is not subject to the minimum command separation period. A separate classifier was used to detect head-shaking gestures. The headshake classifier relied upon the characteristic that headshake gestures contain a larger high frequency component than other gestures. It was implemented by calculating a moving average of the amplitude of the output of a high-pass digital filter with corner frequency 10 Hz. A headshake was detected by the average of the most recent 1s filter outputs being above a threshold of 0.75.

3.1.7. Control Logic

The outputs of the classifier are used to trigger transitions for a finite state machine (FSM). The state machine implements the speed setting of the control system, and contains 5 states, as depicted in

Figure 8 below. The outputs of the state machine set the steady-state forward velocity ($V_{s\text{-}s}$), the maximum turn rate (ω_{max}) and the forward velocity while turning at maximum rate (V_t). These three parameters are used to calculate medial and lateral normalised output voltages to control the speed and turn rate of the wheelchair. These normalised output voltages are then converted to the hardware specific output voltages and sent to the interface hardware.

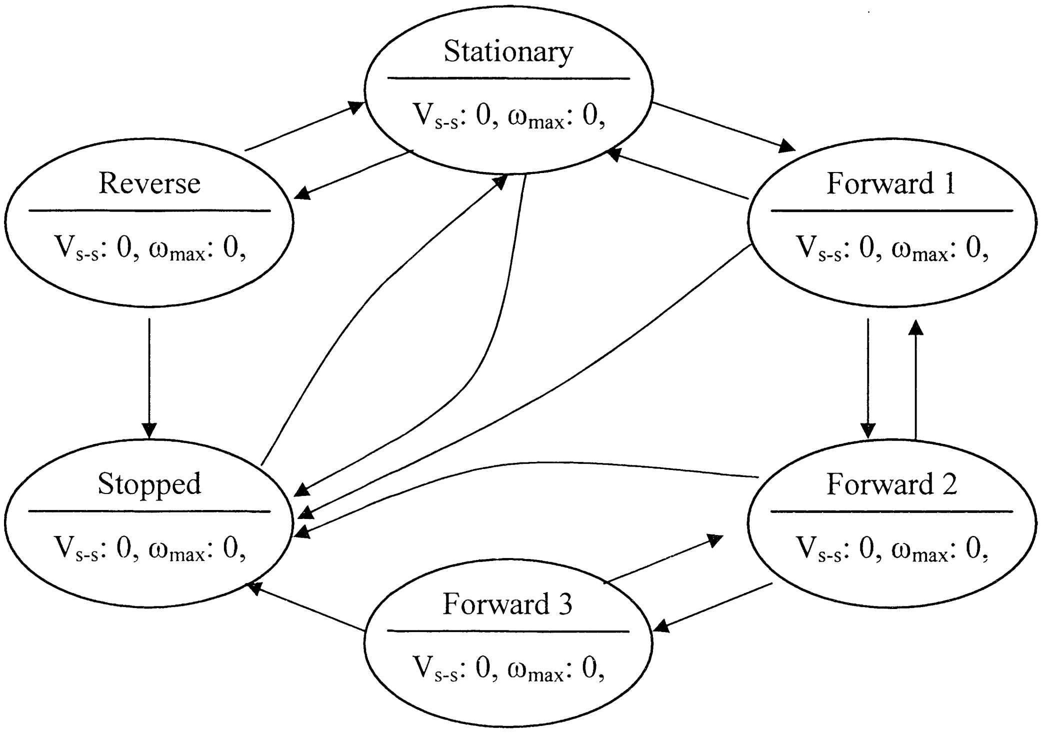

Figure 8 Finite state machine translating classifier output to the control logic parameters

The Left and Right classifier outputs do not affect the state transitions of the FSM. However, the forward speed of the wheelchair is reduced while turning to ensure that the wheelchair corners safely and without applying excessively high centrifugal forces to the occupant of the wheelchair. In a similar fashion to the diamond shaped plates that have been used to limit or guide the movement of joysticks described in the literature, a similar effect is achieved using the three outputs of the FSM.

To allow the user to achieve more precise control over the direction of the wheelchair, the rate of turn is calculated as a function of both the classifier's outputs and the magnitude of the user's lateral flexion. When the no turning command is present, the normalised medial and lateral output voltages are set to be the steady-state forward speed and 0 V respectively. When a turn command is present, and the user's head is laterally flexed to or past the threshold set at their full range of lateral movement, the normalised lateral output voltage is set at the maximum turn and the medial output voltage is set to the forward speed while turning. Where a turn command is present with a smaller magnitude of head tilt, the output voltages are interpolated between these values.

Since the time-response of the wheelchair's electric motors may be fast, abrupt changes of control voltage can result in sudden changes in acceleration and cause the movements of the wheelchair to be abrupt or jerky. This can cause discomfort for the user. To reduce this effect, limits are implemented in software to restrict the rate of change of the output voltage. This slew rate limiting applies a low-pass filter to the output channels and keeps the acceleration of the wheelchair below preset levels. The only exception to the use of slew rate limiting is in the event of an emergency stop, in which case the normalised output voltages are brought to zero immediately.

3.2. Assessing the performance of the prototype classifier

This section presents the methodology employed to evaluate the feasibility of the control system described in Section 3.1. The methods described in the literature for assessing of the performance of ANN classifiers vary widely between applications, and no single methodology has gained universal acceptance. There are two aspects of the real-time performance of the classifier that are of interest: its ability to make classifications correctly, and its ability to make classifications within an acceptable period. These are critical factors in the performance of the wheelchair control system because these factors have a direct effect on the system dependability and usability for highly disabled users. Metrics for real-time classification accuracy and delay are discussed, and the procedure used to measure the performance of the classifier component of the control system by these metrics is described.

The objective of the experiment is to establish whether or not the performance of the ANN classifier can provide appropriate performance to allow disabled people to operate the control system. Further, the experiment also aims to establish whether the performance of the classifier differs between able bodied and disabled subjects, and whether the provision of graphical feedback to users influences any such difference.

Results from this experiment were presented in 2003 (PB Taylor & H Nguyen; PB Taylor & HT Nguyen).

3.2.1. Background

Results from several experiments on the accuracy of ANN head gesture classifiers have appeared in the literature (Joseph & Nguyen 1998; H.T. Nguyen, King & Knight 2004; H. T. Nguyen, Knight & Ekanayke 2003). Accuracies for the classification of discrete patterns collected from able-bodied subjects reported in these papers ranged between 91% and 99%. It was concluded in these papers that such a level of accuracy would be acceptable for a practical wheelchair interface.

The range and strength of movement available to an individual varies widely between different people, particularly for people with high-level disabilities. Since the above papers measured the classifier performance for able-bodied subjects, it remains an open question as to whether an ANN head gesture classifier of this type can provide adequate performance for people with disabilities.

The results presented in this chapter also extend the results of the above papers by examining the classification of gestures in a stream of real-time data, rather than the classification of a single window of sensor data. When a command gesture is performed, sensor data is sampled and passes through the sliding window of the classifier. As the sensor data passes through the sliding window, the input to the ANN may become dissimilar to the patterns used during the training process or that the pattern may validly be classified as more than one type of gesture. It is an open question as to whether or not this will adversely affect the performance of the classifier.

3.2.2. Selection of a metric for classifier accuracy

Many possible metrics can be used to compare the accuracy of two classifiers. This section details the metric chosen and provides a justification for this selection, with respect to the properties of the metric and the alternatives.

The primary metric is the classifier error rate, that is, the proportion of test set movements that are incorrectly classified. This was chosen because it is frequently used in the literature, and is directly relevant to the practical performance of the classifier. Despite most cases found in the literature measuring the error rate on individual patterns, error rate is not limited to this use. It can be applied to streams of real-time data without loss of utility, by considering the classification over a window of data containing one movement. The error rate provides equivalent information to several other metrics described in the literature, such as the percent good metric used by Twomey and Smith (1995) and the percent classification error used by Sexton and Dorsey (2000). Error rate is chosen in preference to the percent good because the classifier is expected to make most classifications correctly, and thus the error rate will reflect changes in performance more clearly.

Sensitivity and specificity are also used as secondary metrics in order to examine the prevalence of different types of misclassification. These are of interest for the wheelchair control system because the existence of classification delay restricts the ability of the user to take corrective action in the event of a misclassification. It is trivial to show that using a different threshold value for the output of the ANN will change the overall error rate and will also alter the ratio of errors between false positives and false negatives, and thus the sensitivity and specificity. Although the nature of the real-time classification task makes the receiver operating characteristic of the classifier nonlinear, the measurement of the sensitivity and specificity for the threshold value used provides insight as to the possible influences on or causes of differences in error rate.

The most common definitions of error rate sensitivity and specificity are formulated for binary classifications as

Equation 43

$$Error\ rate = \frac{N_{FP} + N_{FN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}$$

Equation 44

$$Sensitivity = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

Equation 45

$$Specificity = \frac{N_{TN}}{N_{TN} + N_{FP}}$$

where N_{TP}, N_{TN}, N_{FP} and N_{FN} can be found from the contingency table (see Table 2) as the number of true positive, true negative, false positive and false negative classifications respectively (Fawcett 2004).

Table 2 General form of a contingency table for a binary classifier.

		Actual classification	
		Positive	**Negative**
Expected classification	**Positive**	N_{TP}	N_{FN}
	Negative	N_{FP}	N_{TN}

There is disagreement in the literature as to which method is most valid for the extension of the definition of sensitivity and specificity to multiclass applications (Ferri, Hernándéz-Orallo & Salido 2003; Hand & Till 2001). A method described by Hand and Till (2001) is used for this experiment where all windows classified as having a gesture present are treated as positive classifications. The remaining windows are treated as negative classifications. True classifications are those for which the actual classification made using the ANN matches the expected class, while the remainder are false. The generalised definitions are summarised in Table 3, Equation 46, Equation 48 and Equation 49.

Table 3 General form of a multi-class contingency table.

		Actual classification				
		Forward	**Back**	**Left**	**Right**	**Neutral**
Expected classification	**Forward**	N_{ff}	N_{fb}	N_{fl}	N_{fr}	N_{fn}
	Back	N_{bf}	N_{bb}	N_{bl}	N_{br}	N_{bn}
	Left	N_{lf}	N_{lb}	N_{ll}	N_{lr}	N_{ln}
	Right	N_{rf}	N_{rb}	N_{rl}	N_{rr}	N_{rn}
	Neutral	N_{nf}	N_{nb}	N_{nl}	N_{nr}	N_{nn}

$$N_{TP} = N_{ff} + N_{bb} + N_{ll} + N_{rr}$$ **Equation 46**

$$N_{TN} = N_{nn}$$ **Equation 47**

$$N_{FN} = N_{fn} + N_{bn} + N_{\mathrm{ln}} + N_{rn}$$ **Equation 48**

$$N_{FP} = N_{fb} + N_{fl} + N_{fr} + N_{bf} + N_{bl} + N_{br} + N_{lf} + N_{lb} + N_{lr} \cdots \\ + N_{rf} + N_{rb} + N_{rl} + N_{nf} + N_{nb} + N_{nl} + N_{nr}$$ **Equation 49**

3.2.3. Selection of a metric for classification delay

The classification delay of the classifier is caused by the time taken for sufficient sensor data reflecting that a gesture has been made to be present in the input to the ANN. In order to measure the length of the delay, it is necessary to clearly define the moments when the delay begins and ends.

Ideally, the delay would be measured between the time that the user decided to give a command and the time that the classifier recognises the command that has been given. However, it is impossible to measure or control the moment that the user makes the decision with enough precision to allow meaningful results to be obtained. To avoid this problem, it is assumed that the delay between the moment when the user decides to make a command and the time that the user begins to move is either negligible or constant.

The time of the onset of the movement can be determined by considering all of the sensor data in the recording of a gesture. In considering a recording of sensor data for a single known gesture, it can be assumed that the gesture begins from a neutral position and ends at a neutral position, and between these points, the position of the user's head will deviate from that neutral position. If it is known what type of gesture is present, an approximate method to identify the middle portion of the gesture is to determine the point of maximum deviation from the neutral position. Having found the middle portion of the gesture, the start of the movement can then be found by tracing through the recording to find the closest moment prior to the peak where the user's head is in a neutral position. The detection of a neutral position is performed using threshold-crossing techniques on both axes of the sensor.

Due to the way the ANN output is interpreted by the classifier, the end of the delay can be defined as the first time that the AAN produces a non-neutral classification after the onset of the movement, after ignoring any classification present at the time of the onset of the movement. The justification for this definition of the end of the delay is that any non-neutral classification present at the onset of the gesture is due to the window of data before the gesture, and is therefore due to either a previous gesture or noise.

3.2.4. Methodology for the measurement of delay and accuracy

Data was collected from six adults, aged between 19 and 56, with approval from the UTS Human Research Ethics Committee and informed consent from the volunteers. Of these, two had high-level spinal cord injuries (C4 and C5) and were not able to use a standard joystick to control a wheelchair. The remaining four did not have conditions affecting their movement.

Windowed data was classified using ANN architecture and weights set to those employed by Nguyen, Knight and Ekanayke (2003) as corresponding to one of four types of head movement commands: left, right, forwards or backwards. If the window was not classified as belonging to one of the four classes, it was interpreted as being neutral, or equivalently that no command had been given in that window.

Data for each person was collected in two periods of ten minutes, with the user being prompted to give a specified movement every 6 seconds. Each specified movement was chosen randomly from the following: neutral, forward nod, forward hold, backward nod, backward hold, left nod, left hold, right nod and right hold. The difference between a nod and a hold movement is that a nod moves from the neutral position to a maximum value and immediately begins returning to the neutral, whereas a hold movement pauses briefly at the maximum value before returning. A graphical real-time display of sensor data and classifier output was displayed to the subject for the first period of ten minutes. This was then hidden for the second period and no feedback was provided.

The raw sensor and classifier data was analysed then to produce the results described below. Each 10-minute period was subdivided into the six-second windows corresponding to each specified movement. Windows that contained obvious errors by the subject were removed from the analysis. The start of the movement within each window was determined to be the point where the deviation from the neutral position reached 25% of the maximum value on the relevant axis. The classification of the movement was determined as the first classification made by the ANN after the start of the movement. The delays between the start of the movement and the time of classification were then calculated, and the confusion matrix relating the expected and actual classification updated.

3.2.5. Description of results

The confusion matrices show that the classifier tends to fail to recognize movements at all rather than classify them as another type of movement. This is also reflected in the high positive predictive value and specificity.

As can be seen in the tables, the sensitivity and specificity of the ANN ranged from 81.5% to 96.0% and 90.5% to 99.5%, respectively, and mean classification delays ranged from 948 to 1624 ms.

Table 4 Confusion matrix for real-time classification of gestures performed by able-bodied subjects, with graphical feedback

		Actual classification				
		Forward	**Back**	**Left**	**Right**	**Neutral**
Expected classification	**Forward**	80	0	0	4	6
	Back	0	89	1	0	0
	Left	0	0	85	0	0
	Right	0	0	0	86	1
	Neutral	0	0	0	0	45

Table 5 Confusion matrix for real-time classification of gestures performed by able-bodied subjects, without graphical feedback

		Actual classification				
		Forward	Back	Left	Right	Neutral
Expected classification	Forward	86	0	0	1	2
	Back	0	94	1	0	0
	Left	0	0	91	0	0
	Right	0	0	0	91	0
	Neutral	0	0	0	0	48

Table 6 Confusion matrix for real-time classification of gestures performed by disabled subjects, with graphical feedback

		Actual classification				
		Forward	Back	Left	Right	Neutral
Expected classification	Forward	37	0	0	2	3
	Back	0	42	0	0	0
	Left	0	0	45	0	0
	Right	2	0	0	35	5
	Neutral	0	0	0	0	23

Table 7 Confusion matrix for real-time classification of gestures performed by disabled subjects, without graphical feedback

		Actual classification				
		Forward	**Back**	**Left**	**Right**	**Neutral**
Expected classification	**Forward**	41	0	0	0	3
	Back	0	44	0	0	0
	Left	0	2	38	1	2
	Right	0	1	1	29	11
	Neutral	0	0	0	0	22

Table 8 Confusion matrix for real-time classification of all gestures performed by able-bodied subjects

		Actual classification				
		Forward	**Back**	**Left**	**Right**	**Neutral**
Expected classification	**Forward**	166	0	0	5	8
	Back	0	183	2	0	0
	Left	0	0	176	0	0
	Right	0	0	0	177	1
	Neutral	0	0	0	0	93

Table 9 Confusion matrix for real-time classification of all gestures performed by disabled subjects

		Actual classification				
		Forward	Back	Left	Right	Neutral
Expected classification	Forward	78	0	0	2	6
	Back	0	86	0	0	0
	Left	0	2	83	1	2
	Right	2	1	1	64	16
	Neutral	0	0	0	0	45

Table 10 Sensitivity, specificity and error rate of the artificial neural network, averaged for the 4 possible classifications.

	Specificity	Sensitivity	Error rate
Able-bodied, with GUI	98.0%	90.0%	3.0%
Able-bodied, without GUI	99.5%	96.0%	1.0%
Disabled, with GUI	95.2%	85.2%	6.2%
Disabled, without GUI	90.5%	81.5%	10.8%
All able-bodied	98.7%	93.0%	2.0%
All disabled	92.8%	83.3%	8.5%

Table 11 Mean (Standard deviation) of the delay (ms) between movement and classification.

	Forward	Back	Left	Right
Able-bodied, with GUI	1308 (149)	1278 (98)	996 (100)	983 (165)
Able-bodied, without GUI	1296 (86)	1212 (75)	969 (109)	948 (94)
Disabled, with GUI	1387 (753)	1624 (802)	1119 (199)	1178 (461)
Disabled, without GUI	1378 (383)	1285 (206)	1192 (822)	1260 (663)
All able-bodied	1302 (120)	1244 (93)	982 (105)	965 (135)
All disabled	1382 (589)	1445 (592)	1153 (576)	1215 (557)

3.2.6. Discussion of results

Although the number of subjects involved in this trial is small, the results obtained lead to several useful observations. The performance of the classifier and similarities between the six subjects appear to validate the approach used in developing the prototype. The differences in error rate observed indicate that the ANN is better adapted for the recognition of gestures performed by able-bodied subjects than it is for the recognition of gestures performed by disabled subjects. The classifier error rate for disabled subjects could be considered acceptable in certain controlled operating environments or in conjunction with other control inputs, but is higher than would be considered acceptable in a wheelchair control system for more general operating conditions.

Since the same classifier was used to classify all sampled gestures, the changes in classifier performance imply that differences exist between the characteristics of the gestures. It also implies that some of these characteristics affect the performance of the classifier. This discussion notes several of the more significant observations to be made from the experimental results, suggests several reasons for the differences in gesture characteristics and how these may affect classifier performance, and proposes hypotheses which are to be investigated later in this thesis.

The classifier was shown to perform well for the able-bodied subjects in the experiment, making only 16 errors out of 811 attempts. It was found that the error rate of the classifier was slightly higher for those gestures performed while the GUI was displayed than for those in which it was hidden. Although both sensitivity and specificity are increased, the reduction in error rate following the removal of the GUI is largely associated with an increase in the specificity of the classifier.

The classifier was shown to perform less well for the disabled subjects, making 33 errors out of 389 attempts. Unlike the able-bodied subjects, it was found that the error rate of the classifier was lower for those gestures performed while the GUI was displayed than for those in which it was hidden. The drop in specificity associated with a rise in the number of false negatives following the removal of the GUI is the most significant contribution to the change in error rate.

For both able bodied and disabled subjects, the specificity of the classifier was higher than its sensitivity. That is, the ANN was found to be more likely to fail to detect a gesture than to detect a gesture and incorrectly predict the class of gesture. This is a result of the choice of threshold value for the output nodes of the ANN and is a result of the consequences perceived by Nguyen, Knight and Ekanayke (2003) to be associated with each type of error. It is notable, however, that the approximately 3:5 ratio of sensitivity to specificity between the two groups of users with the GUI displayed is similar to the ratio for the same metrics without the GUI. This supports the conjecture that the receiver operating characteristics are similar in shape, although different in magnitude.

Several possible explanations are suggested by the observed results regarding the influence of the GUI on classifier performance. The key assumptions for these explanations are that the GUI provided feedback on both the recent position of the subject's head and the corresponding ANN classification, that the data was first recorded while the GUI was displayed and subsequently without the GUI, and that classifier error rate without the GUI was superior for gestures performed by able-bodied subjects but inferior for gestures performed by disabled subjects.

The first possible explanation is that the display of the GUI augmented subjects' sense of proprioception and that characteristics of subjects' gestures changed following the removal of this augmentation. The augmentation of the subjects' senses would be of greatest benefit to the disabled subjects, as the physiological nature of their condition reduces some of their own senses. The removal of the GUI could therefore explain the increase in error rate by relatively impairing the ability of disabled subjects to control the position of their head and therefore cause changes in the characteristics of their gestures. The decrease in the error rate for able-bodied subjects in the absence of the GUI may be explained by these subjects responding to reduced information by altering their gestures in a manner that the disabled subjects were unable to do, or that disabled subjects were unaware of the changes in their movement characteristics.

The second possible explanation is that the two groups of subjects reacted to the display of feedback on the classification of their previous gestures and changed the characteristics of subsequent gestures. Since the GUI displayed both positional and classification data, subjects can be expected to have acquired some degree of knowledge, consciously or otherwise, as to the forms of movement which would be classified as belonging to each class. Once acquired, this knowledge could then affect subsequent gestures. The learning of the classifier characteristics by the subjects would generally be expected to result in the same changes for both able bodied and disabled subjects, given that all subjects were given the same briefing and had the same opportunity to observe the GUI. It is therefore unlikely that this influence alone can explain the results observed.

The third possible explanation is that the change in classifier performance is that the change is coincidental and caused by another factor. Temporal factors, such as fatigue, could explain the change in error rate following the removal of the GUI and the difference between the changes observed for able bodied and disabled subjects.

Since there are clear physiological differences between the able bodied and disabled subjects, it would be expected that fatigue would cause different changes in gesture characteristics.

The higher error rate observed for gestures performed by disabled subjects indicates a failure of the ANN to generalise from its training data. Misclassifications can be attributed to the two sources discussed in Section 3.1.3: error due to the difference between the sampled training set and the underlying ideal input-output mapping, and error due to the difference between the ideal input-output mapping for the training set and the function implemented by the ANN. The low classifier error rate for able-bodied subjects indicates that the error between the training data and the function implemented by the ANN is small. The optimisation of the ANN performed by Nguyen, Knight and Ekanayke (2003), from which the classifier implemented in this chapter is derived, is based on samples obtained from able-bodied subjects. Since ANN generalisation is theoretically poorer in regions of the input space where training data is sparse, this opens the question as to whether steps to make the training set more representative of the characteristics of gestures performed by disabled people provide a means by which to improve the performance of the classifier.

The most direct steps to take that should, in theory, make the data more representative of the characteristics of gestures performed by disabled people is to include input-output pairs sampled from such gestures in the training set. Using this approach changes the input-output mapping being approximated by the ANN. Since the ANN tested in this chapter was optimised for a training set containing only data from the gestures of able-bodied people, the results of that optimisation, including not only ANN weights, but also architecture and training algorithm, may no longer be appropriate.

It should be noted in drawing inferences from these results that they are from the testing of a single ANN classifier. The accuracy with which an ANN approximates an input-output mapping is known to be sensitive to the initial weights and training procedure, so it is unknown whether the results observed are typical of ANN classifiers trained on able-bodied data in general.

Comparing the delays between the able bodied and disabled subjects, it appears that although the delay for the disabled group is generally slightly longer, the difference between the two groups is not statistically significant. It is, however, notable that the able-bodied group had a smaller variance in the delay. The implications of this larger variance are not clear, but an increase in sample size may assist in determining this.

A practical system should include a fast acting input to allow the user to react to sudden, unexpected events, for example, to quickly stop the wheelchair after a misclassification or in an emergency. Of all data classified in this experiment, no windows of data which did not contain a gesture were classified as containing one. This combined with the low number of gestures of one class being misclassified as another class give the classifier its high specificity. However, it is clear that a small number of false positives can be expected, and the control interface must be designed to allow for this. Provided that the user is capable of detecting and performing a subsequent corrective gesture, false positive misclassifications do not necessarily prohibit the use of the control system in suitable environments. However, false negative misclassifications increase the effective classifier delay and the classification delays observed indicate that use of the existing ANN alone is not sufficient to provide dependable performance in unstructured environments. Improvements in the accuracy of real-time classification may render the control system more usable in a greater range of mobility applications.

Chapter 4. Advanced head gesture classification

The results presented in Chapter 3 show that an ANN classifier is capable of processing real-time data to recognise gestures, although gestures performed by disabled people are classified less accurately than gestures performed by the able bodied. This chapter investigates methods by which the classifier can be optimised for disabled people, using training set data that is more representative of the gesture characteristics of disabled people. Although the effect of optimisation will differ between individual users due to physiological and behavioural differences, techniques can be applied to optimise the ability of the ANN classifier to accurately detect and classify head gestures for each individual. The training set data is more representative of the gesture characteristics of disabled people by including labelled samples from data containing gestures performed by disabled people.

Specifically, the experimental aim in this chapter is to determine the variation in real-time classifier performance attributable to the network architecture and weight adjustment algorithm, and the effect of changes in the size of the training set for training ANN classifiers using data from gestures performed by disabled people. Further to this, the investigation also aims to determine the effect of these factors on the time taken for the ANN training algorithm to converge or terminate.

The optimisation examined in this chapter has two central contributions to this thesis. The first is that the selection of empirically justifiable parameters for the training of the ANN using data from gestures performed by people with disabilities provides a basis on which further changes to the classifier can be based, and provides a standard to compare measures of classifier performance. This is a necessary step for the development of the classifier, as it is reasonable to anticipate that the use of data from gestures performed by disabled subjects may alter the input-output mapping approximated by the ANN sufficiently to render selections made by Joseph and Nguyen (1998) and Nguyen, Knight and Ekanayke (2003) invalid. The second central contribution of this chapter is to demonstrate the degree of variation in the characteristics of gestures between disabled people. The results of the optimisation are tested using data from gestures performed by two disabled people, independent of the optimisation process, to show the extent to which the optimisation of classifiers generalises to people not included in the training data.

It should be noted at this point that the experimental focus in this chapter differs to that in Chapter 3. Consequently, the results presented in this chapter are not directly comparable to the results presented in Chapter 3. Chapter 3 focused upon the classification performance of a single ANN classifier. This chapter investigates the effect of optimising the training of the ANN, and is concerned more with the expected results of the training process than the performance of a single classifier.

4.1. Background

4.1.1. Training data

Twomey and Smith (1995) argue that the goal of the collected data for training purposes is to obtain a representative sample of the input-output mapping that will produce a reliable estimate of the ideal weights. Following this argument, large sample sizes are desirable to reduce the non-systematic bias that arises from the sampling process. Twomey and Smith also acknowledge that there is an upper limit on the size of a training set. This limit is due to the difficulty of collecting and labelling the data required to make up the training set. The cost of collecting and labelling training data increases as the size of the training set rises, so it is desirable to minimise the number of samples required. Castelli and Cover (1993) argue that, theoretically, the value of additional points of training data decreases exponentially, as some information about the ideal input-output mapping obtained from each additional point can be expected to be redundant.

Although the arguments of Twomey and Smith (1995) and Castelli and Cover (1993) provide an indication of how the theoretically optimal training set size may be determined for ANN input-output mappings whose properties are known, neither the ideal weights nor the utility of a sample of training data are known in advance for the head gesture classification task. Haykin (1999) notes that the necessary size of the training set depends on the network architecture, classification task and accuracy required. The arguments of Haykin suggest that the marginal utility of additional training data can only be determined for a fixed network structure, and since the properties of the classification task are not known in advance, it is necessary to use empirical methods.

It is an open question as to whether there is sufficient practical benefit to be obtained by increasing the number of points for each class in the training set. The effect of additional training points on the classification of new gestures by the people from whom the training data was obtained is one aspect which is of interest in this question, as this will indicate whether the training data set provides a good approximation of the underlying ideal input-output mapping for the people who provided the training data. The effect of additional points of training data for individual end users with disabilities is also of particular interest.

4.1.2. Head gestures in sensor data, classifier window and ANN input

This section discusses the properties of the classification task and inputs to the ANN, expanding on the details described in Section 3.1.5. Examples are provided of gestures as contained in the classifier input stream, and windowed data comprising the corresponding ANN inputs.

The classifier component of the control system has the task of detecting the presence of each type of command gesture in the acquired sensor data. The classifier uses a sliding window of sensor data as the input to an ANN, the outputs of which are used to indicate the presence or absence of each type of command gesture. Each new sample of sensor data is input to the ANN along with other samples acquired in the preceding 2 seconds. The resulting ANN output is used to determine whether a gesture is present, and if present, of which command it represents.

Windows of the input data to the classifier in which a gesture is present are relatively rare when compared to the prevalence of windows that do not contain a gesture. The period between gestures contains windows of data in which no gesture is present.

Further, in the period immediately prior to and following a gesture, the windows of data input to the ANN contain a portion of a gesture.

The classifier considered in this chapter distinguishes between 5 classes of head gesture. These gestures are the forward nod (flexion followed by extension), backward nod (hyperextension followed by extension), left tilt (left lateral flexion and extension), right tilt (right lateral flexion extension) and head shake (oscillatory left and right lateral flexion). While not performing a gesture, subjects were instructed to hold their head approximately level.

The magnitude of signal is normalised to adjust for the initial neutral position held by the subject and the maximum signal recorded on that channel of that subject, such that the normalised signal had a magnitude of 0 when in the neutral position and 1 when the subject's head was positioned at the greatest extension. Separate scaling and offset parameters were determined for the mediolateral and anteroposterior axes, and the recalibrated each time the sensor was fitted to a subject.

The examples in Figure 9 through Figure 13 are typical of gestures performed by able-bodied subjects. More detail on the people from whom data was collected is included in Section 4.3.

Cross-talk between sensor channels is evident in some gestures. One cause for this effect is misalignment between the axes of the sensor and the anatomical axes of the user. Another cause is that the properties of the gesture where the movement of the subject during the gesture may not been restricted to a single plane. An example of the former case can be seen on the mediolateral channel in Figure 9 where it deviates in unison with the anterioposterior channel, but to a lesser degree. An example of the latter is evident in Figure 11, where data in the anteroposterior channel changes during the period that data in the mediolateral channel is changing, but there is relatively little correlation between the channels.

Evidence of movement while the subject is not performing a gesture is often observed on both channels, as can be seen in each of the plots in Figure 9 through Figure 13. These signals are of much lower amplitude than is observed during gestures, and are due to natural variations in head and body position. These variations exhibit both high frequency components, which are evident in each of the plots in Figure 9 through Figure 13, and slower trends, such as the gradual shift in baseline signal following the gesture in Figure 13. These small variations are generally not correlated with the execution of gestures, but in some cases, the average neutral position to which a subject moves following a gesture differs to the neutral position prior to the onset of the gesture. An example of this can be seen in Figure 11. This change in neutral position may be due to the proprioceptive sensation of the subject being elevated while moving to perform the gesture.

