Multi-Frame Event Dependent Locomotion Mode Classification with FIRNNs

by

Norman Arsenault

B.Sc.E., University of New Brunswick, 1994

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science in Engineering

in the Graduate Academic Unit of Electrical and Computer Engineering

Supervisor:	M. Stevenson, PhD, Electrical and Computer Engineering K. Englehart, PhD, Electrical and Computer Engineering
Examining Board:	E. Scheme, PhD, Electrical and Computer Engineering L. Hargrove, PhD, Rehabilitation Institute of Chicago U. Kuruganti, PhD, Kinesiology

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

July, 2015

Abstract

Using electromyography (EMG) data obtained from four transfemoral amputees, Finite Impulse Response Neural Networks (FIRNNs) were investigated for the task of locomotion mode identification. Classification accuracy is found to improve as the duration of the observation interval, presented to the FIRNN, increases. Improvement in classifier accuracy is found to depend on the associated gait event; the more locomotion-mode transitions associated with a gait event, the higher the improvement in the classifier accuracy. Overall, the average classification accuracy on transitions improves by 15.9%. FIRNNs prove much more tolerant to increasing input dimensionality when compared with Linear Discriminant Analysis (LDA) classifiers. When Principal Component Analysis (PCA) is used to reduce input dimensionality, LDA performance is nearly equivalent to FIRNN performance without PCA. A confidence based rejection system is implemented from scaled FIRNN outputs and found to increase classification accuracy and average confidence for the non-rejected patterns. Increasing the observation interval also leads to improved confidence and reduced rejection ratios for fixed decision thresholds.

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature or Abbreviations

EMG	electromyography
FIRNN	finite impulse response neural network
LDA	linear discriminant analysis
PCA	principal component analysis
HS	heel strike
TO	toe-off
MSw	mid-swing
LW	level walking
RA	ramp ascent
RD	ramp descent
SA	stair ascent
SD	stair descent
MAV	mean absolute value
SSC	slope sign changes
WL	waveform length
ZC	zero crossings
RIC	Rehabilitation Institute of Chicago
TMR	targeted muscle reinnervation
DBN	dynamic bayesian network

1. Introduction

1.1. Lower Limb Prosthesis Background

In the United States alone, it's estimated that over 600, 000 people live with a major lower limb amputation [1]. This includes transfemoral, transtibial and hip disarticulated patients, of which transfemoral amputees account for approximately half of the major lower limb amputee population. In addition, approximately 31, 000 new cases of transfemoral amputation occur each year [2]. Existing transfemoral prostheses are incapable of controlling the knee and ankle joints such that a natural gait can be achieved. This can lead to reduced mobility, contribute to development of musculoskeletal conditions and negatively affect quality of life [3]. Given this large amputee population and that existing lower limb prostheses are inadequate in performing multi-joint coordinated movements, an intuitive neural interface control scheme is required.

Transfemoral prostheses can be grouped into three broad different categories: mechanically passive, micro-processor controlled passive devices, and micro-processor controlled powered devices or simply powered devices [1]. The majority of devices in use are of the mechanically passive variety. In mechanically passive devices, the movement of the prosthetic joints relies on the properties of the mechanical components and users must make compensatory

movements with their trunk, pelvis and residual limb to control the prosthesis. Micro-processor controlled passive devices offer varied knee-joint resistance and employ sensors and a micro-processor for closed loop control. The impedance of the knee-joint is then adjusted according to the gait cycle phase which is monitored by the on-board sensors. While this type of prosthesis can be beneficial in smoothing the gait, reducing energy consumption and required hip work, the micro-processor controlled passive device lacks the ability to provide positive net power which hinders the ability to ascend stairs and ramps. Lastly, micro-processor controlled powered devices include the capability to provide net positive power, which when combined with impedance controlled joints, can provide numerous advantages. These include: decreased metabolic energy expenditures, restoration of healthier biomechanics, improved balance and the ability to more easily navigate complex terrain such as ramps and stairs [4].

1.2. Problem Statement

Various approaches have been used to implement control systems for micro-processor controlled powered prostheses. These include: compensatory body movements, key fob or button presses, and echo control [4]. Each of these approaches have significant deficiencies in providing a smooth natural gait capable of coping with locomotion mode transitions. Missing in all these early

methods is the element of intuitive neural control used to derive intent to change locomotion mode. Surface electromyography (or EMG), which has been used for upper-limb prothesis control for many years, offers this neural control element. Despite the non-stationarity of surface EMG on the lower limb, it's proposed that a similar approach can be applied to control powered lower limb prosthesis where intent recognition is implemented by means of pattern recognition of feature vectors derived from EMG and/or mechanical sensors. Past studies using intent recognition have shown that steady state patterns, where the locomotion mode remains constant, can be classified with low error rates; however, classification of locomotion mode transitions, such as changing from level walking to stair ascent, still experience relatively high error rates [2,4,5,6].

1.3. Objectives

The focus of this work was to help improve intent recognition for subjects using a powered lower limb prostheses by making use of the Finite Impulse Response Neural Network (FIRNN) and its temporal processing capabilities for EMG feature classification. As such, the areas of focus were as follows:

a) **Transition Performance** - Improve upon existing transition classification results when using phase-dependent classifiers. Phase-dependent classification refers to performing multiple locomotion mode classification

decisions per gait cycle, as opposed to making a single decision per gait cycle which has proven to be ineffective [7]. The location of the classification decision points is aligned to specific events or phases of the gait cycle.

b) **Gait Cycle** - In order to determine how changes in cyclical EMG patterns might signal a transition, larger sections of time, than which have traditionally been used when dealing with phase dependent classifiers, will be examined. Other than the portion of EMG immediately preceding an event, there may be other aspects of the gait cycle that may be useful in trying to determine intent to change locomotion mode.

c) **FIRNN Confidence** - Apart from the classification accuracy or error rate of the FIRNN classification decisions, can we learn anything from the network outputs which might indicate how certain we are in those decisions?

d) **FIRNN Classifier** - Can the FIRNN structure offer performance improvements over traditional pattern recognition techniques such as Linear Discriminant Analysis (LDA)?

1.4. Significance

Correctly identifying locomotion mode transitions is essential to the intent recognition system in order to provide an intuitive and natural gait [4].

Therefore, improvements in transition recognition have tangible benefits in active lower limb prosthesis control.

Extending the observation interval (the portion of data used for locomotion mode recognition) to encompass an entire gait cycle (or more) may result in performance improvements; most related studies tend to focus on shorter observation intervals.

Unexpected errors can create imbalances or disturbances which can lead to a fall. Rejecting a decision can be preferred over implementing an incorrect one. Establishing a technique to assess FIRNN confidence will provide a basis for deciding to reject a decision.

Given that the human gait is cyclical and time varying in nature, use of the FIRNN may be a suitable application and hasn't been extensively studied for the task of lower-limb EMG pattern recognition.

2. Literature Review

In reviewing past studies regarding intent recognition for lower limb prosthesis control, five main areas of distinction will be highlighted. They are:

1. *Subjects:* Number and type of subjects (able bodied or transfemoral amputee) involved in the study.
2. *Locomotion Modes:* Which types of locomotion mode or movement are contained in the study? These could include movements such as level walking, ramp or stair ascent/descent, standing still, sitting, stepping over an obstacle, ipsilateral or contralateral turning. Does the study also include transitions from one locomotion to another? If so, what type of circuit is navigated and how are the transitions controlled?
3. *Input Data:* What type of input data are used? Data are typically obtained from surface EMG electrodes and/or the prosthesis on-board mechanical sensors. The number and location of EMG electrodes and number and type of mechanical sensors can also vary. Lastly the number and type of features derived from the raw data will also vary.
4. *Frame Arrangement:* The length, number, alignment of frames to specific gait cycle events and possible offset or overlap between successive frames that are used in the study will be also reviewed.

5. *Classifier:* The type and location of the decision points will also be discussed. Common classifier selections include Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and Artificial Neural Networks (ANNs). Location of classifier decision points and input frame arrangement are typically tied to specific gait events in order to compensate for the non-stationarity of the input data. Using smaller duration frames, as compared to a complete gait cycle, allows the data to be considered quasi-stationary [7,8].

2.1. The Gait Cycle

As shown in Figure 2.1, the gait cycle can be separated into two main phases: the Stance Phase, where the foot is making partial or complete contact with the ground and the Swing Phase, where the foot leaves the ground and swings forward. The Stance Phase typically accounts for 60% of the gait-cycle and the Swing Phase accounts for the remaining 40% [9]. The classic gait term definitions recognize six distinct events: 1) Heel Strike (HS), 2) Foot Flat, 3) Mid-Stance, 4) Heel-Off, 5) Toe-Off (TO) and 6) Mid-Swing (MSw). The Stance Phase begins the moment the heel touches the ground with the Heel Strike event. The Stance Phase ends as the Swing Phase begins with Toe-Off, the moment the toes leave the ground. The Swing Phase, can also be sub-divided into acceleration and

deceleration phases. In the acceleration phase, the leg accelerates forward until the Mid-Swing event, where the deceleration phase begins and the leg brakes the velocity of forward movement in order to place the foot down under control. At Mid-Swing, both feet are under the body with heels next to each other [9].

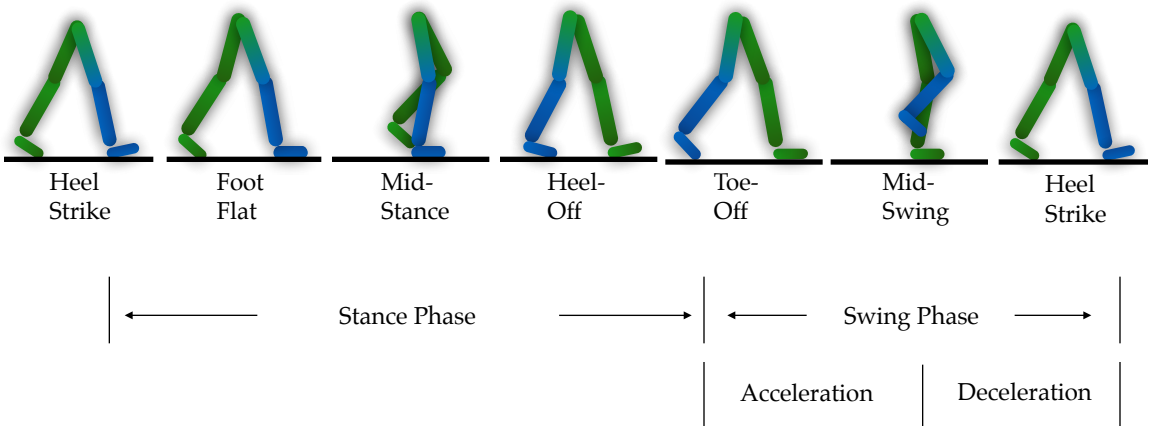

Figure 2.1: Gait cycle overview

2.2. Intent Recognition

In 2009, Hargrove *et al.* [3] introduced an intuitive neural control interface for powered lower limb prosthesis. While EMG based control schemes had been used extensively on upper limb prosthesis for many years it had not been implemented on lower limb prostheses. Until then, control of powered motion was mode based, where the knee and ankle joints operated on pre-determined kinematic profiles for each gait phase and locomotion mode. In order to change locomotion modes, the user needed to inform the prosthesis of the new mode. This type of voluntary control, which can include methods such as: key fob

button presses, echo-control (where the sound leg is instrumented and used for control of the prosthetic limb) and compensatory body movements such as rocking back and forth on the prosthesis are limited in effectiveness and can lead to falls if the wrong mode is selected [4].

As such, a two-tier prosthesis control system was proposed. A *Neural Control Mode* where features derived from EMG data collected on the remaining limb were classified using pattern recognition algorithms to determine intent to change locomotion modes and an *Intrinsic Control Mode* where control of knee and ankle joint parameters such as torque and impedance were governed according to pre-determined adjustable settings based on mode and gait cycle-phase.

Given that EMG information is highly non-stationary over the course of a gait cycle, Huang *et al.* [7] created a phase-dependent locomotion mode classifier such that the EMG can be considered quasi-stationary over short duration time periods. An analysis window was aligned in two ways relative to the HS and TO event: pre-HS, post-HS, pre-TO and post-TO. Separate pattern recognition systems were implemented at each of the four analysis window locations and compared against using EMG from a full-stride cycle with a single pattern recognition classifier. Results demonstrated that a phase-dependent approach to pattern recognition significantly outperformed using a single window spanning

a full-stride cycle; furthermore it demonstrated that the phase-dependent approach is feasible for intuitive mode selection for lower limb prosthesis.

Since then, much of the research work for the *Neural Control Mode* has been advanced by the Rehabilitation Institute of Chicago (RIC) with the prosthesis development and *Intrinsic Control Mode* work done by colleagues at the Center for Intelligent Mechatronics at Vanderbilt University [14-18].

2.2.1. Young *et al.* [5]

Young *et al.* [5] performed a study with four transfemoral amputee subjects using a powered lower limb prosthesis developed by Sup *et al.* [10] from Vanderbilt University. Each subject performed 20 repetitions of two different circuits which contained a total of five different locomotion modes. A ramp circuit contained portions of Level Walking (LW), Ramp Ascent (RA) and Ramp Descent (RD) with the ramp consisting of a 10° slope. For the stair circuit, the 10° ramp was replaced by a four step staircase; so the circuit contained portions of Level Walking, Stair Ascent (SA) and Stair Descent (SD). Types of transitions included in the data, determined by the nature of the circuit, included: transitions from LW to one of the four other classes (RA, RD, SA and SD) as well as transitions to LW from one of the four other classes (RA, RD, SA and SD). All transitions took place at heel strike with the exception of transitions to or from

SA, which occurred at Toe-Off. The prosthesis mode was switched remotely by an observer at the point of transition.

Data from nine EMG channels and 13 mechanical sensors along with gait phase and locomotion mode were recorded. Data were segmented in 300ms frames aligned to HS and TO events. Six features were derived from EMG data, these included: Mean Absolute Value (MAV), Zero Crossing (ZC), Slope Sign Changes (SSC), Waveform Length (WL) and the first two autoregressive coefficients of a sixth order autoregressive model. The four features derived from the mechanical data were: Mean, Standard Deviation, Minimum and Maximum.

Mechanical and EMG features were concatenated and presented to a Dynamic Bayesian Network (DBN) classifier and evaluated at HS and TO using leave-one out cross-validation. This resulted in an average error rate of 1.8% during steady state and 11.3% for transitions when all five locomotion modes were kept as distinct classes. Since the prosthesis parameters for level walking and ramp ascent are very similar, both locomotion modes were combined into a single class with little impact to the test subject. This reduced error rates to 1.3% during steady state and 8% for transitions when four output classes are defined. With the use of a slope estimator, as presented by Sup *et al.* [11], level walking, ramp

ascent and ramp descent can be safely combined into a single class. This resulted in a further error rate reduction of 0.3% for steady state and 0.9% for transitions.

2.2.2. Young *et al.* [2]

In [2], Young *et al.* performed a study with six transfemoral amputee subjects using the same prosthesis and ramp-stair circuits as described in 2.2.1.

Data from 13 mechanical sensors along with gait phase and locomotion mode were recorded. The four features derived from the mechanical data were: Mean, Standard Deviation, Minimum and Maximum.

The length of the analysis window was varied from 50 to 450ms in 50ms increments and aligned to HS and TO events. A separate Linear Discriminant Analysis (LDA) classifier was implemented at HS and TO decision points and tested using leave-one-out cross validation.

Steady state errors decreased from 5.9% at a 50ms analysis window to 3.4% at a 300ms analysis window but did not significantly decrease further for longer windows. Transitional errors were smallest at the 250ms analysis window (approximately 17%).

Next, keeping the window size fixed at 250ms, the number of overlapping analysis windows was increased from 1 to 10 using a frame interval of 20ms. Both transitional and steady state errors increased as additional windows were added, thus subsequent analysis was done using a single 250ms window

The effect of training with data that included transitions was compared to training with data comprised of steady state data only. Transitional error rates were in excess of 80% when transitions were removed from the training data and less than 20% when included in the training data. Steady state errors, however, were reduced from 4% to 1% when transitions were removed from the training data.

2.2.3. Huang *et al.* [12]

In [12], Huang *et al.* performed a study with five transfemoral amputee subjects wearing a hydraulic passive knee. Six locomotion modes were included in the study: Level Walking (LW), Stair Ascent (SA), Stair Descent (SD), Ramp Ascent (RA), Ramp Descent (RD) and stepping over an Obstacle (O). Five transition types were included: LW to SA, LW to RA, LW to O, SD to LW and RD to LW. Tests for each locomotion mode were repeated 15 times. Surface EMG was recorded from 11-13 sites on the residual limb and load measurements obtained from a 6 degree-of-freedom load cell.

EMG features (MAV, ZC, SSC and WL) and mechanical features (Mean, Maximum and Minimum) were fused into a single feature vector for each analysis window. A sliding analysis window of 150ms duration was moved in 12ms increments. Since transitions didn't occur at fixed events, it was decided to predict mode transitions before the occurrence of certain critical gait events.

A phase-dependent classification strategy as previously outlined in [7] was also adopted for this study. Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) classification methods were investigated with leave-one-out cross-validation implemented for training and testing. In evaluating the system, a transitional period was defined as including a full stride and two stance phases of the prosthetic limb.