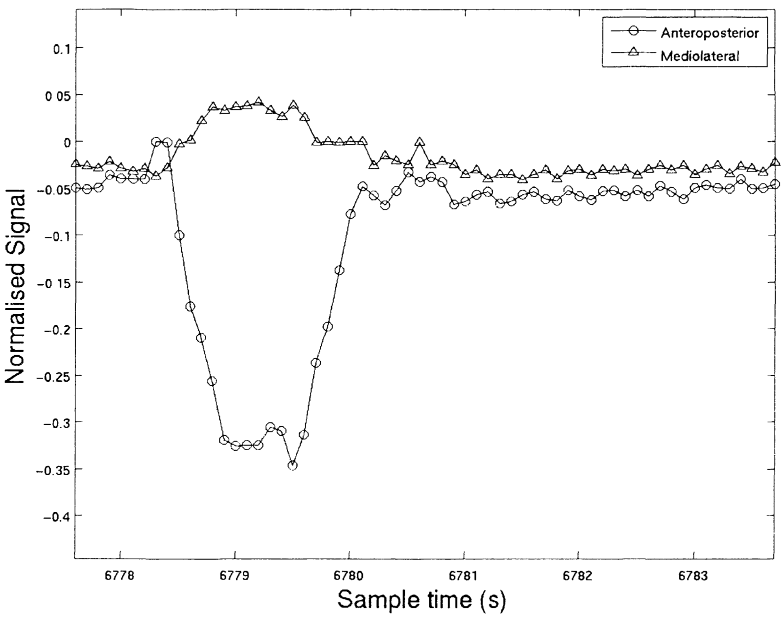

Figure 9 Example backward-nod gesture, performed by Subject 1

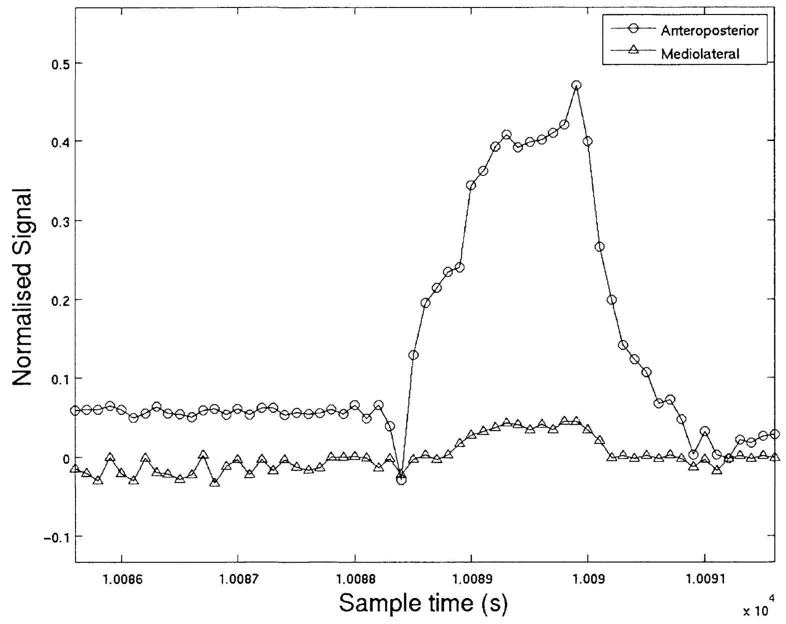

Figure 10 Example forward-nod gesture, performed by Subject 3

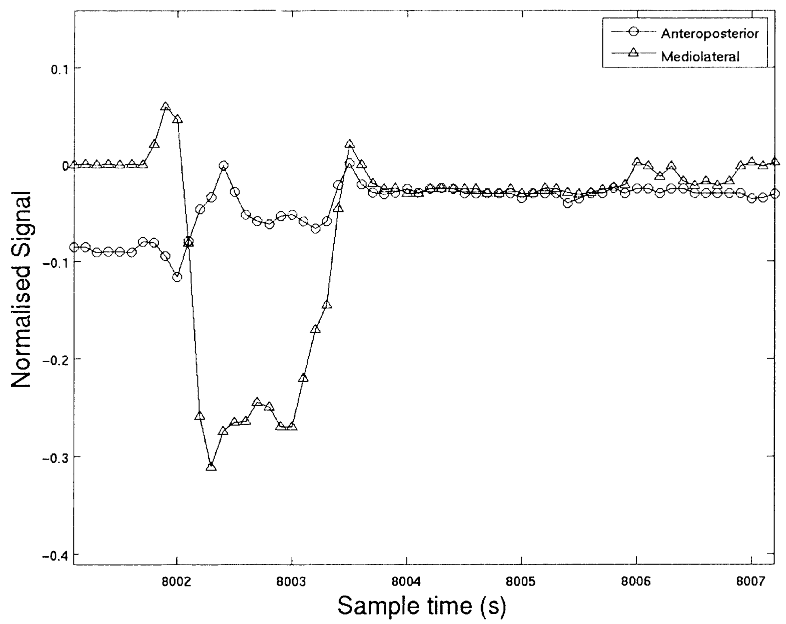

Figure 11 Example left-nod gesture, performed by Subject 2

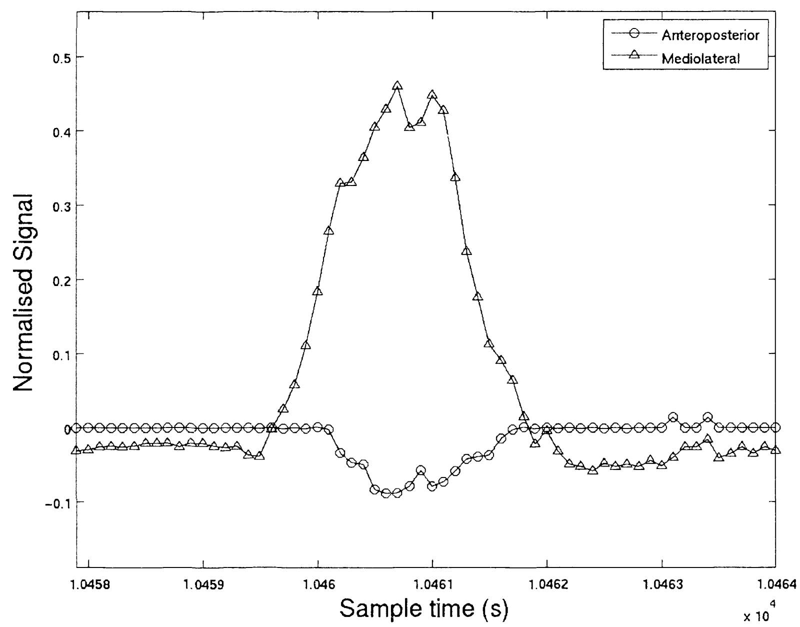

Figure 12 Example right-nod gesture, performed by Subject 4

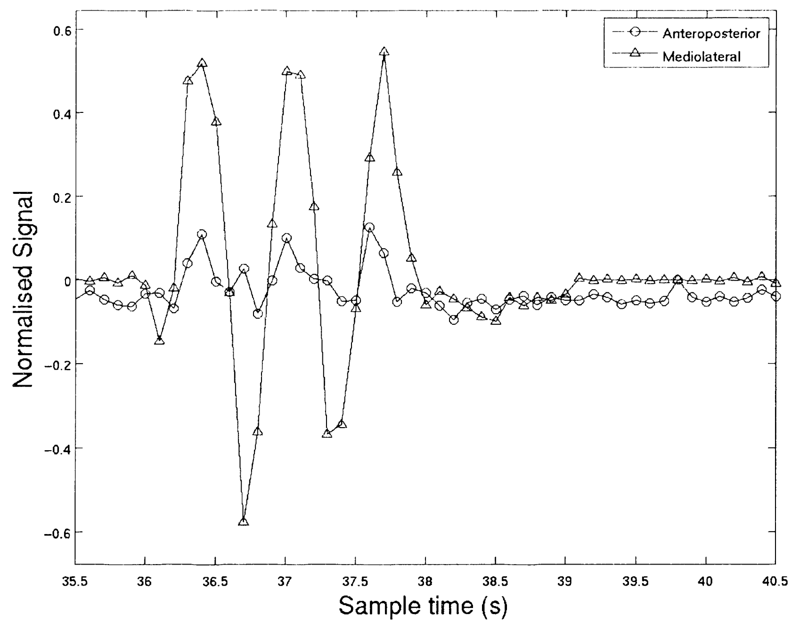

Figure 13 Example shake gesture, performed by Subject 5

As previously discussed, many of the ANN input patterns observed in a stream of sensor data cannot be unequivocally labelled for the head gesture classification task.

In order to compile the training set used in this chapter, one 2s window of data was selected from each recorded gesture in the set of training data. These windows were chosen as being the most representative of the gesture present and were assigned labels accordingly. The windows of data selected from the example gestures in Figure 9 through Figure 13 are shown in Figure 15 through Figure 18to illustrate the effect of this process. The windowed data was converted to vectors of real-valued numbers able to be input to the ANN by appending data from the mediolateral axis to the data from the anteroposterior axis.

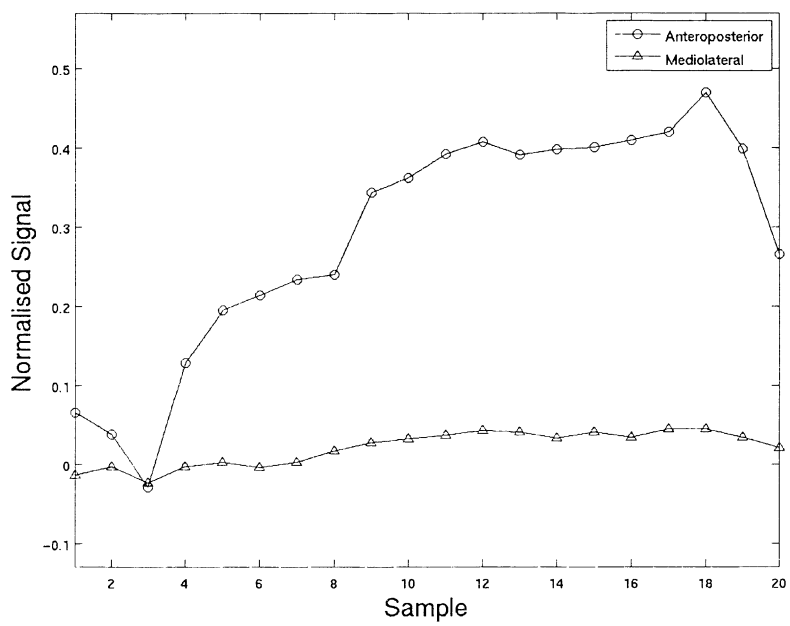

Figure 14 Selected 2s classifier input window corresponding to forward-nod gesture in Figure 10 (Subject 3).

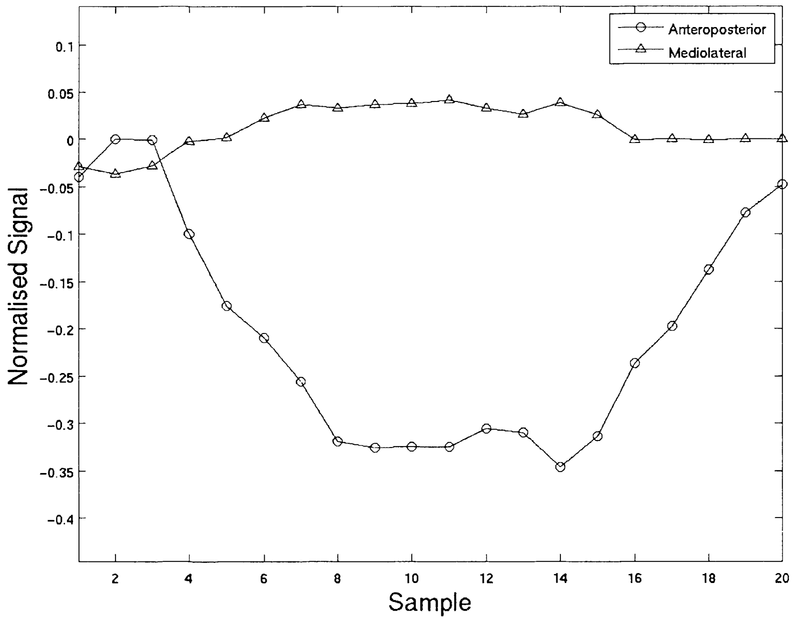

Figure 15 Selected 2s classifier input window corresponding to backward-nod gesture in Figure 9 (Subject 1).

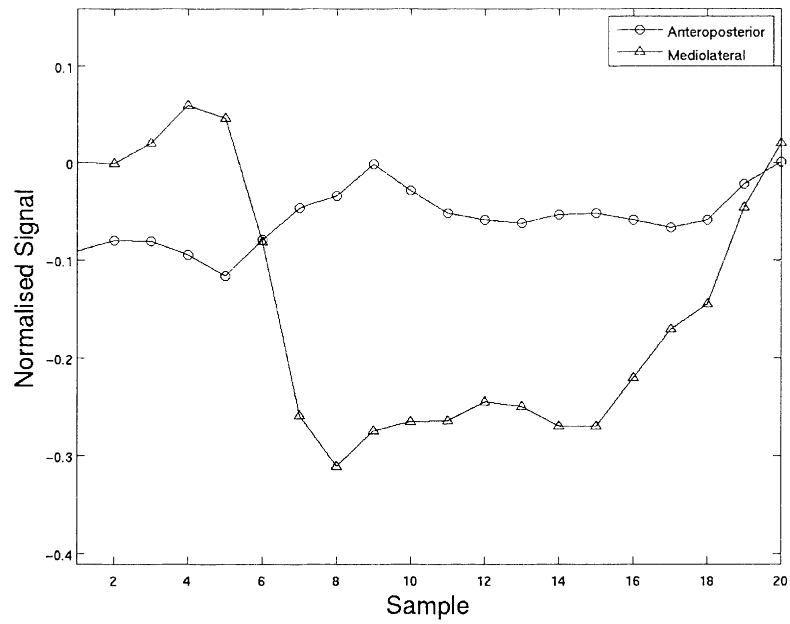

Figure 16 Selected 2s classifier input window corresponding to left-nod gesture in Figure 11 (Subject 2).

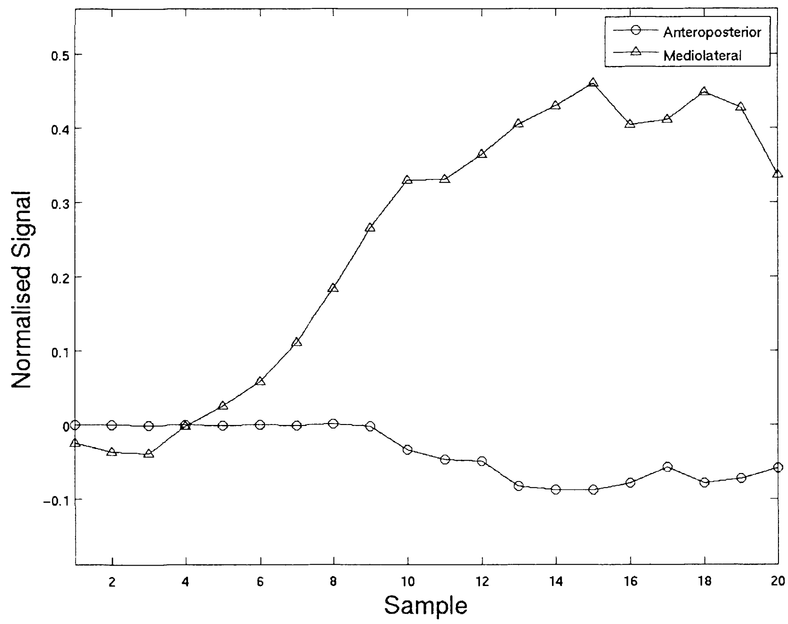

Figure 17 Selected 2s classifier input window corresponding to right-nod gesture in Figure 12 (Subject 4).

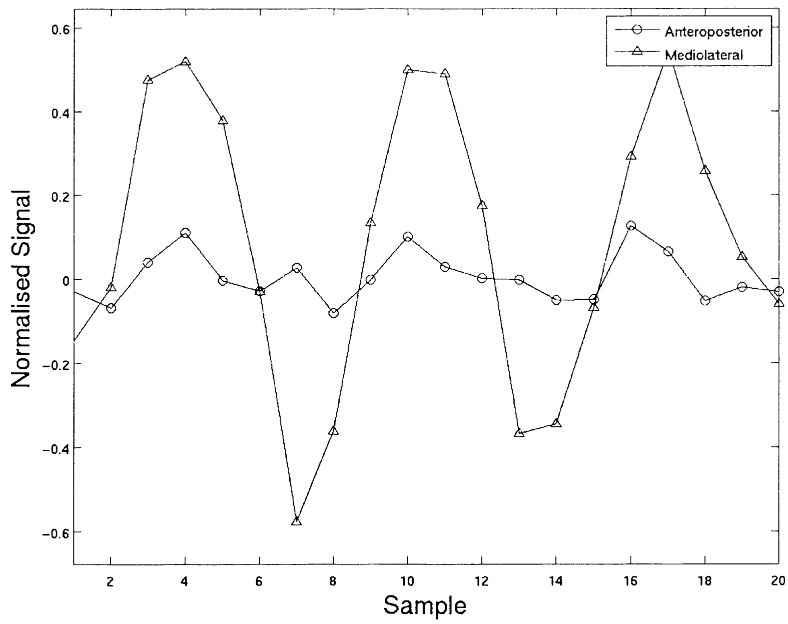

Figure 18 Selected 2s classifier input window corresponding to head-shake gesture in Figure 13 (Subject 5).

4.1.3. ANN Architecture

Selection of an optimal architecture to achieve a practical head gesture classifier is a separate problem from finding the architecture optimal for the input-output mapping of the data in the training set. Determining the optimum number of nodes in a hidden layer using only information that is available a priori is difficult or infeasible in most practical applications. This is due to the complexity of the network mapping and the non-deterministic nature of many training procedures. However, several heuristics, general principles and empirical techniques are often combined to determine the final network architecture (Zurada 1992).

The network architecture for an ANN classifier is often selected empirically for a particular application by comparing the classifier performance of classifiers with different architecture, trained using the same data (Looney 1997). Although techniques have been proposed for determining the optimal architecture through the addition or removal of hidden nodes during training, such as Sietsma & Dow (1988), Guyon (1991) and Reed (1993), there is a lack of evidence to support the use of such methods in the case. Use of these methods would not provide the results necessary to establish whether the optimal ANN architecture for the generic training data is consistent with that for specific disabled end users.

It is an open question as to what the ANN architecture is optimal for the classifier to provide the best performance on the set of people who provided the training data. It is also an open question as to whether the ANN architecture optimised on a general training set will be sufficiently similar to the optimal architecture for disabled end users to provide good generalisation.

4.1.4. Training algorithm

Many different algorithms have been proposed in the literature for the training of ANN, of which some of the most commonly used algorithms are summarised in Section 3.1.4 or found in references such as Widrow and Lehr (1990) and Hagan (1995). Most ANN training algorithms refine an initial estimate of the weights by using an iterative process to minimise a function of the error in terms of the ANN weights, at the exemplar points in the training set. These algorithms are applied on the assumption that the minimum of the error function corresponds to the ANN weights that yield an optimal input-output mapping. These commonly used algorithms use the gradient of the error function at the exemplar points to determine the magnitude and direction of changes to the weights in each iteration.

Algorithms differ in the way in which the magnitude and direction of the change in weights is determined, resulting in algorithms following different paths through the weight space of the ANN. Since the error function may have many minima, differences in the way in which the magnitude and direction of the change in weights is determined in each algorithm result in different algorithms following different trajectories through weight space. The architecture of the ANN and the data in training set determine the properties of the error function to be minimised by the training algorithm and the different trajectories can therefore result in different levels of generalisation.

Although the results of Hornik, Stinchcombe and White (1989) and Irie and Miyake (1988) show that an ANN with a sufficiently large hidden layer can approximate any mapping function with arbitrary accuracy, they do not provide a guarantee that such an ANN can be trained to do so. The use of a sufficiently large hidden layer ensures that weights exist such that the ANN approximates the input-output mapping arbitrarily well, but does not ensure that those weights can be found. The "no free lunch" algorithms of Wolpert and Macready (1992; 1995; 1997) conclude that no single algorithm can exist that provides superior results for all classification tasks. Wolpert and Macready further show that some algorithms may provide better results than others when compared for a particular application. It therefore remains to consider whether characteristics of the head gesture classification task render some algorithms more effective than others.

There are few reports in the literature comparing different training algorithms for head gesture classifiers of the type considered in this chapter. Joseph and Nguyen (1998) showed that the delta-rule gradient descent algorithm described in Section 3.1.4.1 could be used to train real-time head gesture classifiers, but did not compare the results against other algorithms. It was found in Taylor, Nguyen and Craig (2002) that the delta-rule gradient descent outperformed Bilski and Rutkowski's (1998) recursive least squares algorithm on the related task of instantaneous head gesture classification for able-bodied subjects, that is, the classification of windows of data without consideration of real-time effects. These results showed that, on average, the delta-rule gradient descent algorithm achieved a threshold validation set error in fewer iterations, although the difference was not found to be statistically significant (2.030 ± 0.342, 2.943 ± 0.752 respectively). Notably, the delta rule algorithm required significantly less time to perform the calculations for each iteration (0.111 ± 0.002 s/cycle, 2.536 ± 0.023 s/cycle) and was more likely to converge than the RLS algorithm (RLS failed to converge below a threshold validation set error on 12.6% of all attempts, while all attempts to train an ANN using the delta rule passed that threshold). Although suggestive, these results leave open the question as to whether similar differences can be expected for the real-time head gesture classification task for disabled people.

To assess the effect of the choice of training algorithm on classifier performance, three commonly used ANN training algorithms have been selected for further examination: delta rule gradient descent, scaled conjugate gradient and Levenberg-Marquardt. These algorithms are selected based on their frequent use in the ANN literature (Demuth, Beale & Hagan 2005; Hagan 1995). These algorithms were described in more detail in Section 3.1.4.

4.2. Methodology

The experiments described in this chapter aim to address the open questions identified above by determining the effect of the selection of ANN architecture, training algorithm and data set size on the performance of real-time head gesture classifiers. This effect is measured with respect to the ability of the ANN to accurately classify gestures performed by the mixed group of people who provided the training data for the ANN, and also for two people with high level disabilities who were separate from that group. From the results of the experiments, selections are made of those treatments which provide the best results for the general data set, and the performance of classifiers trained using these generic optimal selections are compared to the treatments providing the best results for the separate disabled subjects.

The classification task used in this chapter is the real-time classification of head gestures. As in the classifier used in the prototype control system in Chapter 3, the task requires that an ANN be trained on a set of exemplar patterns derived from recorded gestures. In operation, the ANN is used to detect and classify gestures in a stream of sensor data by using a first-in-first-out (FIFO) buffer. The input to the ANN is the most recent two seconds of sensor data for each channel. The ANN is trained on exemplars from five classes of head gesture: forward nod (flexion followed by extension), backward nod (hyperextension followed by extension), left tilt (left lateral flexion and extension), right tilt (right lateral flexion extension) and headshake (oscillatory left and right lateral flexion).

Sensor data for the experiments in this chapter was obtained from recordings of the gestures used to test the prototype wheelchair control system, as described in Chapter 3. This data was supplemented by data collected from two additional disabled subjects. The protocol for the collection of this additional data is detailed in the following section. For each gesture in those recordings, the portion of the recording commencing several seconds prior to the commencement of the gesture and finishing after the end of the gesture was extracted to create an extracted recording containing a single gesture. From within each single gesture recording, a 2 second portion was selected as the most representative of the gesture by a human expert. One such 2s portion and its associated label is referred to in this chapter as an exemplar pattern.

For the purposes of these experiments, the expected performance of a classifier is optimised by minimising the expected error rate of the classifier. The expected error rate is determined by training multiple ANN classifiers using the same choice of parameters, but with different initial weights. The resulting classifiers produce a distribution of error rates, from which statistics relating to the central tendency and variation of error rate can be derived. The same collection of classifiers is used to produce statistics relating to the central tendency and variation of training time.

Throughout this Chapter, such a collection of ANN classifiers, trained with identical parameters but different initial weights, is termed to be a treatment.

To allow the use of more sensitive statistical tests, a randomised complete block experiment design is used, also known in the statistics literature as blocking. Blocking is an extension of matched pairs techniques, such as the paired t test, to the case of having greater than two treatments. In creating a block, as many parameters as possible are kept constant apart from that which differentiates between the treatments. Each block contains one classifier for each treatment. For example, in comparing between different architectures, all classifiers in one block are trained from identical initial weights (omitting those weights which are omitted by the architectures with fewer hidden nodes), using the same training algorithm and identical allocations of data to the training, validation and test sets.

Error rate, that is, the proportion of test set movements that are incorrectly classified, was used as a metric for the performance of each trained classifier. This was chosen on the grounds that it has been frequently used in the literature, and is directly relevant to the practical performance of the classifier. Although the error rate is measured on individual input patterns in most cases found in the literature, it is not limited to this use and can be applied to the detection and classification of gestures in a real-time application without loss of utility. Error rate provides equivalent information to several other metrics described in the literature, such as the percent good metric used by Twomey and Smith (1995) and the percent classification error used by Sexton and Dorsey (2000).

The null hypothesis for each experiment is that each treatment is of equal effectiveness. Equivalently, the null hypothesis is that the difference between matched observations within a block is zero. For a given level of significance, blocking techniques are noted in the statistics literature as increasing the power of the test by reducing variance of the test statistic, provided that interaction between blocks and treatments can be neglected (Bland 2000; Fisher & Van Belle 1993; Walpole, Myers & Myers 1998).

The Friedman 2-way analysis of variance by ranks is used in this Chapter to determine the significance of difference in the test statistics. As a nonparametric alternative to the commonly used parametric one-way analysis of variance with repeated measures, the Friedman test operates on the rank of each treatment observation, compared to the other observations within that block. The mean rank for each treatment is used to determine whether there is sufficient evidence to conclude that, if there was no difference between treatments, the observed results are unlikely to have occurred by chance alone. The sign test and Spearman rank correlation test can be considered special cases of the Friedman test, where the experiment involves only two blocks or two treatments, respectively (Gibbons & Chakraborti 1992).

The assumptions required for the Friedman test are that all data come from populations having the same continuous distribution, apart from possibly different locations due to treatment and block effects, and that all observations are mutually independent (*Statistics Toolbox User's Guide* 2006). These assumptions satisfied more generally than those required for the equivalent parametric ANOVA, which require that the data come from populations having normal distributions. Whenever the assumptions required for the parametric ANOVA with repeated measures are met, the assumptions of the Friedman test are also satisfied (Gibbons & Chakraborti 1992).

In the event that analysis of variance leads to the rejection of the null hypothesis, such a result indicates that the treatments do not have equal effect but does not indicate which of the treatments is different (Walpole, Myers & Myers 1998). Pair-wise comparisons can be used to determine the significance of differences between individual treatments. However, performing multiple pair-wise comparisons increases the overall risk of a Type I error unless the sensitivity of each individual comparison is reduced. In this chapter, the Tukey-Kramer procedure is used on the block ranks calculated in the Friedman test to determine the significance of differences between each treatment.

4.3. Procedure

4.3.1. Data collection procedures

As in the experiment described in Chapter 3, head position data was measured using an ADXL202EB-232A fitted in a cap. The sensor and processing software were calibrated by observing the pre-processed data values while the subject moved their head circularly to the comfortable limit of their range of movement. The parameters calibrated were the offset present on both axes when the subject's head was in a neutral position, the range of values produced on each axis when the subject extended his/her head to the comfortable limit of their movement in that direction, and the orientation of the ADXL202EB-232A in the hat. The sensor and software were calibrated such that flexion and hyperextension of the subject's head respectively resulted in positive and negative values on the X channel. Right lateral flexion and left lateral flexion respectively resulted in positive and negative values on the Y channel.

As a requirement of the human research ethics approval under which this data was collected, an identifier was generated for each subject so that the data for that subject could be identified but would not be directly traceable to that subject without reference to other records, and data was stored using filenames based on this identifier. The able-bodied subjects from whom data was collected in Chapter 3 will be identified as subjects 1 to 4. Data was also collected from an additional able-bodied subject, identified as subject 5. The two disabled subjects who provided data using the protocol described in Chapter 3 are identified as subjects 6 and 7, and the two disabled subjects who provided data using the protocol below are referred to as subjects 8 and 9. The level of injury of each subject is summarised in Table 12.

Data was collected from two adults, who were affected by high-level spinal cord injuries and were not able to use a standard joystick to control a wheelchair. These two subjects exhibited a greater degree of disability than the disabled subjects from whom data was collected in Chapter 3. Subject 8 was affected by a C4 level spinal cord injury and exhibited a postural tremor. Subject 9 was affected by a C6 level spinal cord injury, with only partial function below C5 and a restricted range of lateral head movement. The protocol for the collection of the data involved a brief explanation to the subject of the steps he/she would be following, the recording of informed consent, the fitting and calibration of instruments, and the recording of sensor data while the subject responded to prompts. Following this collection process, the recorded data was reviewed to ensure that the recorded window edges were appropriately placed and neutral windows were noted.

Table 12 Summary of subjects providing gesture data

Subject	Level of SCI	Sex
1	None	Male
2	None	Female
3	None	Male
4	None	Female
5	None	Male
6	C5	Male
7	C4	Male
8	C4	Female
9	C6	Male

The recording of informed consent was conducted in accordance with the procedures approved by the UTS Human Research Ethics Committee. Subjects were advised that data would be recorded in sessions of between 5 and 10 minutes, and that the duration of each session could be varied according to their preference. Subjects were also advised that they could suspend or stop the recording of data at any time if they wished to do so. Consent was recorded on a form similar to the one shown in Appendix E.

Data for each subject was collected in periods of between 5 minutes and 10 minutes, as directed by the preference of the subject. Each specified gesture was chosen from the following: neutral, forward nod, backward nod, left nod, right nod and headshake. Sensor data was recorded at a sample rate of 50 Hz. Sensor data was displayed to the subjects during the briefing process, but was not shown while gestures were recorded.

Seventy recordings of each gesture were recorded for each subject. In order to avoid instances of the subject being able to pre-empt the prompt, the sequence in which gestures were prompted was randomly permuted each time the 5 gestures were repeated. For the same reason, the delay between each gesture was varied by a random period between 5s and 6s in duration. Details on the selection of the number of samples are reviewed in Appendix D.

As was found in collecting the data for the experiments described in Chapter 3, a portion of the data collected for each subject can be expected to contain discrepancies between the gesture that was prompted and the gesture that the subject actually performed. The major causes of these discrepancies were the subject either becoming distracted or pre-empting (consciously or subconsciously) the movement that would be prompted, even though they were made aware that the prompts were random and unpredictable.

4.3.2. Initialisation of ANN parameters and allocation of data

One set of initial weights was created for each block of matched classifiers. Initial weights were generated using a uniformly distributed variable in the range –0.5 to 0.5, excluding 0. Each classifier used identical input formats and output encoding. Data from each channel was adjusted for the offset introduced by the non-zero angle of the sensor when the subject's head was in a neutral position. A low-pass anti-aliasing filter, with corner frequency 4.75Hz was applied to the data in both channels and the data down-sampled to a sampling rate of 10Hz. Inputs to the ANN were comprised of 20 samples from each channel. A one-hot-one encoding was used as the target ANN output for each of the classes of gesture. That is, the target output of the ANN contained 5 output nodes, exactly one of which had a target value of 1 to indicate the class of the gesture that was present, with the other outputs set to -1. Where no gesture was detected as being present, the target ANN outputs for each output node was –1.

Recorded gestures were allocated from the collected samples to provide training, validation and test data. The allocation of data to the training, validation and test sets was kept constant within each block of classifiers.

The allocation of data to training, validation and test sets for each matched block is summarised in Table 13. For the examination of ANN structure and ANN training algorithm, each matched block was allocated a training dataset comprised 30 exemplar patterns of each class, selected randomly from the data obtained from subjects 1 to 7. The validation dataset comprised 30 recorded gestures of each class, and the corresponding input patterns, selected randomly from the remaining data obtained from subjects 1 to 7. The test set data for the two separate disabled subjects comprised 50 recorded gestures, randomly selected from the data provided by subject 8, and the same number of recorded gestures randomly selected from the data provided by subject 9.

For the examination of the marginal effect of additional training data, each matched block was also allocated 2 supplementary datasets, each containing 15 exemplar patterns obtained from subjects 1 to 7. Each of the supplementary datasets contained exemplar patterns not allocated to any of the other training, validation or supplementary datasets for that block.

Table 13 Composition of training, validation and test sets for experiments in Chapter 4

	Training Set	Supplementary Training Sets	Validation Set	Test Set 1	Test Set 2
Number of samples per class	30	15	30	50	50
Subjects providing data	1, 2, 3, 4, 5, 6, 7	1, 2, 3, 4, 5, 6, 7	1, 2, 3, 4, 5, 6, 7	8	9

4.3.3. Training of artificial neural networks

Each classifier was trained using the Matlab Neural Network Toolbox (*Matlab Neural Network Toolbox* 2006) implementation of the relevant ANN training algorithms, using parameter settings given by Demuth, Beale and Hagan (2005). Training time was measured as the period from the initiation of the training procedure to the termination of that procedure, regardless of whether termination resulted from algorithm reaching the maximum number of epochs or the satisfaction of other stopping criteria.

The criteria for the termination of the ANN training algorithm were:

- Training set error below a threshold, $E < E_{max}$;
- The maximum number of iterations, $k \geq k_{max}$;
- Performance measured on validation set, E_v, increased more than C_v times since the preceding decrease;
- Flat error gradient, (Scaled Congugate gradient only. That is, $\mathbf{r}_k = 0$);
- μ_k exceeds a limit, μ_{max} (Levenberg-Marquardt only)

Where $E_{max} = 0$, $k_{max} = 10^4$, $C_v = 5$, $\mu_{max} = 10^{-10}$.

For each classifier trained, the ANN weights selected from all of the weights visited by the training algorithm were selected according to the weights producing the lowest MSE, as determined using the exemplar patterns in the validation dataset.

4.3.4. Measurement of classifier performance

Classifier performance was measured across the validation or test datasets by applying selected ANN weights to the recorded gestures in the dataset. Each recorded gesture in the data set was classified by applying the ANN to each window of input data in succession, using the same sliding window technique described in Chapter 3. A gesture was considered to be detected when any of the ANN outputs exceeded a threshold value of 0.75. Once detected, the gesture was classified according to the output node having the maximum value. In the event that no output node of the ANN exceeded the threshold value before the end of the recording, the recording was classified as containing no gesture. The error rate was determined for each classifier by determining the proportion of misclassified recordings, out of all the recordings in the dataset.

The secondary metrics used in Chapter 3, sensitivity and specificity, are not used to measure the performance of classifiers in this chapter.

4.3.5. Optimisation of ANN architecture

The effect of the ANN architecture was examined as the number of hidden nodes was increased from an over-determined to an underdetermined network. That is, the number of hidden nodes was varied from having fewer than the number required to be able to partition the input space such that all of the exemplar patterns in the training set were correctly classified for any of the possible weights, to having more free parameters in the network than can be adequately constrained by the exemplar patterns in the training set.

The lower bound on the number of hidden nodes for this experiment was determined using the rule proposed by Looney (1997) that the number of hidden nodes, N_h, in a single layer should be at least $\log_2(M)$, where M is the number of disjoint linearly separable regions in the input space. Since there is at most one output class mapping from a particular input pattern, there must be at least one disjoint linearly separable region in the input space for each of the six output classes. Consequently, the lower bound was chosen to be 2 hidden nodes.

The maximum number of linearly separable regions in the input space is unknown, making the setting of the upper limit more arbitrary. The number of linearly separable regions in the input space could theoretically be as low as one per class, or as high as one per exemplar pattern. Looney (1997) proposes the assumption that there are two linearly separable regions per class, although there is little theoretical or empirical evidence to support this assumption. A greater tolerance was therefore allowed, and the upper limit was set to 30 hidden nodes.

4.3.6. Evaluation of the difference between ANN training algorithms

Three ANN training algorithms are examined in this chapter, being the delta rule, scaled conjugate gradient and Levenberg Marqardt algorithms. Although each of the algorithms is based on gradient descent, the algorithms differ in the assumptions and approximations made regarding the nature of the error surface. Since the algorithms converge along different paths through the ANN weight-space, it is possible that the resulting classifiers approach different minima in the error function.

Each training algorithm is considered as a separate treatment in this experiment. The algorithms were implemented in accordance with the specifications of Demuth, Beale and Hagan (2005). ANN structure was set according to the optimal architecture identified in the previous experiment.

4.3.7. Evaluation of the marginal effect of training data

The effect of the size of the training dataset was evaluated by examining the difference in performance observed from ANN classifiers trained using the algorithm and structure selected in the previous two experiments, against matched classifiers trained using an enlarged training dataset. Three treatments were considered in this experiment. One treatment was a control, using the algorithm and structure selected in the prior experiments. Two active treatments were used, one increasing the size of the training dataset by 50%, thus containing an additional 15 exemplar patterns per class. The other treatment increased the size of the training dataset by 100% to include the exemplar patterns from the control treatment and first active treatment, plus another 15 exemplar patterns per class. The datasets therefore contained 30, 45 and 60 exemplar patterns per class.

4.4. Results

This section presents the results of the experiments described above. The first results presented are those for the evaluation of the effect on the generic data set and the selection of generically optimal treatments based on the observed results. Following this, the results for the classifier performance when measured for each of the two separate highly disabled test subjects are presented.

The results observed do not follow a normal distribution, so non-parametric plots and statistical tests are presented. Results that are represented by absolute metrics are presented as box plots. Comparative results between treatments are shown as rankings. Ranks are used for the comparison between treatments using nonparametric statistical tests, which do not require the assumption of each population having a normal distribution.

Each box in the box plots presents the result for a single treatment. The upper and lower limits of the box indicate the 75th and 25th percentiles respectively. The mark between these levels indicates the median observation for that treatment. The whiskers above and below the box indicate the range of the observed values, excluding outliers. Outliers are those observations that are more than 1.5 interquartile ranges above or below the 75th or 25th quartile respectively, and are marked on the plots individually.

Classifiers were ranked on performance within each block by comparing the error rate of each classifier in the block. The classifier having the lowest error rate in the block was given a rank of 1. In the event that multiple classifiers in a block had the same error rate, each was assigned the rank of the median of the tied values. For example, if 10 classifiers to be ranked had error rates of 0.1, 0.2, 0.2, 0.4, 0.5, 0.5, 0.5, 0.7, 0.8, 0.9, and 1.0, then they would be given ranks of 1, 2.5, 2.5, 4, 6, 6, 6, 7, 8, 9, 10 respectively. Ranks for training time were calculated in a similar manner.

Each row in the rank plots shows the mean rank of the observations in one treatment, where each observation is ranked against observations from other treatments in that block. A treatment that consistently outperforms another would be expected to have a lower rank. The bars on each side in these plots indicate two standard deviations above and below the mean rank. Although the underlying distributions may be non-normal, the rank of these observations is an unbiased, normally distributed estimate (Gibbons & Chakraborti 1992).

The box-whisker plots show the absolute performance of the treatments. The box plots do not allow comparison between blocks of matched classifiers, but show the practical difference between sets of classifiers. The box plots also allow qualitative comparison between experiments. The rank plots allow comparison between treatments within the experiment to which they pertain. They show the relative performance of treatments against each other, and offer the reader with a means to visualise the meaning of the nonparametric statistical tests discussed below.

Friedman's non-parametric analysis of variance by ranks was used to determine whether the variation between any of the treatments was statistically significant at the significance level of 5%. The Friedman's ANOVA test is similar to a parametric ANOVA test, using the rank of each observation rather than the observation itself. Friedman's ANOVA test was first proposed by Friedman (1937). It can be found widely in the literature on nonparametric statistics, and is described by Walpole, Myers & Myers (1998). The null hypothesis of the Friedman's ANOVA test is that there is no difference between each of the treatments. Ranking observations within each block, the average rank of observations from each treatment would be expected to be equal. The null hypothesis is rejected if the average rank of observations for one treatment is significantly different to that of another treatment.

When a significant difference is found to exist using Friedman's ANOVA test, it indicates that at least 2 of the treatments are different, but does not indicate which treatments differ. In the event of a positive result, a Tukey-Kramer pairwise comparison of ranks was used to determine which treatments were different. The ranking of classifiers within each block as performed when using the Tukey-Kramer pairwise comparison was used to select the generic optimal treatment for each experiment. Performing a simple paired test to check for a difference between each pair of treatments has a chance of detecting a difference where no difference exists. The Tukey-Kramer procedure allows comparisons to be made between the treatments while maintaining an upper bound on the likelihood of committing a Type I error. That is, it is a method for performing multiple comparisons that ensures the overall level of significance (Walpole, Myers & Myers 1998).

4.4.1. Results from ANN training and testing on generic datasets

The treatments found to provide the best classification performance on the gestures in the validation set are summarised in Table 14. Further detail on the results follows in Sections 4.4.1, 4.4.2 and 4.4.3.

Table 14 Rank of validation set classifier error rate using best performing treatments

Optimal treatment on generic dataset	**Ranked error rate (mean ± 2 s.d.)**	**Significantly different treatments (p=0.05)**
29 Hidden Nodes	12.8 ± 4.6 of 29 treatments	2, 3, 4, 5, 6, 7, 8, 15, 17, 19, 24
Scaled Conjugate Gradient	1.3 ± 0.6 of 3 treatments	Delta rule, Levenberg-Marquardt
30 training samples per class	1.8 ± 0.6 of 3 treatments	45 samples/class

4.4.1.1. ANN architecture

As is evident in Figure 19, the error rate varies considerably for each architecture. Although the spread of error rates observed for each architecture is considerably larger than the variation between the central tendencies, examining the difference in error rate of the matched classifiers in each block shows a statistically significant difference exists.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that network architecture has no effect on the validation set real-time classifier error rate.

After ranking each classifier by validation set error rate within each block, it was found that classifiers trained with 29 hidden nodes provided the lowest ranked, and therefore most optimal, observed performances. Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between this treatment and architectures with less than 8 hidden nodes. Significant pairwise differences were also found to exist between classifiers with 29 hidden nodes and those with 15, 17, 19 and 24 hidden nodes.

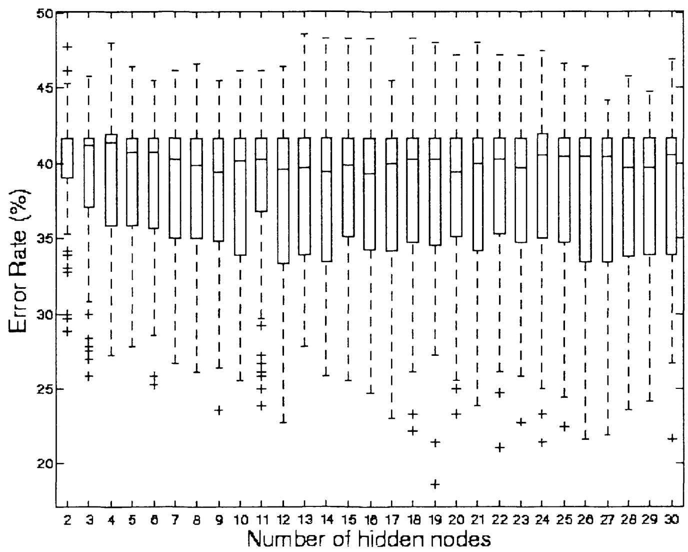

Figure 19 Real-time classification error rate measured on validation set as the network architecture increases in size

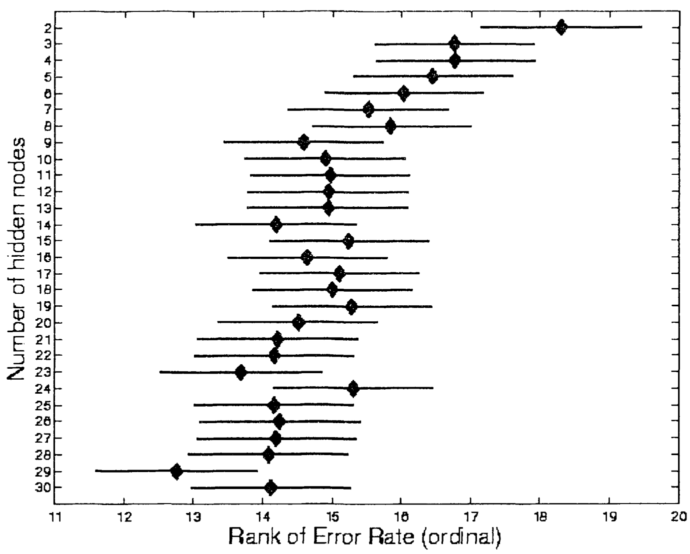

Figure 20 Rank of real-time classification error rate measured on validation set as the network architecture increases in size

As is evident in Figure 21, the time required to train the ANN classifiers varies considerably between architectures but the variation between ANN similar architectures is small.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that network architecture has no effect on the length of time required to train an ANN classifier for this application.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between classifiers with 29 hidden nodes and all other architectures except those with 28 and 30 hidden nodes. Significant differences were found to exist between almost any pair of architectures differing by 3 or more hidden nodes.

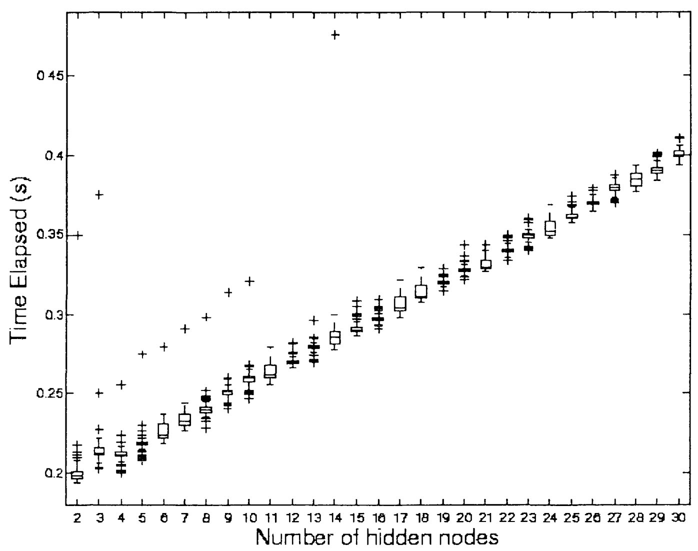

Figure 21 Time elapsed during training as the network architecture increases in size

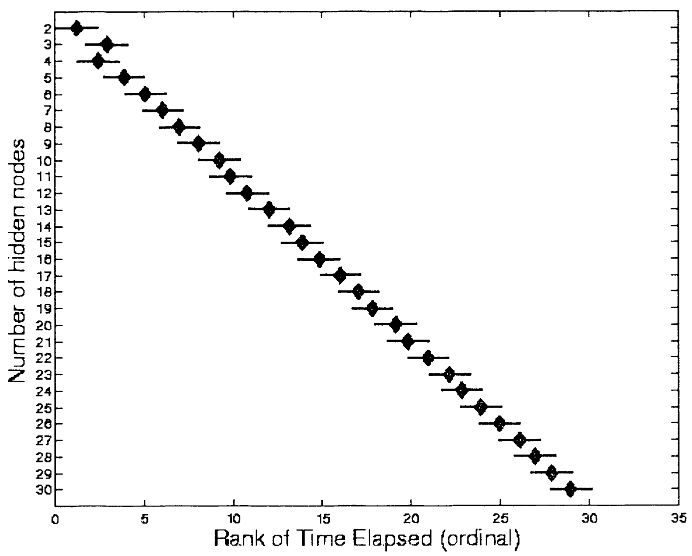

Figure 22 Rank of time elapsed during training as the network architecture increases in size

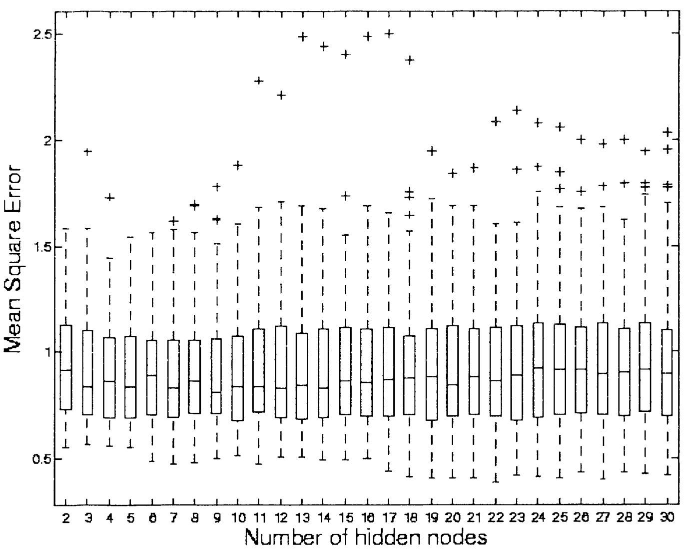

Figure 23 Training set mean squared error at termination of training as the network architecture increases in size

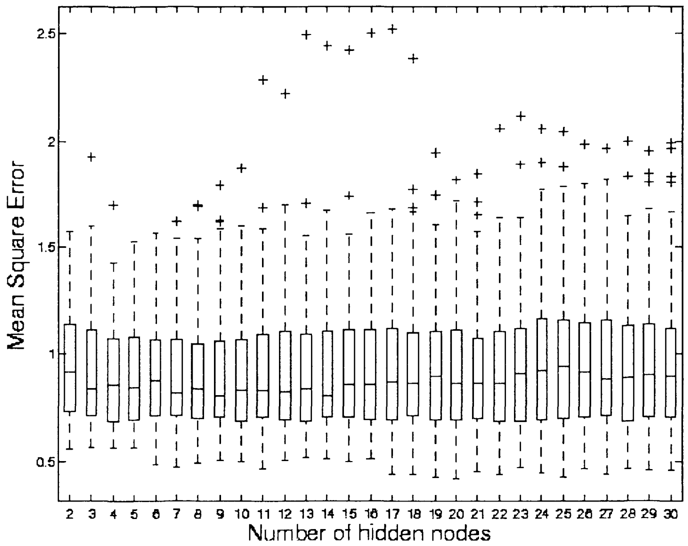

Figure 24 Validation set mean squared error at termination of training as the network architecture increases in size

4.4.1.2. Training algorithm

As is evident in Figure 25, the error rate varies considerably for classifiers trained using the Levenberg- Marquardt algorithm, but is more narrowly distributed for the delta rule and scaled conjugate gradient algorithms. Although the spread of error rates observed for each algorithm is considerably larger than the variation between the central tendencies, examining the difference in error rate of the matched classifiers in each block shows a statistically significant difference exists.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that training algorithm has no effect on the validation set real-time classifier error rate.

After ranking each classifier by validation set error rate within each block, it was found that classifiers trained using the scaled conjugate gradient algorithm provided the lowest ranked, and therefore most optimal, observed performances. Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between all training algorithms.

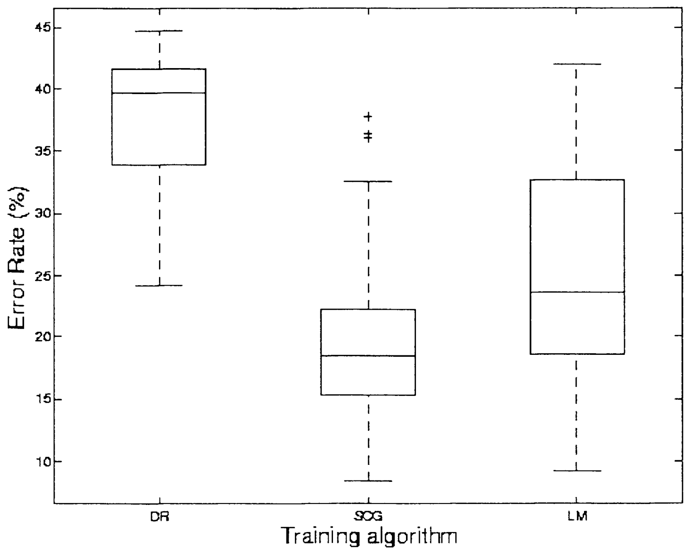

Figure 25 Real-time classification error rate on validation set for classifiers trained using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg-Marquardt ANN training algorithms

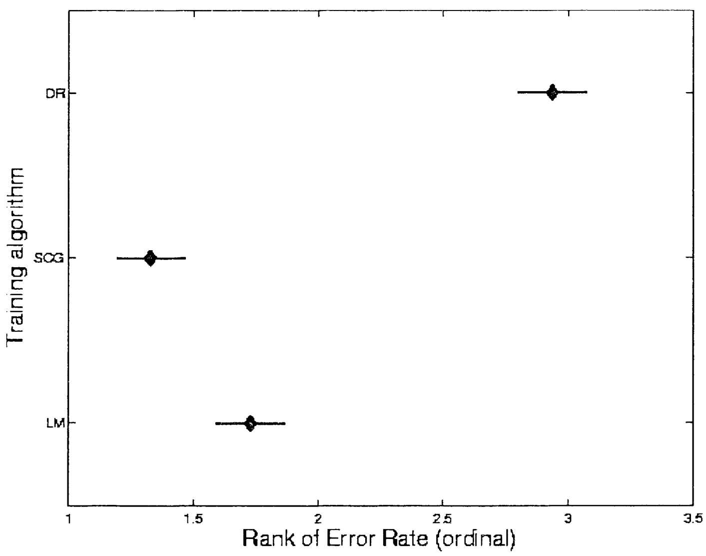

Figure 26 Rank of real-time classification error rate on validation set for classifiers trained using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg- Marquardt ANN training algorithms

As is evident in Figure 27, the time required to train the ANN classifiers varies considerably for the Levenberg-Marquardt algorithm and that the training time using that algorithm is an order of magnitude higher than the delta rule and scaled conjugate gradient algorithms.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the choice of ANN training algorithm has no effect on the length of time required to train an ANN classifier for this application.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between all training algorithms.

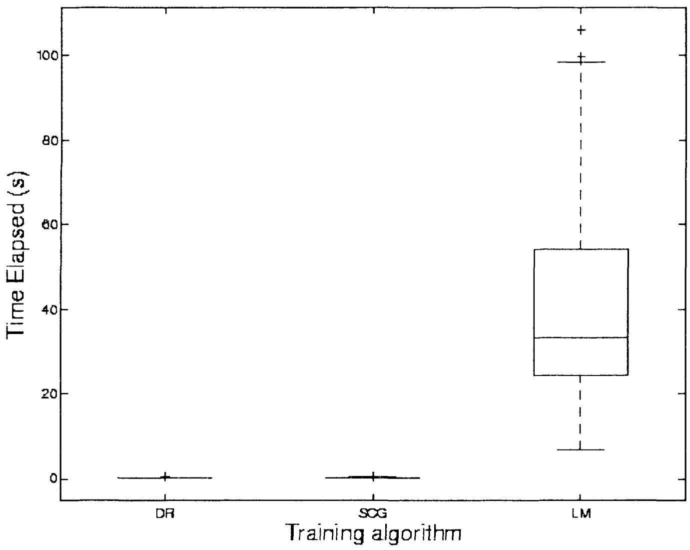

Figure 27 Time elapsed during training of each classifier using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg- Marquardt ANN training algorithms

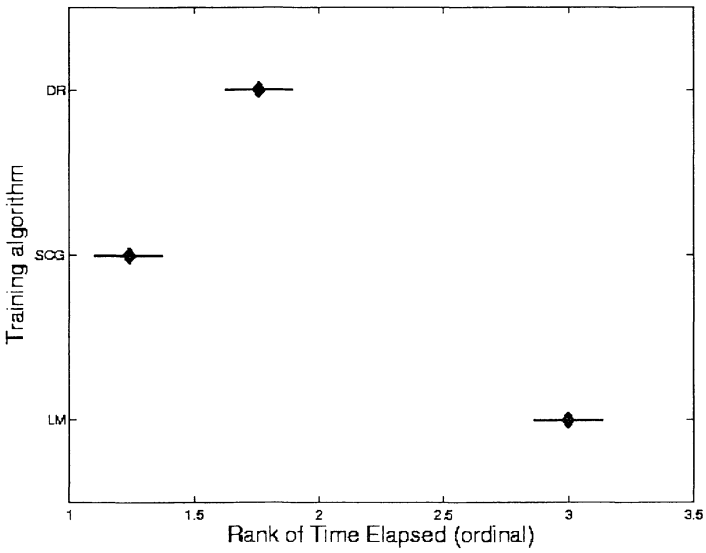

Figure 28 Rank of time elapsed during training of each classifier using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg- Marquardt ANN training algorithms

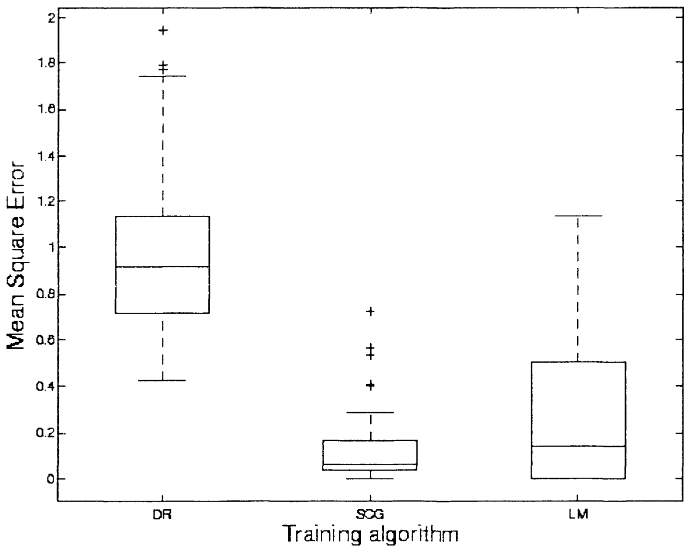

Figure 29 Training set mean squared error at termination of training for classifiers trained using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg- Marquardt ANN training algorithms

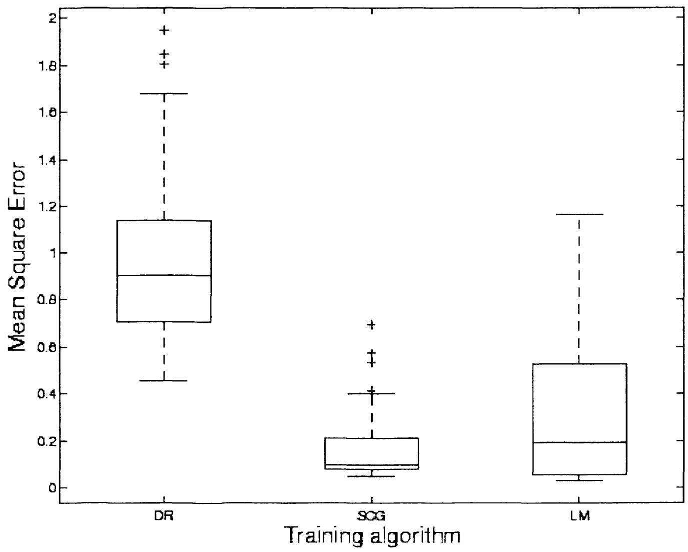

Figure 30 Validation set mean squared error at termination of training for classifiers trained using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg- Marquardt ANN training algorithms

4.4.1.3. Data set size

As is evident in Figure 31, the validation set error rate varies to a considerably greater extent within each treatment than between treatments.

Using the Friedman's ANOVA test, the differences between the mean ranks were not found to be significant at a 0.05 level of significance. These results do not support rejecting the null hypothesis that training set size has no effect on the validation set classifier error rate.

After ranking each classifier by validation set error rate within each block, it was found that classifiers trained with the original data set size of 30 exemplar input-output pairs in the training set provided the lowest ranked, and therefore most optimal, observed performances. Due to the lack of evidence to reject the null hypothesis, pairwise comparison tests following up on the Friedman's ANOVA test results were unnecessary.

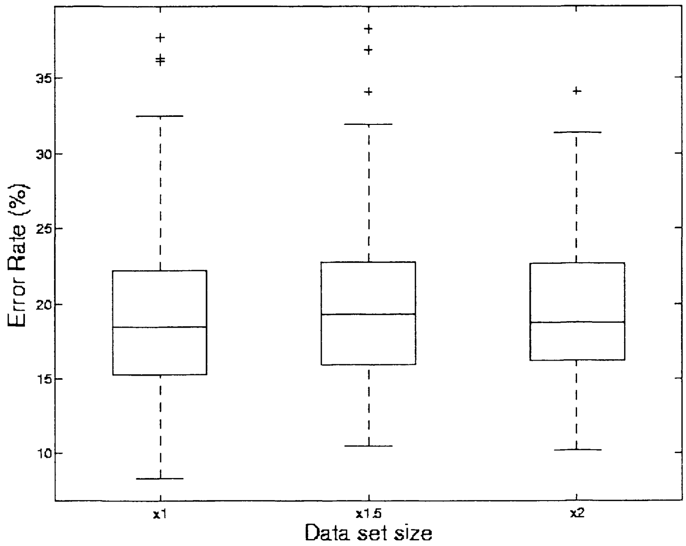

Figure 31 Real-time classification error rate on validation set as size of training set is increased

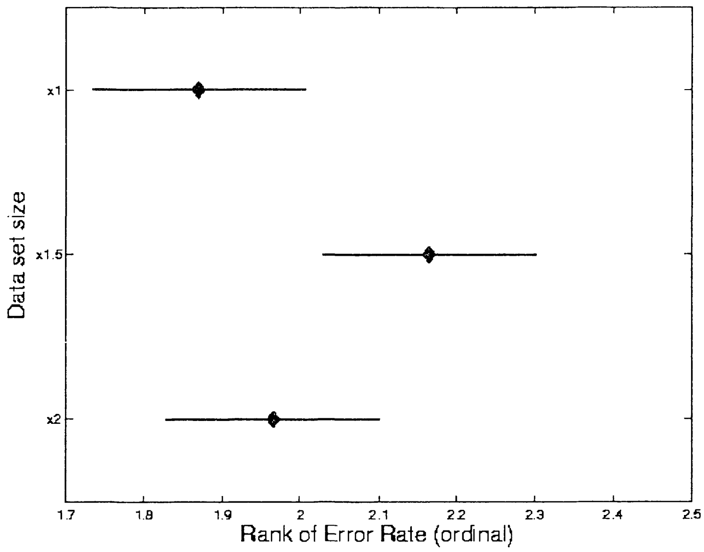

Figure 32 Rank of real-time classification error rate on validation set as size of training set is increased

As is evident in Figure 33, the ANN training time varies to a considerably greater extent within each treatment than between treatments.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that training set size has no effect on the length of time required to train an ANN classifier for this application.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between all training set sizes.

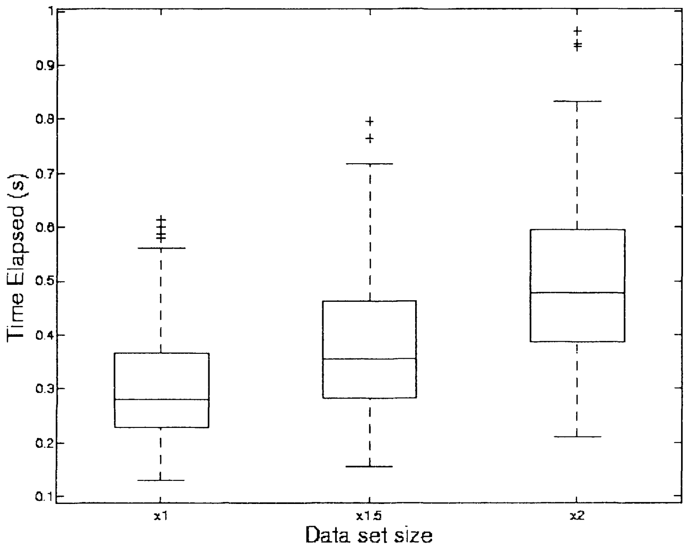

Figure 33 Time elapsed during training of each classifier as size of training set is increased

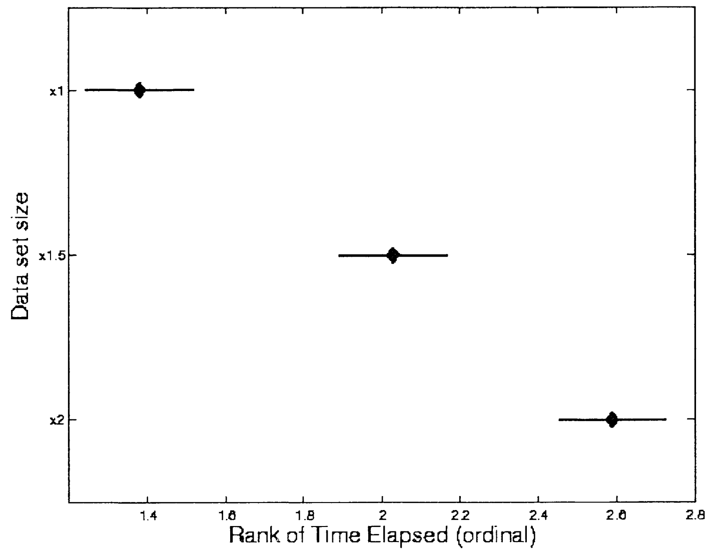

Figure 34 Rank of time elapsed during training of each classifier as size of training set is increased

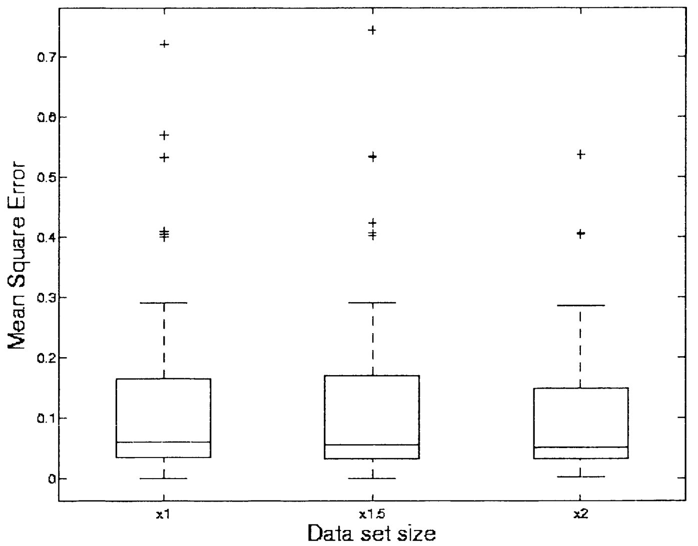

Figure 35 Training set mean squared error at termination of training as size of training set is increased

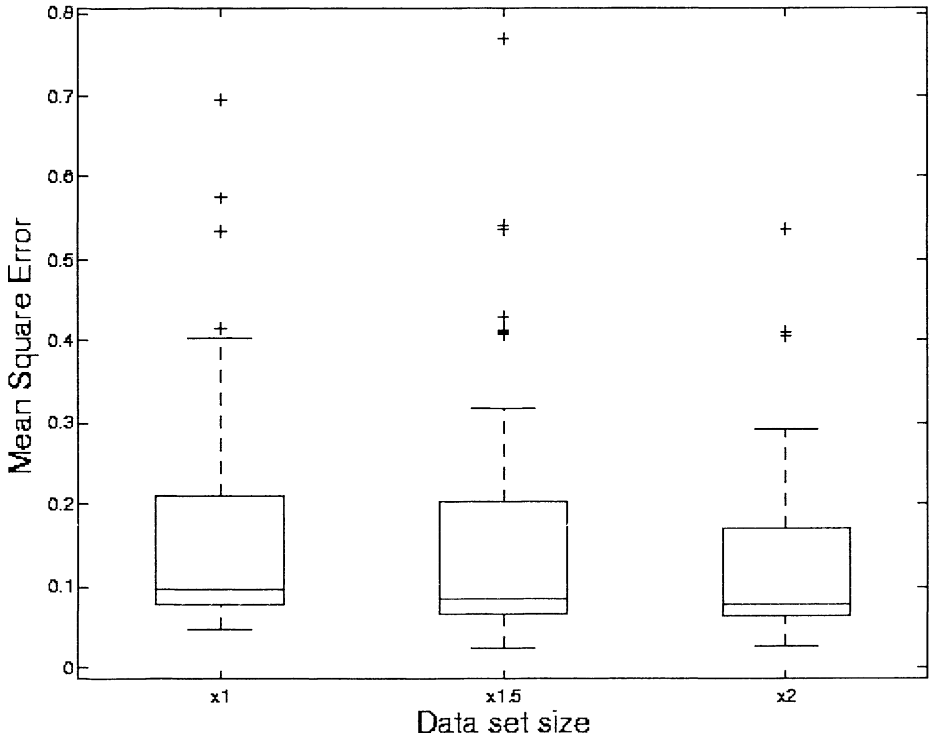

Figure 36 Validation set mean squared error at termination of training as size of training set is increased

4.4.2. Results on subject 8 test set

As is evident in Figure 37, the error rate varies considerably for each architecture. Although the spread of error rates observed for each architecture is considerably larger than the variation between the central tendencies, examining the difference in error rate of the matched classifiers in each block shows a statistically significant difference exists.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that network architectures has no effect on the real-time classifier error rate for Subject 8.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between the Subject

8 test set error rate for classifiers with 29 hidden nodes and architectures with 2, 19 and 20 hidden nodes. The architecture producing the lowest ranked error rates for Subject 8 was 12 hidden nodes, but no significant differences was found to exist between classifiers of this architecture and the architecture containing 29 hidden nodes that was found to result in the optimal the validation set error rate.