The SVM classifier outperformed the LDA classifier by 1.5%-5.9% in the four gait phases (initial double limb stance, single limb stance, terminal double limb stance, and swing phase) however the results were not statistically significant. The SVM classifier produced over 99% accuracy during the stance-phase and 95% accuracy during the swing-phase for steady-state mode identification. During transitional periods, the SVM classifier correctly identified all mode transitions. For all phases, the fusion of EMG and Mechanical features outperformed classifiers using EMG or mechanical features alone. The EMG-based classifier slightly outperformed the mechanical based classifier in three of the four phases , with the mechanical classifier only achieving 50-60% accuracy in phase four since little information is recorded by the load module during the swing phase.

2.3. Confidence

In [13], Scheme *et al.* described a novel myoelectric control scheme for upper-limb prosthesis. Confidence-based rejection was implemented using the outputs of a Linear Discriminant Analysis (LDA) classifier.

The Bayes' classification framework allows for a rejection option if the output doesn't exceed a class specific threshold. For implementation with the LDA classifier, a confidence metric is derived by transforming the LDA outputs back into the linear domain and normalizing by the sum of the class likelihoods such that the resulting class probabilities sum to 1.

If the class with the highest confidence doesn't exceed a class-specific threshold, then the input to the LDA pattern recognition system is rejected and the no-motion class is selected.

Ten able-bodied and three transradial amputee subjects performed a multi-dimensional Fitts law test [26]. Six channels of EMG data were collected from the dominant forearm of the able-bodied subjects and five channels of EMG collected from the area of largest muscle bulk of the affected limb for the amputee subjects. EMG data were segmented into 160ms analysis windows with an overlap of 16ms for feature extraction.

Significantly higher throughput, path efficiency and completion rates were observed for the rejection-based LDA classifier in comparison to a baseline LDA

classifier (with no rejection). The rejection based classifier also outperformed the baseline classifier for all values of rejection threshold.

3. Background and Scope

3.1. Background

3.1.1. RIC Data

In 2012, a series of trials were conducted at the Rehabilitation Institute of Chicago (RIC) with four transfemoral amputees (TF01,TF02,TF09,TF10) using a powered knee and ankle prosthesis developed at Vanderbilt University. All subjects were classified as K3 level community ambulators and had experience using the prosthesis. The prosthesis was fit to each amputee's socket and aligned by a certified prosthetist at the RIC [4].

These trials consisted of 20 repetitions of two different circuits. The first circuit included portions of level walking, ramp ascent and ramp descent. In the second circuit, the ramp portion was replaced by a four step stair case, so it contained portions of level walking, stair ascent and stair descent. Through a partnership with the RIC, these data were shared with the Institute of Biomedical Engineering at UNB and will be the focus of this work.

One of the four subjects, TF09, underwent a surgical procedure called Targeted Muscle Reinnervation (TMR) where residual nerves from the amputated limb are transferred to alternative muscles that are no longer biomechanically functional. The alternative muscles are denervated, and the residual nerves eventually grow

into them. The reinnervated muscles then serve as biological amplifiers of neural commands sent through the residual nerves [3].

Data from nine EMG channels for non-TMR subjects and 10 EMG channels for the TMR subject (eight naturally innervated, two reinnervated) and 13 mechanical sensors along with locomotion mode and gait-cycle phase were recorded for each trial. Surface EMG was collected from nine muscle sites on the residual limb: semitendinosus, biceps femoris, tensor fasciae latae, rectus femoris, vastus lateralis, vastus medialis (not used for TF09), sartorius (non-functional for subject TF02), adductor magnus, and gracilis. Nerves controlling plantar flexion and dorsi flexion were reinnervated to alternative muscles on subject TF09. EMG signals were band-pass filtered with passband between 20 and 420Hz, and sampled at 1000 samples/second. Mechanical data were collected from 13 sensors located on the prosthesis, these included: angle, velocity and current for the knee and ankle, 3 dimensional accelerometer and gyroscope, and load cell. Mechanical signals were sampled at 500 samples/sec and the load cell signal low pass filtered at 20Hz.

There are eight types of transitions contained in the data set, from level walking to the ramp and stair modes (ascent and descent), and from the ramp and stair modes back to level walking. All transitions occur at heel contact with the exception of transitions to/from stair ascent which occur at toe off. Next, if

we look at the locomotion mode distribution, as shown in Figure 3.1, for one of the four subjects (TF01), we find that there is a heavy bias towards level walking with fewer ramp and stair patterns. Given the relatively low number of transitions, it was decided not to create additional classes for each type of transition. All subjects have similar distributions given the fixed nature of the course and equal number of repetitions with slight variances accounting for differences in stride length during the level walking and ramp portions.

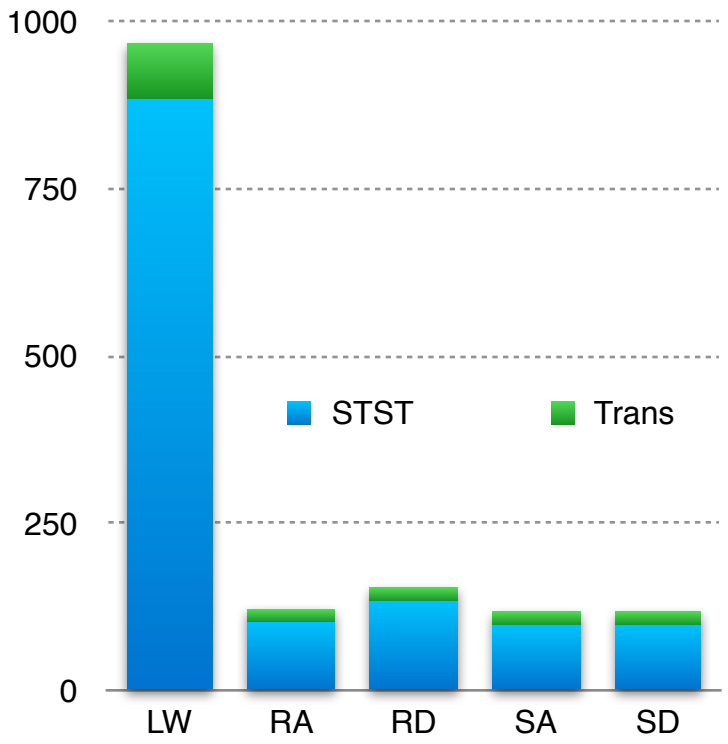

Figure 3.1: TF01 locomotion mode distribution - steady state (STST) and transitions (Trans)

3.1.2. Finite Impulse Response Neural Network (FIRNN) Structure

In 1993, Wan introduced the Finite Impulse Response Neural Network (FIRNN) structure and training algorithm [14]. In the FIRNN, each weight from a static feedforward neural network is replaced by an adaptive FIR filter. The FIRNN results in a formation that is ideally suited to extract time history information from input signals that evolve over time[1]. Figure 3.2 shows a FIR filter of order M, defined as having M delays, M+1 adaptable weights (w_{i0}, w_{i1}, ..., w_{iM}) and the output, $s_i(k)$, being a weighted sum of the current and past inputs, $x_i(k)$, $x_i(k-1)$, …, $x_i(k-M)$.

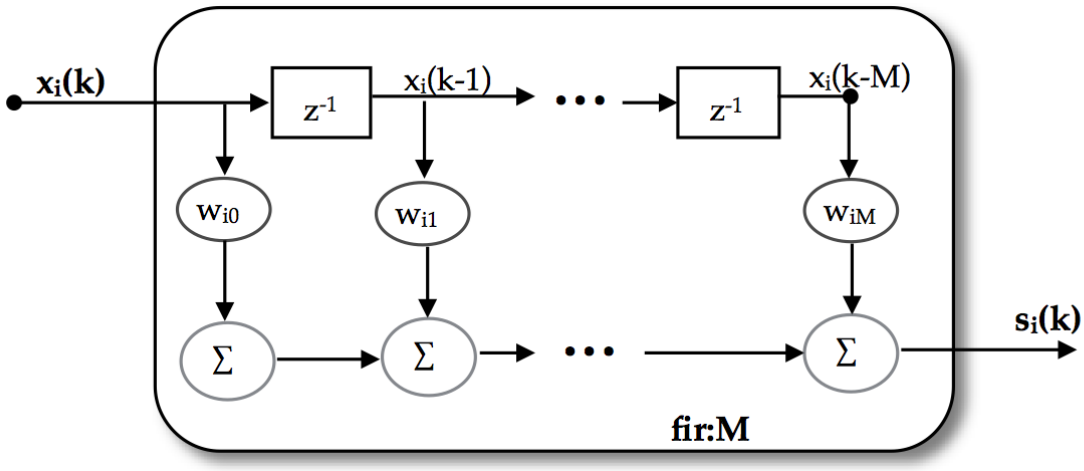

Figure 3.2: FIR filter of order M

Next, a basic processing node, as shown in Figure 3.3, can be constructed using multiple FIR filters. Each input is fed through a separate FIR filter, and

[1] A FIRNN was selected as the winning entry in the 1991 Sante Fe Time Series prediction contest [14].

their outputs, along with an adaptable bias weight, are summed and fed through a transfer function to obtain the node output.

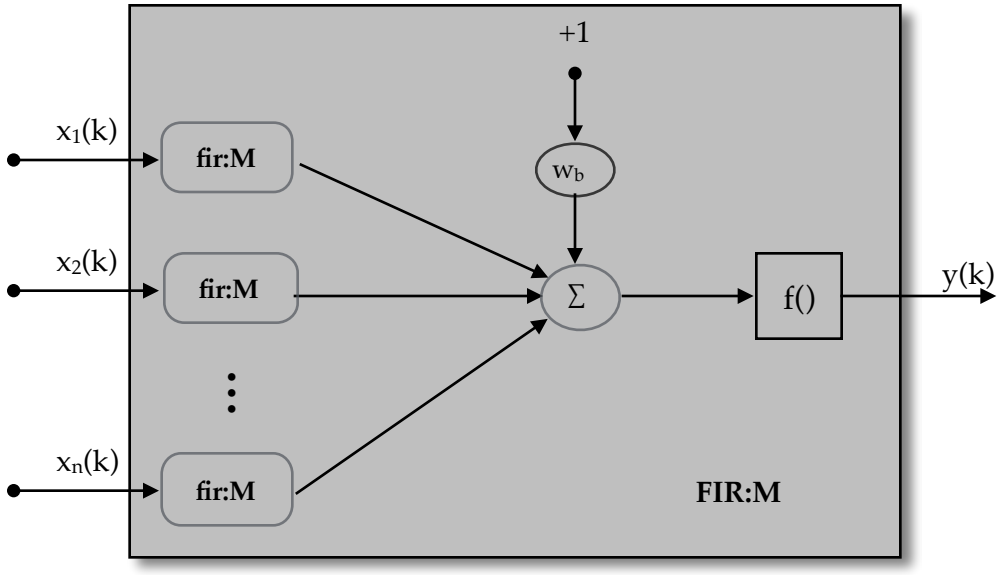

Figure 3.3: FIRNN basic processing node

Then, using multiple processing nodes, we can build a two layer fully - connected FIR neural network as shown in Figure 3.4.

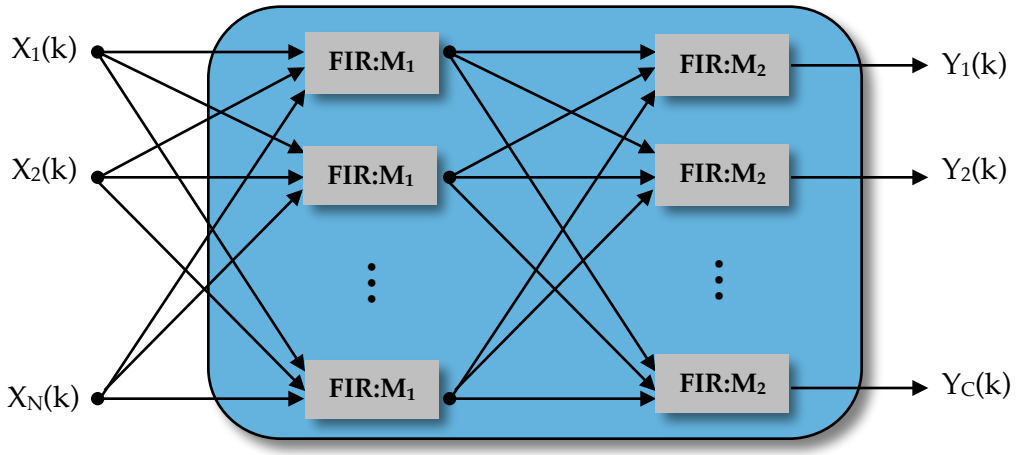

Figure 3.4: Two layer fully connected FIRNN

The number of inputs to the network, N, is determined by the number of elements in the input feature vector and the number of outputs, C, is determined by the number of pattern classes. All of the filters within a single layer are set to the same filter order and the sum of the filter orders across all the layers should equate to the number of frames in the input pattern minus one.

3.1.3. FIRNN Implementation

Training the FIRNN is performed in a supervised manner. A set of data, termed the training set, is presented to the neural network which is initially configured with a random set of initial weights. The resulting network outputs are then compared against a pre-determined set of desired responses and the errors are propagated back through the network to adjust the adaptable weights so that a statistical measure of the error, such as mean square error, is minimized. To monitor training progress, a portion of the input data, termed the test set, is withheld from training and used to test network performance. This iterative process of training and testing continues until a certain performance level is reached, a fixed number of training/testing cycles (or epochs) have been completed, or training is stopped to prevent overtraining. Overtraining can occur if the amount of training data is limited in comparison to the number of adaptable elements in the network and training continues until the network begins to memorize the training data.

For this work, the FIRNN is implemented using Matlab® and the Matlab® Neural Network Toolbox. A separate FIRNN software package developed in C was used to validate the Matlab based implementation. Using a small simulated data set, equivalent network architectures and training algorithms, the weight updates resulting from the Matlab implementation and C package were compared to ensure similarity. Matlab was selected because of its additional training algorithm options, automated stopping criterion, and highly optimized computations which help speed the training process. These options allow for significantly faster training sessions.

Numerous training algorithm options are available. These include:

- *Gradient Descent* - Where the weights and biases are updated in the direction of the steepest negative gradient of the performance or cost function such as Mean Square Error (MSE), Sum Square Error (SSE), Mean Absolute Error (MAE), Sum Absolute Error (SAE).
- *Resilient Backpropagation* - Where only the sign of the derivative is used to indicate the direction of the weight and bias update, the magnitude of the derivative has no effect on the update. The magnitude of the weight update is determined by a parameter which tracks the number of successive weight changes in the same direction. If the derivative is zero, weights and biases remain unchanged.

- *Scaled Conjugate Gradient* - SCG employs second order techniques by making use of the second derivative of the cost function. First order techniques, such as gradient descent, only make use of the first derivative. Second order techniques generally find a better path to cost function minimums but come at higher computational cost.
- *Levenberg-Marquardt* - Designed to approach second-order training speed but without the need to compute the Hessian matrix. Instead the Hessian matrix is approximated using the Jacobian matrix which is much less complex to compute. Levenberg-Marquardt then uses the Hessian approximation to implement a quasi-Newton update scheme [21].

Each of the 4 training algorithms listed above was tested on a limited basis and the Levenberg-Marquardt algorithm was found to offer incrementally improved performance with significantly faster training times, and therefore was used for the remainder of this thesis.

As previously mentioned, choosing when to stop training is an important factor in achieving optimal performance. Stopping too early will prevent peak performance from being achieved and excess training can result in training set memorization and poor generalization on the test set. As such, the following stopping criteria can be employed to terminate training:

- *Epoch Number* - If the number of epochs (or cycles through the training set) reaches a pre-set limit, training is terminated.
- *Minimum Gradient* - Training is terminated when the training set cost function gradient falls below a pre-set threshold.
- *Maximum Failures* - Training is terminated if, after reaching a local minimum on the test set cost function, performance continues to decline for X consecutive epochs, where X is the Max Fail limit.
- *Performance Goal* - If a pre-set performance goal (*e.g.*, MSE = 0) is achieved, then training is terminated.
- *Time* - If a pre-set training time is surpassed, then training is terminated.

In nearly all cases for this work, training was terminated due to the Minimum Gradient or Maximum Failures criteria. Epoch, Performance Goal, and Time were set to values which would not cause them to be activated (*i.e.*, artificially large number of epochs, performance goal set to 0 and infinite training time). While settings for Maximum Failures and Minimum Gradient were varied from default settings, which vary according to training algorithm and performance function, no performance improvements were achieved, so these settings were left at default conditions (Max_Fail = 5, Min_Grad = $1x10^{-10}$).

3.1.4. Frame Segmentation and Data Features

Input data were organized in 250ms non-overlapping frames arranged in a pre-event position, meaning the data were offset such that the analysis window coincides with the period of time immediately preceding the event. For all four test subjects, the analysis windows were aligned to three different gait cycle events: Heel Strike (HS), Toe-Off (TO) and Mid-Swing (MSw). Features were extracted from the raw data, then patterns were labelled based on the upcoming mode and presented to the FIRNN for classification. Separate FIRNN classifiers were implemented at each of the gait cycle events used for frame alignment, as shown in Figure 3.5.