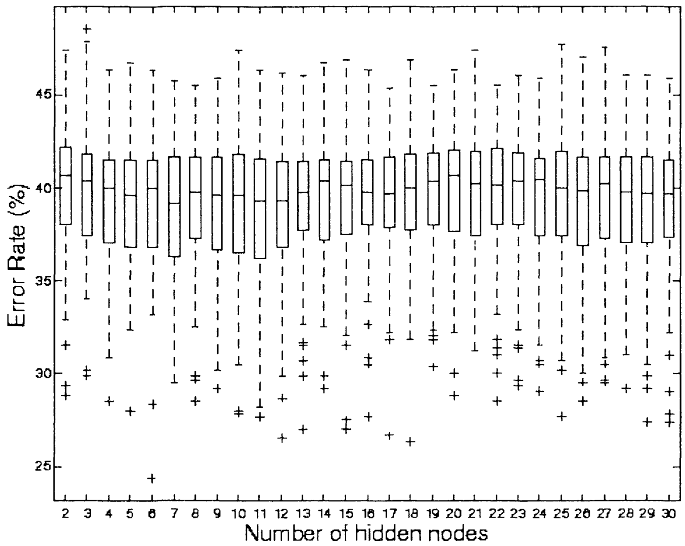

Figure 37 Real-time classification error rates for Subject 8 as the network architecture increases in size

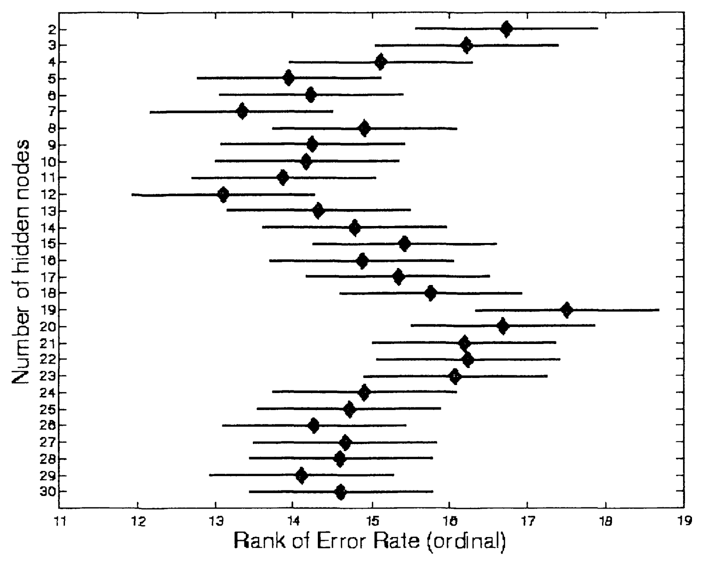

Figure 38 Rank of the real-time classification error rates for Subject 8 as the network architecture increases in size

As is evident in Figure 39, the error rate varies considerably for classifiers trained using the Levenberg- Marquardt algorithm, but is more narrowly distributed for the delta rule and scaled conjugate gradient algorithms, as was the case for the validation set real-time error rate. Although the spread of error rates observed for each algorithm is considerably larger than the variation between the central tendencies, examining the difference in error rate of the matched classifiers in each block shows a statistically significant difference exists.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the choice of ANN training algorithm has no effect on the real-time classifier error rate for Subject 8.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences in the Subject 8 test set error rate exist between all treatments at the 0.05 level of significance. The algorithm producing the lowest ranked error rates for Subject 8 was the scaled conjugate gradient algorithm, matching that found to be optimal on the validation set.

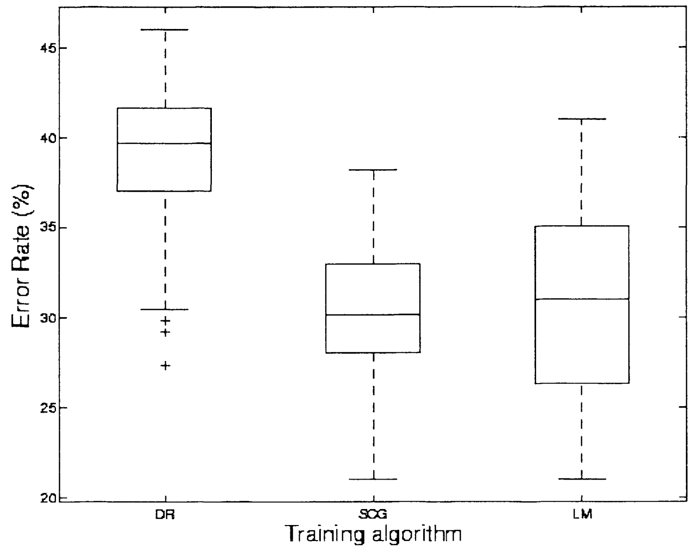

Figure 39 Real-time classification error rate on Subject 8 test set for classifiers trained using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg-Marquardt ANN training algorithms

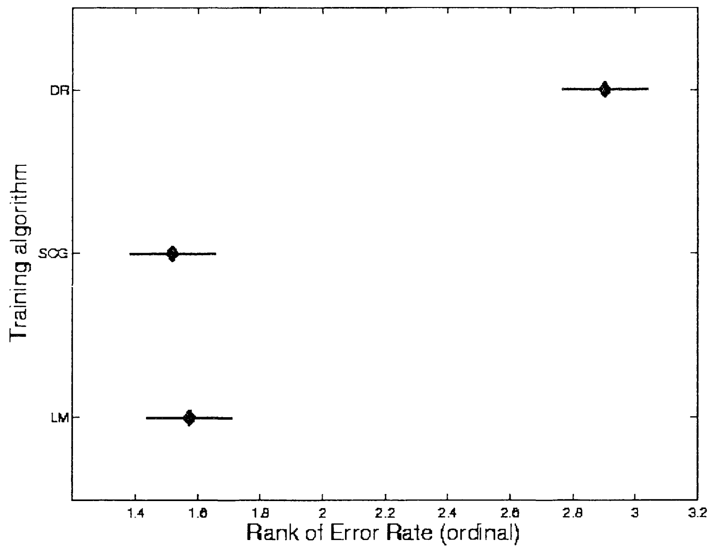

Figure 40 Rank of real-time classification error rate on Subject 8 test set for classifiers trained using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg- Marquardt ANN training algorithms

As is evident in Figure 41, the Subject 8 test set error rate varies to a considerably greater extent within each treatment than between treatments. Although the spread of error rates observed for each algorithm is considerably larger than the variation between the central tendencies, examining the difference in error rate of the matched classifiers in each block shows a statistically significant difference exists.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the size of the training set has no effect on the real-time classifier error rate for Subject 8.

After ranking each classifier by validation set error rate within each block, it was found that classifiers trained with the original data set size of 30 exemplar input-output pairs in the training set provided the highest ranked, and therefore least optimal, observed performances for this test set. Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences in the Subject 8 test set error rate exist between the classifiers trained using the original data set size of 30 exemplar input-output pairs and the largest data-set at the 0.05 level of significance. Although the largest training set provided the lowest ranked error rate on this test set, the difference between this treatment and the training set size of 45 exemplar input-output pairs in the training set was not found to be statistically significant.

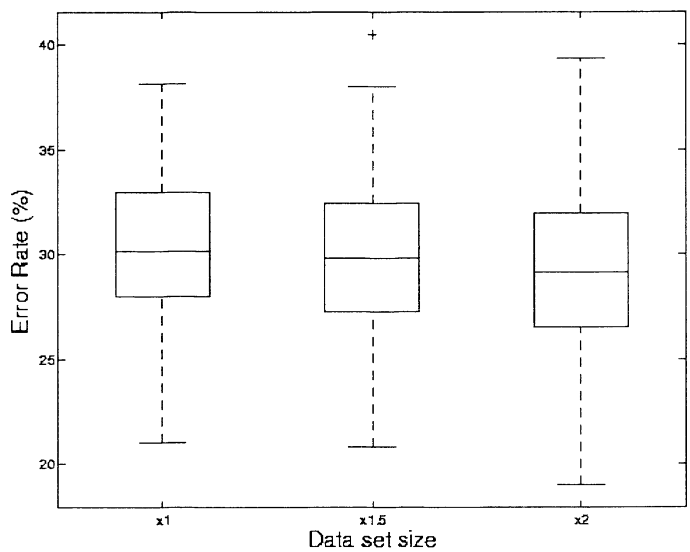

Figure 41 Real-time classification error rate on Subject 8 test set as size of training set is increased

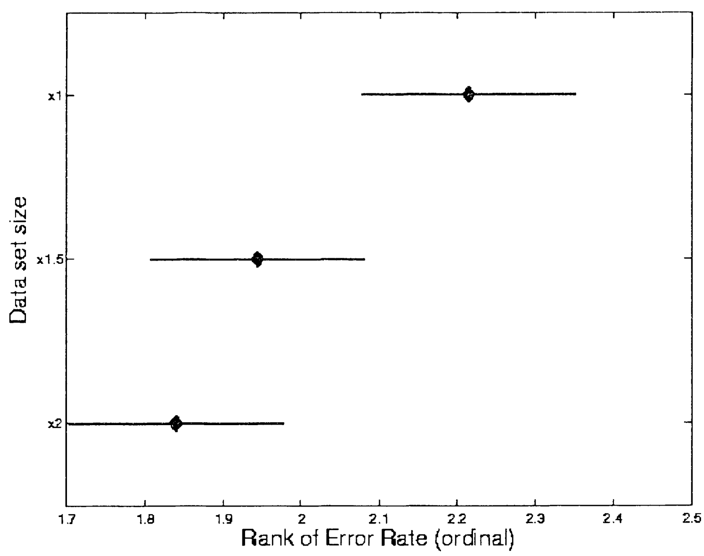

Figure 42 Rank of real-time classification error rate on Subject 8 test set as size of training set is increased

4.4.3. Results on subject 9 test set

As is evident in Figure 43, the error rate varies considerably for each architecture. Although the spread of error rates observed for each architecture is considerably larger than the variation between the central tendencies, examining the difference in error rate of the matched classifiers in each block shows a statistically significant difference exists.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that network architectures has no effect on the real-time classifier error rate for Subject 9.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between the Subject 9 test set error rate for classifiers with 29 hidden nodes and architectures with 2, 3 or 4 hidden nodes. The architecture producing the lowest ranked error rates for Subject 9 was 14 hidden nodes, but no significant differences was found to exist between classifiers this architecture and the architecture containing 29 hidden nodes that was found to result in the optimal the validation set error rate.

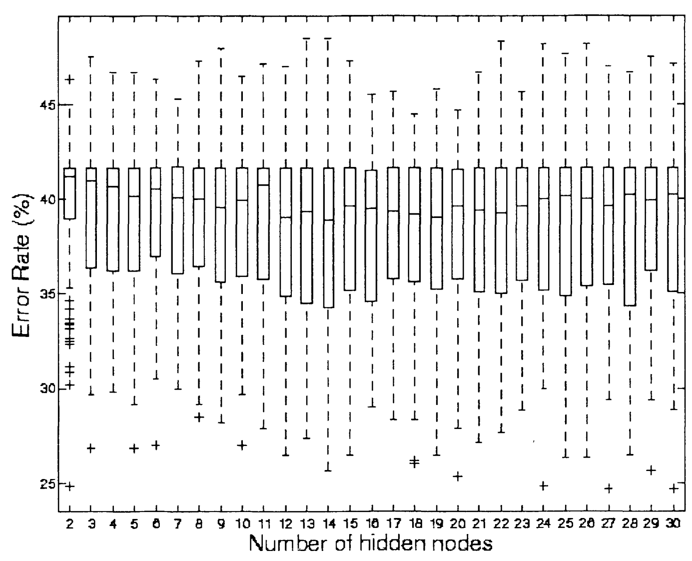

Figure 43 Real-time classification error rates for Subject 9 as the network architecture increases in size

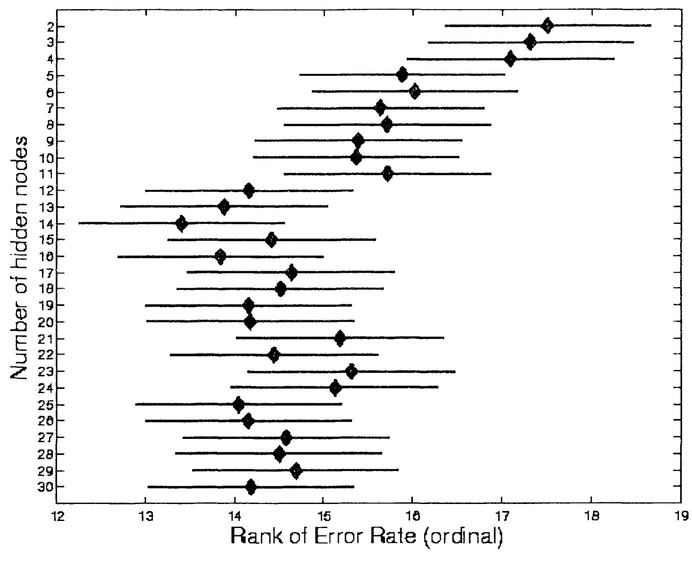

Figure 44 Rank of real-time classification error rates for Subject 9 as the network architecture increases in size

As is evident in Figure 45, the error rate varies considerably for classifiers trained using the Levenberg- Marquardt algorithm, but is more narrowly distributed for the delta rule and scaled conjugate gradient algorithms, as was the case for the validation set real-time error rate. Although the spread of error rates observed for each algorithm is considerably larger than the variation between the central tendencies, examining the difference in error rate of the matched classifiers in each block shows a statistically significant difference exists.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the choice of ANN training algorithm has no effect on the real-time classifier error rate for Subject 9.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences in the Subject 9 test set error rate exist between all treatments at the 0.05 level of significance. The algorithm producing the lowest ranked error rates for Subject 9 was the scaled conjugate gradient algorithm, matching that found to be optimal on the validation set.

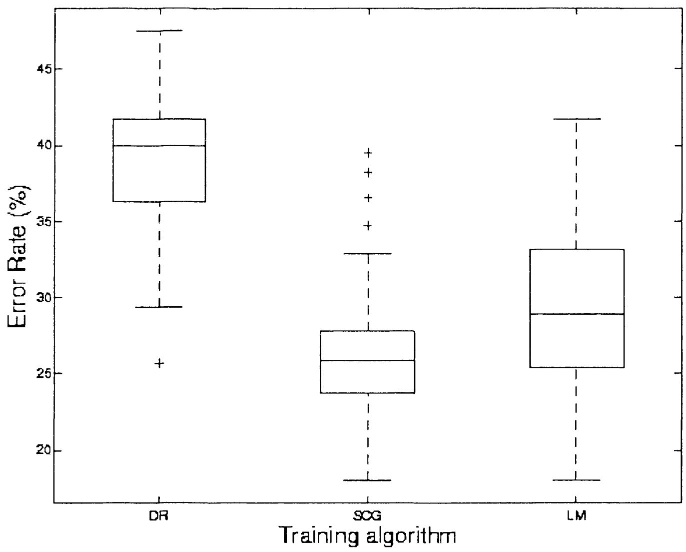

Figure 45 Real-time classification error rate on Subject 9 test set for classifiers trained using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg-Marquardt ANN training algorithms

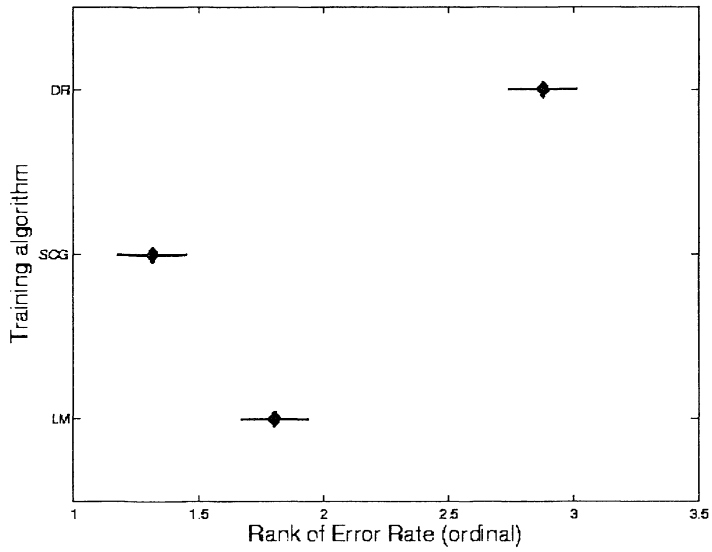

Figure 46 Rank of real-time classification error rate on Subject 9 test set for classifiers trained using delta rule (DR) , scaled conjugate gradient (SCG) and Levenberg- Marquardt ANN training algorithms

As is evident in Figure 47, the Subject 9 test set error rate varies to a considerably greater extent within each treatment than between treatments. Although the spread of error rates observed for each algorithm is considerably larger than the variation between the central tendencies, examining the difference in error rate of the matched classifiers in each block shows a statistically significant difference exists.

Using the Friedman's ANOVA test, the differences between the mean ranks were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the size of the training set has no effect on the real-time classifier error rate for Subject 9.

After ranking each classifier by validation set error rate within each block, it was found that classifiers trained with the original data set size of 30 exemplar input-output pairs in the training set provided the lowest ranked observed performances for this test set, matching the optimal treatment found on the validation set. Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences in the Subject 9 test set error rate exist between the classifiers trained using the original training set size of 30 exemplar input-output pairs and training set size of 45 exemplar input-output pairs at the 0.05 level of significance.

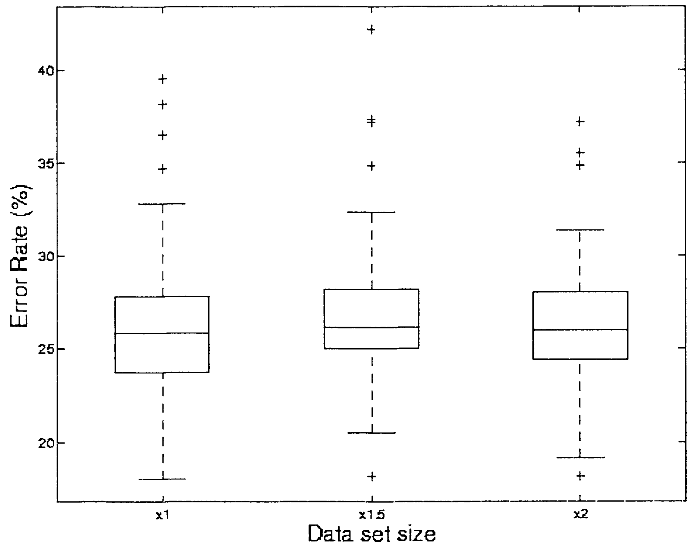

Figure 47 Real-time classification error rate on Subject 9 test set as size of training set is increased

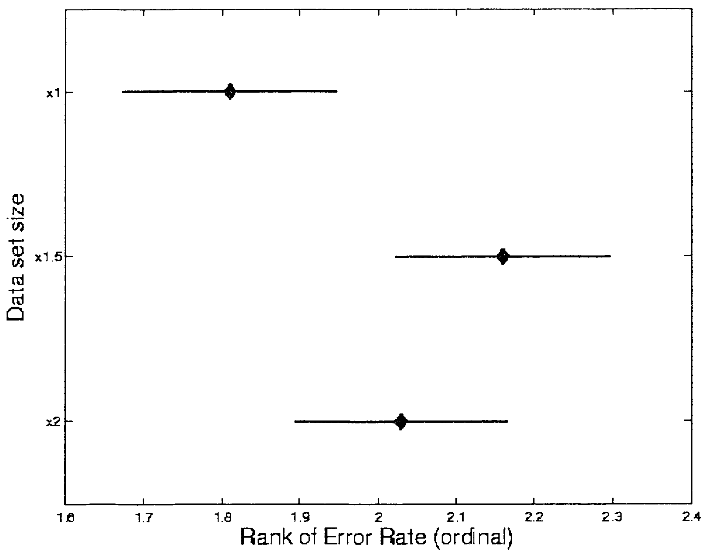

Figure 48 Rank of real-time classification error rate on Subject 9 test set as size of training set is increased

4.5. Discussion

The results for ANN architecture show that although selection of architecture makes a significant difference to the error rate of the resulting classifiers, the marginal effect of each additional hidden node is small. As a trend, increasing the number of hidden nodes was associated with an improvement in the validation set error rate. The difference in performance is statistically significant when comparing the smallest architectures against larger ones, but the improvement in error rate decreases with each additional hidden node, so the significance of the difference between classifiers having more than 9 and less than 30 hidden nodes was small. No architecture was found where the performance a classifier was statistically significantly different to that of a classifier having one more or one fewer hidden node. The architecture containing 29 hidden nodes found to be optimal in terms of real-time validation set error rate produces performances on the validation set and both user specific test sets that show little practical or statistical difference to almost all other architectures. The architecture of 29 hidden nodes would not have been selected as optimal on the user specific tests sets, but the performance was not significantly different to the architecture that would.

Apart from small architectures, the large variation in results for each treatment suggests that the impact of architecture on classifier performance is related to its effect on the error surface and the ability of the training algorithm to find appropriate weights, rather than the ability of the ANN to implement the necessary input-output mapping. The relatively high values for training set and validation set MSE also suggest that the training algorithm often fails to find weights that yield a good approximation of the input-output mappings in the training set. The consistent performance as the number of hidden nodes increases suggests that overfitting of the training data through the use of an undetermined ANN is not a major influence.

Training time is clearly affected by network architecture. This result is a consequence of the additional computations required for each iteration of the training algorithm. The linear trend correlating increasing number of hidden nodes with longer training times is consistent with that which would be expected for the delta rule algorithm that was used to train the classifiers.

The scaled conjugate gradient algorithm is clearly associated with superior results on validation set real-time classification performance. Results for the experiment comparing between algorithms show that ANN trained using the SCG produced performances significantly superior to the delta rule algorithm for all three measures of real-time classification performance. The Levenberg-Marquardt algorithm was also found to produce superior results when compared to the delta-rule algorithm, but was inferior to SCG on two measures. Levenberg-Marquardt was also found to take an order of magnitude longer to train each ANN. The significantly lower training and validation set MSE observed for the SCG and Levenberg-Marquardt algorithms suggest that these algorithms are more successful than the delta rule algorithm at finding weights that yield a good approximation of the input-output mappings in the training set.

Much of the ANN literature does not consider the choice of a training algorithm as a major influence on the performance of ANN in classifier applications, focussing on the effect of architecture and training data. These results show that the choice of algorithm has a statistically and practically significant effect on real-time head gesture classifier performance. In this minimax comparison, the results show that the SCG algorithm is the most optimal of the algorithms considered for this application. The implication of these results is that there may be further improvements in classifier performance possible through the analysis of the properties of the head gesture classification task and determining how the SCG algorithm produces these superior performances in this application.

Although the training set size of 30 exemplar patterns was found to produce the superior validation set real-time classification performances, it is not clear that this is of practical significance. The small effect size with respect to the variation within each treatment suggests that the marginal benefit of each additional sample from a general population of subjects is low. It would be expected from the theoretical arguments discussed in Section 4.1 that, all other things being equal, larger training sets would produce equal or superior results. Although several explanations for these results are possible, the small effect size and lack of a consistent trend in the three measures of performance suggest that the benefit of such an approach to the further optimisation of the head gesture classifier would be limited.

Chapter 5. Adaptive training of head gesture classifiers

The results presented in Chapter 3 and Chapter 4 have shown that the real-time classification performance of the proposed wheelchair control system's classifier component is inferior for disabled users. This may occur for either or both of the reasons that the input-output mapping of the training set did not sufficiently well approximate the underlying ideal input-output mapping for the head gestures performed by the disabled users, or the function implemented by the ANN did not sufficiently well approximate the input-output mapping in the training set.

The minimum error rates and mean squared error results presented in Chapter 4 show that an ANN is able to implement a function approximating the input-output mapping of a training set containing gestures performed by both able-bodied and disabled people to achieve good generalisation for those people. When using the treatments observed to produce the optimal real-time classification set performance for the group of people who provided the training data, disabled people whose gestures had not been included in the training set were observed to have an error rate approximately 5% to 10% higher.

It was also shown in Chapter 4 that increasing the quantity of training data from the same set of people who provided the original data did not produce significantly improved results for separate disabled people. This indicates that the use of this type of data is not an effective way in which to make the training set more representative of the ideal input-output mappings. These results, however, do not indicate whether data obtained from the specific end user can be used to improve the performance of the classifier.

It remains an open question whether the use of data from a specific user can be associated with a change in the classifier performance. Since a wheelchair is controlled by a single user, one option to improve the performance of the classifier is to adapt the classifier to that particular user. The degree to which classifier performance can be altered by the use of user specific training data is important to the development of the wheelchair control system.

The previous discourse on the way in which the distribution of training data affects the generalisation of ANN classifiers (see Sections 3.1.5 and 4.1.1) is also relevant to this Chapter. Using generic data for the training of the head gesture classifier has practical benefits for the implementation of a wheelchair control system, such as the relative ease of the collection of a large data set and the ability to optimise the classifier for that dataset in a centralised process. However, if a classifier trained on a generic dataset is only able to adequately generalise to a small, homogeneous population of potential users then the usefulness of the system will be marginalised by its cost. As the population of potential users becomes larger and more diverse, there is greater potential for overlap between the sets of input points that would ideally be classified as belonging to each class. Consequently, it is relevant to consider the way in which gestures performed by highly disabled people differ from the gestures performed by the group of able-bodied and disabled people whose data provided the generic training data used in Chapter 4. A review of data obtained from highly disabled people is presented in Section 5.1.

The cost of collecting, labelling and training each sample of user specific data for the implementation of a wheelchair control system is considerably higher than for generic data, as it must be carried out for each new user of the control system. It is therefore important to use the labelled data as efficiently as possible. Since there is clearly some utility in the use of generic data in training ANN head gesture classifiers, despite the lower performance for highly disabled people, it is therefore desirable to obtain the benefit of such data while also improving the performance for the end user. To achieve this, this chapter examines the effect of retraining an ANN trained on generic data. The ANN is adapted for the specific user by including exemplar input-output mappings observed from gestures performed by that person in the training set during this retraining.

Two adaptive retraining procedures are considered: one using a training set comprising a mixture of generic and user specific data, the other containing only user specific data. The details of this retraining are presented in Section 5.2. The two procedures are used to provide insight as to whether there is a sustained benefit in including generic data during the adaptive retraining. It is an open question as to whether inclusion of generic data in the adaptive retraining has any effect on the performance of the resulting classifier, since the overlap between classes over different people is unknown.

In order for user-specific data to improve the real-time classification performance in this way, training on user-specific data must alter the function implemented by the ANN in some way, thereby altering the resulting input-output mapping. The overlap between the set of ANN input points mapping to one output for a particular person and the set of input points mapping to a different output for another user is unknown, but possibly non-empty. Consequently, it is unknown whether the use of user-specific data in the ANN training will also affect the performance of the classifier for other users, or for users whose condition varies over time.

In evaluating the effect of using user specific data in ANN training, it is therefore necessary to consider the change in performance for the specific person for whom the ANN has been trained, for the population who provided the generic dataset and for other people whom no data has been included in the training set.

The methodology and procedure of an experiment testing the effect of the proposed adaptive algorithm are detailed in Sections 5.2 and 5.3, followed by the results and conclusions drawn from the experiment in Sections 5.4 and 5.5.

5.1. Head gestures by people with disabilities

Examples of gestures performed by disabled subjects are shown in Figure 53. Gestures performed by people with disabilities have clear similarities to those performed by people without disabilities shown in Section 4.1.2, but there are also notable differences. The magnitude of maximum deviation from the neutral position is often smaller in gestures performed by people with disabilities, even after normalisation for range of head movement. This is evident particularly in gestures involving movement in the mediolateral plane.

The magnitude of crosstalk between channels and non-gesture deviations in head position is often greater in data recorded from people with disabilities. This is apparent in Figure 50, Figure 51, Figure 56 and Figure 57, and can be contrasted to that seen in Figure 9 and Figure 12. The increase in the magnitude of non-gesture deviations is partly due to the effect of normalisation, as the people with disabilities from whom data was collected exhibited smaller ranges of movement, but may also be associated with a lesser degree of fine control of movement.

Gestures performed by disabled people also appear to exhibit more pronounced deviations from the gesture prescribed action. An example of such a deviation can be seen at the beginning of the gestures in Figure 50 and Figure 57, where the movement in the primary axis of the gesture first deviates from the neutral position away from the direction of the gesture, before this is reversed a short time later. Although similar artefacts can be observed in gestures performed by people without disabilities, as can be seen in Figure 10 and Figure 11, the magnitude and prevalence of these artefacts is less common.

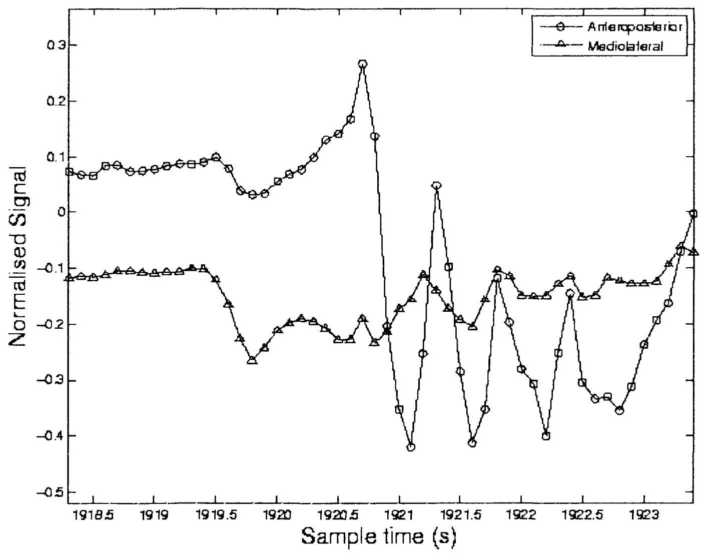

Figure 49 Example backward nod gesture, performed by Subject 8 (C4)

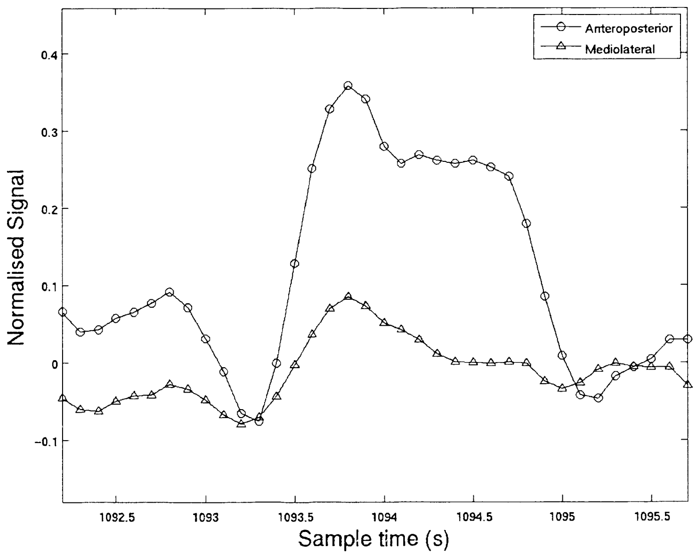

Figure 50 Example forward nod gesture, performed by Subject 8

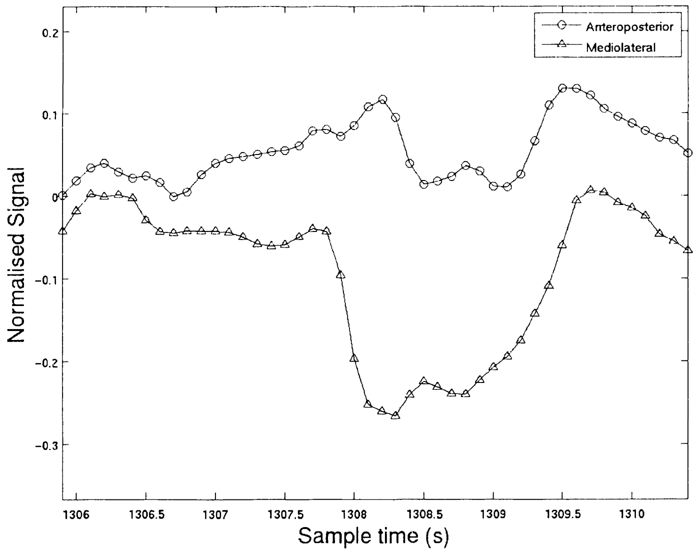

Figure 51 Example left nod gesture, performed by Subject 8

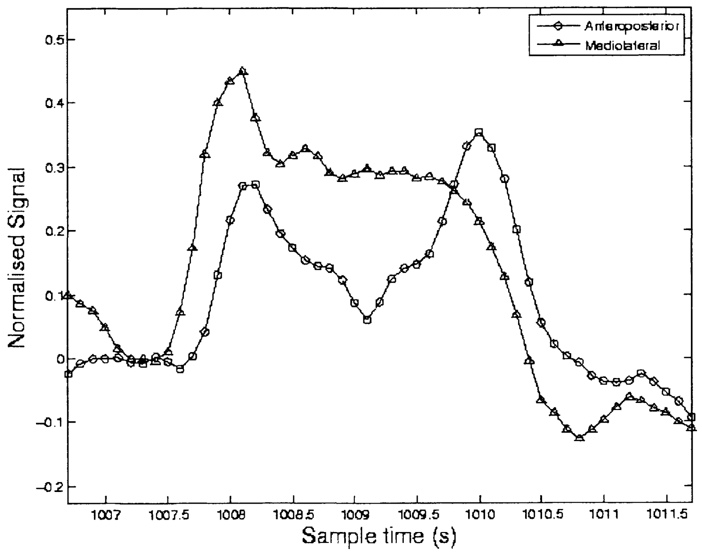

Figure 52 Example right nod gesture, performed by Subject 8

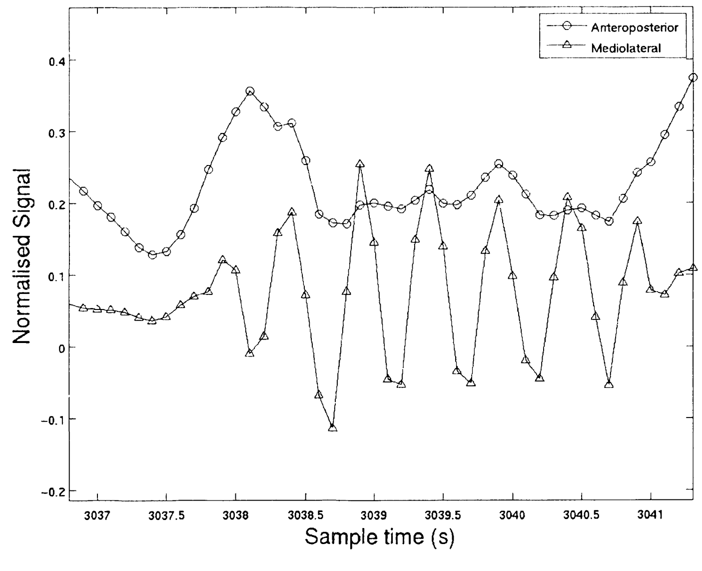

Figure 53 Example shake gesture, performed by Subject 8

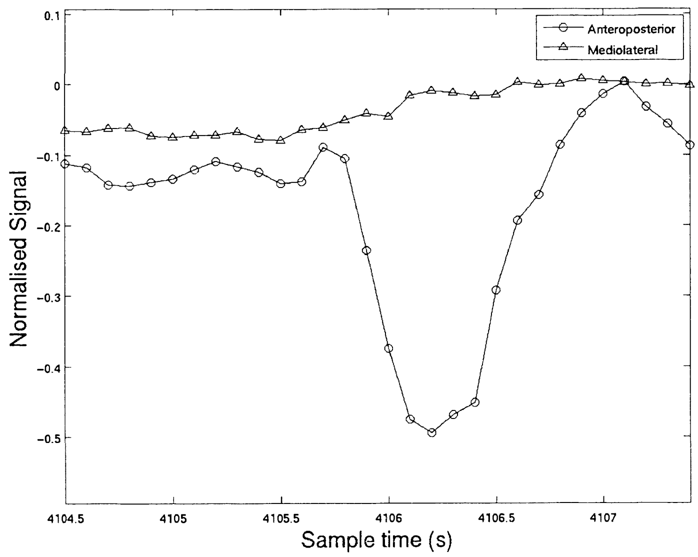

Figure 54 Example backward nod gesture, performed by Subject 9 (C6)

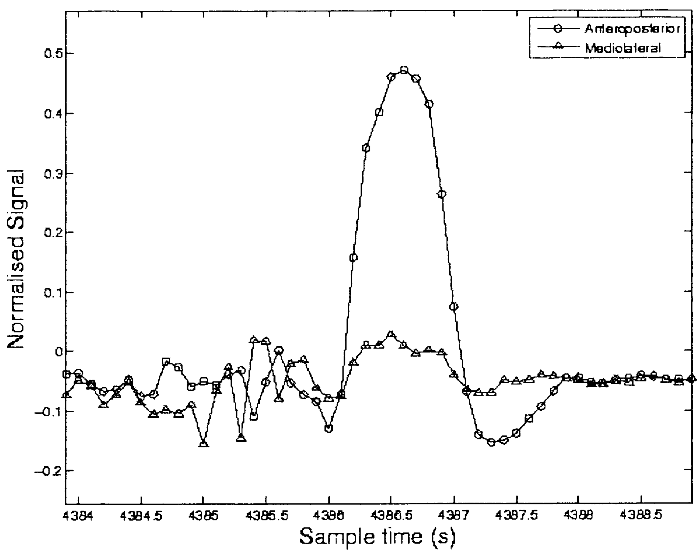

Figure 55 Example forward nod gesture, performed by Subject 9

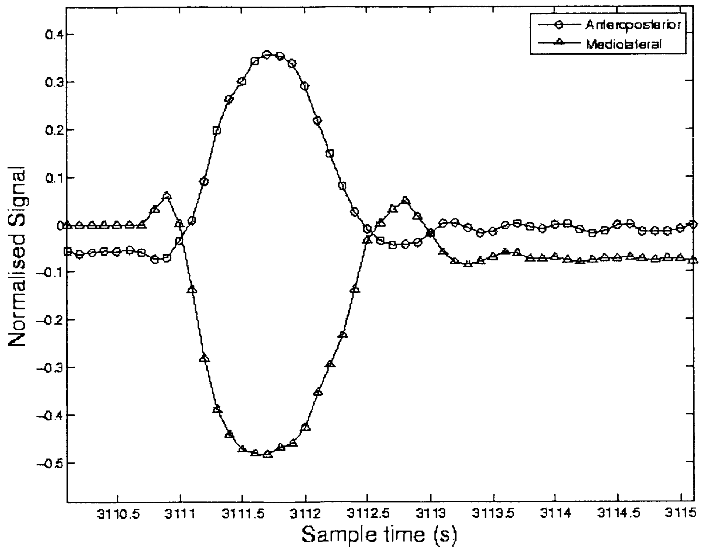

Figure 56 Example left nod gesture, performed by Subject 9

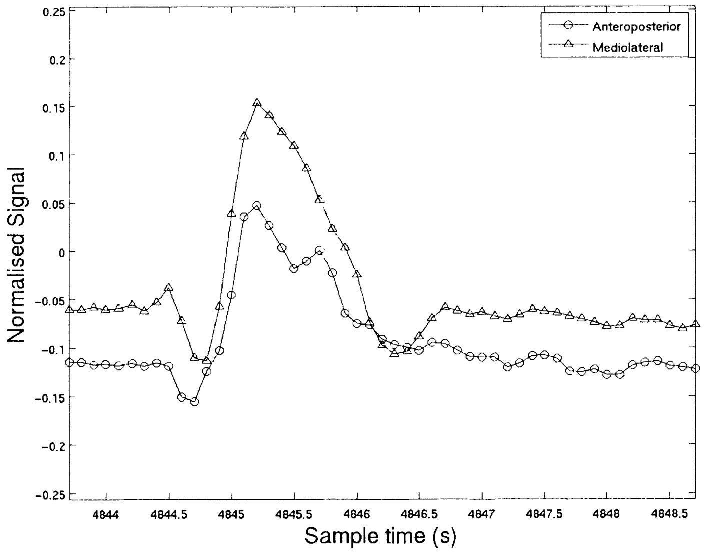

Figure 57 Example right nod gesture, performed by Subject 9

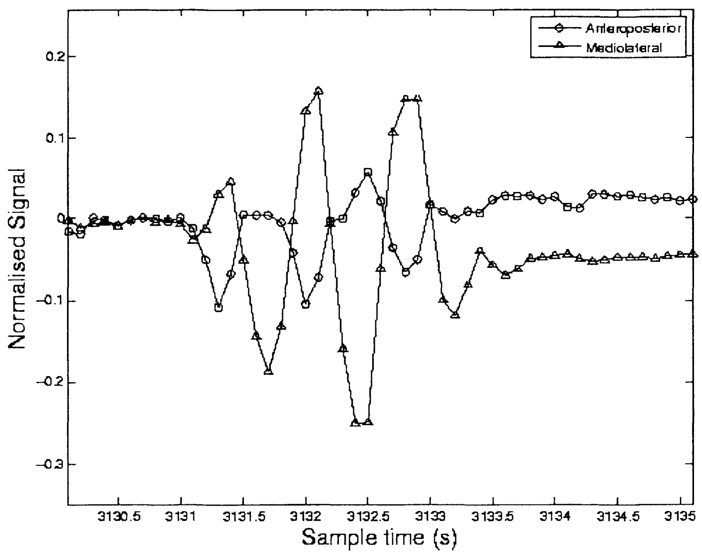

Figure 58 Example shake gesture, performed by Subject 9

5.2. Methodology for assessing the utility of the adaptive algorithm

The aim of the experiment described in this section is to determine whether adaptive retraining of an ANN classifier with user specific data can be expected to improve the performance of the classifier. Two control treatments are used to provide benchmarks to which the two adaptive retraining procedures can be compared. These control treatments are selected to offer insight into the reasons for any observed differences in performance.

One of the control treatments, summarised in Section 5.2.1, can be considered a negative control. This treatment allows the classifiers trained using the adaptive algorithm to be compared to classifiers trained without the use of user specific data and without an adaptive retraining phase. This control treatment is equivalent to the generic optimal training procedure found in Chapter 4 without using any user specific data or retraining. This algorithm is referred to in the procedure, results and discussion as Control treatment.

The two adaptive retraining procedures, summarised in Sections 5.2.2 and 5.2.3, differ only in the composition of the training set during the retraining phase. The adaptive retraining procedure summarised in Section 5.2.2 combines generic data and user specific data in the training and validation sets during the retraining phase, and is referred to as the Combined Adaptive retraining treatment. The other adaptive retraining procedure, summarised in Section 5.2.3, uses only user specific data in the training and validation sets during the retraining phase, and is referred to as the Specific Adaptive retraining procedure.

The other control treatment, summarised in Section 5.2.4, can be considered a positive control. The algorithm described in Section 5.2.4 trains ANN solely with user specific data, without generic data and without a retraining phase. This algorithm is referred to in the procedure, results and discussion as Custom training treatment, as the resulting classifier can be considered to be specifically trained for that user.

The methodology and procedure used in this chapter are intentionally similar to those used in Chapter 4. The aim of this similarity is to simplify the making of comparisons between the results. As was the case in previous chapters, classifier performance is tested using recorded gestures from two subjects with high-level spinal cord injuries. Data from each of these subjects was treated separately, comprising two independent test sets. Classifier performance is measured on these test sets according to the sliding window error rate observed on recorded gestures that had not been used in the training of that classifier.

A randomised complete block experiment design is used, with blocks formed by training classifiers using the adaptive algorithm and three control treatments from the same initial conditions. Each of the classifiers in a block shares the same small, random initial weights and the same allocation of data to the training, validation, retraining and test sets. Since the generic training phase of the adaptive algorithm is identical to the Control treatment, the final ANN weights from negative control in a block is identical to the intermediate ANN weights observed in the corresponding adapted ANN. This does not apply to the Custom treatment, as there is no generic training in this treatment.

5.2.1. Algorithm for the Control training procedure.

Given:

- A set, D_g, of exemplar input-output pairs observed in data from recorded gestures provided by a general population of users (i.e. generic data);
- Positive integers N_{gt}, N_{gv},

1. Define $T_g \subseteq D_g$ such that T_g contains N_{gt} input-output pairs for each class.
2. Define $V_g \subseteq D_g$ such that V_g contains N_{gv} input-output pairs for each class and $T_g \cap V_g = \varnothing$.
3. Set the initial weights of the ANN, W_0, to small, random values
4. Define W_g to be the final weights resulting from training the ANN from W_0, using training set T_g and validation set V_g.
5. Return W_g

5.2.2. Algorithm for the Combined Adaptive retraining procedure.

Given:

- A set, D_g, of exemplar input-output pairs observed in data from recorded gestures provided by a general population of users (i.e. generic data);
- A set, D_u, of exemplar input-output pairs observed in data from recorded gestures provided by the specific end user (i.e. user specific data);
- Positive integers N_{gt}, N_{gv}, N_{ut}, N_{uv}

1. Define $T_g \subseteq D_g$ such that T_g contains N_{gt} input-output pairs for each class.
2. Define $V_g \subseteq D_g$ such that V_g contains N_{gv} input-output pairs for each class and $T_g \cap V_g = \varnothing$.

3. Set the initial weights of the ANN, W_0, to small, random values

4. Define W_g to be the final weights resulting from training the ANN from W_0, using training set T_g and validation set V_g.

5. Define $T_u \subseteq D_u$ such that T_u contains N_{ut} input-output pairs for each class.

6. Define $V_u \subseteq D_u$ such that V_u contains N_{uv} input-output pairs for each class and $T_u \cap V_u = \varnothing$.

7. Define $T_c = T_g \cup T_u$

8. Define $V_c = V_g \cup V_u$

9. Define W_{ac} to be the final weights resulting from retraining the ANN from W_g, using training set T_c and validation set V_c.

10. Return W_{ac}

5.2.3. Algorithm for the Specific Adaptive retraining procedure.

Given:

- A set, D_g, of exemplar input-output pairs observed in data from recorded gestures provided by a general population of users (i.e. generic data);
- A set, D_u, of exemplar input-output pairs observed in data from recorded gestures provided by the specific end user (i.e. user specific data);
- Positive integers N_{gt}, N_{gv}, N_{ut}, N_{uv}

1. Define $T_g \subseteq D_g$ such that T_g contains N_{gt} input-output pairs for each class.
2. Define $V_g \subseteq D_g$ such that V_g contains N_{gv} input-output pairs for each class and $T_g \cap V_g = \varnothing$.
3. Set the initial weights of the ANN, W_0, to small, random values
4. Define W_g to be the final weights resulting from training the ANN from W_0, using training set T_g and validation set V_g.
5. Define $T_u \subseteq D_u$ such that T_u contains N_{ut} input-output pairs for each class.
6. Define $V_u \subseteq D_u$ such that V_u contains N_{uv} input-output pairs for each class and $T_u \cap V_u = \varnothing$.
7. Define W_{as} to be the final weights resulting from retraining the ANN from W_g, using training set T_u and validation set V_u.
8. Return W_{as}.

5.2.4. Algorithm for Custom training procedure.

Given:

- A set, D_u, of exemplar input-output pairs observed in data from recorded gestures provided by the specific end user (i.e. user specific data);
- Positive integers N_{ut}, N_{uv}

1. Define $T_u \subseteq D_u$ such that T_u contains N_{ut} input-output pairs for each class.
2. Define $V_u \subseteq D_u$ such that V_u contains N_{uv} input-output pairs for each class and $T_u \cap V_u = \emptyset$.
3. Set the initial weights of the ANN, W_0, to small, random values
4. Define W_c to be the final weights resulting from training the ANN from W_0, using training set T_u and validation set V_u.
5. Return W_c

5.3. Procedure for assessment of the adaptive training algorithm

In this experiment, the performance of classifiers trained using the two adaptive retraining procedures is compared to classifiers trained using the control treatments. For each of the treatments, classifiers are trained from small random initial weights and real-time classification error rate is used to measure the performance of each classifier. For each treatment, 100 ANN were trained. The performance of each resulting classifier was determined by measuring the real-time classification error rate on a test set containing recorded gestures performed by Subject 8, a test set containing recorded gestures performed by Subject 9 and a test set containing recorded gestures corresponding to those generic input-output pairs in the validation set used to train that ANN. These real-time classification error rates observed on these three sets are referred to as the Subject 8, Subject 9 and validation set real-time classification error rates, respectively.

Each of the 100 ANN trained for each treatment was matched to corresponding ANN trained using the other treatments. Each member of a block was trained from the same initial weights and used the same allocation of data to each of the datasets. Matched classifiers differed only in the procedure used to train and retrain the ANN. That is, T_g, V_g, T_u, V_u and W_0 were equal for each ANN in a particular block. N_{gt}, N_{gv}, N_{ut}, N_{uv} were constant parameters for all ANN trained.

Initial weights were generated using a uniformly distributed variable in the range –0.5 to 0.5, excluding 0. Recorded gestures and input-output pairs were allocated from the gestures recorded by Subjects 1-7 for the generic data sets. The Subject 8 data sets and Subject 9 data sets each contained only gestures and input-output pairs from those subjects. The collection of recorded gestures from which these sets were derived contained the same recordings as those used in Chapter 4, and used the same labelling of the recorded gestures and exemplar patterns. The number of recorded gestures of each class that were allocated to each set is listed in Table 15.

Table 15 Size of datasets for control and adaptive retraining procedures.

Dataset	Parameter	Number of input-output pairs per class
Generic training set	N_{gt}	30
Generic validation set	N_{gv}	30
User specific training set	N_{ut}	15
User specific validation set	N_{uv}	10

The training and retraining of each ANN was performed using generic optimal settings found in Chapter 4. That is, network architecture containing 29 hidden nodes, trained using the scaled conjugate gradient algorithm and a generic training set size of 30 input-output pairs per class. Classifier performance was measured by determining the sliding window classification error rate on each test set, using the same procedure employed in Chapter 4.

For the purposes of statistical hypothesis testing, this chapter investigates the null hypothesis that there is no difference between the error rates of ANN classifiers that have been trained using any of the four procedures.

5.4. Results

The first results presented are those for the evaluation of the effect of user specific adaptation for Subject 8, followed by the evaluation of the effect of user specific adaptation for Subject 9. Each set of results contains a box plot of the real-time classification error rates observed for validation, Subject 8 test and Subject 9 test sets. By design, the techniques used for the analysis of experimental results are similar to those used in Chapter 4. The reader is referred to Section 4.4 for descriptions of the statistical tests applied to these results and the presentation of the results.

Ranking results for the user specific adaptation are summarised in Table 16. Further details on the results obtained are reviewed in 5.4.1 and 5.4.2.

Table 16 Rank of adapted classifier error rates on user specific test sets

Subject	Ranked error rate using Adapted - Combined algorithm (mean ± 2 s.d.)	Ranked error rate using Adapted - Specific algorithm (mean ± 2 s.d.)
8	2.3 ± 0.7 of 4 treatments	1.6 ± 0.7 of 4 treatments
9	2.4 ± 0.7 of 4 treatments	2.4 ± 0.7 of 4 treatments

5.4.1. Adaptive retraining using data from Subject 8 (C4)

As is evident in Figure 59, the error rate measured on the gestures in the validation set varies considerably for each procedure. Using the Friedman's ANOVA test, the differences between the mean ranks of the error rate of each procedure were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the training or retraining algorithm used has no effect on the Subject 8 set real-time classifier error rate.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between the Control procedure and the Specific Adaptive and Custom procedures. Pairwise comparison tests also show that significant differences exist at the 0.05 level of significance between the Custom procedure and both the Combined and Specific Adaptive procedures, and between the Combined Adaptive and Specific Adaptive procedures.

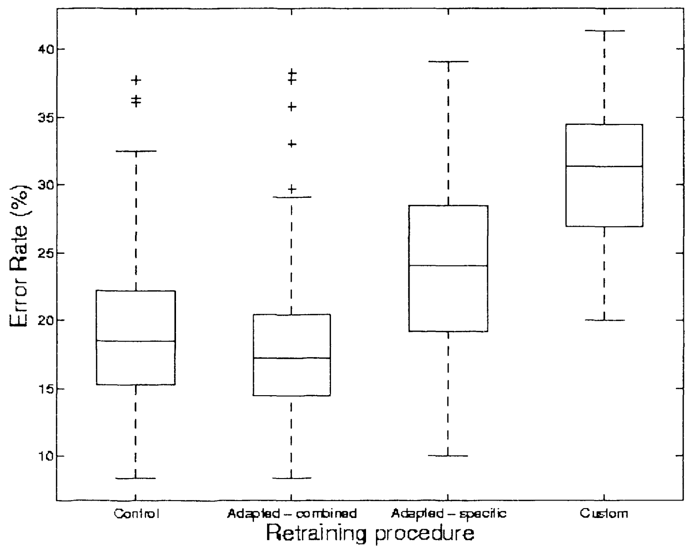

Figure 59 Real-time classification error rate on validation set for classifiers adapted using data from Subject 8 and control treatments

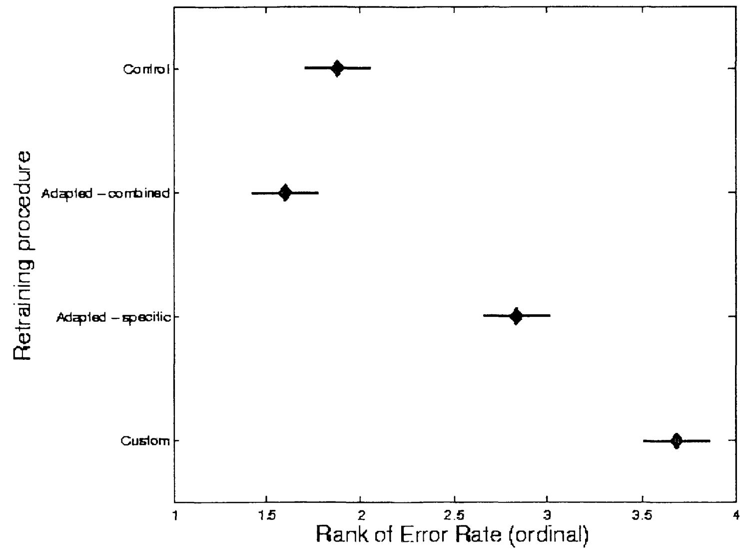

Figure 60 Rank of real-time classification error rate on validation set for classifiers adapted using data from Subject 8 and control treatments

As is evident in Figure 61, the error rate measured on gestures in the user specific test set varies considerably for each procedure. The spread of error rates for each treatment observed on the Subject 8 test set is similar to that for the error rate measured on the validation set, while the central tendencies are dissimilar. Using the Friedman's ANOVA test, the differences between the mean ranks of the error rate of each procedure were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the training or retraining algorithm used has no effect on the user-specific test set real-time classifier error rate.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between the Control procedure and both the Specific Adaptive and Custom procedures. Pairwise comparison tests also show that significant differences exist at the 0.05 level of significance between the Combined Adaptive procedure and both the Specific Adaptive and Custom procedures, and also between the Specific Adaptive and the Custom procedures. No significant difference was found between the Combined Adaptive and Control procedures.

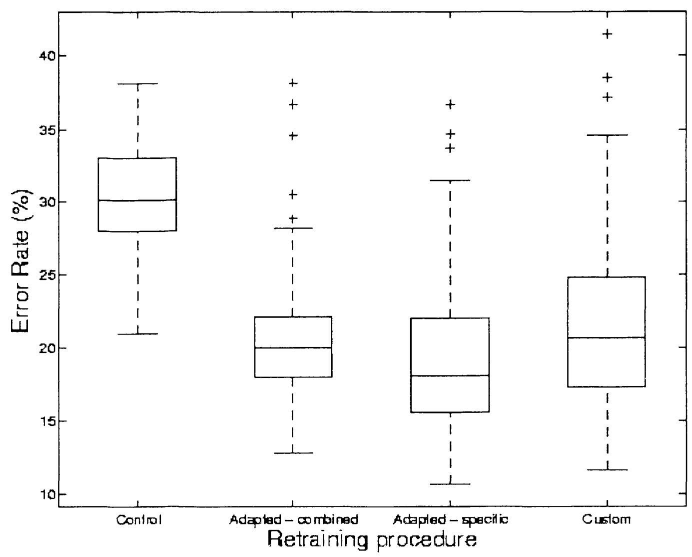

Figure 61 Real-time classification error rate on Subject 8 test set for classifiers adapted using data from Subject 8 and control treatments

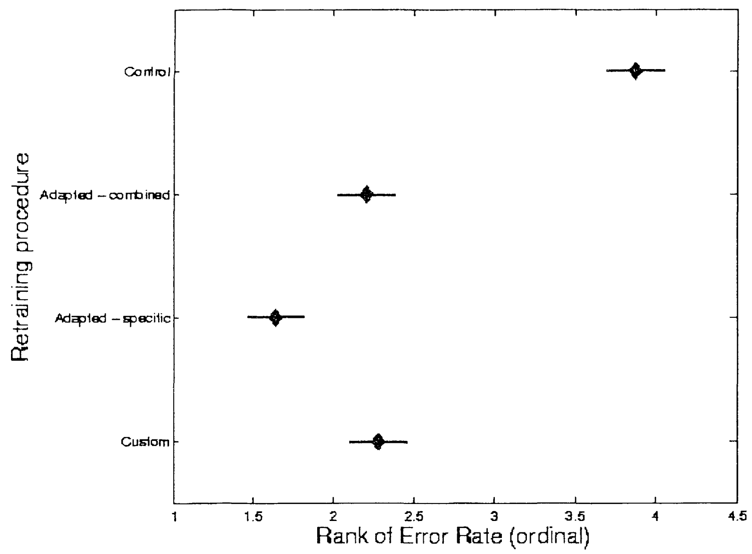

Figure 62 Rank of real-time classification error rate on Subject 8 test set for classifiers adapted using data from Subject 8 and control treatments

As is evident in Figure 64, the error rate measured on gestures in the Subject 9 test set varies considerably for each procedure. The spread of results for each treatment is greater than that observed for the validation and user-specific test sets. Using the

Friedman's ANOVA test, the differences between the mean ranks of the error rate of each procedure were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the training or retraining algorithm used has no effect on the validation set real-time classifier error rate.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between the Control procedure and the Specific Adaptive and Custom procedures. Pairwise comparison tests also show that significant differences exist at the 0.05 level of significance between the Custom procedure and both the Combined and Specific Adaptive procedures, and between the Combined Adaptive and Specific Adaptive procedures

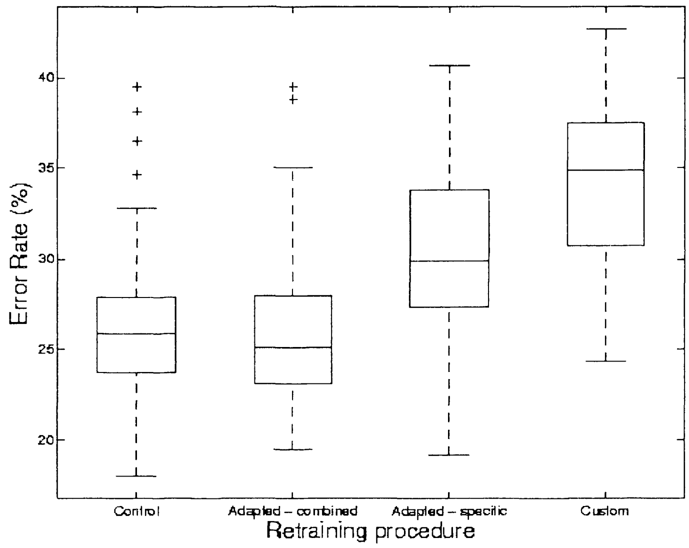

Figure 63 Real-time classification error rate on Subject 9 test set for classifiers adapted using data from Subject 8 and control treatments

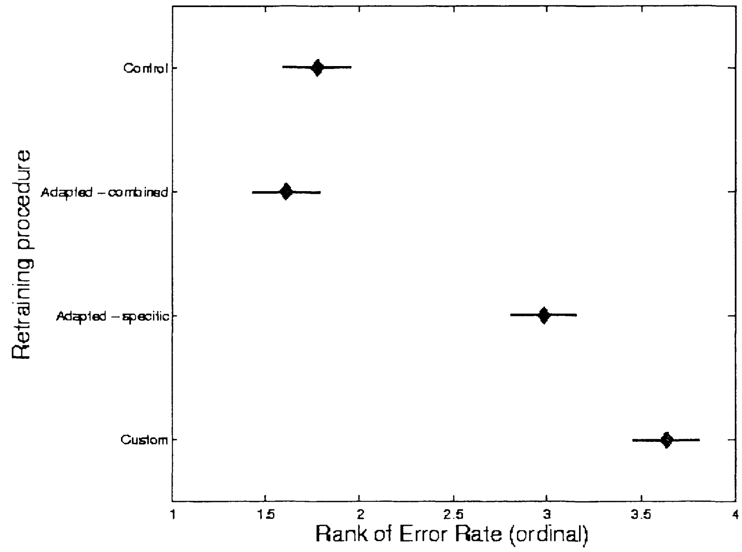

Figure 64 Rank of real-time classification error rate on Subject 9 test set for classifiers adapted using data from Subject 8 and control treatments

5.4.2. Adaptive retraining using data from Subject 9 (C6)

As is evident in Figure 65, the error rate measured on the gestures in the validation set varies considerably for each procedure. The spread of error rates observed for Subject 9 are similar to the spread of error rates observed on the validation set of classifiers adapted for Subject 8. Using the Friedman's ANOVA test, the differences between the mean ranks of the error rate of each procedure were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the training or retraining algorithm used has no effect on the Subject 9 set real-time classifier error rate.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between all procedures.

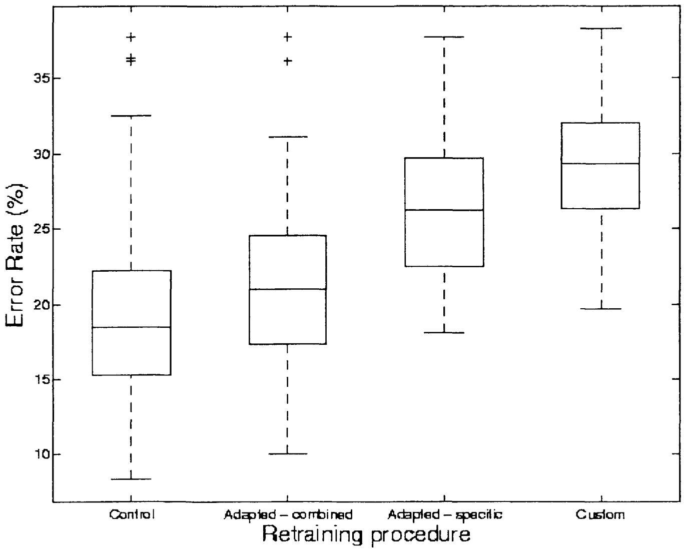

Figure 65 Real-time classification error rate on validation set for classifiers adapted using data from Subject 9 and control treatments

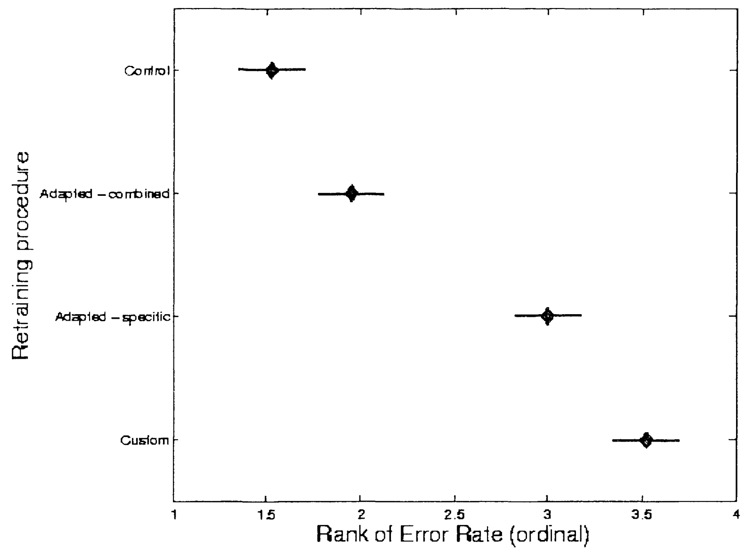

Figure 66 Rank of real-time classification error rate on validation set for classifiers adapted using data from Subject 9 and control treatments

As is evident in Figure 67, the error rate measured on gestures in the Subject 8 test set varies considerably for each procedure. The spread of error rate for each treatment observed on this test set is larger than for the spread of error rate observed on the validation set, although there is less difference in the central tendencies. Using the Friedman's ANOVA test, the differences between the mean ranks of the error rate of each procedure were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the training or retraining algorithm used has no effect on the Subject 8 test set real-time classifier error rate.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between the Custom procedure and all other procedures. No other statistically significant pairwise differences were found.