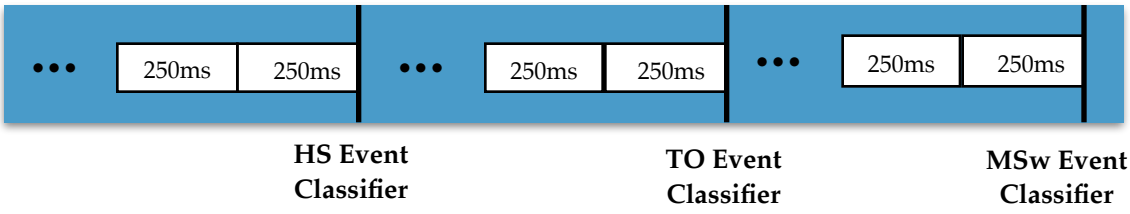

Figure 3.5: Gait cycle frame segmentation

As first presented by Hudgins in [8], four time domain features were derived for each 250ms frame: Mean Absolute Value (MAV), Slope Sign Changes (SSC), Zero Crossings (ZC) and Waveform Length (WL).

3.1.5. Performance Evaluation

For all subjects, each of the three event classifiers (HS, TO, MSw) were tested using four-fold cross-validation where 3/4 of the data were used for training and the remaining 1/4 were used for testing. Each cross-validation fold contained different test data such that all the input data could be included in the testing process. Each cross-validation fold was then evaluated with five different sets of random initial weights to allow for variance within the training process for different initial conditions. These five simulations were then averaged to determine the performance for each fold and the results from the four cross-validation folds were then averaged to determine the performance for each subject-event classifier pairing.

3.1.6. FIRNN Structure Evaluation

Before beginning a thorough investigation with the FIRNN, some elements of its structure (as seen in 3.2.2) must first be defined. For all work in this thesis, the number of layers in the FIRNN was limited to two fully connected layers. The number of inputs to the first layer is determined by the number of elements in each frame. Given nine channels of EMG data and four features/channel, the corresponding network would have 36 inputs to the first layer. The number of processing nodes in the first layer is variable and user defined. The number of

processing nodes in the second layer is determined by the number of output classes. For this work the number of output classes corresponds to the number of locomotion modes. The total filter order of the FIRNN, which is the filter order of each layer summed together, is set to the number frames in the input pattern minus 1. How the total filter order of the FIRNN is distributed among the two layers is up to the developer.

Beginning with two 250ms non-overlapping frames, the number of processing nodes in the first layer was varied. Since there are two frames in each input pattern, the total filter order of the FIRNN was set to 1. This delay can be assigned to filters on the first or second layer. Figure 3.6, shows the overall performance averaged across the three event classifiers, for all four subjects when the number of processing nodes in the first layer was varied with the delay assigned to either the first or second layer.

From Figure 3.6, we see that both configurations (delay on the first or second layer) yielded nearly identical results. Since the arrangement of placing the delay on the second layer results in significantly fewer network weights, the arrangement of placing the entire total filter order on the second layer was kept going forward. Keeping the number of weights low aids in both network generalization and training time. Results for all four subjects plateaued after

seven processing nodes in the first layer, so that arrangement was also kept going forward.

Three of the test subjects behaved very similarly but TF09 outperformed the other subjects by approximately 3%. As previously discussed, this subject underwent the TMR surgical procedure, so their EMG can be considered somewhat enhanced over the other 3 subjects.

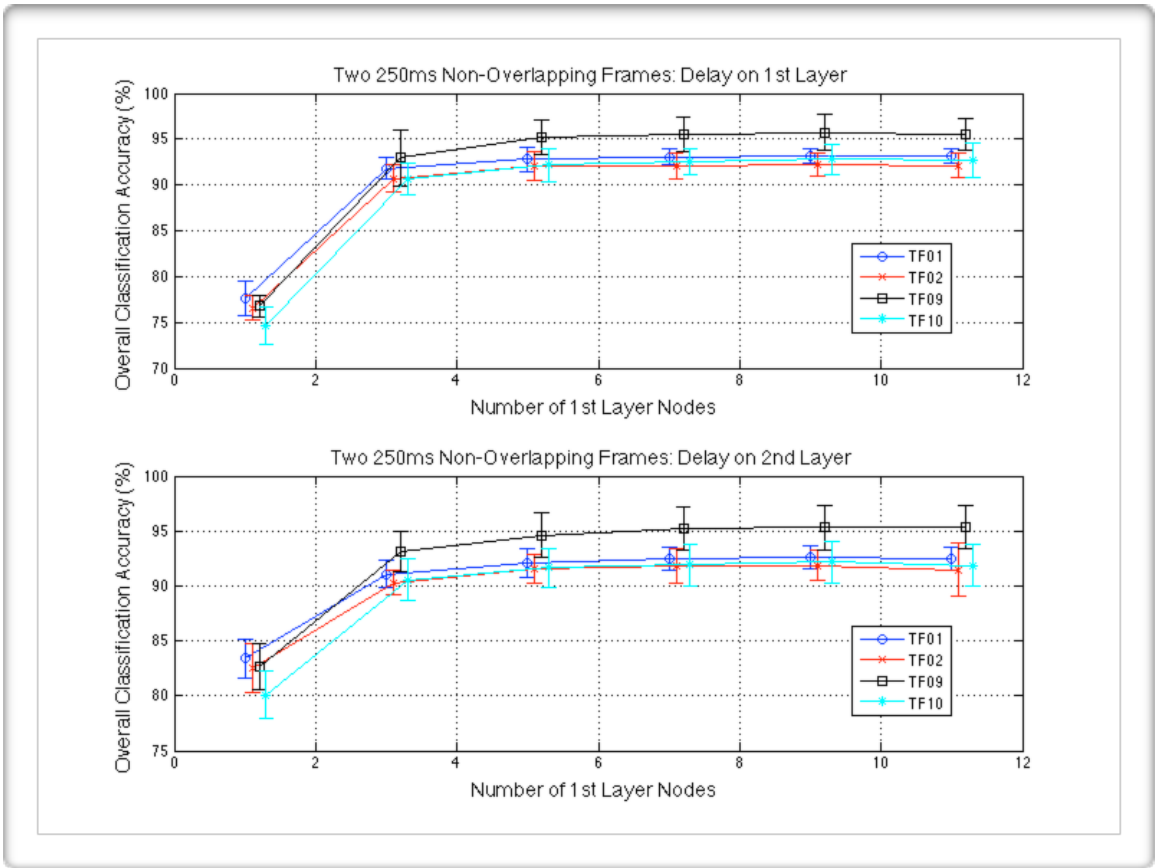

Figure 3.6: Subject performance for varying number of 1st layer nodes

3.2. Scope

The scope of this thesis consists of an investigation of the neural control mode for lower-limb microprocessor controlled powered prostheses, using data provided by the RIC from four subjects who participated in a series of trials using a prosthesis developed at Vanderbilt University. Each subject completed a series of ramp and stair trials over the course of a single day and the prosthesis wasn't doffed or re-donned over the course of the trials. Two of the subjects had data available from a second testing session, however only a single day session was used for these subjects in this analysis.

Data features consisted of 4 time domain EMG features: MAV, SSC, ZC and WL.

Pattern recognition algorithm selection included FIRNN and LDA classifiers with and without the addition of Principal Component Analysis (PCA) for dimensionality reduction. For all evaluations, separate classifiers were implemented at HS, TO and MSw gait events. Frame length was kept fixed at 250ms and apart from a limited investigation which used overlapping frames with a 50ms frame interval, successive frames did not overlap.

4. Multi-Frame Analysis

In this chapter, the performance of the FIRNN classifier will be evaluated for various configurations of overlapping and non-overlapping 250ms frames. Performance of the FIRNN will also be compared to a Linear Discriminant Analysis (LDA) classifier.

4.1. FIRNN Results: Non-Overlapping Frames

Figure 4.1 shows the overall classification accuracy rate for all subjects as the number of frames was varied from 1 to 8 when using the FIRNN classifier. Eight 250ms non-overlapping frames span an observation interval of 2.0 seconds. Given the short and fixed nature of the stair (4 steps) and ramp (staircase replaced by 10^{o} slope) circuits, an observation interval which spanned two or more gait cycles could potentially capture more than one transition. For all subjects used in this study, an observation interval of 2.0 seconds represents 65% - 112% of the average gait cycle duration. In order to account for gait cycles which may be shorter than the average duration, the observation interval was limited to 2.0 seconds.

From Figure 4.1, we can see the general trend of increasing overall accuracy rate with an increasing number of frames. Improvements from one to eight

frames in overall accuracy range between 3.2% (TF09) and 7.6% (TF10) with an average improvement of 4.9% across all subjects.

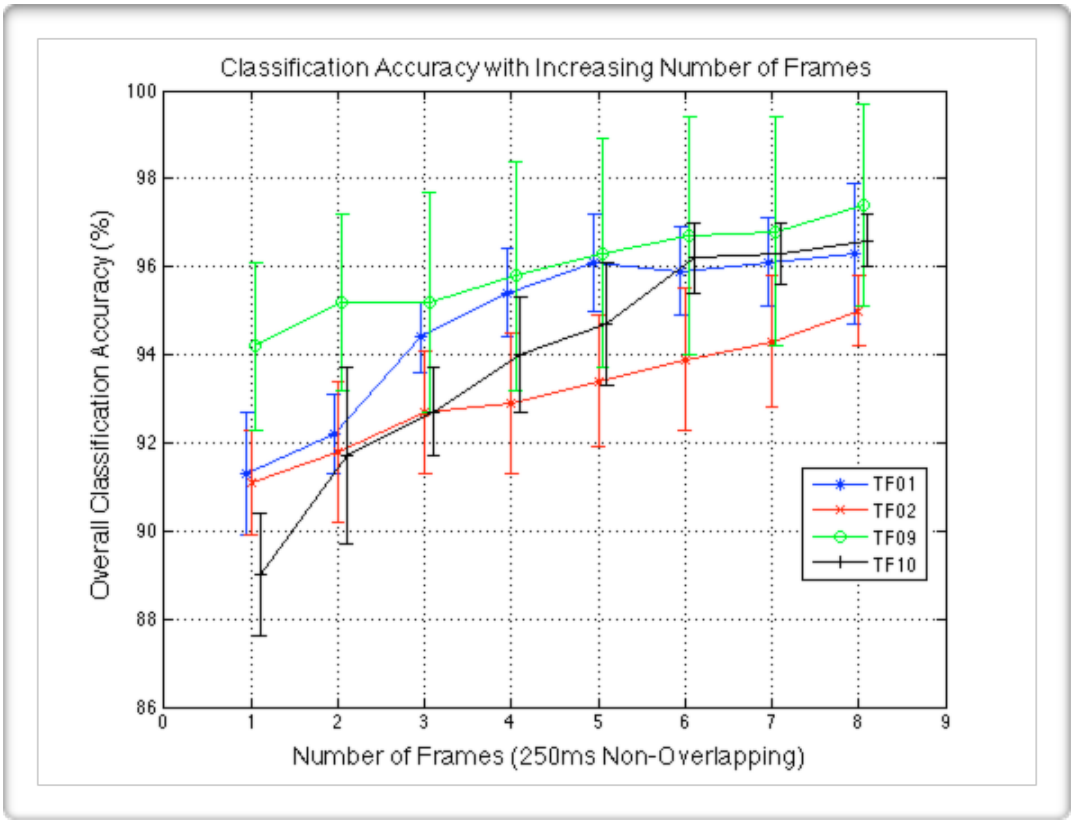

Figure 4.1: FIRNN overall classification accuracy for all subjects with increasing number of frames

Also noteworthy, is that the initial performance advantage of the TMR subject (TF09) demonstrated in the first few frames, decreases as additional frames are added to the input pattern. For four frames and greater, TF02 had the highest error rate of the four subjects. This may be attributed to the fact that TF02 also had one less functional EMG channel (sartorius) during the recording trials.

Next, the results of individual gait event classifiers which make up the overall accuracy rate will be examined. Figure 4.2 shows the FIRNN accuracy rate as a function of the number of frames for the Heel Strike (HS), Toe-Off (TO) and Mid-Swing (MSw) event classifiers for each subject. As seen from the figure, the trend was for all three event classifiers to improve as additional frames were added to the input pattern across all four subjects. For all subjects, the MSw classifier achieved the highest classification accuracy, followed by the TO and HS classifiers respectively. The TO and MSw event classifiers for subject TF09 slightly countered this trend but their performance never differed by more than 1%. Since the majority of transitions occur at the HS event, with only those leading to/from SA taking place at TO, and none take place at the MSw event, the performance level of each event classifier isn't completely unexpected. The largest improvement in accuracy rate from one to eight frames is seen by the HS classifier at an average of 8.3% across all four subjects; followed by the TO classifier at 3.8% and MSw classifier at 2.6%.

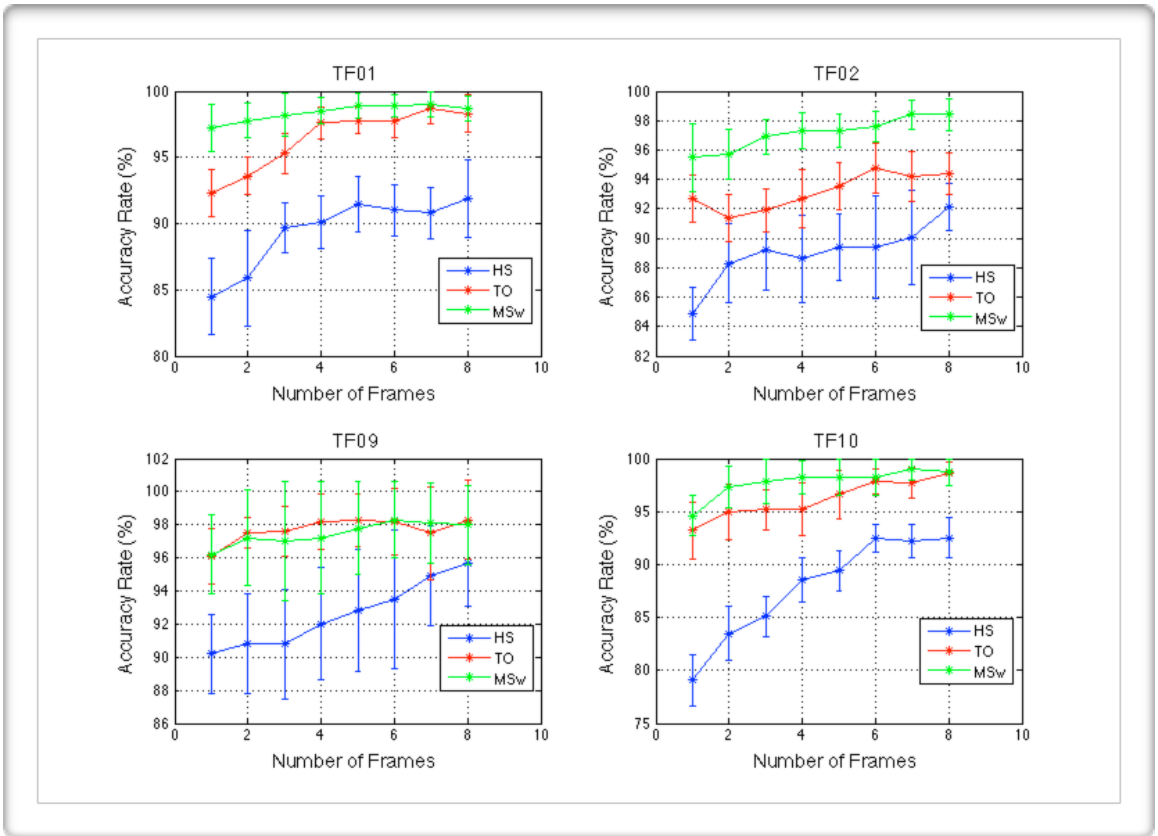

Figure 4.2: Individual FIRNN gait event classifier performance

Next, the results for each subject were broken down into steady state and transition patterns in an attempt to identify what's driving the larger improvements on the HS and TO classifiers. Figure 4.3 shows the classification accuracy results of each subject broken down into steady state and transition patterns.

From these results, we can see each subject experiences incremental improvement in steady state performance and significant improvement in transition performance as additional frames are added to the input pattern. While starting at a much higher level, average steady state improvement in

classification accuracy rate was 3.6% with an average steady state classification accuracy of 97.5% at eight frames. The average improvement in transition performance was 15.9% and the average transition classification accuracy rate at eight frames was 86.4%.