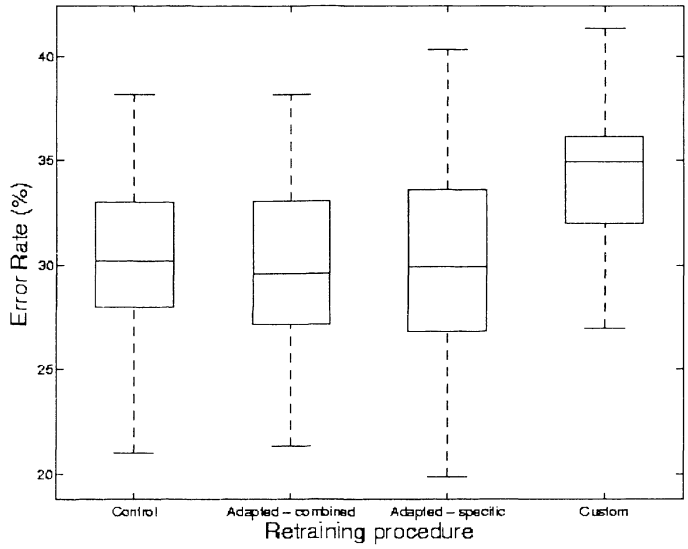

Figure 67 Real-time classification error rate on Subject 8 test set for classifiers adapted using data from Subject 9 and control treatments

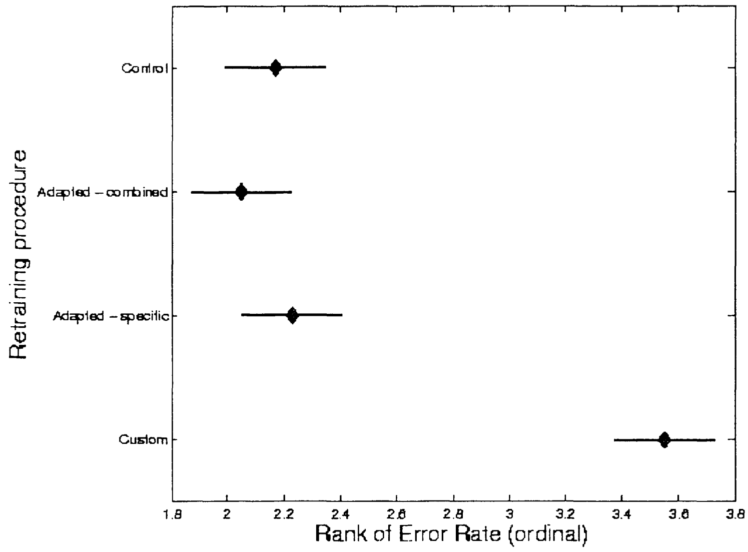

Figure 68 Rank of real-time classification error rate on Subject 8 test set for classifiers adapted using data from Subject 9 and control treatments

As is evident in Figure 69, the error rate measured on the user-specific test set varies considerably for each procedure. Using the Friedman's ANOVA test, the differences between the mean ranks of the error rate of each procedure were found to be significant at a 0.05 level of significance. These results support rejecting the null hypothesis that the training or retraining algorithm used has no effect on the validation set real-time classifier error rate.

Pairwise comparison tests following up on the Friedman's ANOVA test results show that significant differences at the 0.05 level of significance exist between the Custom procedure and all other procedures. No other statistically significant pairwise differences were found.

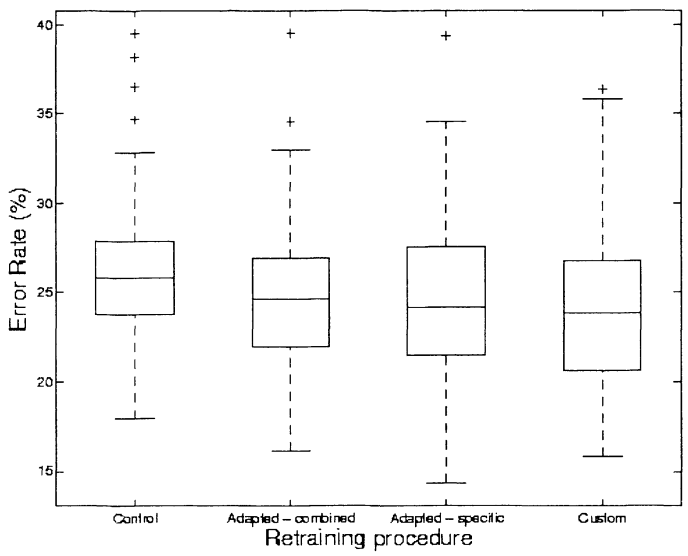

Figure 69 Real-time classification error rate on Subject 9 test set for classifiers adapted using data from Subject 9 and control treatments

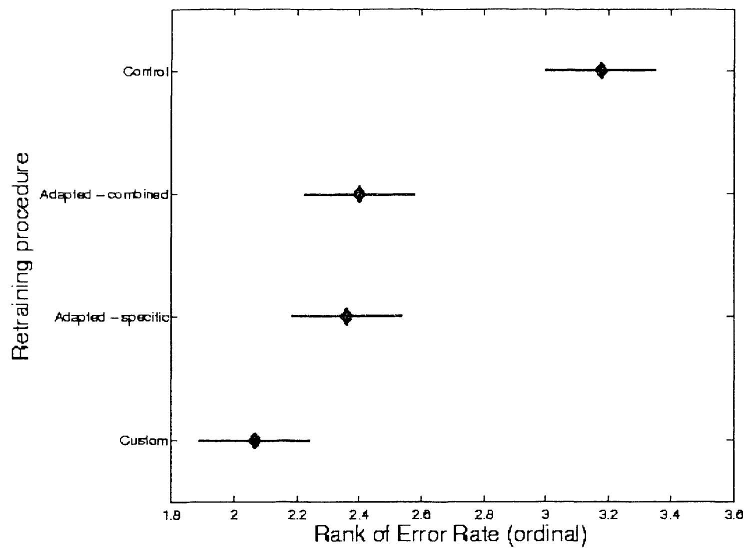

Figure 70 Rank of real-time classification error rate on Subject 9 test set for classifiers adapted using data from Subject 9 and control treatments

5.5. Discussion

These results show that the use of user specific data to train or retrain an ANN classifier for the head gesture classifier can improve the performance of the classifier for that user. For both subjects for whom classifiers were adapted, the use of data from the end user produced significantly lower error rates when compared to the negative control, which was trained on generic data only. Of the adaptive retraining procedures, the Specific Adaptive procedure was found to provide lower real-time classification error rates for the person for whom the ANN was adapted than the Combined adaptive procedure, but this difference was only significant for Subject 8. The Combined Adaptive procedure was found to provide similar results for both specific end users to the Custom procedure.

The real-time classification error rate results reported above show that the use of user specific training data does not improve the classifier performance for other people, and can be detrimental. This is particularly evident in the results for the Specific Adaptive procedure and Custom procedure. In each of the results where classifiers were retrained using data from a different person, validation set real-time classification rate was significantly higher than the control treatment. The Combined Adaptive algorithm does not show this effect, and instead provides results similar to the control treatment when tested on a new user. The Combined Adaptive retraining procedure appears to produce the benefit of improved performance for the specific end user, without the detrimental effect of reduced performance for gestures drawn from other sources of data.

This result suggests that the set of input-output pairs in the user-specific data partially overlaps with the set of input-output pairs in the original training set. The improvement in performance produced by the Combined Adaptive algorithm demonstrates that the training set was expanded to include points from a region of input space for which the ANN generalised poorly from the original training set. The degradation in performance for other users when using the Specific Adaptive algorithm and Custom algorithm shows that the points sampled from the specific, highly disabled user do not enclose the whole region of input space represented by the original training set. Significantly, the increased performance on the user specific test set of the Specific Adaptive and Custom algorithms for Subject 9 imply that there may have been a degree of overlap between regions of input space belonging to different classes in the user-specific and original training sets. Although this may also be due to other factors, it raises an important open question for the further development of the control system, as to what extent variation in the gesture characteristics of highly disabled people such that the same region of input space should be mapped to different command classes limits the maximum performance of classifiers trained using generic data.

The larger spread of results observed for the Specific Adaptive algorithm can be attributed to the smaller number of input-output pairs in the training set during retraining, and similarly for training using the Custom algorithm. This observation does not, by itself, indicate that these methods should not be used to train ANN classifiers for a specific user. Rather, since the cost of creating a training set containing user-specific data is largely determined by the cost of identifying and labelling suitable patterns, the larger spread of results observed highlights the need to determine what quantity of user-specific data can be expected to be optimal to train such classifiers.

Although the number of subjects is small and care must be taken in attempting to draw generally applicable conclusions from this sample, the results show that user specific data can be used to improve classifier performance for at least a subset of the population of potential users.

Chapter 6. Conclusions

The central aim of this research has been to develop a wheelchair control system for highly disabled people, providing a means for satisfying some of the mobility needs of people who have difficulty or are unable to achieve mobility through existing assistive technologies. The control system proposed focuses on the population of people with high level spinal cord injury as potential users.

The main objectives of this thesis have been:

- to propose the design of a wheelchair control system and demonstrate its feasibility for use by disabled people;
- to show that the proposed control system's performance for disabled users can be improved by using a mixture of data from people with and without disabilities.
- To show that the performance of the proposed control system can be further improved for disabled users by adapting the control system using data from the particular end user.

The specific aims of the experiments presented in Chapter 3, Chapter 4 and Chapter 5 have been:

- To determine whether the proposed wheelchair control system can feasibly recognise commands given by disabled people in real-time, in a manner that provides sufficient performance for dependable operation;
- To investigate techniques for improving the performance of the proposed control system using generic data, by optimising the ability of the system to recognise command gestures performed by highly disabled people;

- To propose and test a procedure for improving the performance of the proposed control system for a user by adapting the system using data from that particular user.

6.1. Findings

A wheelchair control system was proposed in Chapter 3, advancing from designs previously described in the literature. The ability of this control system to recognise command gestures performed by members of the population of potential users was tested to demonstrate its feasibility. While demonstrating the feasibility of the proposed control system, these results showed that the performance of the system was lower for people with disabilities than for able-bodied users. This difference is suspected to arise from different characteristics in the movement patterns of a person with a high level disability, due to restriction in range of movement, poorer motor skills, medication or other interfering factors.

The experimental results in Chapter 4 show optimisation of the head gesture classifier using generic data can significantly improve the ability of the system to recognise commands performed by highly disabled people. This conclusion supports the feasibility of the control system architecture presented in Chapter 3. It was found that selection of the ANN architecture, training algorithm and training set size all had a significant effect of some degree on the ability of the classifier component to recognise command gestures by people with disabilities in real time. It was found that selection of ANN architecture and quantity of training data have a small but statistically significant effect on classifier performance for people with high level disabilities. The selection of ANN training algorithm was found to have the largest effect.

The results presented in Chapter 4 show that although conventional ANN optimisation techniques can improve classifier performance, the real-time classification performance for people with disabilities is still inferior to that for able-bodied people. The results presented in Chapter 5 show that using data from a specific end user to train the ANN can significantly improve classifier performance, over the improvement achieved by the use of generic techniques. It was found that retraining an ANN, originally trained on generic data, using a combination of user specific and generic data could improve the performance of the classifier for that end user while minimising or avoiding any reduction in classifier performance for other people. It was also found that retraining such an ANN with user specific data alone improves the performance of the classifier for that end user but is detrimental to the classification performance for other people.

6.2. Contributions

The head gesture wheelchair control system proposed in Chapter 3 advances on the design proposed by Joseph and Nguyen (1998) and differs from others previously proposed in the literature. The experimental results involving people with disabilities presented in Chapter 3 demonstrate the feasibility of the control system.

Methods for improving the performance of the control system using generic data were identified and investigated in Chapter 4. Optimal selections for the training of ANN for head gesture classification using generic data have been found empirically and tested for people with high-level disabilities. Prior results in the literature do not sufficiently predict the effectiveness of these methods and selections in this application.

The utility of user specific data in improving the ability of an ANN to classify head gestures by people with high level disabilities in real-time is not addressed in the literature. The adaptive algorithms proposed in Chapter 5 can be considered relatively minor contributions, being extensions of conventional ANN training techniques. The application of these algorithms to the real-time head gesture classification task, or similar applications, is not addressed in previous publications.

6.3. Limitations

The population of potential users for the wheelchair control system proposed in Chapter 3 is restricted to those people who have some degree of voluntary, controlled head movement. This restriction prevents the system from meeting the mobility needs of people with the highest levels of disability. This restriction is necessary, as there is a wide range of levels of disability and it is unlikely that a control system meeting the needs of some of the potential users without excluding others.

The experimental results presented in Chapters 3, 4 and 5 are based on gestures recorded in a protocol involving prompted gestures by a stationary person. To further assess the feasibility of the control system, it is necessary to demonstrate the ability of the system to accurately classify gestures performed while directing the motion of the wheelchair. This was not performed in this thesis due to safety reasons at this stage of development.

The number of subjects from whom data was collected was limited by the availability of suitable volunteers. Conclusions can be made regarding the classification performance for subjects from whom data was collected, but the sample size is insufficient to make strong conclusions extrapolating to broader populations.

There are many strategies that can be applied to generic optimisation, of which the factors investigated in Chapter 4 are a small subset of those that could be used. The generic optimal selections made based on the observed results in Chapter 4 are therefore not to be considered globally optimal, but rather are optimal settings from those tested. The scope of the optimisation examined in Chapters 4 and 5 was restricted to the ANN component of the classifier, and therefore omits other factors which may have a significant effect, such as pre-processing of the sensor data to transform the ANN input space into one in which the classes of gesture are more easily separated. This restriction was necessary to render the investigation tractable given the resources available.

The adaptive retraining procedures proposed in Chapter 5 require a labour intensive collection and labelling of user specific data, which would not necessarily be viable in a practical wheelchair control system for wide use. This will be addressed in Section 6.4.

6.4. Future directions

There are several avenues of research that can be considered natural extensions of the results and conclusions presented in this thesis.

For the further development of a wheelchair control system for highly disabled people, the utility of the control system could be enhanced by allowing for the recognition of a greater number of command gestures. This would allow users to communicate with the control system more quickly and more flexibly, making it more useful in meeting the mobility needs of people with disabilities.

Similarly, head gesture recognition could be combined with other control interface methods. For example, commands or control signals in the form of speech, facial gesture, EOG and EEG could be used in conjunction with head gestures. In addition to allowing faster or more flexible control interfaces, the fusion of multiple techniques could be used to provide redundancy for fail-safe or more reliable control. In order to pursue this avenue, it will be necessary to consider the potential for interference between interface techniques, such as EMG signals generated by a person performing a head gesture obscuring EEG potentials measured at the same time.

Since it has been shown that training the ANN using user specific data improves the performance of the classifier, this leads to a range of opportunities for more advanced procedures for adaptive retraining. One barrier to the widespread use of the adaptive algorithm proposed in Chapter 6 is that it requires a human expert to review and label sufficient data for each potential user. The development of more loosely supervised techniques, such as reinforcement learning (Sutton 1998), or techniques which increase the marginal utility of each labelled piece of data, such as active or query learning (Krogh & Vedelsby 1995), may make it more viable to collect sufficiently large datasets for large numbers of users.

The input space of the classifier, in its current form, is large compared to the output space and the region of input space corresponding to one of the commands being present is small compared to the portion of input space not corresponding to any command gesture. Points from some regions of the classifier input space are also rarely encountered in sensor data. As a result, it is difficult to ensure that the training data provides adequate coverage of the input space to provide reliable performance, particularly if the classifier is to be adapted to the specific end user. Alternative pre-processing techniques could be used to reduce the size of the input space. For example, time-frequency transformations could re-represent the sensor data in a more compressed form. A study of the biomechanics of head movement in highly-disabled people may lead to more sophisticated pre-processing methods.

Since the techniques developed for this project can also be applied to a wide range of real-time classification tasks, other applications in assistive technologies may also be investigated. Use of the same techniques has already been examined for environmental control units (H. T. Nguyen et al. 2002). Applications such as human-computer interfaces and communication devices are also natural extensions of the control system examined in this thesis.

References

ABLEDATA 2003, *Peachtree Proportional Head Control System*, National Institute on Disability and Rehabilitation Research, viewed 1/11/2007 2007, http://www.abledata.com.

Adachi, Y, Kuno, Y, Shimada, N & Shirai, Y 1998, 'Intelligent wheelchair using visual information on human faces', *Proceedings of the 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 354-9.

Adaptive Switch Laboratories Inc 2004, *PASL105IVC Proximity Head Array Package for Invacare Electronics*, viewed 1/11/2007 2007, http://www.asl-inc.com/.

Analog Devices 1999a, *ADXL202/ADXL210: Low Cost ±2g / ±10 g Dual Axis iMEMS® Accelerometers with Digital Output*, Analog Devices, viewed 10/8/2004 2004, http://www.analog.com.

Analog Devices 1999b, *ADXL202EB-232A: ADXL202 Evaluation Board with RS-232 Interface and Datalogging.*, Analog Devices, viewed 10/8/2004 2004, http://www.analog.com.

Analog Devices 2000, *ADXL202E: Low-Cost ±2 g Dual-Axis Accelerometer with Duty Cycle Output*, Analog Devices, viewed 10/8/2004 2004, http://www.analog.com.

Anderson, CW, Devulapalli, SV & Stolz, EA 1995, 'EEG signal classification with different signal representations', *Proceedings of the 1995 IEEE Workshop Neural Networks for Signal Processing*, pp. 475-83.

Anderson, CW, Stolz, EA & Shamsunder, S 1998, 'Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks', *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 3, pp. 277-86.

Bahi, DH, LaCourse, JR & Hludik, FC 1991, 'Use of a discrete electro-oculographic control system with and without oscillatory motion of the head', *Proceedings of the 1991 IEEE Seventeenth Annual Northeast Bioengineering Conference*, pp. 129-30.

Barea, R, Boquete, L, Mazo, M & Lopez, E 2002, 'System for assisted mobility using eye movements based on electrooculography', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 1534-4320, pp. 209-18.

Barea, R, Boquete, L, Mazo, M, Lopez, E & Bergasa, LM 2000, 'EOG guidance of a wheelchair using neural networks', *Proceedings of the 15th International Conference on Pattern Recognition*, vol. 4, pp. 668-71.

Barg, J 1959, 'Typing By Head', *Toomeyville Jr. Gazette*, 1959, p. 5, http://www.disabilitymuseum.org.

Barnard, E 1992, 'Optimization for training neural nets', *IEEE Transactions on Neural Networks*, vol. 3, no. 2, pp. 232-40.

Bergasa, LM, Mazo, M, Gardel, A, Barea, R & Boquete, L 2000, 'Commands generation by face movements applied to the guidance of a wheelchair for handicapped people', *Proceedings of the 15th International Conference on Pattern Recognition*, vol. 4, pp. 660-3.

Bergasa, LM, Mazo, M, Gardel, A, Garcia, JC, Ortuno, A & Mendez, AE 1999, 'Guidance of a wheelchair for handicapped people by face tracking', *Proceedings of the 7th IEEE International Conference on Emerging Technologies and Factory Automation*, vol. 1, pp. 105-11.

Bilski, J & Rutkowski, L 1998, 'A fast training algorithm for neural networks', *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 6, pp. 749-53.

Bland, M 2000, *An introduction to medical statistics*, 3rd edn, Oxford medical publications., Oxford University Press, Oxford.

Brandt, Å, Iwarsson, S & Ståhle, A 2004, 'Older people's use of powered wheelchairs for activity and participation', *Journal of Rehabilitation Medicine*, vol. 26, no. 2, pp. 70-7.

Castelli, V & Cover, TM 1993, 'The relative value of labeled and unlabeled samples in pattern recognition', *Proceedings of the 1993 IEEE International Symposium on Information Theory*, p. 355.

Chakraborty, D & Pal, NR 2003, 'A novel training scheme for multilayered perceptrons to realize proper generalization and incremental learning', *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 1-14.

Chakraborty, G, Shiratori, N & Noguchi, S 1993, 'Optimization of overtraining and overgeneralization', *Proceedings of the 1993 International Joint Conference on Neural Networks*, vol. 3, Nagoya, pp. 2257-62.

Chan, L-W 1990, 'Efficacy of different learning algorithms of the back-propagation network', *IEEE Region 10 Conference on Computer and Communication Systems*, vol. 1, Hong Kong, pp. 23-7.

Chauhan, S, Sharma, P, Singh, HR, Mobin, A & Agrawal, SS 2000, 'Design and development of voice-cum-auto steered robotic wheelchair incorporating reactive fuzzy scheme for anti-collision and auto routing', *Proceedings of TENCON 2000*, vol. 1, pp. 192-5.

Chen, S-H, Chen, Y-L, Chiou, Y-H, Tsai, J-C & Kuo, T-S 2003, 'Head-controlled device with M3S-based for people with disabilities', *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 1587-9.

Chen, Y-L 2001, 'Application of tilt sensors in human-computer mouse interface for people with disabilities', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 9, no. 3, pp. 289-94.

Chen, Y-L, Chen, S-C, Chen, W-L & Lin, J-F 2003, 'A head orientated wheelchair for people with disabilities', *Disability & Rehabilitation*, vol. 25, no. 6, pp. 249 -53.

Chen, Y-L, Chen, W-L & Lai, J-S 2002, 'A head orientated electric wheelchair for people with disabilities', *Proceedings of the Second Joint MBS/BMES Conference, 2002.*, vol. 3, pp. 2371-2.

Compaq, Intel, Microsoft & NEC 1998, *Universal Serial Bus Specification*, http://www.usb.org/.

Cooper, RA 1995, 'Intelligent control of power wheelchairs', *Engineering in Medicine and Biology Magazine, IEEE*, vol. 14, no. 4, pp. 423-31.

Cooper, RA, Jones, DK, Fitzgerald, S, Boninger, ML & Albright, SJ 2000, 'Analysis of position and isometric joysticks for powered wheelchair driving', *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 902-10.

Cooper, RA, Widman, LM, Jones, DK, Robertson, RN & Ster, JF, III 2000, 'Force sensing control for electric powered wheelchairs', *IEEE Transactions on Control Systems Technology*, vol. 8, no. 1, pp. 112-7.

Corbett, D & Martinez, P 1998, 'Toward learning individual characteristics in a hybrid neuro-fuzzy wheelchair controller', *International Journal of Intelligent Systems*, vol. 13, no. 6, pp. 561-70.

Coyle, D, Prasad, G & McGinnity, TM 2005, 'A time-series prediction approach for feature extraction in a brain-computer interface', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 4, pp. 461-7.

Coyle, ED 1995, 'Electronic wheelchair controller designed for operation by hand-operated joystick, ultrasonic noncontact head control and utterance from a small word-command vocabulary', *IEE Colloquium on New Developments in Electric Vehicles for Disabled Persons*, pp. 3/1-3/4.

Cripps, RA 2003, *Spinal cord injury, Australia 2001-02*, Report Number INJCAT 58, Australian Institute of Health and Welfare, Canberra.

Cripps, RA 2004, *Spinal cord injury, Australia 2002-03*, Report Number INJCAT 64, Australian Institute of Health and Welfare, Canberra.

del R. Milan, J, Mourino, J, Marciani, MG, Babiloni, F, Topani, F, Canale, I, Heikkonen, J & Kaski, K 1998, 'Adaptive brain interfaces for physically-disabled people', *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 4, pp. 2008-11.

Demuth, H, Beale, M & Hagan, M 2005, *Neural Network Toolbox User's Guide*, 4 edn, Mathworks, Natick, USA.

Dewsbury, G, Sommerville, I, Clarke, K & Rouncefield, M 2003, *A Dependability Model for Domestic Systems*, 2788 edn, Lecture Notes in Computer Science.

Dewsbury, G, Taylor, B & Edge, M 2001, 'Designing Safe Smart Home Systems for Vulnerable People', *Dependability and Healthcare Informatics Workshop*, eds R Proctor & M Rouncefield, Edinburgh, pp. 55-70.

Dewsbury, G, Taylor, B & Edge, M 2002, 'Designing Dependable Assistive Technology Systems for Vulnerable People', *Health Informatics Journal*, vol. 8, no. 2, pp. 104-10.

Ding, D, Cooper, RA & Spaeth, D 2004, 'Optimized joystick controller', *Proceedings of the 26th Annual International Conference of the Engineering in Medicine and Biology Society*, vol. 7, pp. 4881-3.

Evans, DG, Drew, R & Blenkhorn, P 2000, 'Controlling mouse pointer position using an infrared head-operated joystick', *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 1, pp. 107-17.

Fawcett, T 2004, 'ROC Graphs: Notes and Practical Considerations for Researchers', *Machine Learning (Submitted)*,

Fehr, L, Langbein, WE & Skaar, SB 2000, 'Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey', *Journal of Rehabilitation Research and Development*, vol. 37, no. 3, pp. 353-60.

Ferri, C, Hernándéz-Orallo, J & Salido, MA 2003, *Volume under the ROC Surface for Multi-class Problems*, 2837 edn, Lecture Notes in Computer Science.

Fisher, L & Van Belle, G 1993, *Biostatistics : a methodology for the health sciences*, Wiley science paperback series., Wiley, New York ; Chichester.

Friedman, M 1937, 'The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance', *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675-701.

'Gangway!: A New Style Vehicle Comes to Warm Springs.' 1932, *The Polio Chronicle*, June 1932, p. 10, www.disabilitymuseum.org.

Garrett, D, Peterson, DA, Anderson, CW & Thaut, MH 2003, 'Comparison of linear, nonlinear, and feature selection methods for EEG signal classification', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 141-4.

'Gasoline Custer Chair' 1933, *The Polio Chronicle*, July 1933, p. 8, www.disabilitymuseum.org.

Gibbons, JD & Chakraborti, S 1992, *Nonparametric statistical inference*, 3rd edn, M. Dekker, New York.

Gope, C, Kehtarnavaz, N & Nair, D 2005, 'Neural network classification of EEG signals using time-frequency representation', *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks*, vol. 4, pp. 2502-7.

Griss, P, Enoksson, P, Tolvanen-Laakso, HK, Merilainen, P, Ollmar, S & Stemme, G 2001, 'Micromachined electrodes for biopotential measurements', *Journal of Microelectromechanical Systems*, vol. 10, no. 1, pp. 10-6.

Guo, S, Cooper, RA, Boninger, ML, Kwarciak, A & Ammer, B 2002, 'Development of power wheelchair chin-operated force-sensing joystick', *Proceedings of the Second Joint EMBS/BMES Conference*, vol. 3, pp. 2373-4.

Guyon, I 1991, 'Neural networks and applications tutorial', *Physics Reports*, vol. 207, no. 3-5, pp. 215-59.

Hagan, MT 1995, *Neural network design*, PWS Pub., Boston.

Hagan, MT & Menhaj, MB 1994, 'Training feedforward networks with the Marquardt algorithm', *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989-93.

Hamill, J & Knutzen, KM 2003, *Biomechanical Basis of Human Movement*, 2nd edn, Lippincott Williams & Wilkins.

Han, J-S, Zenn Bien, Z, Kim, D-J, Lee, H-E & Kim, J-S 2003, 'Human-machine interface for wheelchair control with EMG and its evaluation', *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 1602-5.

Hand, DJ & Till, RJ 2001, 'A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems', *Machine Learning*, vol. 45, no. 2, pp. 171-86.

Hansen, M, Haugland, MK & Sinkjaer, T 2004, 'Evaluating robustness of gait event detection based on machine learning and natural sensors', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 1, pp. 81-8.

Haykin, SS 1999, *Neural networks : a comprehensive foundation*, 2nd edn, Prentice Hall, Upper Saddle River, N.J.

Hiraiwa, A, Shimohara, K & Tokunaga, Y 1989, 'EMG pattern analysis and classification by neural network', *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 1113-5.

Hornik, K, Stinchcombe, M & White, H 1989, 'Multilayer feedforward networks are universal approximators', *Neural Networks*, vol. 2, no. 5, pp. 359-66.

Irie, B & Miyake, S 1988, 'Capabilities of three-layered perceptrons', *IEEE International Conference on Neural Networks*, vol. 1, Osaka, Japan, pp. 641-8.

Jonkers, I, Nuyens, G, Seghers, J, Nuttin, M & Spaepen, A 2004, 'Muscular effort in multiple sclerosis patients during powered wheelchair manoeuvres', *Clinical Biomechanics*, vol. 19, no. 9, pp. 929-38.

Joseph, T & Nguyen, H 1998, 'Neural network control of wheelchairs using telemetric head movement', *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 5, pp. 2731-3.

King, LM, Nguyen, HT & Taylor, PB 2005, 'Hands-free Head-movement Gesture Recognition using Artificial Neural Networks and the Magnified Gradient Function', *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society*, Shanghai, China, pp. 2063-6.

Knight, G 1999, 'Telemetric head movement control of a powered wheelchair', Undergraduate thesis, University of Technology Sydney, Sydney.

Komiya, K, Morita, K, Kagekawa, K & Kurosu, K 2000, 'Guidance of a wheelchair by voice', *Proceedings of the 26th Annual Confjerence of the IEEE Industrial Electronics Society*, vol. 1, pp. 102-7.

Krogh, A & Vedelsby, J 1995, 'Neural Network Ensembles, Cross Validation, and Active Learning', in G Tesauro, D Touretzky & T Leen (eds), *Advances in Neural Information Processing Systems*, vol. 7, The MIT Press, pp. 231-8.

Kruse, D & Schur, L 2003, 'Employment of People with Disabilities Following the ADA', *Industrial Relations*, vol. 42, no. 1, pp. 31-66.

Kuno, Y, Murashima, T, Shimada, N & Shirai, Y 2000, 'Interactive gesture interface for intelligent wheelchairs', *Proceedings of the IEEE International Conference on Multimedia and Expo*, vol. 2, pp. 789-92.

Kuno, Y, Murashina, T, Shimada, N & Shirai, Y 2000, 'Intelligent wheelchair remotely controlled by interactive gestures', *Proceedings of the 15th International Conference on Pattern Recognition*, vol. 4, pp. 672-5.

Kuno, Y, Shimada, N & Shirai, Y 2003, 'Look where you're going [robotic wheelchair]', *Robotics & Automation Magazine, IEEE*, vol. 10, no. 1, pp. 26-34.

LaCourse, JR & Hludik, FC, Jr. 1990, 'An eye movement communication-control system for the disabled', *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 12, pp. 1215-20.

Law, CKH, Leung, MYY, Xu, Y & Tso, SK 2002, 'A cap as interface for wheelchair control', *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and System*, vol. 2, pp. 1439-44.

Lin, C-T & Lee, CSG 1996, *Neural fuzzy systems*, Prentice Hall, Upper Sandle River, NJ, USA.

Looney, CG 1997, *Pattern recognition using neural networks : theory and algorithms for engineers and scientists*, Oxford University Press, New York.

LoPresti, EF & Brienza, DM 2004, 'Adaptive software for head-operated computer controls', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 1, pp. 102-11.

Maiorescu, VA, Serban, M & Lazar, AM 2003, 'Classification of EEG signals represented by AR models for cognitive tasks - a neural network based method', *International Symposium on Signals, Circuits and Systems*, vol. 2, pp. 441-4.

Martinez-Coll, A, Papacosta, C & Nguyen, H 2003a, 'Feasibility of Bilateral Sternocleidomastoid (SCM) Surface Electromyography (sEMG) for Variable Control of Powered Wheelchair using Head movement Technology', *World Congress on Medical Physics and Biomedical Engineering*, Sydney, p. 2853 (CDROM).

Martinez-Coll, A, Papacosta, C & Nguyen, H 2003b, 'Surface Electromyography (sEMG) of the Sterocleidomastoid (SCM) Muscle for Variable Control using Head Movement Technology', *25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Cancun, Mexico, pp. 1598-601.

Matlab Neural Network Toolbox 2006, computer program, The MathWorks, Inc, Natick, MA, Version V5.0.1.

Matsumoto, Y, Ino, T & Ogsawara, T 2001, 'Development of intelligent wheelchair system with face and gaze based interface', *Proceedings of the 10th IEEE International Workshop on Robot and Human Interactive Communication*, pp. 262-7.

Matsumura, Y, Mitsukura, Y, Fukumi, M, Akamatsu, N, Yamamoto, Y & Nakaura, K 2002, 'Recognition of EMG signal patterns by neural networks', *Proceedings of the 9th International Conference on Neural Information Processing*, vol. 2, pp. 750-4.

Mazo, M 2001, 'An integral system for assisted mobility [automated wheelchair]', *Robotics & Automation Magazine, IEEE*, vol. 8, no. 1, pp. 46-56.

Mazo, M, Garcia, JC, Rodriguez, FJ, Urena, J, Lazaro, JL & Espinosa, F 2002, 'Experiences in assisted mobility: the SIAMO project', *Proceedings of the 2002 International Conference on Control Applications*, vol. 2, pp. 766-71.

McCulloch, WS & Pitts, W 1943, 'A logical calculus of the ideas immanent in nervous activity', *Bulletin of Mathematical Biology*, vol. 5, no. 4, pp. 115-33.

McIsaac, P, Craig, A, Tran, Y & Boord, P 2002, 'The mind switch environmental control system :remote hands free control for the severely disabled', *Technology and Disability*, vol. 14, pp. 15-20.

Miller, DP & Slack, MG 1994, 'Increasing access with a low-cost robotic wheelchair', *Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems*, vol. 3, pp. 1663-7.

Min, J-W, Lee, K, Lim, S-C & Kwon, D-S 2002, 'Human-friendly interfaces of wheelchair robotic system for handicapped persons', *IEEE/RSJ International Conference on Intelligent Robots and System*, vol. 2, pp. 1505-10.

Minsky, ML & Papert, SA 1969, *Perceptrons*, MIT Press, Cambridge, Mass.

Møller, MF 1993, 'A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning', *Neural Networks*, vol. 6, pp. 525-33.

Møller, MF 1997, 'Efficient training of feed-forward neural networks', Ph.D thesis, University of Aarhus, Aarhus, Denmark.

Moon, I, Lee, M, Ryu, J & Mun, M 2003, 'Intelligent robotic wheelchair with EMG-, gesture-, and voice-based interfaces', *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 4, pp. 3453-8.

Morimoto, C, Yacoob, Y & Davis, L 1996, 'Recognition of head gestures using Hidden Markov Models', *International Conference on Pattern Recognition*, Vienna, Austria, pp. 461-5.

Newell, J, Paul H. & Barr, JS 1971, 'Voice-operated powered devices', *Journal of Biomechanics*, vol. 4, no. 1, pp. 45-6.

Nguyen, HT, King, LM & Knight, G 2004, 'Real-time head movement system and embedded Linux implementation for the control of power wheelchairs', *Proceedings of the 26th Annual International Conference of the Engineering in Medicine and Biology Society*, vol. 2, pp. 4892-5.

Nguyen, HT, Knight, G & Ekanayke, R 2003, 'Telemetric head movement control of a power wheelchair', *World Congress on Medical Physics and Biomedical Engineering*, Sydney, pp. 2516 (CD-ROM).

Nguyen, HT, Legaspi, S, Knight, G, Ekanayke, R, Taylor, PB & Martinez-Coll, A 2002, 'A head movement system for environmental control units', *Engineering and Physical Sciences in Medicine*, Rotorua, New Zealand, p. 157.

Nguyen, HT, Nguyen, ST, Taylor, PB & Middleton, J 2007, 'Head Direction Command Classification using an Adaptive Optimal Bayesian Neural Network', *International Journal of Factory Automation, Robotics and Soft Computing*, vol. In Press, no. Paper ID 365,

Nguyen, S, Nguyen, H & Taylor, P 2004, 'Hands-free control of power wheelchairs using Bayesian neural network classification', *IEEE Conference on Cybernetics and Intelligent Systems*, vol. 2, Singapore, pp. 746-50.

Nguyen, ST, Nguyen, HT & Taylor, P 2006, 'Improved Head Direction Command Classification using an Optimised Bayesian Neural Network', *IEEE International Conference of the Engineering in Medicine and Biology Society*, New York, USA.

Nguyen, ST, Nguyen, HT & Taylor, PB 2006, 'Bayesian Neural Network Classification of Head Movement Direction using Various Advanced Optimisation Training Algorithms', *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, Pisa, Italy, pp. 1014-9.

Noyes, J & Starr, A 1996, 'Use of automatic speech recognition: current and potential applications', *Computing & Control Engineering Journal*, vol. 7, no. 5, pp. 203-8.

O'Connor, PJ 2001, *Spinal cord injury, Australia, 1999-00*, Report Number INJCAT 40, Australian Institute of Health and Welfare, Canberra.

O'Connor, PJ 2003a, *Spinal cord injury, Australia 2000–01*, Report Number INJCAT 50, Australian Institute of Health and Welfare, Canberra.

O'Connor, PJ 2003b, *Trends in spinal cord injury, Australia 1986-1997*, Report Number INJCAT 52, Australian Institute of Health and Welfare, Canberra.

O'Connor, PJ 2004, 'Prevalence of spinal cord injury in Australia', vol. 43, no. 1, pp. 42-6.

Pruski, A & Bourhis, G 1992, 'The VAHM project: a cooperation between an autonomous mobile platform and a disabled person', *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 268-73.

Ragupathy, SC, Kumar, DK, Polus, B & Kamei, K 2004, 'Electrocardiogram removal from electromyogram of the muscles', *Proceedings of the 26th Annual International Conference of the Engineering in Medicine and Biology Society*, vol. 1, pp. 243-6.

Rao, RS, Seliktar, R & Rahman, T 1999, 'Evaluation of joysticks in cursor positioning task by individuals with cerebral palsy', *Proceedings of the First Joint BMES/EMBS Conference*, vol. 1, p. 666.

Reischl, M, Groll, L & Mikut, R 2004, 'EMG-control of prostheses by switch signals: extraction and classification of features', *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, pp. 94-6.

Rockland, RH & Reisman, S 1998, 'Voice activated wheelchair controller', *Proceedings of the IEEE 24th Annual Northeast Bioengineering Conference*, pp. 128-9.

Rosenberg, R 1998, 'The biofeedback pointer: EMG control of a two dimensional pointer', *Digest of Papers from the Second International Symposium on Wearable Computers*, pp. 162-3.

Routhier, F, Vincent, C, Desrosiers, J & Nadeau, S 2003, 'Mobility of wheelchair users: a proposed performance assessment framework', *Disability & Rehabilitation*, vol. 25, no. 1, pp. 19 - 34.

Rumelhart, DE, Hinton, GE & Williams, RJ 1986 'Learning internal representations by error propagation ', in *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations* MIT Press, pp. 318-62

Salganicoff, M, Jayachandran, V, Pine, D, Rahman, T, Mahoney, R, Chen, S, Kumar, V, Harwin, W & Gonzalez, JG 1995, 'A virtual head-stick rehabilitation robot system', *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2413-8.

Schur, LA 2003, 'Barriers or Opportunities? The Causes of Contingent and Part-Time Work Among People with Disabilities', *Industrial Relations*, vol. 42, no. 4, pp. 589-622.

Serby, H, Yom-Tov, E & Inbar, GF 2005, 'An improved P300-based brain-computer interface', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 89-98.

Sexton, RS & Dorsey, RE 2000, 'Reliable classification using neural networks: a genetic algorithm and backpropagation comparison', *Decision Support Systems*, vol. 30, no. 1 SU -, pp. 11-22.

Sietsma, J & Dow, RJF 1988, 'Neural net pruning-why and how', *IEEE International Conference on Neural Networks*, vol. 1, pp. 325-33.

Simpson, RC & Levine, SP 1997, 'Adaptive shared control of a smart wheelchair operated by voice control', *Proceedings of the 1997 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, Ann Arbor, MI, USA, pp. 622-6.

Spinal Cord Injury: Facts and Figures at a Glance. 2004, National Spinal Cord Injury Statistical Center, University of Alabama at Birmingham, viewed 21/9/2005 2005, www.spinalcord.uab.edu.

Statistics Toolbox User's Guide 2006, 5 edn, Mathworks, Natick, USA.

Sutton, RS 1998, *Reinforcement learning : an introduction*, MIT Press, Cambridge, Mass.

Takami, O, Irie, N, Kang, C, Ishimatsu, T & Ochiai, T 1996, 'Computer interface to use head movement for handicapped people', *Proceedings of IEEE TENCON 96 on Digital Signal Processing Applications*, vol. 1, pp. 468-72.

Tanaka, K, Matsunaga, K & Wang, HO 2005, 'Electroencephalogram-Based Control of an Electric Wheelchair', *IEEE Transactions on Robotics*, vol. 21, no. 4, pp. 762-6.

Taylor, PB & Nguyen, H 2003, 'Neural network classification of head-movement for wheelchair control', *World Congress on Medical Physics and Biomedical Engineering*, Sydney, Australia, pp. 3953 (CD-ROM).

Taylor, PB, Nguyen, H & Craig, A 2002, 'Head Movement Recognition for Power Wheelchair Control', *Engineering and Physical Sciences in Medicine*, Rotorua, New Zealand, p. 135.

Taylor, PB & Nguyen, HT 2003, 'Performance of a head-movement interface for wheelchair control', *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 1590-3.

Thieffry, R, Monacelli, E, Henaff, P & Delaplace, S 2003, 'Configuring sensors by user learning for a locomotion aid interface', *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 1509-14.

Thompson, B, Picton, P & Jones, NB 1996, 'A comparison of neural network and traditional signal processing techniques in the classification of EMG signals', *IEE Colloquium on Artificial Intelligence Methods for Biomedical Data Processing*, pp. 8/1-8/5.

Thorkildsen, R 1994, *Research Synthesis on Quality and Availability of Assistive Technology Devices*, Report Number 7, Center for Persons with Disabilities, Utah State University.

Tordoff, B & Mayol, W 2001, *Articulated Humanoid Body*, The MathWorks, Inc, viewed 17/6/2006 2006, http://www.mathworks.com/matlabcentral/fileexchange/.

Twomey, JM & Smith, AE 1995, 'Performance measures, consistency, and power for artificial neural network models', *Mathematical and Computer Modelling*, vol. 21, no. 1-2 SU, pp. 243-58.

Tzafestas, SG 2001, 'Research on autonomous robotic wheelchairs in Europe [Guest Editorial]', *Robotics & Automation Magazine, IEEE*, vol. 8, no. 1, pp. 4-6.

van der Zwaag, BJ, Corbett, D & Jain, L 1999, 'Minimising tremor in a joystick controller using fuzzy logic', *Proceedings of the Third International Conference on Knowledge-Based Intelligent Information Engineering Systems*, pp. 5-8.

Vidal, JJ 1977, 'Real-time detection of brain events in EEG', *Proceedings of the IEEE*, vol. 65, pp. 633-64.

Walpole, RE & Myers, RH 1993, *Probability and statistics for engineers and scientists*, 5th edn, Prentice Hall, Englewood Cliffs, N.J.

Walpole, RE, Myers, RH & Myers, SL 1998, *Probability and statistics for engineers and scientists*, 6th edn, Prentice Hall, Upper Saddle River, NJ.

Warren, CG 1990, 'Powered mobility and its implications', *Journal of Rehabilitation Research and Development*, vol. Clinical Supplement No.2; Choosing a Wheelchair System, pp. 74-85.

Warren, L, Wrigley, JM, Yoels, WC & Philip R. Fine, P 1996, 'Factors associated with life satisfaction among a sample of persons with neurotrauma', *Journal of Rehabilitation Research and Development*, vol. 33, no. 4, pp. 404-8.

Wattenberg, T 2004, 'Beyond legal compliance: Communities of advocacy that support accessible online learning', *The Internet and Higher Education*, vol. 7, no. 2, pp. 123-39.

Werbos, PJ 1974, 'Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences', PhD thesis, Harvard University, Cambridge MA, USA.

Widrow, B & Lamego, MM 2002, 'Neurointerfaces', *IEEE Transactions on Control Systems Technology*, vol. 10, no. 2, pp. 221-8.

Widrow, B & Lehr, MA 1990, '30 years of adaptive neural networks: perceptron, Madaline, and backpropagation', *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415-42.

Williams, M & Kirsch, R 2004, 'Evaluation of command sources for a high tetraplegia neural prosthesis', *Proceedings of the 26th Annual International*

Conference of the Engineering in Medicine and Biology Society, vol. 7, pp. 4803-6.

Wolpaw, JR, Birbaumer, N, Heetderks, WJ, McFarland, DJ, Peckham, PH, Schalk, G, Donchin, E, Quatrano, LA, Robinson, CJ & Vaughan, TM 2000, 'Brain-computer interface technology: a review of the first international meeting', *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164-73.

Wolpert, DH 1992, 'Stacked Generalization', *Neural Networks*, vol. 5, pp. 241-59.

Wolpert, DH & Macready, WG 1995, *No Free Lunch Algorithms for Search*, Report Number SFI TR 95-02-010, The Santa Fe Institute.

Wolpert, DH & Macready, WG 1997, 'No free lunch theorems for optimization', *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67-82.

Woods, K & Bowyer, KW 1997, 'Generating ROC curves for artificial neural networks', *IEEE Transactions on Medical Imaging*, vol. 16, no. 3, pp. 329-37.

Yom-Tov, E & Inbar, GF 2002, 'Feature selection for the classification of movements from single movement-related potentials', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 3, pp. 170-7.

Ziskind, A 1959, 'Typing By Electric Eye', *Toomeyville Jr. Gazette*, p. 2, http://www.disabilitymuseum.org.

Zola, IK 1988, 'The Oration: Ageing And Disability: Toward A Unifying Agenda', *Australian Disability Review*, vol. 1, no. 3, pp. 3-12.

Zurada, JM 1992, *Introduction to artificial neural systems*, West, St. Paul.

Appendix A. Background on the high-level spinal cord injury

Research presented in this thesis contributes to the development of a wheelchair control system for highly disabled individuals. The aim of developing the wheelchair control system is to provide a method for the needs of people who are unable to operate a power wheelchair using existing techniques, due to high-level spinal cord injuries or conditions resulting similar motor impairments.

The level of a spinal cord injury is often measured by the vertebrae corresponding with the lowest level of full function. A working definition for quadriplegia, or tetraplegia as it is more commonly referred to in the relevant literature, is that the lowest level of full function is between C1 and C8, which are the eight cervical vertebrae, located in the neck (*Spinal Cord Injury: Facts and Figures at a Glance.* 2004). High-level tetraplegia is defined as an injury from C1 to C4. Low-level tetraplegia is defined as being from C5 to C8. Potential users of the control system considered in this thesis are more likely to be high-level tetraplegics, as low-level tetraplegics are more likely to be able to use a joystick interface. Figure 71 gives an approximate guide to the level of function corresponding to each level.

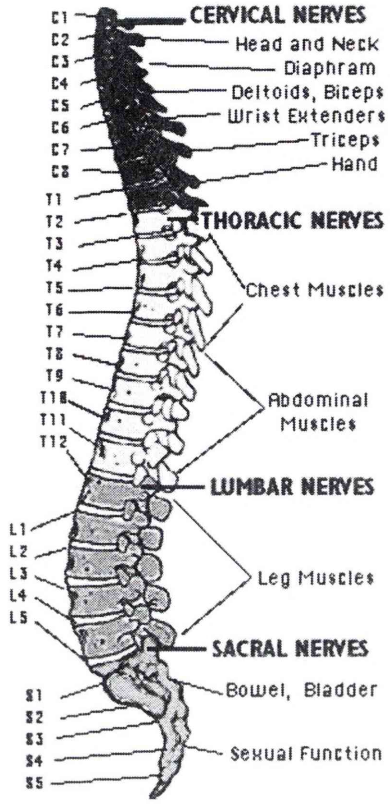

Figure 71 Approximate mapping of level of injury to remaining function (*Spinal Cord Injury: Facts and Figures at a Glance.* 2004)

It should be noted that approximately two thirds of SCI's are classed as incomplete, meaning that at there is least some preservation of sensory or motor functions found below the neurological level or the injury and includes the lowest sacral segment (O'Connor 2001). Main requirement of interfacing techniques is that they must use the abilities of the user to allow them to communicate their intentions.

Appendix B. Artificial neural network activation functions

As mentioned in Section 4.1, the following functions are used in the literature as activation functions for neurons in ANN. Each of the functions mentioned is defined below, where v is the weighted sum of the inputs to the neuron and a is a positive constant.

Heaviside

$$f(v) = \begin{cases} 1, & v > 0 \\ 0.5, & v = 0 \\ 0, & v < 0 \end{cases}$$

Signum

$$f(v) = \begin{cases} +1, & v > 0 \\ 0, & v = 0 \\ -1, & v < 0 \end{cases}$$

Sigmoid

$$f(v) = \frac{1}{1+e^{-av}}$$

Bipolar sigmoid

$$f(v) = \frac{1 - e^{-av}}{1 + e^{-av}} = \frac{2}{1 + e^{-av}} - 1$$

Appendix C. On alternative classifier performance metrics

Many of the metrics used throughout the literature lack the flexibility of being applicable to both instantaneous classification and the real-time detection and classification application considered in this thesis. For example, normalised RMS error has been frequently used in the ANN classifier literature, but can only be validly calculated where a precise expected output can be provided for every input pattern. In this experiment, such information is not available, as although each recorded gesture can be given a label, the individual input patterns that arise as the gesture signal passes across the sliding window of the classifier input do not necessarily have such precise labels. Such patterns can be a representative of one particular class or representative of a transition between two classes. In the latter case, the pattern could conceivably be labelled by an expert to be an example of either class, both or neither. Twomey and Smith (1995) indicate that the normalised RMS error carries different information to the error rate. However, no satisfactory metric has been found in the literature that can be validly calculated in the conditions found in this experiment. Other metrics subject to this problem include the mean absolute error and mean squared error.

Metrics that relate to the training process, such as cycles to convergence, are not useful in the measurement of classifier performance. Such metrics do not provide information relating to the performance of the training algorithm with regard to the performance of the resulting classifiers. These metrics are more significant where the efficiency of the training algorithm is of interest.

Receiver operating characteristic metrics, such as area under curve, provide insight into the trade off between sensitivity and specificity. While useful for choosing between two classifiers, these metrics have a number of drawbacks. Fisher and Van Belle (1993) state that the ROC curve does not provide enough information to discriminate between classifiers as it does not depend on the prevalence of each class. Calculation of the properties of the ROC curve also presents problems. Plotting an ROC curve requires that the sensitivity be found at a range of specificities. For

ANN classifiers, this is typically done by varying the threshold of the output nodes, although Woods and Bowyer (1997) varied the bias input to the hidden layer nodes. This adds significant computational cost, as it requires that the error of the classifier must be found and analysed many times. For these reasons, receiver operating characteristic analysis will not be performed as part of these experiments.

Appendix D. On the selection of sample sizes

Sample sizes and resampling techniques

The literature reviewed in Chapter 2 does not reveal a set number of samples that have been be gathered in similar research that can be expected to give statistically significant results. However, it is possible to use methods described in the statistics literature to estimate the number of samples required based on the results described in Chapter 3. Based on the results presented in Taylor and Nguyen (2003), the generic classifiers can be expected to correctly classify approximately 90% of movements for disabled users, or approximately 98% for able-bodied users.

It is reasonable to expect that the actual performance of the classifiers is independent of the process used to measure the performance, and as stated by Bland (2000) when two random variables are added, "the mean of the sum is the sum of the means, and if the two variables are independent, the variance of the sum is the sum of their variances". Consequentially, the total variance observed can be expressed as:

$$\sigma_{total}^2 = \sigma_{sampling}^2 + \sigma_{classifier}^2$$

Equation 50

There are three sample sizes that are relevant in this section:

- The number of samples in the test set of each classifier
- The number of different classifiers
- The number of subjects

Number of movements in the test set of each classifier

The critical issue in selecting the number of movements in each test set is that there must be sufficient samples in each set to estimate the statistics about the performance metrics with sufficient precision and reduce the variance introduced by the sampling process. If the sample size is insufficient, then the resolution at which these statistics can be estimated will be too coarse and the error due to the sampling process will be too great.

Walpole and Myers (1993) state that the variance of a point estimate of a proportion, p, for sufficiently large sample size, n, is given by

$$\sigma^2 = \frac{p(1-p)}{n}$$

Equation 51

Walpole and Myers (1993) further states that the (1-α) 100% confidence interval for a point estimate, $\hat{p}$, of a proportion, p, can then be expressed as

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Equation 52

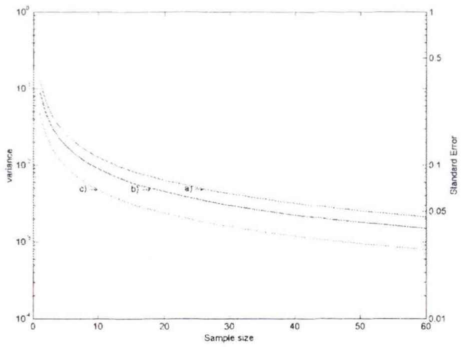

Figure 72 Variance and Standard Error caused by sampling, where a) p = 0.85, b) p = 0.9, c) p = 0.95, as estimated by Equation 51

From the results presented in Chapter 3, an accuracy of approximately 90% is expected for generic classifiers. Using Equation 51 with a sample size of 36, this leads to an estimated variance of 0.0025, and corresponding standard error of 0.05.

Since the variance of the classifier accuracy is expected to be in the range from 0.01 to 0.1, this level of variance from the sampling process would be acceptable.

Number of different classifier pairs

In determining the number of classifier pairs required in order to be able to expect significant results, the approach taken was to estimate number of samples required for a comparison of means using methods described widely in the statistics literature. According to both Walpole, Myers and Myers (1998) and Fisher and Van Belle (1993), the number of different classifier pairs required for the testing means can be estimated by:

σ^2, the variance of the population

α, the probability of a Type I error, that is, rejecting the null hypothesis when it is actually true. The significance level is equal to $1-\alpha$.

β, the probability of a Type II error, that is, accepting the null hypothesis when it is actually false. The power of the test is $1-\beta$.

$\mu-\mu_0$, the magnitude of the difference in means to be detected.

In this case, μ and σ^2 are the mean and variance of the difference between the paired classifiers. Fisher and Van Belle (1993) and Walpole, Myers and Myers (1998) propose the following equations to estimate the sample size required for a one-sample test or paired-sample test.

Fisher and Van Belle (1993) state that sample sizes are calculated as a function of the standardised difference between the population and the hypothesised population. For a one-sample test, this standardised difference is defined by Equation 53. Extending this to the paired sample case, μ is taken to be the expected difference between the members of each pair and μ_0 is the expected difference between each pair under the null hypothesis.

The estimated sample size required for a two-sided test, as stated by Fisher and Van Belle (1993), is shown in Equation 54. Equation 55 shows the estimated sample size required for a single-sided test.

$$\Delta = \frac{|\mu - \mu_0|}{\sigma}$$

Equation 53

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2}$$

Equation 54

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{\Delta^2}$$

Equation 55

Although the equations are expressed in different formats in the two references, it can be shown that the two forms are mathematically equivalent. Fisher and Van Belle (1993) note that where the variance is not known, that it can be estimated using a previous study or example. Fisher and Van Belle continue, stating that there are no explicit formulae that apply in this case, but that it is widely recommended to add between 2 and 4 observations to each of the groups.

Estimates of the magnitude of a difference that can be detected as significant for a given number of samples are shown in Figure 73 to Figure 75 below. The estimates of the sample size required to detect a difference between several absolute accuracies are displayed in Figure 76 through to Figure 78.

Figure 73 shows that to detect a difference of 5%, with a significance level and power of 95%, a sample size of approximately 217 pairs is necessary if the variance is 0.05. It also shows that the variance of the data has a large effect on the sample size required. For example, reducing the variance to 0.01 reduces the sample size required by an order of magnitude to approximately 43 pairs. A contrasting example is that increasing the variance to 0.1 leads to an increase in the estimated sample size of 433 pairs. Figure 74 and Figure 75 show the effect of reducing the power and significance level of the test. These three figures also highlight the diminishing return obtained by obtaining larger samples, in terms of the increase in resolution obtained for an increase in sample size. Figure 76, Figure 77 and Figure 78 are included to allow comparison with the corresponding figures for the comparison of proportions.

The number of classifier pairs used is limited by the availability of computing resources rather than the ability to obtain data from human sources, so this sample size may be varied at a later stage if the variance is found to be larger than expected. With these issues considered, it appears that a sample size of 250 classifier pairs would be appropriate.

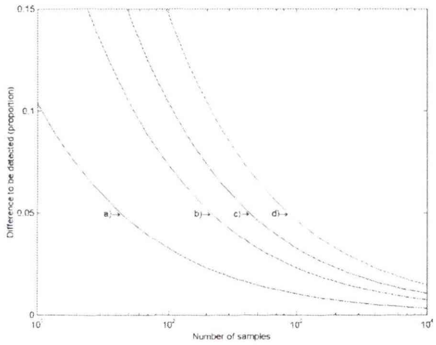

Figure 73 Difference detectable for given sample size (number of classifier pairs), estimated by Equation 54, where $\alpha = 0.05$, $\beta = 0.05$ and a) $\sigma^2 = 0.01$, b) $\sigma^2 = 0.05$, c) $\sigma^2 = 0.1$, d) $\sigma^2 = 0.2$

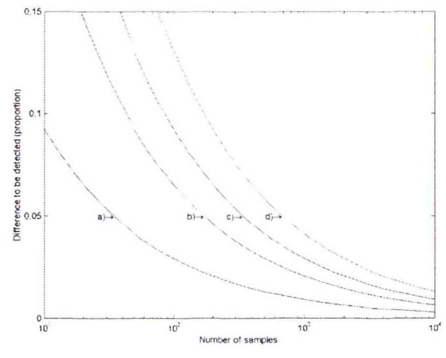

Figure 74 Difference detectable for given sample size (number of classifier pairs), estimated by Equation 54, where $\alpha = 0.05$, $\beta = 0.10$ and a) $\sigma^2 = 0.01$, b) $\sigma^2 = 0.05$, c) $\sigma^2 = 0.1$, d) $\sigma^2 = 0.2$

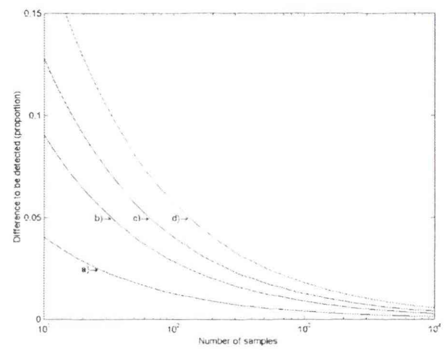

Figure 75 Difference detectable for given sample size (number of classifier pairs), estimated by Equation 54, where $\alpha = 0.10$, $\beta = 0.50$ and a) $\sigma^2 = 0.01$, b) $\sigma^2 = 0.05$, c) $\sigma^2 = 0.1$, d) $\sigma^2 = 0.2$

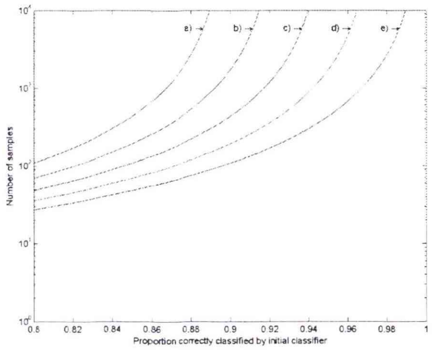

Figure 76 Estimates of the number of classifier pairs required for the testing of means for $\alpha = 0.05$, $\beta = 0.05$, $\sigma^2 = 0.1$ and mean accuracy a) 0.9, b) 0.925, c) 0.95, d) 0.975, e) 1.0, estimated by Equation 54

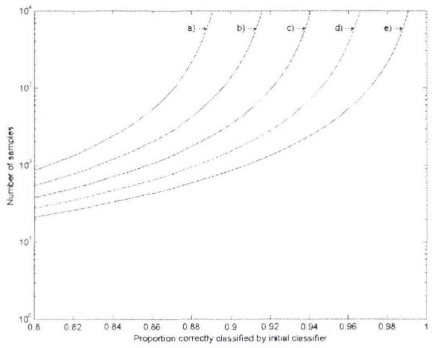

Figure 77 Estimates of the number of classifier pairs required for the testing of means for $\alpha = 0.05$, $\beta = 0.10$, $\sigma^2 = 0.1$ and mean accuracy a) 0.9, b) 0.925, c) 0.95, d) 0.975, e) 1.0, estimated by Equation 54

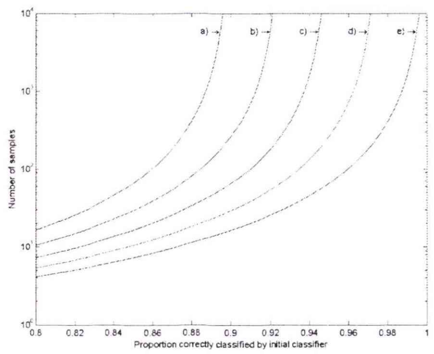

Figure 78 Estimates of the number of classifier pairs required for the testing of means for $\alpha = 0.10$, $\beta = 0.50$, $\sigma^2 = 0.1$ and mean accuracy a) 0.9, b) 0.925, c) 0.95, d) 0.975, e) 1.0, estimated by Equation 54

Number of Subjects

The objective for the experiments presented in this thesis is not to draw statistically significant conclusions about the entire population of potential wheelchair users. Rather, it is intended to show that there are some cases where statistically significant changes in classifier performance can be attributed to the parameters examined in the experiments. In similar research that has been published recently, the number of subjects involved in experiments has varied from one subject (Hansen, Haugland & Sinkjaer 2004; Kuno, Murashima et al. 2000; Kuno, Murashina et al. 2000; Law et al. 2002; Thieffry et al. 2003) to as many as 7 disabled subjects (Y.-L. Chen, Chen & Lai 2002) or 11 subjects of mixed ability levels (5 able-bodied, 3 manual wheelchair users and 2 powered wheelchair users) (Kuno, Shimada & Shirai 2003). Yom-Tov and Inbar (2002) preformed experiments with 5 able-bodied subjects, while Barea et al. (2002) involved 5 subjects who were powered wheelchair users but with impairment levels slightly below that at for which the device being tested had been designed.

The main limiting factors on the number of subjects involved in the experiments in these chapters are the availability of suitable volunteers and resources required for the collection, labelling and processing of data for each subject. Considering the sample sizes used in the literature, it would appear reasonable to use a sample size of between two and six subjects.

Appendix E. UTS Human Research Ethics Committee Approval

Research & Development Office
PO Box 123
Broadway NSW 2007
Australia
Fax +61 2 9514 1244
www.uts.edu.au/research/index.html
UTS CRICOS Provider Code: 00099F

U|T|S
University of Technology, Sydney

10 May 2004

Professor Hung Nguyen
Faculty of Engineering
KURC in Health Technology
Level 5, Building 2
Broadway Campus

Dear Hung

UTS HREC 2004-039 - NGUYEN, Professor Hung, CRAIG, Professor Ashley (for TAYLOR, Mr Philip - PhD student) – "Integration of powered wheelchairs and environmental control systems (ECS) by head movement technology"

Thank you for your response to my letter dated 27 April 2004. Your response satisfactorily addresses the concerns and questions raised by the Committee, and I am pleased to inform you that ethics clearance is now granted.

Your clearance number is UTS HREC 2004-039A.

Please note that the ethical conduct of research is an on-going process. The *National Statement on Ethical Conduct in Research Involving Humans* requires us to obtain a report about the progress of the research, and in particular about any changes to the research which may have ethical implications. This report form must be completed at least annually, and at the end of the project (if it takes more than a year), or in the event of any changes to the research as referred to above, in which case the HREC Secretariat should be contacted beforehand. The Ethics Secretariat will contact you when it is time to complete your first report.

I also refer you to the AVCC guidelines relating to the storage of data, which require that data be kept for a minimum of 5 years after publication of research. However, in NSW, longer retention requirements are required for research on human subjects with potential long-term effects, research with long-term environmental effects, or research considered of national or international significance, importance, or controversy. If the data from this research project falls into one of these categories, contact University Records for advice on long-term retention.

If you have any queries about your ethics clearance, or require any amendments to your research in the future, please do not hesitate to contact the Ethics Secretariat at the Research and Commercialisation Office, on 02 9514 9615.

Yours sincerely,

Production Note:
Signature removed prior to publication.
Associate Professor Jane Stein-Parbury
Chairperson, UTS Human Research Ethics Committee

Office: City campus, No 1 Broadway, Sydney NSW

UNIVERSITY OF TECHNOLOGY, SYDNEY
CONSENT FORM

I ______________________ agree to participate in the research project "Integration of powered wheelchairs and environmental control systems (ECS) by head movement technology", being conducted by Dr. Hung Nguyen (B2, Level 5, ph: 9514 2451), Dr. Ashley Craig (B8, Level 2, ph: 9514 1358) and Mr. Philip Taylor (B1, Level 24, ph: 9514 2317) of the University of Technology, Sydney as part of a doctoral research project.

I understand that the purpose of this study is develop a single wireless control device to operate both powered wheelchairs as well as the immediate home environment (turning on/off lights, appliances, opening/closing of doors/windows, etc...).

I understand that my participation in this research will involve a single session of no more than 2 hours during which I will be seated in a powered wheelchair and instrumented with various monitors (all externally and painlessly attached). I will be asked to follow a computer prompt and perform appropriate actions by moving my head slightly (as in nodding).

I understand that the data collected will be stored and used by the listed researchers and others in the research group to develop and test systems for movement recognition.

I have been informed that I may experience some minor discomfort associated with the wearing of the sensors, as well as some mild fatigue from performance of the head movement protocol.

I am aware that I can contact Dr. Hung Nguyen, Dr. Ashley Craig or Mr. Philip Taylor if I have any concerns about the research. I also understand that I am free to withdraw my participation from this research project at any time I wish and without giving a reason.

I agree that all my questions have been answered fully and clearly.

I agree that the research data gathered from this project may be published in a form that does not identify me in any way.

______________________________ ____/____/____

Signed by

______________________________ ____/____/____

Witnessed by

NOTE:

This study has been approved by the University of Technology, Sydney Human Research Ethics Committee. If you have any complaints or reservations about any aspect of your participation in this research which you cannot resolve with the researcher, you may contact the Ethics Committee through the Research Ethics Officer, Ms Susanna Davis (ph: 02 - 9514 1279, Susanna.Davis@uts.edu.au). Any complaint you make will be treated in confidence and investigated fully and you will be informed of the outcome.

Appendix F. Friedman's 2-way analysis of variance by ranks of experimental results

The tables in this appendix pertain to the Friedman's 2-way analysis of variance by ranks for the experiments performed in Chapter 4 and Chapter 5.The P value in the final column gives the probability that the null hypothesis is true, given the observed data.