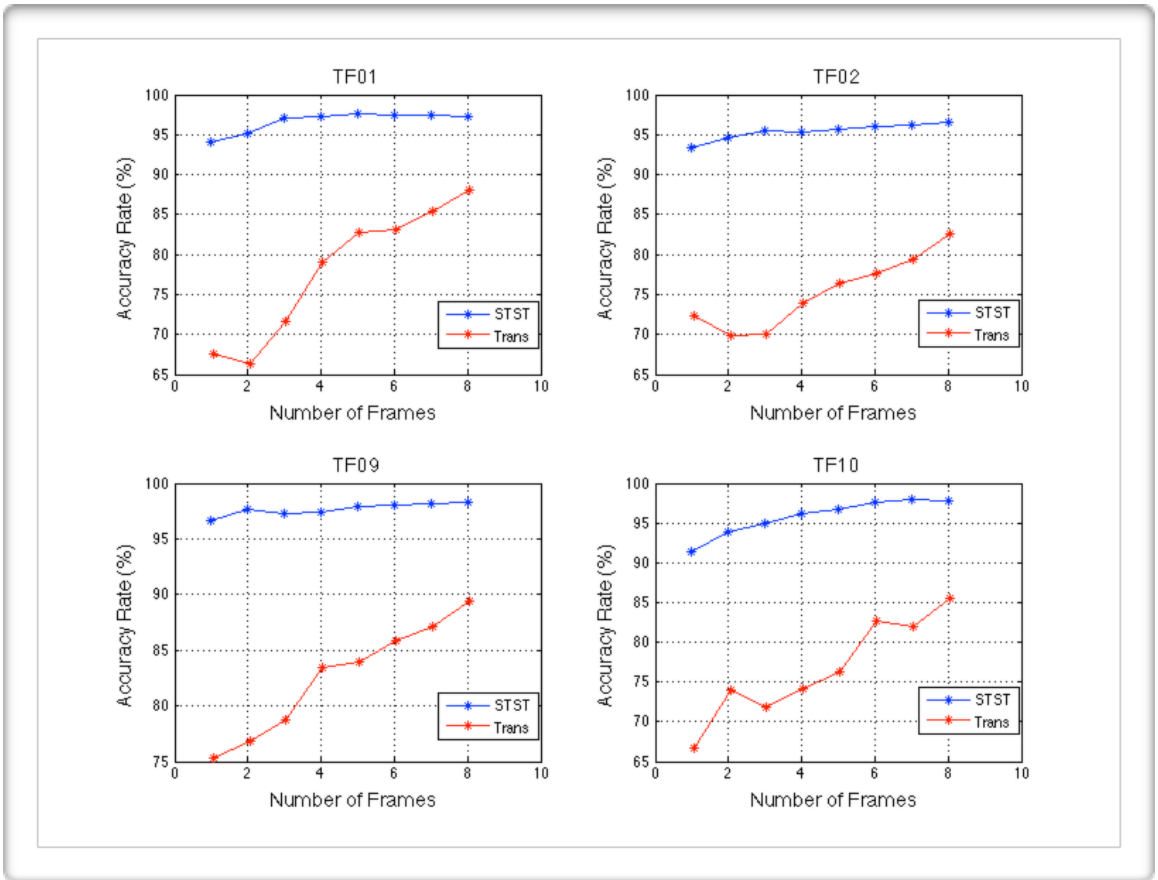

Figure 4.3: Steady state and transition performance for each subject

4.2. FIRNN Results: Overlapping Frames

Tables 4.1 and 4.2 contain the results of a classification accuracy (%) comparison between the use of overlapping and non-overlapping frames. The arrangement of eight non-overlapping frames was compared to other

configurations that would result in either an equivalent FIRNN architecture having the same number of adaptable elements or an equivalent observation interval where the number of frames and overlap were adjusted to obtain the same total 2000ms observation interval.

Subject	8 Frames 0 ms Overlap 2000ms Obs. Int.	8 Frames 50 ms Overlap 1650 Obs. Int.	8 Frames 100 ms Overlap 1300ms Obs. Int.	8 Frames 200 ms Overlap 600ms Obs. Int.
TF01	96.3 ± 1.6	95.8 ± 1.0	95.8 ± 1.3	93.9 ± 1.3
TF02	95.0 ± 0.8	93.9 ± 1.4	93.4 ± 1.3	92.3 ± 1.2
TF09	97.4 ± 2.3	96.2 ± 3.5	95.8 ± 2.7	95.1 ± 1.5
TF10	96.6 ± 0.6	96.5 ± 0.7	95.6 ± 0.7	93.4 ± 1.1

Table 4.1: FIRNN classification accuracy (%) for equivalent FIRNN architectures and differing observation intervals (Obs. Int.)

Subject	8 Frames 0 ms Overlap 2000ms Observation Interval	11 Frames 75 ms Overlap 2000ms Observation Interval*
TF01	96.3 ± 1.6	96.0 ± 1.3
TF02	95.0 ± 0.8	94.1 ± 1.9
TF09	97.4 ± 2.3	96.9 ± 3.3
TF10	96.6 ± 0.6	96.8 ± 0.8

Table 4.2: FIRNN classification accuracy (%) for equivalent observation intervals and differing FIRNN architectures
*** Number of simulations reduced from five to one due to extended training time**

From Table 4.1, we see that for equivalent FIRNN architectures, the use of overlapping frames does not offer any performance improvements and classification accuracy improves with increasing observation intervals. From Table 4.2, the classification accuracy for equivalent observation intervals appears

similar but the additional number of adaptable elements in the FIRNN, due to the increased number of frames in the input pattern, result in significantly more training time. From these results, the investigation on the use of overlapping frames was not pursued.

4.3. FIRNN Weight Comparison

A comparison between the number of adaptable elements in the FIRNN when:

i) the input pattern consists of 2x250ms non-overlapping frames with variable number of processing nodes in the first layer, and

ii) the input patterns consists of a variable number of frames with seven nodes in the first layer;

could indicate if increases in classification accuracy are due to: *a)* the increasing number of adaptable elements or network capacity or *b)* the capability of the FIRNN to extract temporal cues from the input data. Figure 4.4 compares the performance and the number of FIRNN adaptable elements (weights and biases) for i) and ii). This can be considered as the data from Figure 4.1 superimposed on Figure 3.6 with the x-axis changed to show the number of of adaptable elements in the FIRNN instead of the number of first layer nodes or number of frames.

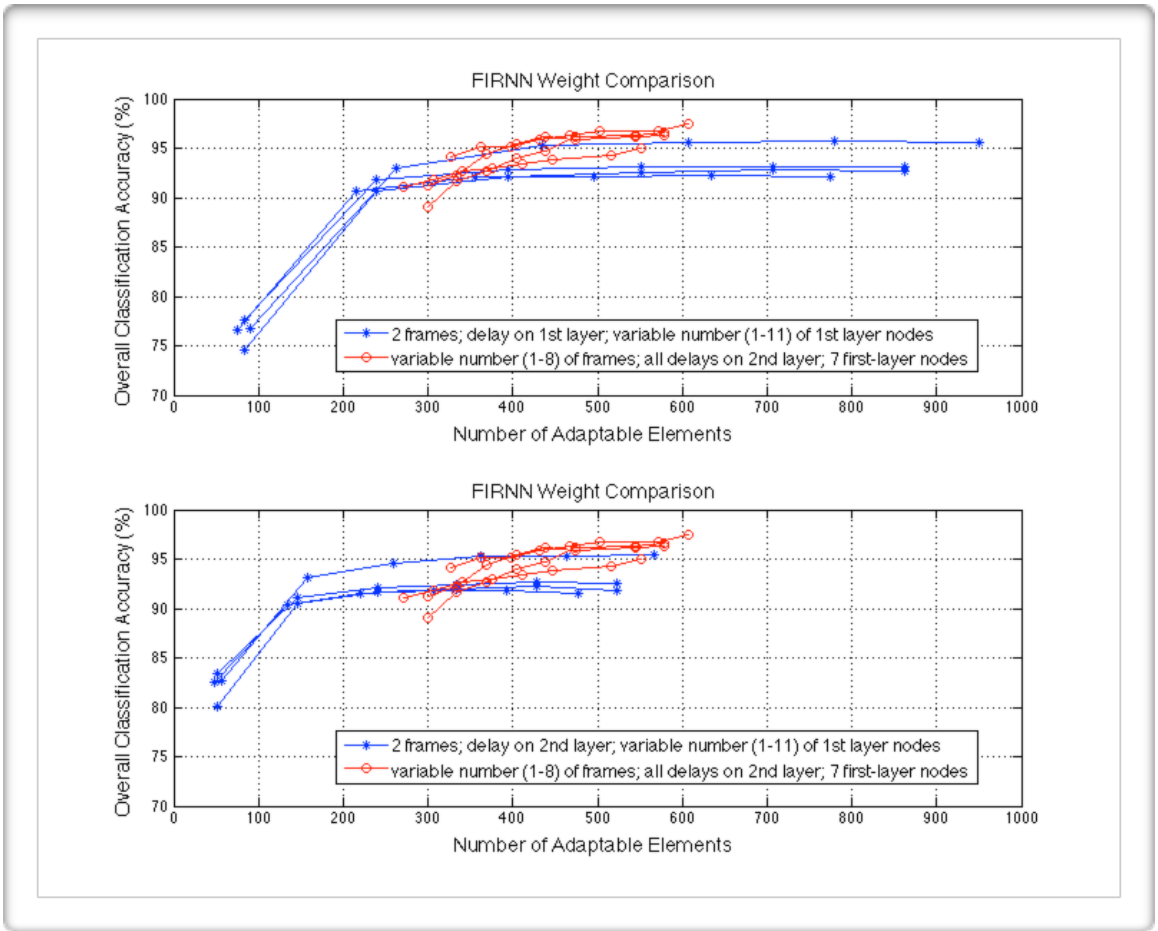

Figure 4.4: FIRNN weight/performance comparison between varying number of first layer node (blue curves) and varying the number of frames in the observation interval (red curves)

In the top sub-plot of Figure 4.4, the blue curves represent the performance of the four subjects as the number of first layer nodes is varied between 1 and 11 using 2x250ms non-overlapping frames with the single delay assigned to the first layer. The red curves indicate the performance of the four subjects when the number of frames is varied between one and eight with the corresponding number of delays assigned to processing nodes in the second layer.

In the bottom sub-plot of Figure 4.4, the blue curves again represent the performance of the 4 subjects as the number of first layer nodes is varied between 1 and 11 using 2x250ms non-overlapping frames but here the single delay is assigned to the second layer. The red curves represent the same as mentioned above.

From the Figure, we can see that increasing the number of frames as opposed to increasing the number of processing nodes in the first layer is more effective and efficient in use of the available network elements. This indicates the FIRNN is achieving performance improvements by extracting temporal cues in input data as opposed to increasing network capacity.

4.4. LDA Comparison

4.4.1. LDA Classifiers [27]

From Bayes Theorem, the posterior probability of class C_i, given observation x is defined as:

$$p(C_i \mid x) = \frac{p(x \mid C_i)\ p(C_i)}{p(x)} \tag{4.1}$$

where $p(x \mid C_i)$ is the conditional probability density of obtaining observation x from class C_i, $p(C_i)$ is the prior probability of class C_i, and $p(x)$ is the

unconditional probability density for observation x. The Bayes classifier assigns observation x to the class with the highest posterior probability $p(C_i \mid x)$.

If we note that $p(x)$ is common for all classes, we can omit the denominator from (4.1). Assuming Gaussian probability density distribution for all classes and that all classes share a common covariance matrix, $\sum_i = \sum, \forall_i$, the classification function can be rewritten as:

$$f(x) = \underset{i}{\operatorname{argmax}} \left\{ x^T \Sigma^{-1} \mu_i - \frac{1}{2} \mu_i^T \Sigma^{-1} \mu_i + \ln[p(C_i)] \right\} \tag{4.2}$$

where the covariance matrix $\sum$ and class mean vectors μ_i can be estimated from the training set. This forms the basis of the LDA classifier used in this work. As with the FIRNN classifier, four fold cross-validation was used during testing to divide the training and test data and ensure all patterns were cycled through the test process.

4.4.2. LDA Comparison Results

Using the same 250ms non-overlapping frames aligned to HS, TO and MSw gait events; the performance of the LDA classifier was compared to that of the FIRNN classifier for each subject.

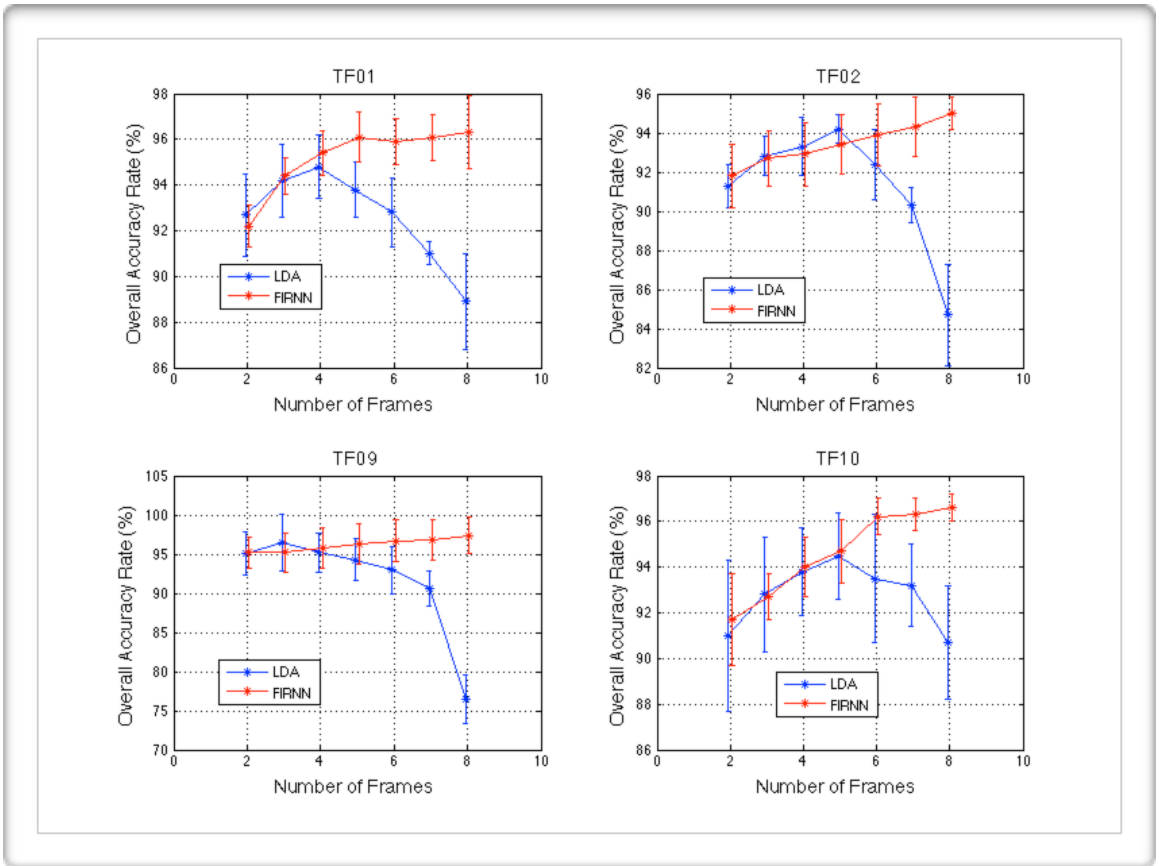

Figure 4.5: Overall classification accuracy rates for LDA and FIRNN classifiers using 250ms non-overlapping frames

Results for the LDA were obtained using previously developed software from the Institute of Biomedical Engineering at UNB and Horton [6].

From Figure 4.5, we can see LDA performance begins to degrade in the region of 3-5 frames while the performance of the FIRNN continues to improve. For all subjects the maximum overall accuracy rate of the FIRNN exceeds that of the LDA classifier.

The LDA performance drop-off may be explained by the curse-of-dimensionality. First termed by Bellman in 1961 [15], the curse-of-dimensionality

refers to a phenomena that arises when analyzing and organizing data in high-dimensional spaces that does not occur in low-dimensional settings. When the dimensionality of the input increases, the volume of the input space increases so much that available input data becomes sparse. The curse-of-dimensionality is not specifically a problem of high dimensional data, but a joint problem between the data and the algorithm being applied. In this case, the FIRNN appears unaffected by the curse while the LDA classifier is more sensitive to the increasing dimensionality of the input feature space. Given that the number of first layer nodes in the FIRNN is set to seven, the number of input features (36), which are equal to the number of EMG channels (9) times the number of features per channel (4), are reduced (from 36 to 7) before being forwarded to the second layer for temporal processing. For the LDA classifier, the number of input features (288) are equal to the number of EMG channels (9) times the number of features per channel (4) times the number of frames (8). The dimension of the covariance matrix then grows significantly with each additional frame; in the case of 8 frames, the size of the covariance matrix has grown to 288x288 = 82,944 elements.

Supplemental figures for the individual gait event accuracy rates of the FIRNN and LDA classifiers are shown in Appendix A.

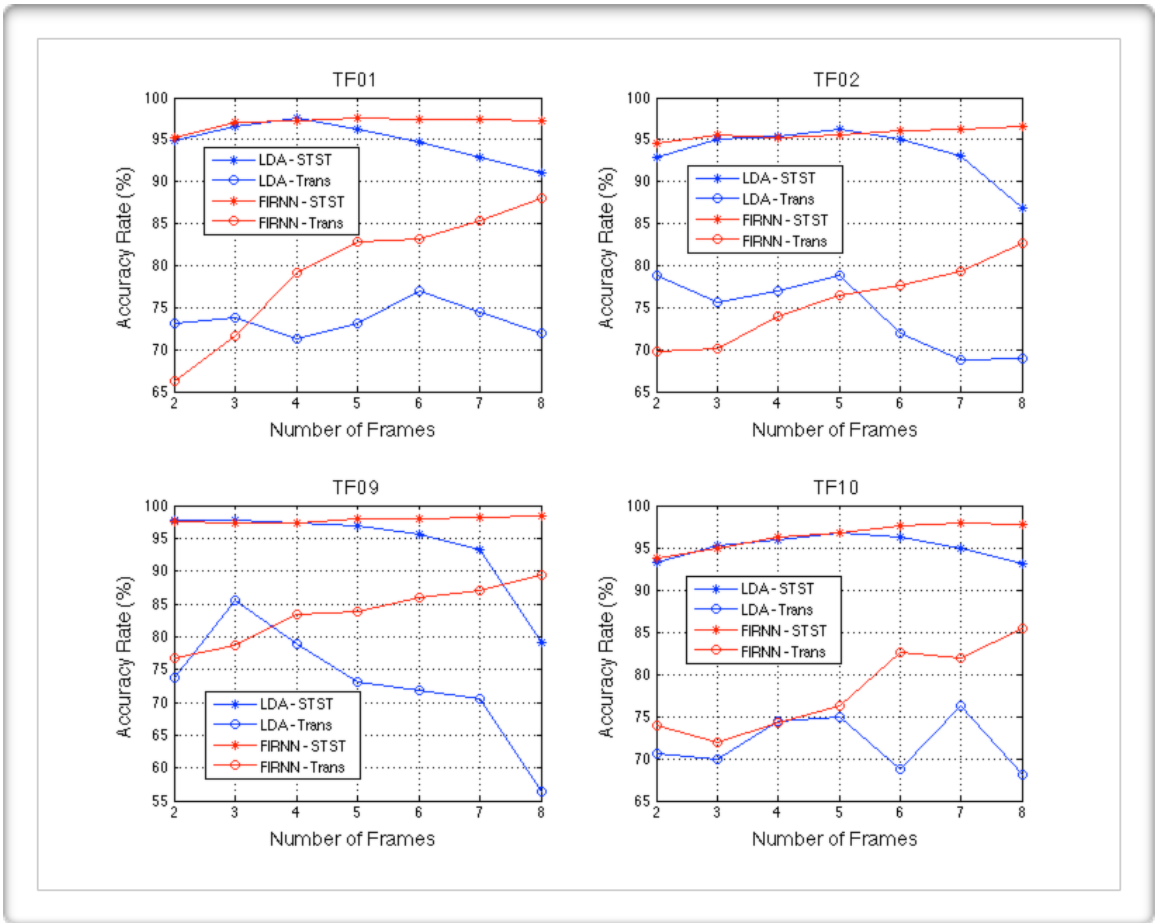

Figure 4.6: Steady state (STST) and transition (Trans) performance for LDA and FIRNN classifiers using 250ms non-overlapping frames

Next, comparing the steady state and transition performance of the two algorithms, we find the results as shown in Figure 4.6.

As is shown, peak FIRNN transition classification accuracy exceeded peak LDA transition classification accuracy by an average of 7% across all subjects. For steady state patterns, FIRNN and LDA peak performance were much closer with a 0.6% average difference in favour of the FIRNN classifier. The curse-of-dimensionality affecting the LDA classifier was more evident on steady state

patterns; however, the LDA's transition performance at 8 frames was lower than that at 2 frames for all subjects.

4.5. Principal Component Analysis

In order to reduce the effect of increasing dimensionality on the input data set, a statistical procedure entitled principal component analysis (PCA) was employed. PCA uses an orthogonal transformation to convert a data set of potentially correlated variables into a set of linearly uncorrelated variables called principal components [16]. The number of principal components is always less than or equal to the number of original variables. The first principal component explains the largest possible variance and each successive component describes the next highest possible variance under the constraint that it is orthogonal to all preceding components.

The principal components are found by calculating the eigenvectors and eigenvalues of the input data covariance matrix. The eigenvector with the largest eigenvalue represents the direction of greatest variation, the one with the second largest eigenvalue, the direction with the highest variation in the subspace orthogonal to the first eigenvector, and so on.

By projecting the *N*-dimensional feature vector onto the first *K* eigenvectors, the dimensionality of the input data set is reduced from *N* to *K* until the curse of dimensionality can be avoided.

4.5.1. PCA Results

Beginning with the input pattern constructed from eight frames, principal component analysis was applied to the input feature set and the corresponding LDA and FIRNN results are shown in Figure 4.7 for various number of principal components. The maximum number of components was 36 for subjects TF01 and TF10, 40 for subject TF09 and 32 for subject TF02. The dashed blue line indicates the LDA performance with 8 frames when no PCA was applied. Similarly, the dashed red line represents the performance level of the FIRNN at eight frames without PCA. It also corresponds to the peak performance level without PCA (as the best FIRNN performance was obtained at 8 frames for all subjects). The dashed green line represents the peak performance of the LDA classifier without PCA. This occurred in the vicinity of 3 - 5 frames, as shown in Figure 4.5.

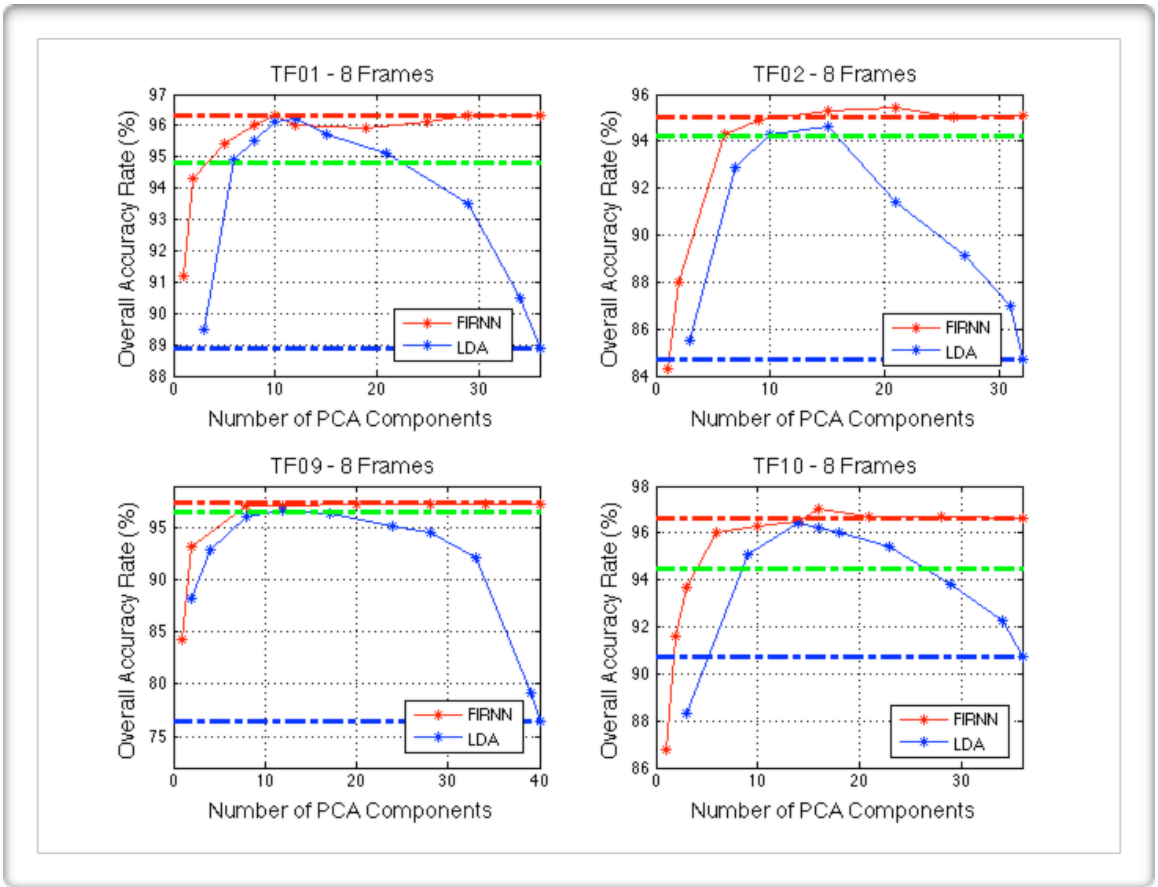

Figure 4.7: PCA performance results for LDA and FIRNN with 8 frames (dashed blue line = LDA performance with 8 frames no PCA, dashed red line = FIRNN performance with 8 frames no PCA, dashed green line = peak LDA performance without PCA)

Beginning with the LDA results, we can note the following:

- When the number of principal components was equal to the number of input features before PCA is applied, the performance for PCA and non-PCA data inputs was the same. This is because all the signal variance is explained by the PCA components. As we can see from the Figure 4.7, this was the case for all 4 subjects where the far right data point on the blue curve overlays the dashed blue line.

- The approximate shape of the curve as the number of principal components is varied resembles an upside down U. When the number of principal components was low, those few components didn't have the capability to accurately represent the richness of the original data set so classification performance suffered. As the number of principal components increased, performance improved until too many components remained and the curse-of-dimensionality returned.
- For all subjects, LDA performance at 8 frames with PCA offered a significant advantage over non-PCA use (the dashed blue line). The average improvement of the peak performance with PCA over the non-PCA performance level was 10.8%, though the size of improvement varied based on the decrease in performance due to the curse-of-dimensionality. If PCA was repeated with the input patterns constructed from a fewer number of frames, the average improvement would likely fall until PCA no longer offered any improvement over the non-PCA data set.
- For all subjects, using PCA with 8 frames exceeded the LDA peak performance level without PCA (the dashed green line), though the improvement was slight at an average of 1%. Finding the optimal LDA performance with PCA would require repeating the analysis with fewer

number of frames and determining the performance for all variations of principal components.

- For all subjects, using PCA with eight frames approaches the FIRNN performance level (dashed red line) but never exceeded it. On average, the peak LDA with PCA performance level was 0.4% below that of the FIRNN without PCA.

Next, when Principal Component Analysis was applied to the FIRNN input data set, we can note the following:

- When the number of PCA components was equal to the number of features before PCA was applied, the performance remained largely unchanged. The far right data point on the red curve matches very closely to the dashed red line with slight differences resulting from training with the PCA feature set.
- Since the FIRNN structure appears much more tolerant than the LDA to the curse-of-dimensionality, use of PCA offered little to no improvement over non-PCA classification rates, however, a training time reduction was realized from having fewer input features. As with the LDA principal component analysis, when the number of components fell below what was needed to adequately represent the input data set, performance began to

drop significantly. When the number of components increased to a sufficient level, performance closely tracked the non-PCA level.

- Given the arrangement of the FIRNN with no delays in the first layer, we could say that it approximates its own PCA analysis by determining which input features are important for classification, as determined by the training process, and forwards that summary to the second layer for temporal processing.

Steady state and transition results for each subject when PCA was applied for both the FIRNN and LDA classifiers are shown in Figure 4.8.

When comparing the steady state performance of the FIRNN and LDA classifiers, we see that the peak performance of both classifiers was virtually identical. The only distinguishing factor was that the FIRNN maintained that peak performance level as additional principal components were added to the input, where as LDA performance began to fall with increasing input dimensionality (between 10-20 principal components).

On transition patterns, peak LDA performance equaled FIRNN transition performance for subjects TF02 and TF10. For subjects TF01 and TF09, the FIRNN transition peak performance exceeded the LDA peak performance by 1.9% and 5.7% respectively.

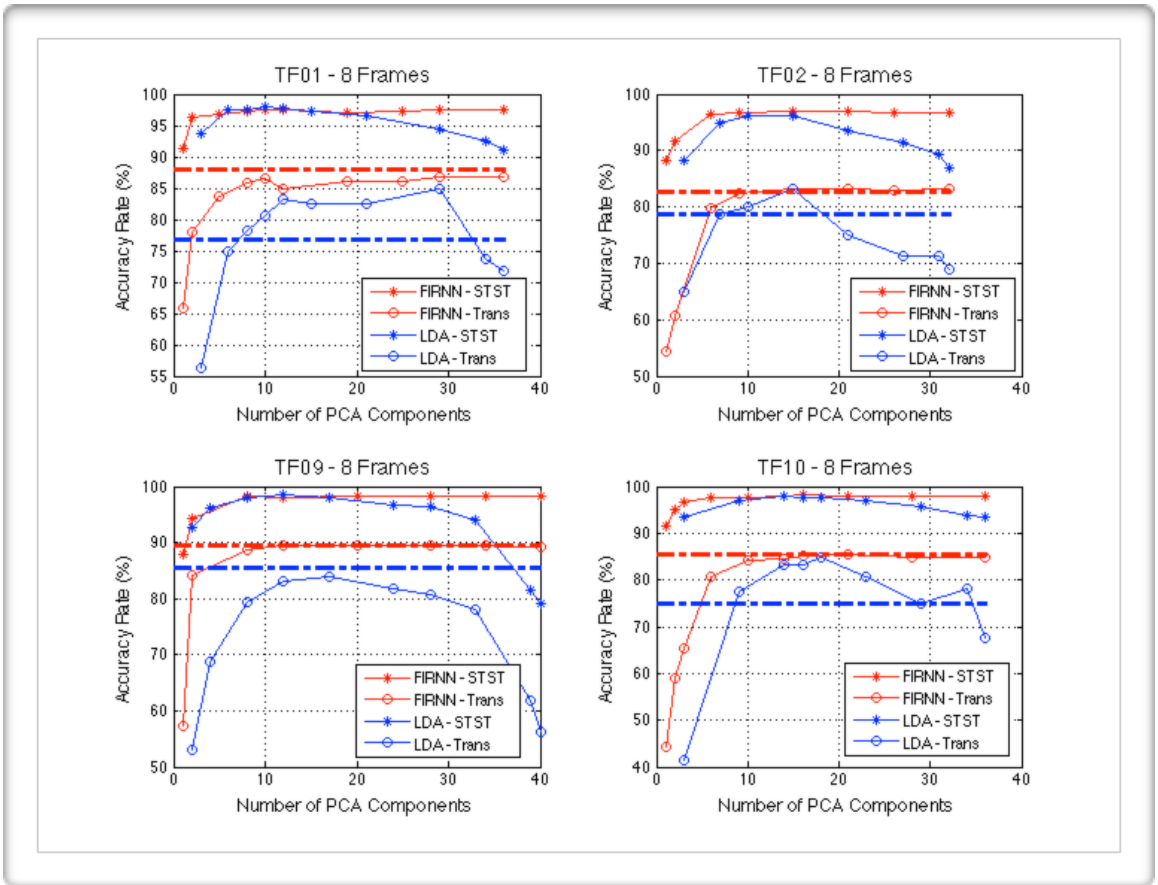

Figure 4.8: Steady state and transition results when PCA is used with FIRNN and LDA classifiers (dashed blue line = LDA peak transition performance without PCA, dashed red line = FIRNN peak transition performance without PCA)

For each subject, the dashed red and blue lines represent the peak transition performance of the FIRNN and LDA classifiers without PCA. From the Figure 4.8, we see that PCA offered virtually no improvement or change in transition performance for the FIRNN. However, when combined with the LDA classifier, PCA offered an average subject transition improvement of 5.2%.

4.6. Discussion

From the results for non-overlapping 250ms frames, shown in Figures 4.1 - 4.3, we saw that the overall classification accuracy rate improved when additional frames were added to the input pattern. Classifiers located at the gait cycle events where most of the transitions occurred showed the most improvement (HS followed by TO then MSw). This occurred due to significant transition performance improvement in classification accuracy while steady state improvement was limited.

This appears to indicate that there is additional useful information available to the FIRNN in the grouping of preceding frames other than the EMG content immediately preceding the events where transitions take place. The FIRNN is then using this temporal pattern in EMG information for improved transition performance. As a subject approaches a transition, *e.g.*, from level walking to stair descent, it's natural to assume there will be slight changes in stride length and cadence over the last few cycles to ensure that the foot is placed safely and securely in the correct location. These changes in the pattern exhibited by gait accelerations and decelerations were more effectively captured if the overall analysis window extended back far enough.

For equivalent FIRNN architectures, the use of overlapping frames did not offer any performance improvements, as classification accuracy generally

improved with increasing observation interval. For equivalent observation intervals, the classification accuracy remained similar but the training time increased significantly as the overlap increased. (Note that as the frame overlap increased, the number of frames required to maintain the same observation interval increased, thus resulting in an increased number of adaptable weights and additional training time.)

To compare between: a) when the input pattern consisted of two non-overlapping frames and the number of hidden nodes was varied and b) when the input pattern consisted of a fixed number of processing nodes and the number of frames was varied, the classification accuracy and number of adaptable elements in the FIRNN was evaluated. The results from Figure 4.4 suggest three important outcomes: 1) that the performance of the FIRNN isn't simply based on the number of adaptable elements in the network; 2) equal performance is realizable with fewer weights when delays are placed in the second layer processing nodes; 3) temporal cues allow for increased classification accuracy with fewer weights than simply increasing network capacity with additional first layer processing nodes.

When comparing the FIRNN and LDA classifiers, it was found that the overall classification accuracy rates were quite similar for up to four or five frames (depending on the subject). Afterwards, the classification accuracy of the LDA

classifier began to degrade in increasing amounts. The LDA classifier behaviour was consistent with the curse-of-dimensionality, where due to the rapidly increasing dimensionality of the input feature space, the available data became increasingly sparse. To combat this effect, principal component analysis was used to transform the data into a set of linearly uncorrelated variables or components. From Figure 4.7, we can see that PCA had significant positive effect on the classification accuracy of the LDA; when used with the FIRNN, PCA had little to no effect. Overall classification accuracy of the LDA classifier with PCA approached the overall classification accuracy of the FIRNN without PCA. The results also indicate that the FIRNN appears less affected by the curse-of-dimensionality, which may be due to the organization of the FIRNN. With no delays in the first layer processing nodes, we could say the first layer nodes approximate the principal components by determining which combination of input features are passed on to the second layer for temporal evaluation.