Table 17 Friedman's 2-way analysis of variance by ranks of the validation set real-time classification error rate as network architecture increases in size

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H_0 \mid \chi^2)$
Treatments	3434.43	28	122.658	49.2383	0.00785547
Error	191869	2772	69.2168		
Total	195304	2899			

Table 18 Friedman's 2-way analysis of variance by ranks on the training time as network architecture increases in size

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H_0 \mid \chi^2)$
Treatments	200864	28	7173.73	2770.54	0
Error	2135.5	2772	0.770382		
Total	203000	2899			

Table 19 Friedman's 2-way analysis of variance by ranks of the validation set real-time classification error rate for ANN trained with different algorithms

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H_0 \mid \chi^2)$
Treatments	140.54	2	70.27	140.54	0
Error	59.46	198	0.300303		
Total	200	299			

Table 20 Friedman's 2-way analysis of variance by ranks of the training time of ANN classifiers trained with different algorithms

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H_0 \mid \chi^2)$
Treatments	163.52	2	81.76	163.52	0
Error	36.48	198	0.184242		
Total	200	299			

Table 21 Friedman's 2-way analysis of variance by ranks of the validation set real-time classification error rate as training set increases in size

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H_0 \mid \chi^2)$
Treatments	4.535	2	2.2675	4.62755	0.0988872
Error	191.465	198	0.966995		
Total	196	299			

Table 22 Friedman's 2-way analysis of variance by ranks on the training time of ANN classifiers as training set increases in size

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	73.34	2	36.67	73.34	1.11022e-16
Error	126.66	198	0.639697		
Total	200	299			

Table 23 Friedman's 2-way analysis of variance by ranks of the Subject 8 test set real-time classification error rate as network architecture increases in size

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	3156.07	28	112.717	43.7731	0.0292561
Error	198726	2772	71.6905		
Total	201882	2899			

Table 24 Friedman's 2-way analysis of variance by ranks of the Subject 8 test set real-time classification error rate for ANN trained with different algorithms

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	123.005	2	61.5025	123.935	0
Error	75.495	198	0.381288		
Total	198.5	299			

Table 25 Friedman's 2-way analysis of variance by ranks of the Subject 8 test set real-time classification error rate as training set increases in size

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H_0 \mid \chi^2)$
Treatments	7.485	2	3.7425	7.59898	0.0223821
Error	189.515	198	0.957146		
Total	197	299			

Table 26 Friedman's 2-way analysis of variance by ranks of the Subject 9 test set real-time classification error rate as network architecture increases in size

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H_0 \mid \chi^2)$
Treatments	3079.97	28	109.999	43.9214	0.028287
Error	193269	2772	69.7219		
Total	196349	2899			

Table 27 Friedman's 2-way analysis of variance by ranks of the Subject 9 test set real-time classification error rate for ANN trained with different algorithms

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H_0 \mid \chi^2)$
Treatments	128.165	2	64.0825	129.134	0
Error	70.335	198	0.355227		
Total	198.5	299			

Table 28 Friedman's 2-way analysis of variance by ranks of the Subject 9 test set real-time classification error rate as training set increases in size

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	6.26	2	3.13	6.40409	0.0406789
Error	189.24	198	0.955758		
Total	195.5	299			

Table 29 Friedman's 2-way analysis of variance by ranks of the validation set real-time classification error rate for classifiers retrained with user specific data from Subject 8

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	271.085	3	90.3617	164.627	0
Error	222.915	297	0.750556		
Total	494	399			

Table 30 Friedman's 2-way analysis of variance by ranks of the Subject 8 test set real-time classification error rate for classifiers retrained with user specific data from Subject 8

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	276.565	3	92.1883	167.277	0
Error	219.435	297	0.738838		
Total	496	399			

Table 31 Friedman's 2-way analysis of variance by ranks of the Subject 8 test set real-time classification error rate for classifiers retrained with user specific data from Subject 8

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	283.635	3	94.545	171.553	0
Error	212.365	297	0.715034		
Total	496	399			

Table 32 Friedman's 2-way analysis of variance by ranks of the retraining time for classifiers retrained with user specific data from Subject 8

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	274.22	3	91.4067	164.532	0
Error	225.78	297	0.760202		
Total	500	399			

Table 33 Friedman's 2-way analysis of variance by ranks of the validation set real-time classification error rate for classifiers retrained with user specific data from Subject 9

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	255.375	3	85.125	156.512	0
Error	234.125	297	0.7883		
Total	489.5	399			

Table 34 Friedman's 2-way analysis of variance by ranks of the Subject 8 test set real-time classification error rate for classifiers retrained with user specific data from Subject 9

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	148.68	3	49.56	90.383	0
Error	344.82	297	1.16101		
Total	493.5	399			

Table 35 Friedman's 2-way analysis of variance by ranks of the Subject 9 test set real-time classification error rate for classifiers retrained with user specific data from Subject 9

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	67.445	3	22.4817	40.9585	6.67308e-09
Error	426.555	297	1.43621		
Total	494	399			

Table 36 Friedman's 2-way analysis of variance by ranks of the retraining time for classifiers retrained with user specific data from Subject 9

Source of variability	Sum of Squares, S	Degrees of freedom, n	Mean Squares, S/n	Friedman's χ^2 statistic	P-value, $P(H^0 \mid \chi^2)$
Treatments	232.2	3	77.4	139.32	0
Error	267.8	297	0.901684		
Total	500	399			

Appendix G. Pairwise comparison of ranks between treatments

Table 37 Pairs of treatments with different ANN architectures having statistically significant differences in validation set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
2	7	2.780	0.465	5.095
2	8	2.460	0.145	4.775
2	9	3.720	1.405	6.035
2	10	3.410	1.095	5.725
2	11	3.335	1.020	5.650
2	12	3.365	1.050	5.680
2	13	3.365	1.050	5.680
2	14	4.110	1.795	6.425
2	15	3.060	0.745	5.375
2	16	3.670	1.355	5.985
2	17	3.200	0.885	5.515
2	18	3.300	0.985	5.615
2	19	3.020	0.705	5.335
2	20	3.790	1.475	6.105
2	21	4.090	1.775	6.405
2	22	4.140	1.825	6.455
2	23	4.615	2.300	6.930
2	24	2.990	0.675	5.305
2	25	4.140	1.825	6.455
2	26	4.055	1.740	6.370
2	27	4.105	1.790	6.420
2	28	4.215	1.900	6.530
2	29	5.540	3.225	7.855
2	30	4.180	1.865	6.495
3	14	2.570	0.255	4.885
3	21	2.550	0.235	4.865
3	22	2.600	0.285	4.915
3	23	3.075	0.760	5.390
3	25	2.600	0.285	4.915
3	26	2.515	0.200	4.830
3	27	2.565	0.250	4.880
3	28	2.675	0.360	4.990
3	29	4.000	1.685	6.315
3	30	2.640	0.325	4.955
4	14	2.580	0.265	4.895
4	21	2.560	0.245	4.875
4	22	2.610	0.295	4.925
4	23	3.085	0.770	5.400
4	25	2.610	0.295	4.925
4	26	2.525	0.210	4.840
4	27	2.575	0.260	4.890
4	28	2.685	0.370	5.000

4	29	4.010	1.695	6.325
4	30	2.650	0.335	4.965
5	23	2.765	0.450	5.080
5	28	2.365	0.050	4.680
5	29	3.690	1.375	6.005
5	30	2.330	0.015	4.645
6	23	2.345	0.030	4.660
6	29	3.270	0.955	5.585
7	29	2.760	0.445	5.075
8	29	3.080	0.765	5.395
15	29	2.480	0.165	4.795
17	29	2.340	0.025	4.655
19	29	2.520	0.205	4.835
24	29	2.550	0.235	4.865

Table 38 Pairs of treatments trained using different algorithms having statistically significant differences in validation set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
DR	SCG	1.610	1.333	1.887
DR	LM	1.210	0.933	1.487
SCG	LM	-0.400	-0.677	-0.123

Table 39 Pairs of treatments trained with different training set size having statistically significant differences in validation set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
x1	x1.5	-0.295	-0.569	-0.021

Table 40 Pairs of treatments with different ANN architectures having statistically significant differences in training time

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
2	5	-2.630	-4.990	-0.270
2	6	-3.850	-6.210	-1.490
2	7	-4.810	-7.170	-2.450
2	8	-5.750	-8.110	-3.390

2	9	-6.830	-9.190	-4.470
2	10	-8.010	-10.370	-5.650
2	11	-8.630	-10.990	-6.270
2	12	-9.560	-11.920	-7.200
2	13	-10.790	-13.150	-8.430
2	14	-11.930	-14.290	-9.570
2	15	-12.650	-15.010	-10.290
2	16	-13.610	-15.970	-11.250
2	17	-14.790	-17.150	-12.430
2	18	-15.810	-18.170	-13.450
2	19	-16.590	-18.950	-14.230
2	20	-17.900	-20.260	-15.540
2	21	-18.600	-20.960	-16.240
2	22	-19.710	-22.070	-17.350
2	23	-20.930	-23.290	-18.570
2	24	-21.620	-23.980	-19.260
2	25	-22.700	-25.060	-20.340
2	26	-23.730	-26.090	-21.370
2	27	-24.870	-27.230	-22.510
2	28	-25.740	-28.100	-23.380
2	29	-26.670	-29.030	-24.310
2	30	-27.740	-30.100	-25.380
3	7	-3.130	-5.490	-0.770
3	8	-4.070	-6.430	-1.710
3	9	-5.150	-7.510	-2.790
3	10	-6.330	-8.690	-3.970
3	11	-6.950	-9.310	-4.590
3	12	-7.880	-10.240	-5.520
3	13	-9.110	-11.470	-6.750
3	14	-10.250	-12.610	-7.890
3	15	-10.970	-13.330	-8.610
3	16	-11.930	-14.290	-9.570
3	17	-13.110	-15.470	-10.750
3	18	-14.130	-16.490	-11.770
3	19	-14.910	-17.270	-12.550
3	20	-16.220	-18.580	-13.860
3	21	-16.920	-19.280	-14.560
3	22	-18.030	-20.390	-15.670
3	23	-19.250	-21.610	-16.890
3	24	-19.940	-22.300	-17.580
3	25	-21.020	-23.380	-18.660
3	26	-22.050	-24.410	-19.690
3	27	-23.190	-25.550	-20.830
3	28	-24.060	-26.420	-21.700
3	29	-24.990	-27.350	-22.630
3	30	-26.060	-28.420	-23.700
4	6	-2.650	-5.010	-0.290
4	7	-3.610	-5.970	-1.250
4	8	-4.550	-6.910	-2.190
4	9	-5.630	-7.990	-3.270
4	10	-6.810	-9.170	-4.450
4	11	-7.430	-9.790	-5.070
4	12	-8.360	-10.720	-6.000
4	13	-9.590	-11.950	-7.230
4	14	-10.730	-13.090	-8.370

4	15	-11.450	-13.810	-9.090
4	16	-12.410	-14.770	-10.050
4	17	-13.590	-15.950	-11.230
4	18	-14.610	-16.970	-12.250
4	19	-15.390	-17.750	-13.030
4	20	-16.700	-19.060	-14.340
4	21	-17.400	-19.760	-15.040
4	22	-18.510	-20.870	-16.150
4	23	-19.730	-22.090	-17.370
4	24	-20.420	-22.780	-18.060
4	25	-21.500	-23.860	-19.140
4	26	-22.530	-24.890	-20.170
4	27	-23.670	-26.030	-21.310
4	28	-24.540	-26.900	-22.180
4	29	-25.470	-27.830	-23.110
4	30	-26.540	-28.900	-24.180
5	8	-3.120	-5.480	-0.760
5	9	-4.200	-6.560	-1.840
5	10	-5.380	-7.740	-3.020
5	11	-6.000	-8.360	-3.640
5	12	-6.930	-9.290	-4.570
5	13	-8.160	-10.520	-5.800
5	14	-9.300	-11.660	-6.940
5	15	-10.020	-12.380	-7.660
5	16	-10.980	-13.340	-8.620
5	17	-12.160	-14.520	-9.800
5	18	-13.180	-15.540	-10.820
5	19	-13.960	-16.320	-11.600
5	20	-15.270	-17.630	-12.910
5	21	-15.970	-18.330	-13.610
5	22	-17.080	-19.440	-14.720
5	23	-18.300	-20.660	-15.940
5	24	-18.990	-21.350	-16.630
5	25	-20.070	-22.430	-17.710
5	26	-21.100	-23.460	-18.740
5	27	-22.240	-24.600	-19.880
5	28	-23.110	-25.470	-20.750
5	29	-24.040	-26.400	-21.680
5	30	-25.110	-27.470	-22.750
6	9	-2.980	-5.340	-0.620
6	10	-4.160	-6.520	-1.800
6	11	-4.780	-7.140	-2.420
6	12	-5.710	-8.070	-3.350
6	13	-6.940	-9.300	-4.580
6	14	-8.080	-10.440	-5.720
6	15	-8.800	-11.160	-6.440
6	16	-9.760	-12.120	-7.400
6	17	-10.940	-13.300	-8.580
6	18	-11.960	-14.320	-9.600
6	19	-12.740	-15.100	-10.380
6	20	-14.050	-16.410	-11.690
6	21	-14.750	-17.110	-12.390
6	22	-15.860	-18.220	-13.500
6	23	-17.080	-19.440	-14.720
6	24	-17.770	-20.130	-15.410

6	25	-18.850	-21.210	-16.490
6	26	-19.880	-22.240	-17.520
6	27	-21.020	-23.380	-18.660
6	28	-21.890	-24.250	-19.530
6	29	-22.820	-25.180	-20.460
6	30	-23.890	-26.250	-21.530
7	10	-3.200	-5.560	-0.840
7	11	-3.820	-6.180	-1.460
7	12	-4.750	-7.110	-2.390
7	13	-5.980	-8.340	-3.620
7	14	-7.120	-9.480	-4.760
7	15	-7.840	-10.200	-5.480
7	16	-8.800	-11.160	-6.440
7	17	-9.980	-12.340	-7.620
7	18	-11.000	-13.360	-8.640
7	19	-11.780	-14.140	-9.420
7	20	-13.090	-15.450	-10.730
7	21	-13.790	-16.150	-11.430
7	22	-14.900	-17.260	-12.540
7	23	-16.120	-18.480	-13.760
7	24	-16.810	-19.170	-14.450
7	25	-17.890	-20.250	-15.530
7	26	-18.920	-21.280	-16.560
7	27	-20.060	-22.420	-17.700
7	28	-20.930	-23.290	-18.570
7	29	-21.860	-24.220	-19.500
7	30	-22.930	-25.290	-20.570
8	11	-2.880	-5.240	-0.520
8	12	-3.810	-6.170	-1.450
8	13	-5.040	-7.400	-2.680
8	14	-6.180	-8.540	-3.820
8	15	-6.900	-9.260	-4.540
8	16	-7.860	-10.220	-5.500
8	17	-9.040	-11.400	-6.680
8	18	-10.060	-12.420	-7.700
8	19	-10.840	-13.200	-8.480
8	20	-12.150	-14.510	-9.790
8	21	-12.850	-15.210	-10.490
8	22	-13.960	-16.320	-11.600
8	23	-15.180	-17.540	-12.820
8	24	-15.870	-18.230	-13.510
8	25	-16.950	-19.310	-14.590
8	26	-17.980	-20.340	-15.620
8	27	-19.120	-21.480	-16.760
8	28	-19.990	-22.350	-17.630
8	29	-20.920	-23.280	-18.560
8	30	-21.990	-24.350	-19.630
9	12	-2.730	-5.090	-0.370
9	13	-3.960	-6.320	-1.600
9	14	-5.100	-7.460	-2.740
9	15	-5.820	-8.180	-3.460
9	16	-6.780	-9.140	-4.420
9	17	-7.960	-10.320	-5.600
9	18	-8.980	-11.340	-6.620
9	19	-9.760	-12.120	-7.400

9	20	-11.070	-13.430	-8.710
9	21	-11.770	-14.130	-9.410
9	22	-12.880	-15.240	-10.520
9	23	-14.100	-16.460	-11.740
9	24	-14.790	-17.150	-12.430
9	25	-15.870	-18.230	-13.510
9	26	-16.900	-19.260	-14.540
9	27	-18.040	-20.400	-15.680
9	28	-18.910	-21.270	-16.550
9	29	-19.840	-22.200	-17.480
9	30	-20.910	-23.270	-18.550
10	13	-2.780	-5.140	-0.420
10	14	-3.920	-6.280	-1.560
10	15	-4.640	-7.000	-2.280
10	16	-5.600	-7.960	-3.240
10	17	-6.780	-9.140	-4.420
10	18	-7.800	-10.160	-5.440
10	19	-8.580	-10.940	-6.220
10	20	-9.890	-12.250	-7.530
10	21	-10.590	-12.950	-8.230
10	22	-11.700	-14.060	-9.340
10	23	-12.920	-15.280	-10.560
10	24	-13.610	-15.970	-11.250
10	25	-14.690	-17.050	-12.330
10	26	-15.720	-18.080	-13.360
10	27	-16.860	-19.220	-14.500
10	28	-17.730	-20.090	-15.370
10	29	-18.660	-21.020	-16.300
10	30	-19.730	-22.090	-17.370
11	14	-3.300	-5.660	-0.940
11	15	-4.020	-6.380	-1.660
11	16	-4.980	-7.340	-2.620
11	17	-6.160	-8.520	-3.800
11	18	-7.180	-9.540	-4.820
11	19	-7.960	-10.320	-5.600
11	20	-9.270	-11.630	-6.910
11	21	-9.970	-12.330	-7.610
11	22	-11.080	-13.440	-8.720
11	23	-12.300	-14.660	-9.940
11	24	-12.990	-15.350	-10.630
11	25	-14.070	-16.430	-11.710
11	26	-15.100	-17.460	-12.740
11	27	-16.240	-18.600	-13.880
11	28	-17.110	-19.470	-14.750
11	29	-18.040	-20.400	-15.680
11	30	-19.110	-21.470	-16.750
12	14	-2.370	-4.730	-0.010
12	15	-3.090	-5.450	-0.730
12	16	-4.050	-6.410	-1.690
12	17	-5.230	-7.590	-2.870
12	18	-6.250	-8.610	-3.890
12	19	-7.030	-9.390	-4.670
12	20	-8.340	-10.700	-5.980
12	21	-9.040	-11.400	-6.680
12	22	-10.150	-12.510	-7.790

12	23	-11.370	-13.730	-9.010
12	24	-12.060	-14.420	-9.700
12	25	-13.140	-15.500	-10.780
12	26	-14.170	-16.530	-11.810
12	27	-15.310	-17.670	-12.950
12	28	-16.180	-18.540	-13.820
12	29	-17.110	-19.470	-14.750
12	30	-18.180	-20.540	-15.820
13	16	-2.820	-5.180	-0.460
13	17	-4.000	-6.360	-1.640
13	18	-5.020	-7.380	-2.660
13	19	-5.800	-8.160	-3.440
13	20	-7.110	-9.470	-4.750
13	21	-7.810	-10.170	-5.450
13	22	-8.920	-11.280	-6.560
13	23	-10.140	-12.500	-7.780
13	24	-10.830	-13.190	-8.470
13	25	-11.910	-14.270	-9.550
13	26	-12.940	-15.300	-10.580
13	27	-14.080	-16.440	-11.720
13	28	-14.950	-17.310	-12.590
13	29	-15.880	-18.240	-13.520
13	30	-16.950	-19.310	-14.590
14	17	-2.860	-5.220	-0.500
14	18	-3.880	-6.240	-1.520
14	19	-4.660	-7.020	-2.300
14	20	-5.970	-8.330	-3.610
14	21	-6.670	-9.030	-4.310
14	22	-7.780	-10.140	-5.420
14	23	-9.000	-11.360	-6.640
14	24	-9.690	-12.050	-7.330
14	25	-10.770	-13.130	-8.410
14	26	-11.800	-14.160	-9.440
14	27	-12.940	-15.300	-10.580
14	28	-13.810	-16.170	-11.450
14	29	-14.740	-17.100	-12.380
14	30	-15.810	-18.170	-13.450
15	18	-3.160	-5.520	-0.800
15	19	-3.940	-6.300	-1.580
15	20	-5.250	-7.610	-2.890
15	21	-5.950	-8.310	-3.590
15	22	-7.060	-9.420	-4.700
15	23	-8.280	-10.640	-5.920
15	24	-8.970	-11.330	-6.610
15	25	-10.050	-12.410	-7.690
15	26	-11.080	-13.440	-8.720
15	27	-12.220	-14.580	-9.860
15	28	-13.090	-15.450	-10.730
15	29	-14.020	-16.380	-11.660
15	30	-15.090	-17.450	-12.730
16	19	-2.980	-5.340	-0.620
16	20	-4.290	-6.650	-1.930
16	21	-4.990	-7.350	-2.630
16	22	-6.100	-8.460	-3.740
16	23	-7.320	-9.680	-4.960

16	24	-8.010	-10.370	-5.650
16	25	-9.090	-11.450	-6.730
16	26	-10.120	-12.480	-7.760
16	27	-11.260	-13.620	-8.900
16	28	-12.130	-14.490	-9.770
16	29	-13.060	-15.420	-10.700
16	30	-14.130	-16.490	-11.770
17	20	-3.110	-5.470	-0.750
17	21	-3.810	-6.170	-1.450
17	22	-4.920	-7.280	-2.560
17	23	-6.140	-8.500	-3.780
17	24	-6.830	-9.190	-4.470
17	25	-7.910	-10.270	-5.550
17	26	-8.940	-11.300	-6.580
17	27	-10.080	-12.440	-7.720
17	28	-10.950	-13.310	-8.590
17	29	-11.880	-14.240	-9.520
17	30	-12.950	-15.310	-10.590
18	21	-2.790	-5.150	-0.430
18	22	-3.900	-6.260	-1.540
18	23	-5.120	-7.480	-2.760
18	24	-5.810	-8.170	-3.450
18	25	-6.890	-9.250	-4.530
18	26	-7.920	-10.280	-5.560
18	27	-9.060	-11.420	-6.700
18	28	-9.930	-12.290	-7.570
18	29	-10.860	-13.220	-8.500
18	30	-11.930	-14.290	-9.570
19	22	-3.120	-5.480	-0.760
19	23	-4.340	-6.700	-1.980
19	24	-5.030	-7.390	-2.670
19	25	-6.110	-8.470	-3.750
19	26	-7.140	-9.500	-4.780
19	27	-8.280	-10.640	-5.920
19	28	-9.150	-11.510	-6.790
19	29	-10.080	-12.440	-7.720
19	30	-11.150	-13.510	-8.790
20	23	-3.030	-5.390	-0.670
20	24	-3.720	-6.080	-1.360
20	25	-4.800	-7.160	-2.440
20	26	-5.830	-8.190	-3.470
20	27	-6.970	-9.330	-4.610
20	28	-7.840	-10.200	-5.480
20	29	-8.770	-11.130	-6.410
20	30	-9.840	-12.200	-7.480
21	24	-3.020	-5.380	-0.660
21	25	-4.100	-6.460	-1.740
21	26	-5.130	-7.490	-2.770
21	27	-6.270	-8.630	-3.910
21	28	-7.140	-9.500	-4.780
21	29	-8.070	-10.430	-5.710
21	30	-9.140	-11.500	-6.780
22	25	-2.990	-5.350	-0.630
22	26	-4.020	-6.380	-1.660
22	27	-5.160	-7.520	-2.800

22	28	-6.030	-8.390	-3.670
22	29	-6.960	-9.320	-4.600
22	30	-8.030	-10.390	-5.670
23	26	-2.800	-5.160	-0.440
23	27	-3.940	-6.300	-1.580
23	28	-4.810	-7.170	-2.450
23	29	-5.740	-8.100	-3.380
23	30	-6.810	-9.170	-4.450
24	27	-3.250	-5.610	-0.890
24	28	-4.120	-6.480	-1.760
24	29	-5.050	-7.410	-2.690
24	30	-6.120	-8.480	-3.760
25	28	-3.040	-5.400	-0.680
25	29	-3.970	-6.330	-1.610
25	30	-5.040	-7.400	-2.680
26	29	-2.940	-5.300	-0.580
26	30	-4.010	-6.370	-1.650
27	30	-2.870	-5.230	-0.510

Table 41 Pairs of treatments trained using different algorithms having statistically significant differences in training time

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
DR	SCG	0.520	0.243	0.797
DR	LM	-1.240	-1.517	-0.963
SCG	LM	-1.760	-2.037	-1.483

Table 42 Pairs of treatments trained with different training set size having statistically significant differences in training time

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
x1	x1.5	-0.650	-0.927	-0.373
x1	x2	-1.210	-1.487	-0.933
x1.5	x2	-0.560	-0.837	-0.283

Table 43 Pairs of treatments with different ANN architectures having statistically significant differences in Subject 8 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
2	5	2.785	0.431	5.139
2	6	2.500	0.146	4.854
2	7	3.390	1.036	5.744
2	9	2.480	0.126	4.834
2	10	2.560	0.206	4.914
2	11	2.860	0.506	5.214
2	12	3.625	1.271	5.979
2	13	2.405	0.051	4.759
2	26	2.460	0.106	4.814
2	29	2.620	0.266	4.974
3	7	2.880	0.526	5.234
3	12	3.115	0.761	5.469
4	19	-2.390	-4.744	-0.036
5	19	-3.565	-5.919	-1.211
5	20	-2.745	-5.099	-0.391
6	19	-3.280	-5.634	-0.926
6	20	-2.460	-4.814	-0.106
7	18	-2.425	-4.779	-0.071
7	19	-4.170	-6.524	-1.816
7	20	-3.350	-5.704	-0.996
7	21	-2.850	-5.204	-0.496
7	22	-2.895	-5.249	-0.541
7	23	-2.730	-5.084	-0.376
8	19	-2.600	-4.954	-0.246
9	19	-3.260	-5.614	-0.906
9	20	-2.440	-4.794	-0.086
10	19	-3.340	-5.694	-0.986
10	20	-2.520	-4.874	-0.166
11	19	-3.640	-5.994	-1.286
11	20	-2.820	-5.174	-0.466
11	22	-2.365	-4.719	-0.011
12	18	-2.660	-5.014	-0.306
12	19	-4.405	-6.759	-2.051
12	20	-3.585	-5.939	-1.231
12	21	-3.085	-5.439	-0.731
12	22	-3.130	-5.484	-0.776
12	23	-2.965	-5.319	-0.611
13	19	-3.185	-5.539	-0.831
13	20	-2.365	-4.719	-0.011
14	19	-2.725	-5.079	-0.371
16	19	-2.630	-4.984	-0.276
19	24	2.600	0.246	4.954
19	25	2.790	0.436	5.144
19	26	3.240	0.886	5.594
19	27	2.840	0.486	5.194
19	28	2.910	0.556	5.264

19	29	3.400	1.046	5.754
19	30	2.900	0.546	5.254
20	26	2.420	0.066	4.774
20	29	2.580	0.226	4.934

Table 44 Pairs of treatments trained using different algorithms having statistically significant differences in Subject 8 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower	Upper)
DR	SCG	1.385	1.109	1.661
DR	LM	1.330	1.054	1.606

Table 45 Pairs of treatments trained with different training set size having statistically significant differences in Subject 8 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower	Upper)
x1	x2	0.375	0.100	0.650

Table 46 Pairs of treatments with different ANN architectures having statistically significant differences in Subject 9 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower	Upper)
2	12	3.345	1.024	5.666
2	13	3.630	1.309	5.951
2	14	4.100	1.779	6.421
2	15	3.090	0.769	5.411
2	16	3.665	1.344	5.986
2	17	2.870	0.549	5.191
2	18	2.990	0.669	5.311
2	19	3.350	1.029	5.671
2	20	3.330	1.009	5.651
2	21	2.325	0.004	4.646
2	22	3.065	0.744	5.386
2	24	2.380	0.059	4.701
2	25	3.460	1.139	5.781
2	26	3.355	1.034	5.676

2	27	2.925	0.604	5.246
2	28	3.005	0.684	5.326
2	29	2.815	0.494	5.136
2	30	3.320	0.999	5.641
3	12	3.160	0.839	5.481
3	13	3.445	1.124	5.766
3	14	3.915	1.594	6.236
3	15	2.905	0.584	5.226
3	16	3.480	1.159	5.801
3	17	2.685	0.364	5.006
3	18	2.805	0.484	5.126
3	19	3.165	0.844	5.486
3	20	3.145	0.824	5.466
3	22	2.880	0.559	5.201
3	25	3.275	0.954	5.596
3	26	3.170	0.849	5.491
3	27	2.740	0.419	5.061
3	28	2.820	0.499	5.141
3	29	2.630	0.309	4.951
3	30	3.135	0.814	5.456
4	12	2.935	0.614	5.256
4	13	3.220	0.899	5.541
4	14	3.690	1.369	6.011
4	15	2.680	0.359	5.001
4	16	3.255	0.934	5.576
4	17	2.460	0.139	4.781
4	18	2.580	0.259	4.901
4	19	2.940	0.619	5.261
4	20	2.920	0.599	5.241
4	22	2.655	0.334	4.976
4	25	3.050	0.729	5.371
4	26	2.945	0.624	5.266
4	27	2.515	0.194	4.836
4	28	2.595	0.274	4.916
4	29	2.405	0.084	4.726
4	30	2.910	0.589	5.231
5	14	2.475	0.154	4.796
6	14	2.615	0.294	4.936

Table 47 Pairs of treatments trained using different algorithms having statistically significant differences in Subject 9 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
DR	SCG	1.565	1.289	1.841
DR	LM	1.075	0.799	1.351
SCG	LM	-0.490	-0.766	-0.214

Table 48 Pairs of treatments trained with different training set size having statistically significant differences in Subject 9 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower		Upper)
x1	x1.5	-0.350	-0.624	-0.076

Table 49 Pairs of treatments with different retraining procedure, adapted for Subject 8, having statistically significant differences in validation set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower		Upper)
Control	Adapted - specific	-0.955	-1.311	-0.599
Control	Custom	-1.805	-2.161	-1.449
Adapted - combined	Adapted - specific	-1.235	-1.591	-0.879
Adapted - combined	Custom	-2.085	-2.441	-1.729
Adapted - specific	Custom	-0.850	-1.206	-0.494

Table 50 Pairs of treatments with different retraining procedure, adapted for Subject 8, having statistically significant differences in Subject 8 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower		Upper)
Control	Adapted - combined	1.670	1.314	2.026
Control	Adapted - specific	2.235	1.879	2.591
Control	Custom	1.595	1.239	1.951
Adapted - combined	Adapted - specific	0.565	0.209	0.921
Adapted - specific	Custom	-0.640	-0.996	-0.284

Table 51 Pairs of treatments with different retraining procedure, adapted for Subject 8, having statistically significant differences in Subject 9 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
Control	Adapted - specific	-1.205	-1.561	-0.849
Control	Custom	-1.860	-2.216	-1.504
Adapted - combined	Adapted - specific	-1.370	-1.726	-1.014
Adapted - combined	Custom	-2.025	-2.381	-1.669
Adapted - specific	Custom	-0.655	-1.011	-0.299

Table 52 Pairs of treatments with different retraining procedure, adapted for Subject 8, having statistically significant differences in retraining time

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
Control	Adapted - combined	0.440	0.082	0.798
Control	Adapted - specific	2.130	1.772	2.488
Control	Custom	1.390	1.032	1.748
Adapted - combined	Adapted - specific	1.690	1.332	2.048
Adapted - combined	Custom	0.950	0.592	1.308
Adapted - specific	Custom	-0.740	-1.098	-0.382

Table 53 Pairs of treatments with different retraining procedure, adapted for Subject 9, having statistically significant differences in validation set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower		Upper)
Control	Adapted - combined	-0.425	-0.779	-0.071
Control	Adapted - specific	-1.475	-1.829	-1.121
Control	Custom	-2.000	-2.354	-1.646
Adapted - combined	Adapted - specific	-1.050	-1.404	-0.696
Adapted - combined	Custom	-1.575	-1.929	-1.221
Adapted - specific	Custom	-0.525	-0.879	-0.171

Table 54 Pairs of treatments with different retraining procedure, adapted for Subject 9, having statistically significant differences in Subject 8 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower		Upper)
Control	Custom	-1.380	-1.736	-1.024
Adapted - combined	Custom	-1.500	-1.856	-1.144
Adapted - specific	Custom	-1.320	-1.676	-0.964

Table 55 Pairs of treatments with different retraining procedure, adapted for Subject 9, having statistically significant differences in Subject 9 test set real-time classification error rate

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower		Upper)
Control	Adapted - combined	0.775	0.419	1.131
Control	Adapted - specific	0.815	0.459	1.171
Control	Custom	1.110	0.754	1.466

Table 56 Pairs of treatments with different retraining procedure, adapted for Subject 9, having statistically significant differences in retraining time

Treatment pair		Mean difference in rank	Confidence interval for mean rank (Lower | Upper)	
Control	Adapted - combined	0.580	0.222	0.938
Control	Adapted - specific	2.040	1.682	2.398
Control	Custom	1.260	0.902	1.618
Adapted - combined	Adapted - specific	1.460	1.102	1.818
Adapted - combined	Custom	0.680	0.322	1.038
Adapted - specific	Custom	-0.780	-1.138	-0.422