5. Confidence

In this Chapter, we'll examine the topic of confidence and how it can be derived from the FIRNN outputs and used in the decision making process. As described by Scheme *et al.* [13], it can be more beneficial to reject a potentially correct decision and do nothing, than to implement an incorrect one. While that particular study involved upper-limb prostheses, the idea of rejecting potentially incorrect decisions is perhaps even more important in the use of lower-limb prosthesis. The wearer of the lower-limb prosthesis relies on the device for support and balance; incorrect mode selections can trigger imbalances potentially leading to a fall.

Unlike upper-limb prostheses where the no-motion class can be selected for rejected inputs, lower-limb prostheses are in continuous movement during locomotion and any sudden stops are inadvisable. The exact protocol to follow under such circumstances is outside the scope of this thesis, however possible outcomes could include: assessment on the likelihood of transition from the current mode, if unlikely maintain current locomotion mode, if likely then the incorporation of a hybrid safe mode could be investigated. While not ideal, the hybrid safe mode may offer enough functionality over a range of modes to at least prevent critical disturbances.

5.1. FIRNN Outputs

Outputs of the FIRNN final layer nodes are used to make classification decisions. When all frames of the input pattern have been shifted into the FIRNN such that the entire pattern is contained within the FIRNN structure, the chosen class is determined by selecting the final layer node with the largest output.

A hyperbolic tangent transfer function implemented according to: output = {2/(1 + exp(-2 x input)) - 1}, was used for both layers such that all final layer outputs were constrained between +1 and -1. Final layer outputs were then linearly mapped from the interval [-1,1] to the interval [0,1] according to: new output value = (original output value + 1)/2.

Given a sufficiently rich training set, a complex network architecture and training success in minimizing the mean square error; it can be shown that a neural network approximates a Bayes classifier and that the output of each final layer node represents the posterior probability that the current pattern belongs to the associated class and thus the output probabilities should sum to +1 [17].

Examining the final layer sum of outputs across all subjects, all gait event classifiers, all four cross-validations and all initial weight configurations when the input patterns consisted of 8 frames, we obtained the distribution as shown in Figure 5.1.

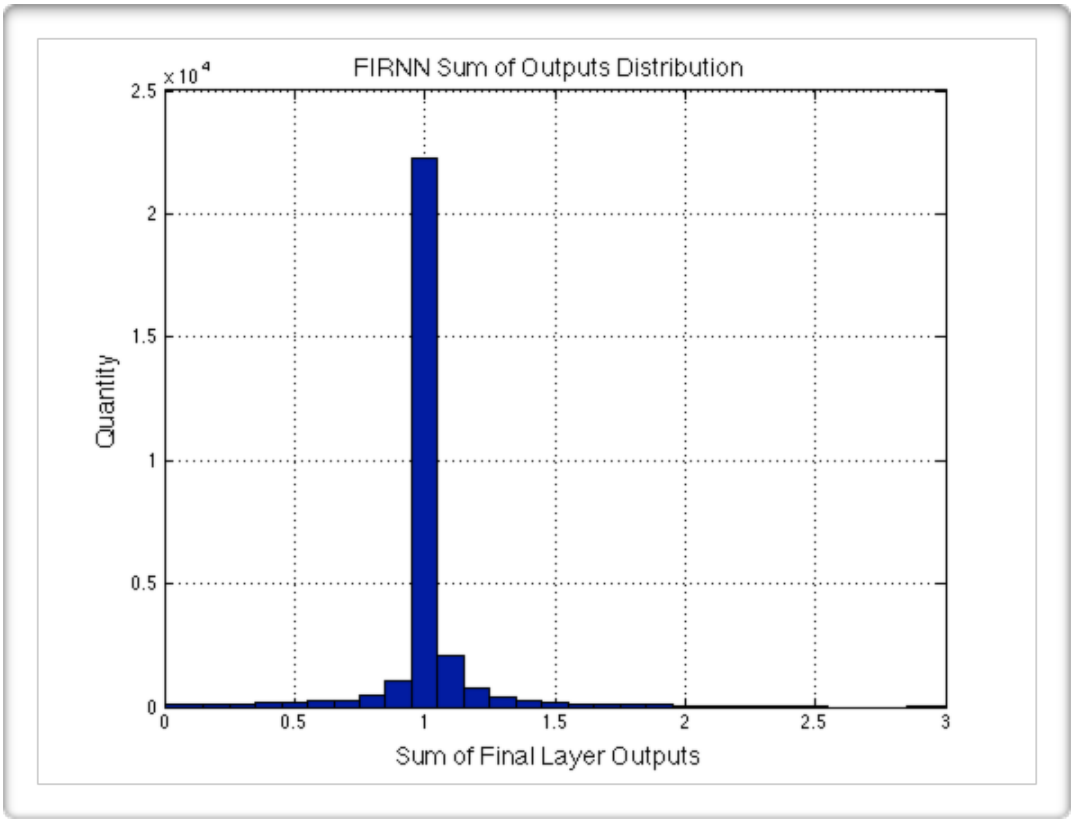

Figure 5.1: FIRNN sum of final layer outputs distribution over all subjects, all event classifiers, all cross-validations and all initial weight configurations for 8 frames

Given that the sum of outputs distribution is highly clustered in the bin centered at 1.0 (bin width of 0.1), we can determine that the initial assumptions regarding neural network Bayes classifier approximation held true in this case.

5.2. Confidence Based Rejection

The Bayes classifier framework allows for a rejection option if the probability of the current pattern belonging to the selected class does not exceed a specified decision threshold. The goal of such an approach is that by rejecting patterns

with low confidence, improvements in accuracy rate and confidence can be obtained for the remaining non-rejected patterns.

Two confidence metrics, D0 and D1, were defined as follows:

- D0 is the value of the largest final layer output at the time of classification (*i.e.,* when the input pattern was fully contained within the FIRNN) and represents the probability that the presented pattern belongs to the associated class.
- D1 is the difference between the value of the final layer output with the highest output and the value of the output node with the next highest output.

Conceptually, we can think of confidence based rejection with metric D0 as rejecting patterns with a low probability of the presented pattern belonging to the selected class. Likewise, we can think of confidence based rejection with metric D1 as rejecting patterns where there is a high likelihood of confusion between two classes.

Using this confidence based rejection approach, the accuracy rate, average confidence and rejection ratio (the ratio of patterns rejected to total number of patterns in the test set) were determined for each confidence metric. Results are given for all subjects and shown for both steady state and transition patterns.

The decision threshold was varied between 0 and +1 in increments of 0.001 and the input patterns consisted of eight 250ms non-overlapping frames.

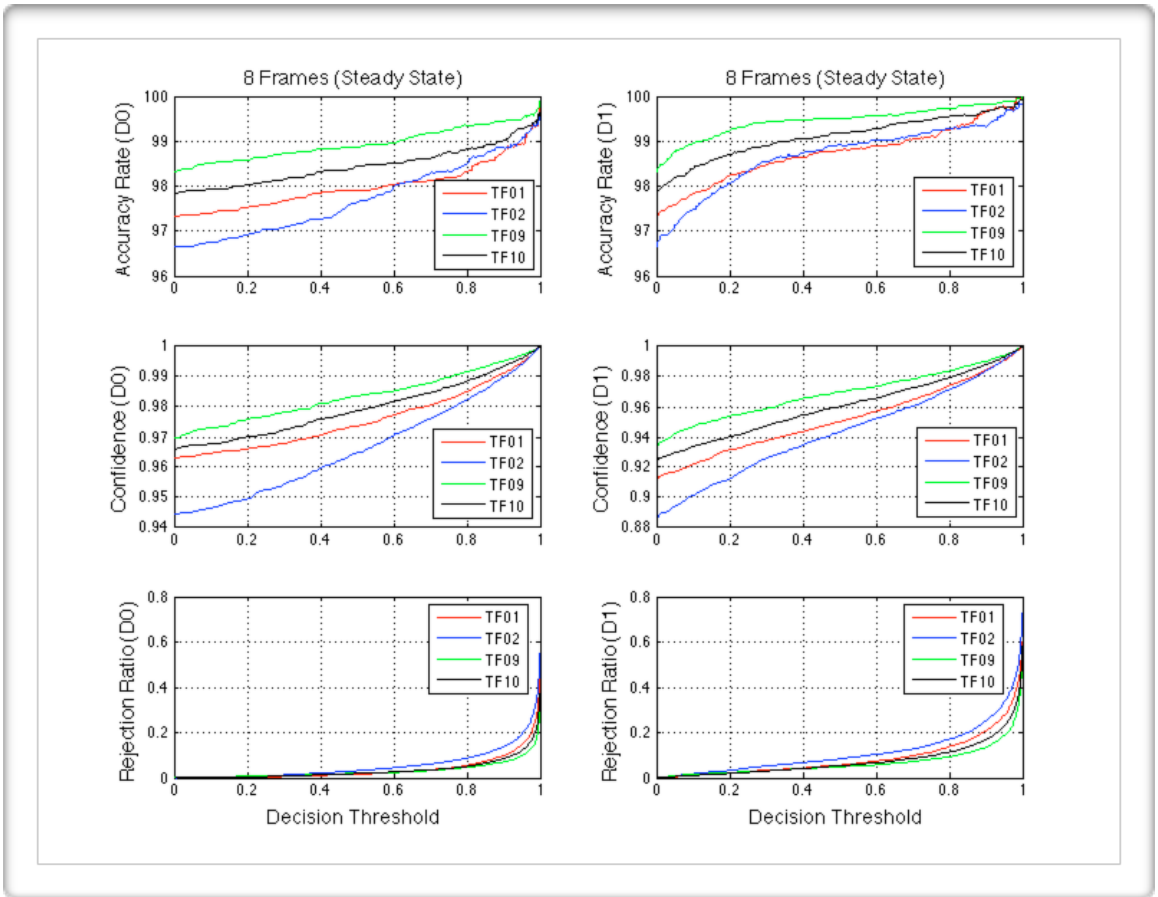

Figure 5.2: Confidence based rejection for steady state patterns (8 frames)

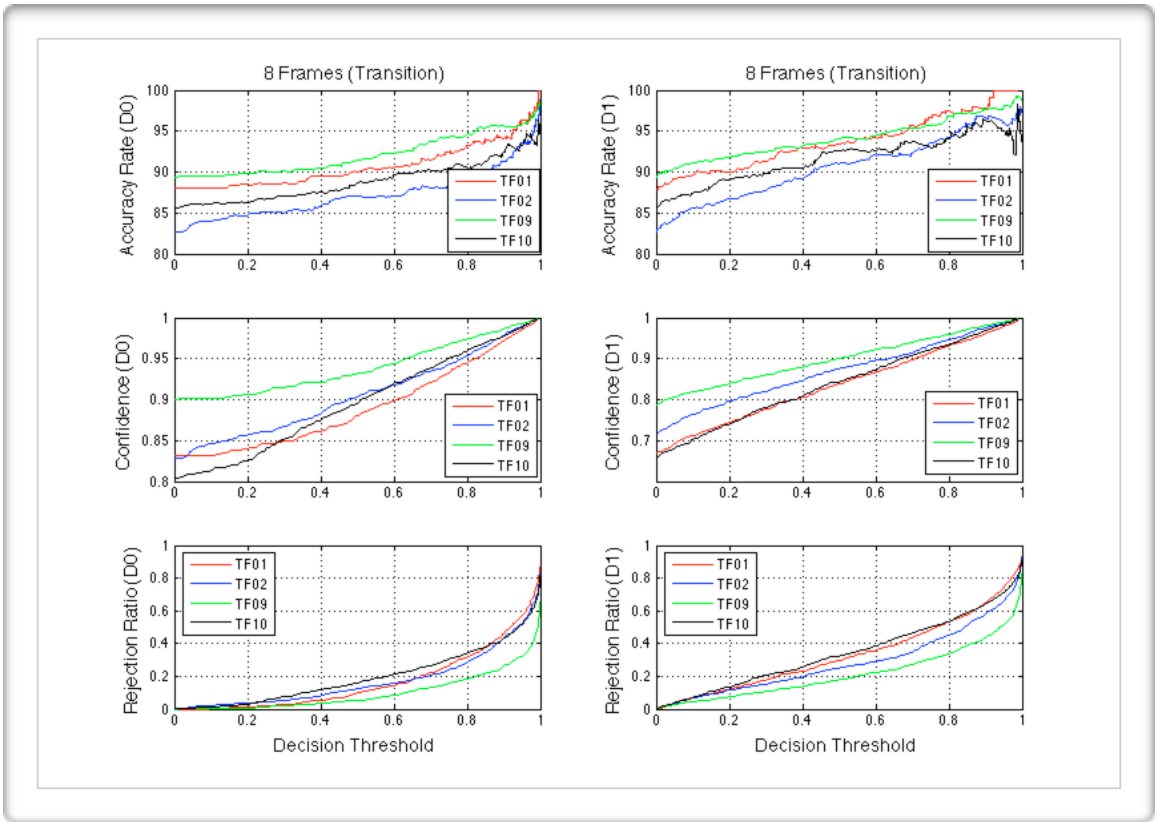

Figure 5.3: Confidence based rejection for transition patterns (8 frames)

From Figures 5.2 and 5.3 we can note the following:

- increasing accuracy rates and average confidence with increasing decision threshold.
- for a given decision threshold, confidence metric D1 appears to offer improved accuracy and confidence though it may come at a higher rejection rate.
- improvements are realized for both steady state and transition patterns
- for a given decision threshold, proportionally fewer steady state patterns are rejected when compared to transition patterns.

Plotting accuracy rate against rejection ratio allows for a more efficient comparison between confidence metrics. The desired outcome is for the curves to reach as close as possible to the top left corner as this region represents high classification accuracy with low rejection. The point on the curve in the lower left represents a decision threshold of 0 (no rejection) and the curve progresses in decision threshold increments of 0.001. The curve end point in the top right region represents a decision threshold of 0.999.

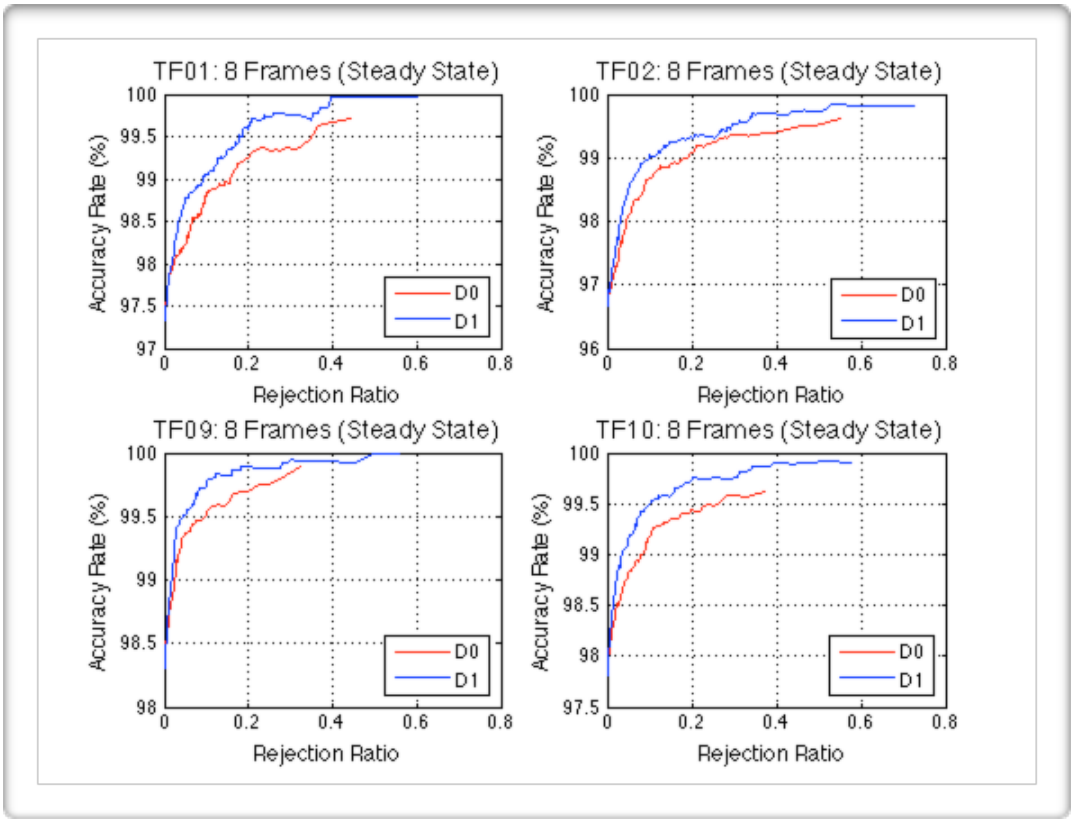

Figure 5.4: Accuracy rate vs rejection ratio for steady state patterns (8 frames)

Of note, the rejection ratio at the curve end point, RR_{ep}, would indicate a confidence above 0.999 for (1 - RR_{ep}) fraction of patterns tested. Therefore, the

desired outcome is for the curve to terminate as soon as possible such that the end point would be at a low rejection ratio.

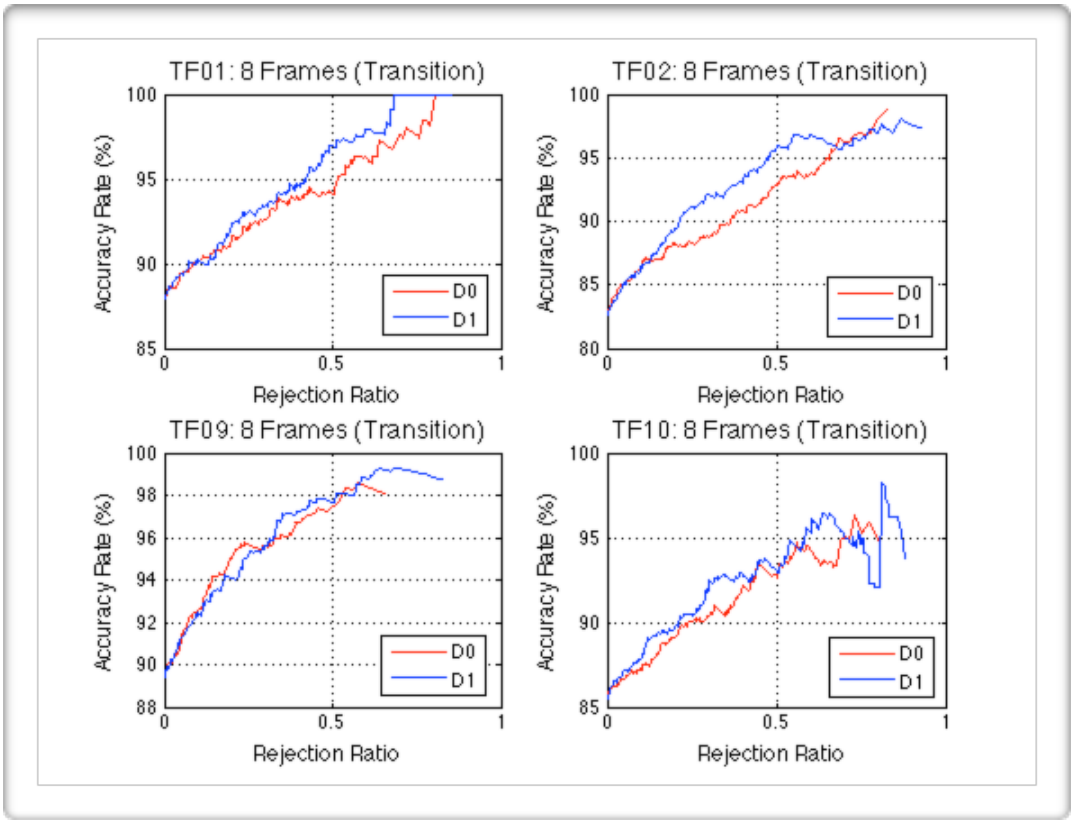

Figure 5.5: Accuracy rate vs rejection ratio for transition patterns (8 frames)

For steady state patterns, confidence metric D1 outperformed D0 for all subjects. On transition patterns, the performance of the two metrics was similar for rejection ratios smaller than approximately 0.2, after which metric D1 appears to offer a slight performance advantage. Given that the two metrics are not completely unrelated, the result is perhaps not surprising. Any pattern with low confidence according to metric D0, will by definition also have low confidence according to metric D1. Therefore, metric D1 can be thought of as not only

rejecting patterns that have a high likelihood of confusion between 2 classes, but also rejecting patterns that have a low probability of belonging to the selected class.

Also noteworthy from Figures 5.4 and 5.5 is the slope of the red and blue curves indicating that for relatively low rejection rates, measurable improvements in classification accuracy can be obtained.

To demonstrate how specific locomotion mode classification decisions might be affected by confidence based rejection using metric D1, Tables 5.1 and 5.2 display a combined confusion and confidence matrix for subject TF10 having an observation interval of eight non-overlapping frames. Table 5.1 displays the confusion and confidence data when no rejection was applied (*i.e.,* the decision threshold was set to 0.0) and Table 5.2 displays the confusion and confidence data for the remaining patterns after a decision threshold of 0.5 was applied. The integer in each cell represents the traditional confusion matrix data and the number in parentheses represents the average confidence, according to metric D1, for all decisions located within that cell.

From Table 5.1, we see that the largest source of confusion is between the level walking (LW) and ramp classes (RA, RD); this has also been noted in previous studies using similar data [5,6]. Also evident from Table 5.1, the average confidence of correct decisions (the diagonal elements of the matrix where the

chosen class equals the target class) is notably higher than the average confidence of incorrect decisions.

		Target Class				
		LW	*RA*	*RD*	*SA*	*SD*
Chosen Class	*LW*	4995 (0.94)	67 (0.56)	58 (0.29)	16 (0.39)	8 (0.25)
	RA	49 (0.45)	832 (0.87)	2 (0.08)	5 (0.29)	0 (n/a)
	RD	39 (0.33)	1 (0.10)	835 (0.86)	3 (0.26)	5 (0.14)
	SA	7 (0.06)	0 (n/a)	0 (n/a)	576 (0.89)	0 (n/a)
	SD	10 (0.20)	0 (n/a)	5 (0.68)	0 (n/a)	572 (0.86)

Table 5.1: Combined confusion and confidence matrix for all TF10 classification decisions (decision threshold = 0.0, confidence metric D1, n/a indicates not applicable, eight frames)

		Target Class				
		LW	*RA*	*RD*	*SA*	*SD*
Chosen Class	*LW*	4817 (0.97)	34 (0.88)	13 (0.69)	6 (0.72)	1 (0.59)
	RA	23 (0.76)	743 (0.94)	0 (n/a)	2 (0.63)	0 (n/a)
	RD	9 (0.70)	0 (n/a)	738 (0.93)	0 (n/a)	0 (n/a)
	SA	0 (n/a)	0 (n/a)	0 (n/a)	542 (0.93)	0 (n/a)
	SD	0 (n/a)	0 (n/a)	4 (0.78)	0 (n/a)	507 (0.93)

Table 5.2: Combined confusion and confidence matrix for non-rejected TF10 classification decisions (decision threshold = 0.5, confidence metric D1, n/a indicates not applicable, eight frames)

When confidence based rejection was applied to the FIRNN outputs, as shown in Table 5.2, we note that confusion between level walking and the ramp classes was significantly reduced and the overall classification accuracy increased from

96.6% to 98.8%. Confidence in correct decisions increased incrementally, however the confidence of the remaining incorrect classification decisions also significantly increased. While not a preferred outcome, this is readily explained by the fact that all incorrect decisions having low confidence were rejected, thus by definition, the remaining non-rejected incorrect decisions will have much higher confidence figures.

In this case, the decision threshold of 0.5 was somewhat arbitrarily selected to demonstrate the effect of using confidence based rejection on specific locomotion mode classification decisions. A real-world implementation would likely need to involve a tuning process to assess acceptable rejection ratios and subject/ prosthesis response to rejected decisions.

Previously, the confidence based rejection figures for classification accuracy, average confidence and rejection ratio were shown for a fixed number of frames with a variable decision threshold. Figures 5.6 and 5.7 show the same measures of classification accuracy, average confidence and rejection ratio for a fixed decision threshold (0.5) and a variable number for frames.

As shown in Figures 5.6 and 5.7, using a fixed decision threshold of 0.5, measures of classification accuracy, average confidence and rejection ratio improve with supplementary frames added to the input pattern.

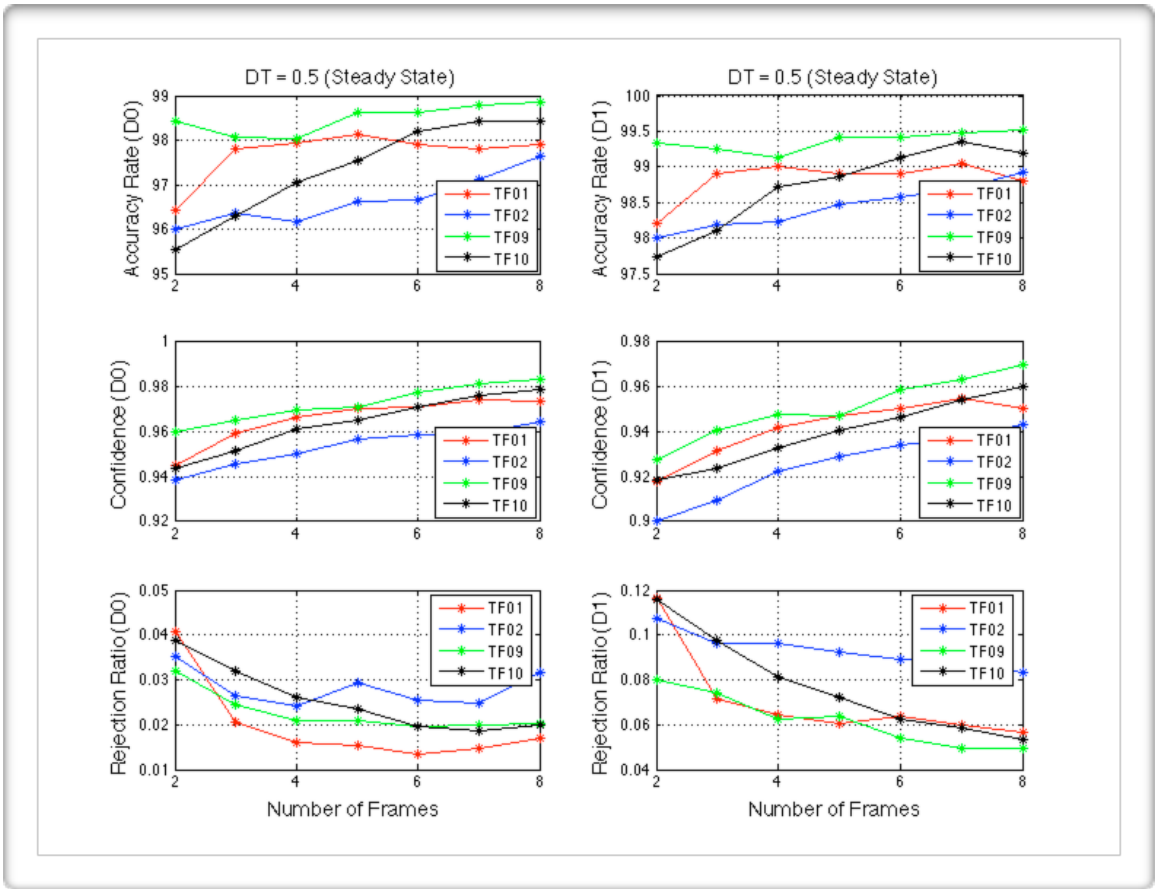

Figure 5.6: Confidence based rejection for fixed decision threshold (0.5) - steady state

As the number of frames increases from two to eight frames, while applying a rejection threshold of 0.5 to the D1 confidence metric, the transition performance improved an average of 11.5% in classification accuracy, 7% in confidence and 11% in rejection ratio. While similar improvements in transition performance were obtained with metric D0, the overall performance level using metric D1 was, on average, higher by 3.5% in terms of classification accuracy at eight frames, although an average of 15.5% additional patterns were rejected. Improvements in steady state performance were notably smaller; however, given that the starting performance was much higher, this is to be expected.

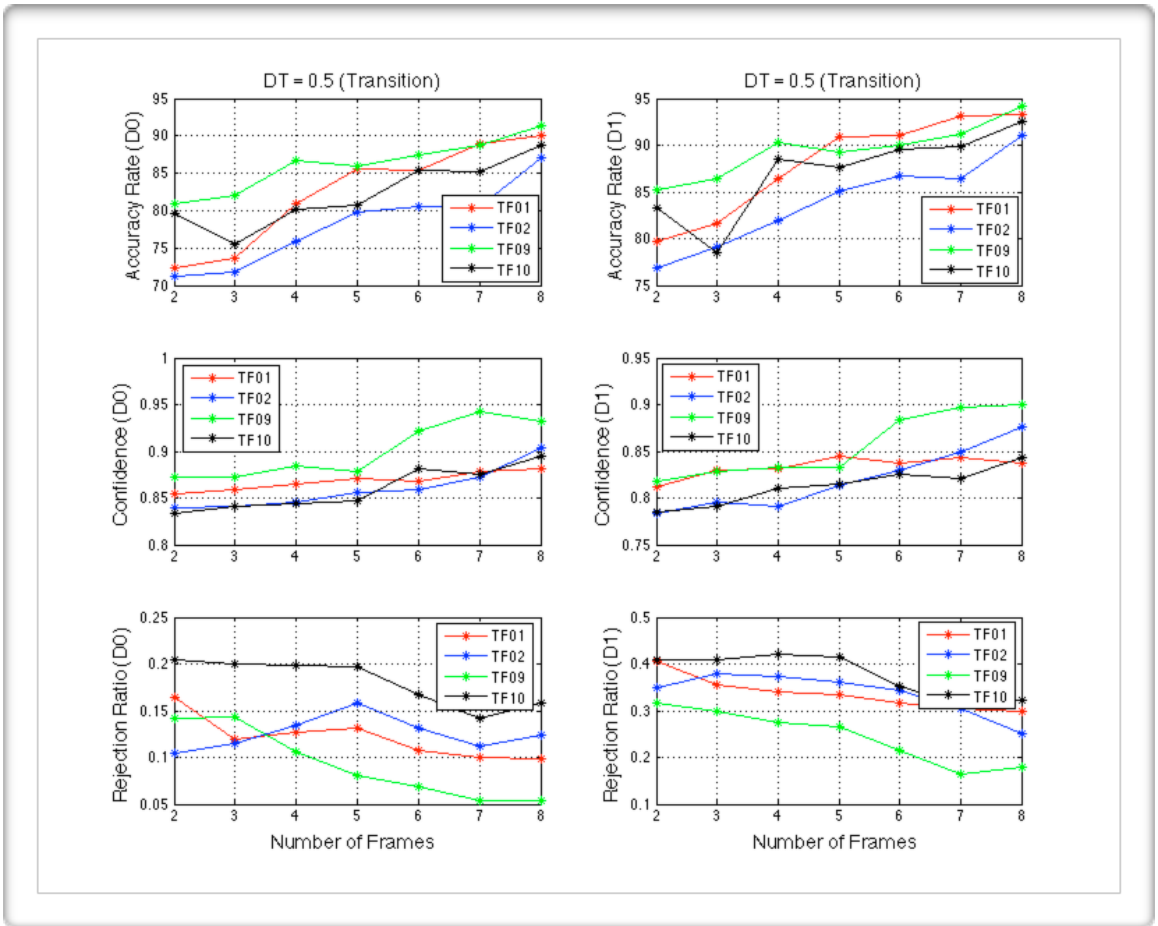

Figure 5.7: Confidence based rejection for fixed decision threshold (0.5) - transition

5.3. Discussion

Using the property that the normalized output of each FIRNN final layer node represent the probability that the current pattern belongs to each class and the sum of these output probabilities should sum to +1, two confidence metrics were accessed over a range of decision thresholds and number of input frames. Confidence metric D0 was intended to reject patterns with a low probability of the presented pattern belonging to the selected class and metric D1 was intended

to reject patterns where there is a high likelihood of confusion between two classes.

For both steady state and transition patterns, classification accuracy and confidence increased with an increasing rejection threshold. Proportionally more transition patterns were rejected over steady state patterns for a given rejection threshold. Plotting accuracy against the rejection ratio showed that confidence metric D1 outperformed D0 for both steady state and transition patterns.

In a past study using the same RIC data set, Young *et al.* [5] showed that level walking and ramp classes are easily confused; the results reported in Table 5.1 are consistent with those findings. While Young *et al.* took the approach of combining the level walking, ramp ascent and ramp descent into a single class with the use of a slope estimator, the use of a rejection based confidence metric may offer an alternate solution without class combination or additional hardware as indicated by the results shown in Table 5.2. Finally, as the number of frames was increased, Figures 5.6 and 5.7 showed increasing classification accuracy, increasing confidence and declining rejection ratios for all four subjects.

6. Conclusions and Future Work

6.1. Conclusions

As discussed in section 1.3, the main goal of this work was to improve intent recognition for subjects using powered lower limb prosthesis by making use of the FIRNN and its temporal processing capabilities. Within that main goal, four objectives were identified.

1. *Transition Performance: Improve upon existing transition classification results when using phase-dependent classifiers.* In [6], Horton achieved a 87.9% overall accuracy rate for subject TF01 in a preliminary investigation using FIRNNs. That study utilized the same RIC data with the same EMG features but was limited to six 250ms overlapping frames in 50ms increments (resulting in an observation interval of 500ms). In this work, by expanding the observation interval to 2000ms with eight 250ms non-overlapping frames, a 96.3% overall accuracy rate was achieved for subject TF01 using a two second analysis window. In Young *et al.* [5] a 11.3% transition error rate was obtained with a Dynamic Bayesian Network when all five locomotion modes were kept in distinct classes. From the results identified in Figure 4.3, a transition classification accuracy of 86.4% (or 13.6% error rate) was obtained with the FIRNN when using eight non-

overlapping 250ms frames. While this error rate is higher than the error rate achieved by Young *et al.* some key difference existed between the two studies. In this work, the input data features consisted of four EMG features (MAV, SSC, ZC, WL) only whereas Young *et al.* used the same four EMG features plus two autoregressive coefficients and four features derived from the 13 mechanical sensors. Given that the addition of autoregressive EMG features and mechanical features have both shown to aid in classification, the results obtained in this work with only four EMG features indicate reasonable likelihood of further improvements with the addition of mechanical and autoregressive EMG features.

2. *Gait Cycle: Examination of longer analysis windows than typically used with phase-dependent classifiers.* From the results identified in Figure 4.1, we can see that for an increasing number of non-overlapping 250 ms frames, which also increases the observation interval, overall classification accuracy increases. Table 1 also indicates that despite the same number of 250ms frames, when overlap is increased, causing the overall analysis window to shrink, classification performance suffers. Assuming a frame length of 250ms, classification accuracy with use of overlapping frames is also comparable to an equivalent analysis window made up of non-overlapping frames. Based on these results, we can say that there is additional

information available from the gait cycle that is useful for classification as the overall analysis window extends as far back as a complete gait-cycle or more.

3. *FIRNN Confidence: Apart from using FIRNN outputs for classification decisions, can useful information be obtained to aid in classification and assess confidence?* From Chapter 5, we saw that by implementing a confidence based rejection system, both classification accuracy and confidence in executed decisions can be improved with increasing decision threshold when using a fixed number of frames in the analysis window. For a fixed decision threshold, we observe increasing classification accuracy, increasing confidence and decreasing rejection ratios as more frames are included in the analysis window. Confidence metric D1, which determines the probability difference between the two largest outputs, outperformed confidence metric D0 which represents the probability of the presented pattern belonging to the class with the largest output. Past studies [5,6] have identified errors between level walking and ramp classes as the main source of confusion; confidence based rejection with a metric D1 may offer an alternative over merging locomotion mode classes and addition of extra hardware.

4. *FIRNN Classifier: Can the FIRNN structure offer performance improvements over more common techniques such as Linear Discriminant Analysis.* From Chapter

4, we see that the FIRNN structure is effective in extracting information from non-overlapping frames as they evolve over time and use that information for locomotion mode identification. When comparing FIRNN classification accuracy against the LDA classifier, it was found that the LDA classification performance suffered due to the curse-of-dimensionality and that the FIRNN appears immune to this phenomena. With the addition of Principal Component Analysis to LDA inputs, classification performance reached nearly the same level as the FIRNN classifier without application of PCA. Using PCA with the FIRNN had negligible effect on its peak performance. Given the way in which the delays are allocated within the FIRNN, where all the delays are assigned to the second layer processing nodes, we could say that the first layer approximates PCA by reducing the number of input components passed on to the second layer for temporal processing.

Recent results obtained by the RIC found significant correlation between offline and and real-time classification error-rates [18]. While offline error-rates were found to be lower than real-time error-rates, the most accurate control system (Dynamic Bayesian Network) included time history information and EMG signals. Given that the FIRNN architecture implemented for this work

incorporates both time history information and EMG data, the offline results reported here, in conjunction with the recent RIC results [18], may offer useful insight into how a FIRNN based neural control system might react in a real-time implementation.

6.2. Suggestions for Future Work

Suggestions for future work topics include:

1. Repeating the multi-frame analysis on FIRNN with use of non-overlapping frames constructed from features using both additional EMG features as well as features derived from the mechanical sensors.
2. With the receipt of additional subject data, some of which have data over multiple days, the multi-frame analysis with the FIRNN could be expanded over additional subjects. Testing could also be expanded to include data from the second day with training performed with data from the first day. This would allow for assessment of FIRNN resilience to EMG disturbances created as a result of doffing/re-donning the prosthesis [19].
3. Since the data are largely populated with steady state patterns, use of error scaling may assist with transition performance. This could allow for increased emphasis on errors obtained from transition patterns during the training process.

Bibliography

[1] L.J. Hargrove, A.M. Simon, R.D. Lipschutz, S.B. Finucane, T.A. Kuiken, "Non-weight-bearing neural control of a powered transfemoral prosthesis," *Journal of NeuroEngineering and Rehabilitation*, Vol. 10, No. 62, 2013

[2] A.J. Young, A.M. Simon, L.J. Hargrove, "A Training Method for Locomotion Mode Prediction Using Powered Lower Limb Prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 22, No. 3, May 2014

[3] L.J. Hargrove, H. Huang, A.E. Schultz, B.A. Lock, R.D. Lipschutz, T. A. Kuiken, "Toward the development of a neural interface for lower limb prosthesis control," *Proc. 31st Int'l Conf. IEEE Eng. Med. Biol. Soc.*, Minneapolis, MN, USA, Sept. 2-6, 2009

[4] A.J. Young, A.M. Simon, N.P Fey, L.J. Hargrove, "Intent Recognition in a Powered Lower Limb Prosthesis Using Time History Information," *Annals of Biomedical Engineering*, Vol. 42, No. 3, March 2014

[5] A.J. Young, A.M. Simon, L.J. Hargrove, "An Intent Recognition Strategy for Transfemoral Amputee Ambulation Across Different Locomotion Modes," *35th Annual International Conference of the IEEE EMBS*, Osaka, Japan, 3 - 7 July 2013

[6] C. Horton, "Enhancing the Use of EMG in Event-Dependent Mode Classification for Powered Lower Limb Prosthesis Control," M.Sc.Eng Thesis, University of New Brunswick, February 2014

[7] H. Huang, T.A. Kuiken, R.D. Lipschutz, "A Strategy for Identifying Locomotion Modes Using Surface Electromyography," *IEEE Transactions on Biomedical Engineering*, Vol. 56, No. 1, January 2009

[8] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82-94, Jan. 1993

[9] http://www.physio-pedia.com/Gait_Cycle

[10] F. Sup, A. Bohara, M. Goldfarb, “Design and control of a powered transfemoral prosthesis,” Int. J. Robot. Res., Vol. 27, 2008

[11] F. Sup, H.A. Varol, M. Goldfarb, "Upslope Walking with a Powered Knee and Ankle Prosthesis: Initial Results with an Amputee Subject,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 19, No. 1, Feb. 2011

[12] H. Huang, F. Zhang, L.J. Hargrove, Z. Dou, D.R. Rogers, K.B. Englehart, “Continuous Locomotion-Mode Identification for Prosthetic Legs Based on Neuromuscular-Mechanical Fusion,” *IEEE Transactions on Biomedical Engineering*, Vol. 58, No. 10, October 2011

[13] E.J. Scheme, B.S. Hudgins, K.B. Englehart, “Confidence-Based Rejection for Improved Pattern Recognition Myoelectric Control,” *IEEE Transactions on Biomedical Engineering*, Vol. 60, No. 6, June 2013

[14] E. Wan, “Finite Impulse Response Neural Networks with Application in Time Series Prediction,” Ph.D. Dissertation, Stanford University, Stanford, CA, Nov. 1993

[15] R. Bellman, “Adaptive Control Processes: A Guided Tour,” 5th Edition, *Princeton University Press*, 1961

[16] H. Hotelling, ”Analysis of a complex statistical variables into principal components,” *Journal of Educational Psychology,* 1933, vol. 24, p. 417-441

[17] M.D. Richard and R.P. Lippmann, “Neural Network Classifiers Estimate Bayesian a Posteriori Probabilities,” *Neural Computation*, 1991, vol. 3, p. 461-483

[18] L. J. Hargrove, A. J. Young, A. M. Simon, N. P. Fey, R. D. Lipschutz, S. B. Finucane, E. G. Halsne, K. A. Ingraham, and T. A. Kuiken, “Intuitive control of a powered prosthetic leg during ambulation: A randomized clinical trial,” *The Journal of the American Medical Association,* June, 2015

[19] J. Spanias, E. J. Perreault, L. J. Hargrove, "Detection and Compensation for EMG Disturbances for Powered Lowing Limb Prosthesis Control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* TNSRE-2014-00169.R1

[20] L.J. Hargrove, A.M. Simon, A.J. Young, R.D. Lipschutz, S. B. Finucane, D.G. Smith, T.A. Kuiken, "Robotic Leg Control with EMG Decoding in an Amputee with Nerve Transfers," *The New England Journal of Medicine*, 2013 Sept 26

[21] M.T. Hagan, M.B Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE Transactions on Neural Networks,* Vol. 5, No. 6, November 1994

[22] K. B. Englehart, B. S. Hudgins, M. Stevenson, P. A. Parker, "Myoelectric signal classification using a finite impulse response neural network," *Proc. 69 16th Int'l Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 28 - Sept. 1, 1994, Vol. 2.

[23] A.M. Huff, B.E. Lawson, M. Goldfarb, "A running controller for a powered transfemoral prosthesis," *Proc. 34th Int'l Conf. IEEE Eng. Med. Biol. Soc.*, San Diego, CA, USA, Aug. 28 - Sept. 1, 2012

[24] B.E. Lawson, H.A. Varol, M. Goldfarb, "Standing stability enhancement with an intelligent powered transfemoral prosthesis," *IEEE Transactions Biomedical Engineering*, Vol. 58, No. 9, Sept. 2011

[25] H.A. Varol, F. Sup, M. Goldfarb, "Multiclass real time intent recognition of a powered lower limb prosthesis," *IEEE Transactions Biomedical Engineering*, Vol. 57, No. 3, Mar. 2010

[26] P. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Exp. Psychol.,* vol. 47, pp. 381-391, 1954

[27] R. Duda, P.E. Hart, D.G. Stork, "Pattern Classification," Second Edition, Wiley-Interscience, 2001

Appendix A

Supplemental Figures

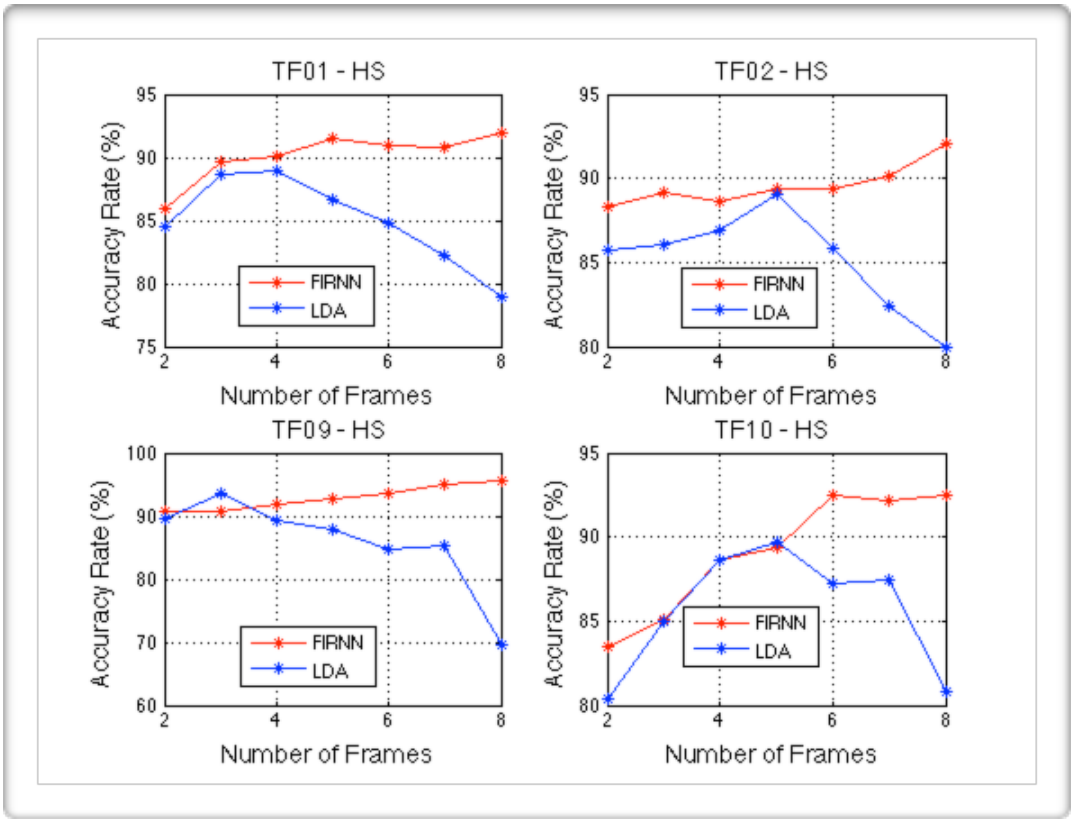

Figure A.1: HS accuracy rates for LDA and FIRNN classifiers using 250ms non-overlapping frames

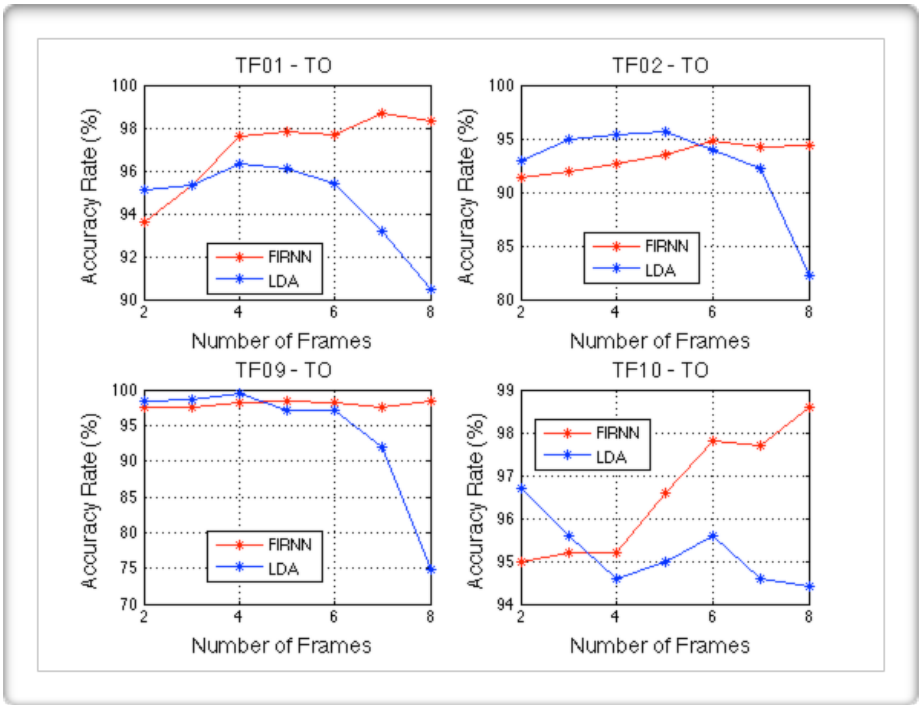

Figure A.2: TO accuracy rate for LDA and FIRNN classifiers using 250ms non-overlapping frames

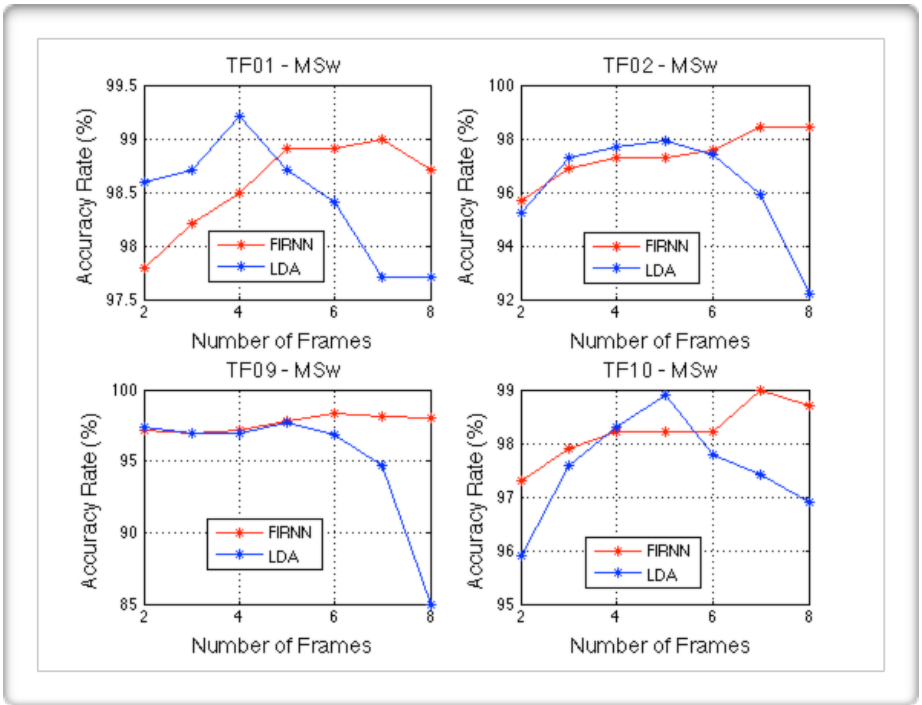

Figure A.3: MSw accuracy rates for LDA and FIRNN classifiers using 250ms non-overlapping frames

Curriculum Vitae

Candidate's full name: Norman J. Arsenault

Universities attended: University of New Brunswick, BScE 1994